

ARTICLE

Received 11 Feb 2015 | Accepted 16 Jul 2015 | Published 1 Sep 2015

DOI: 10.1038/ncomms9101

OPEN

# Cas9-Assisted Targeting of Chromosome segments CATCH enables one-step targeted cloning of large gene clusters

Wenjun Jiang<sup>1</sup>, Xuejin Zhao<sup>2</sup>, Tslil Gabrieli<sup>3</sup>, Chunbo Lou<sup>2</sup>, Yuval Ebenstein<sup>3</sup> & Ting F. Zhu<sup>1</sup>

The cloning of long DNA segments, especially those containing large gene clusters, is of particular importance to synthetic and chemical biology efforts for engineering organisms. While cloning has been a defining tool in molecular biology, the cloning of long genome segments has been challenging. Here we describe a technique that allows the targeted cloning of near-arbitrary, long bacterial genomic sequences of up to 100 kb to be accomplished in a single step. The target genome segment is excised from bacterial chromosomes *in vitro* by the RNA-guided Cas9 nuclease at two designated loci, and ligated to the cloning vector by Gibson assembly. This technique can be an effective molecular tool for the targeted cloning of large gene clusters that are often expensive to synthesize by gene synthesis or difficult to obtain directly by traditional PCR and restriction-enzyme-based methods.

<sup>1</sup>School of Life Sciences, Center for Synthetic and Systems Biology, MOE Key Laboratory of Bioinformatics, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Tsinghua University, Beijing 100084, China. <sup>2</sup>CAS Key Laboratory of Microbial Physiological and Metabolic Engineering, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China. <sup>3</sup>Raymond and Beverly Sackler Faculty of Exact Sciences, School of Chemistry, Tel Aviv University, Tel Aviv 6997801, Israel. Correspondence and requests for materials should be addressed to C.L. (email: louchunbo@gmail.com) or to Y.E. (email: uv@post.tau.ac.il) or to T.F.Z. (email: tzhu@biomed.tsinghua.edu.cn).

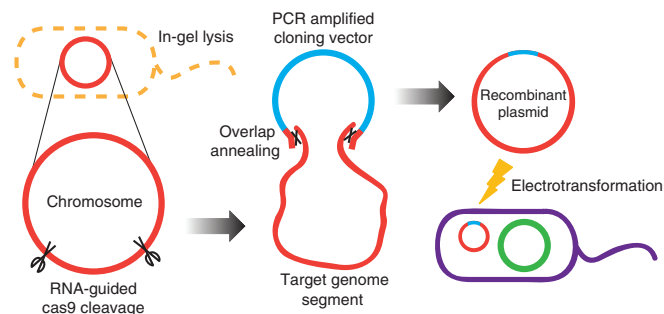
Cloning long genome segments of large gene clusters is an imperative step in engineering organisms for the purposes of producing high-added-value biomolecules such as pharmaceuticals and biofuels<sup>1–6</sup>. Traditional PCR-based cloning methods are often limited by the sequence length and GC content of the DNA template: standard PCR reactions routinely yield fragments of up to 10 kb, while longer PCR products require tedious optimization of reaction conditions and, even under ideal conditions for special cases, are typically limited to 35 kb (ref. 7). Alternatively, one may generate long genomic sequences of interest through the assembly of multiple short fragments, such as overlapping PCR products or chemically synthesized DNA oligos, although such methods tend to be time-consuming and expensive, particularly for obtaining sequences longer than 50 kb (which typically require three to five stages, each containing multiple assembly events)<sup>8,9</sup>. Another route to obtain long genomic sequences is by restriction enzyme digestion of genomic DNA. However, being a non-targeted approach, selecting a specific sequence of interest from a vast number of restriction digest products can be intensely challenging and cumbersome<sup>10</sup>. Certain techniques such as transformation-associated recombination (TAR)<sup>11,12</sup> and single-strand overlapping annealing<sup>13</sup> have been developed to clone-specific, large bacterial gene clusters following restriction enzyme digestion. Nevertheless, the utility of these techniques remains limited because they rely on the availability of unique restriction sites that flank the target genomic region and often the presence of selection markers in the target sequence. To facilitate advancements in biotechnology and synthetic biology, it is crucial to develop a general approach to clone near-arbitrary, long genomic sequences that are difficult to obtain using conventional methods.

The CRISPR-Cas9 endonuclease can be directed by guide RNAs to cleave specific sequences of DNA targets<sup>14–17</sup>. In the case of the *Streptococcus pyogenes* Cas9 (spCas9), the guide RNA system consists of a CRISPR RNA (crRNA) and a trans-activating crRNA (tracrRNA), which can be fused to form a single guide RNA (sgRNA)<sup>18</sup>. The target sequence is composed of a crRNA-complementary protospacer sequence and a protospacer adjacent motif (PAM)<sup>18</sup>. Here the use of long (~20 bp) and programmable protospacer sequences as recognition sites offers the spCas9 endonuclease system higher targeting specificity and versatility than traditional restriction enzymes (with fixed recognition sites limited to 4–8 bp). This has motivated extensive development of Cas9-based tools for genome-editing, sequence-specific control of gene expression, as well as imaging *in vivo*<sup>19–21</sup>. In contrast, the potential applications of the Cas9 system *in vitro* have not yet been explored; instead, they mainly focused on testing the enzyme's cleavage efficiency and sequence-recognition specificity<sup>22,23</sup>.

Here we show that in addition to being a versatile genome-editing tool *in vivo*, the *in vitro* application of spCas9 brings unparalleled efficiency and simplicity to the cloning of large gene clusters, allowing the targeted cloning of near-arbitrary, long bacterial genomic sequences of up to 100 kb to be accomplished in a single step. The technique described below can be an effective molecular tool for the targeted cloning of large gene clusters that are often expensive to synthesize by gene synthesis or difficult to obtain directly by traditional PCR and restriction-enzyme-based methods.

## Results

**Design of the CATCH cloning method.** In our cloning method by Cas9-Assisted Targeting of CHromosome segments (CATCH; Fig. 1), bacterial chromosomes after cell lysis are digested by RNA-guided Cas9 at designated target sites in agarose gel. The



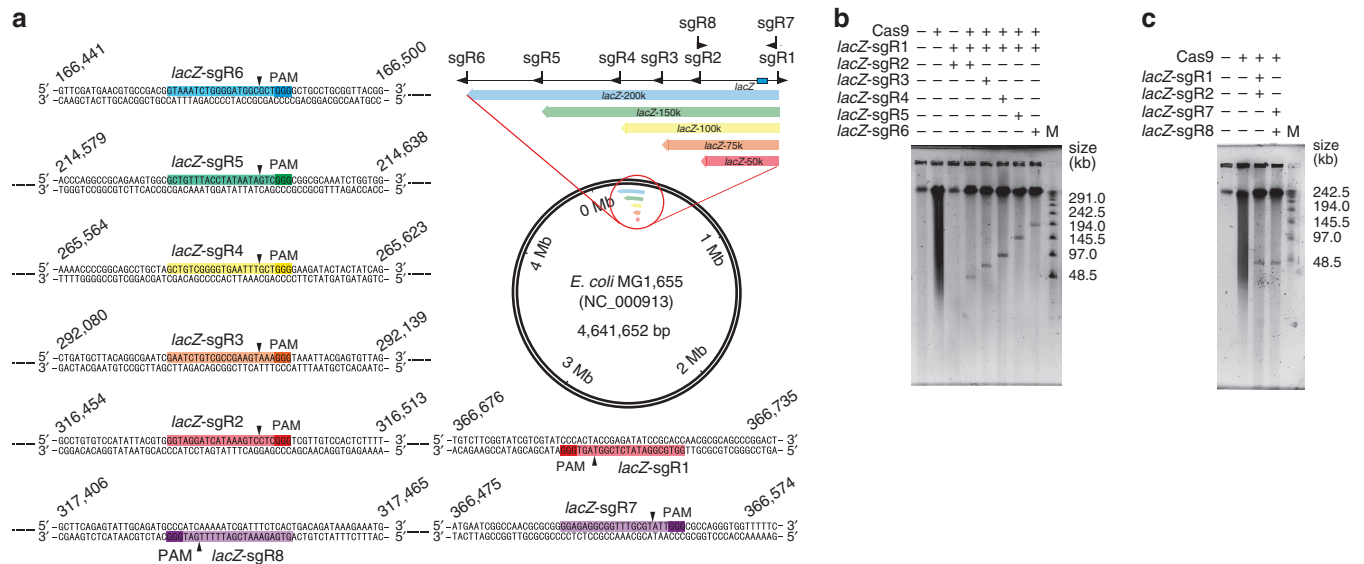
**Figure 1 | One-step large-gene-cluster cloning by CATCH.** After in-gel lysis of bacterial cells, the chromosomes are cleaved by RNA-guided Cas9 at the designated target sites. A cloning vector (length not to scale) that shares 30-bp terminal sequence overlaps (black cross) with the target DNA at both ends is ligated to the target segment in a Gibson assembly mix. The recombinant plasmid is then electrotransformed into a cloning host.

cloning vectors are designed so that they share terminal sequence overlaps (30 bp) with the target DNA at both ends, and are ligated to the target DNA through sequence complementarity in a Gibson assembly mix<sup>9</sup>. The recombinant plasmids are then electrotransformed into a cloning host. The procedure takes ~8 h of bench time over 1–2 days to accomplish using standard laboratory equipment at reasonable costs, which drastically simplifies and accelerates efforts to clone large bacterial genomic sequences.

## Cleavage and isolation of long genome segments by CATCH.

To test the nuclease activity of Cas9 in agarose gel and its effectiveness in isolating long DNA sequences, we designed five sgRNA pairs to target segments of different lengths (50, 75, 100, 150 and 200 kb guided by *lacZ*-sgRNAs 1–6, respectively; see Methods) in the *Escherichia coli* genome, all containing a *lacZ* gene (Fig. 2a). After being embedded in low-melting-temperature agarose gel plugs, bacterial cells were treated by lysozyme and proteinase K, and washed by buffer successively to remove cellular components, leaving behind the genomic DNA. The intact chromosomes were protected by the agarose matrix, allowing for further manipulations with minimal mechanical shearing. The plugs were soaked in a reaction buffer containing pre-assembled Cas9 with the corresponding sgRNA pair and incubated at 37 °C for 2 h to allow for sufficient enzyme diffusion and digestion of genomic DNA in agarose gel. After digestion, 1/3 of a plug was cut out for pulsed-field gel electrophoresis (PFGE) to assess the cleavage efficiency (Fig. 2b). A clear band at the expected length was observed in each of the five lanes, while the control lanes showed either no band or heavy smear, suggesting sufficient cleavage specificity of the RNA-guided Cas9 in agarose gel.

To optimize the RNA-guided Cas9 in-gel digestion and to understand factors affecting the cleavage efficiency, we prepared agarose gel plugs of different volumes, each containing bacterial cells of different concentrations, and treated with different concentrations of Cas9 and sgRNA at different incubation time durations (Supplementary Fig. 1). We showed that bacterial cell concentration at  $5 \times 10^8 \text{ ml}^{-1}$  was optimal for both cleavage and visualization on PFGE gel. We also showed that final concentrations of 0.02–0.1  $\text{mg ml}^{-1}$  Cas9 was optimal for in-gel digestion (note that the enzyme activity may vary depending on the expression and purification methods used to obtain the enzyme). Notably, using a final concentration of sgRNA  $>30 \text{ ng } \mu\text{l}^{-1}$  appeared to be essential for the in-gel digestion and Cas9 cleavage, and thus careful preparation of sufficient sgRNA is



**Figure 2 | CATCH cleavage and cloning of long genomic sequences of various lengths.** (a) A total of six sgRNA pairs were designed to target genome segments of different lengths (50, 75, 100, 150 and 200 kb, respectively, including one pair targeting 50 kb but with the PAM sequences falling on the flanking sequences instead of the target genome segment) in the *E. coli* str. K-12 substr. MG1655 genome, all containing a *lacZ* gene. The genome location, target sequences, PAM sequences and cleavage sites are detailed and highlighted. (b) *E. coli* chromosomes in agarose gel plugs were digested by Cas9 with the corresponding sgRNA pairs and analysed using PFGE. (c) A 50-kb segment on the *E. coli* chromosomes was targeted by two sgRNA pairs of opposite orientations, and the cleavage products were analysed using PFGE. M, marker.

**Table 1 | Cloning efficiency of different target sequences and vectors.**

Target	Vector	Number of positive colonies	Number of total colonies	Average positive rate ± s.d. (%)
<i>E. coli-lacZ</i> 50 kb	pCC1BAC	28 ± 4	42 ± 8	65.8 ± 4.0
<i>E. coli-lacZ</i> 50 kb (R)	pCC1BAC	26 ± 8	41 ± 7	62.2 ± 9.6
<i>E. coli-lacZ</i> 75 kb	pCC1BAC	8 ± 3	23 ± 8	35.1 ± 1.6
<i>E. coli-lacZ</i> 100 kb	pCC1BAC	5 ± 3	28 ± 15	21.5 ± 11.2
<i>E. coli-lacZ</i> 150 kb	pCC1BAC	1*	51*	1.9*
<i>E. coli-sfGFP</i> 50 kb	pCC1BAC	25 ± 6	47 ± 12	55.2 ± 12.4
<i>E. coli-sfGFP</i> 75 kb	pCC1BAC	12 ± 4	29 ± 5	39.2 ± 6.4
<i>E. coli-sfGFP</i> 100 kb	pCC1BAC	7 ± 3	35 ± 12	21.6 ± 7.4
<i>B. subtilis-pks</i> 78 kb	pCC1BAC	4 ± 1	34 ± 9	12.2 ± 3.7
<i>S. venezuelae-jad</i> 36 kb	p15A	60 ± 4	66 ± 10	90.5 ± 6.1
<i>S. aureofaciens-ctc</i> 32 kb	p15A	70 ± 3	78 ± 6	90.8 ± 3.7

All experiments were performed in triplicates.

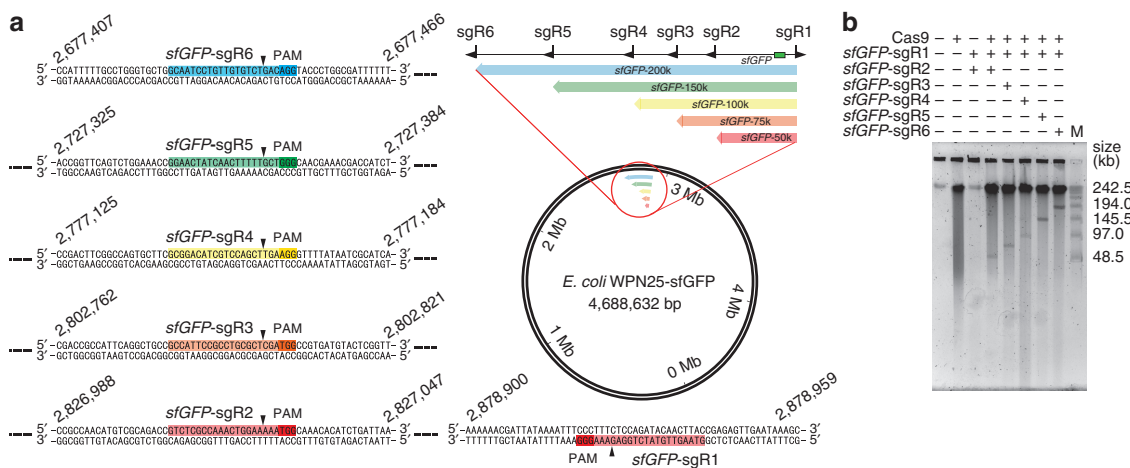
\*Only one positive colony was obtained in three independent trials.

critical to the success of CATCH cloning. In addition, an incubation time of 1–2 h was sufficient for Cas9 diffusion and cleavage in gel plugs.

Furthermore, we tested whether the orientation of sgRNA targeting could influence the cleavage and cloning efficiency, potentially affected by the post-cleavage exonucleolytic processing activity of Cas9 (ref. 18). We designed the *lacZ*-sgRNAs 7 and 8 to target the same 50-kb genome segment containing a *lacZ* gene as *lacZ*-sgRNAs 1 and 2 do, but with the PAM sequences falling on the flanking sequences instead of the target genome segment, leaving the two exonucleolytically processed ends on the target sequence (Fig. 2a). The cleavage and cloning experiments suggested that no apparent difference of cleavage efficiency (Fig. 2c) and cloning efficiency as a result of different sgRNA-targeting orientation was observed (Fig. 4a ‘*E. coli-lacZ* 50 kb (R)’, Table 1; see results below). To test whether the location of target genome segments could affect the cleavage and cloning efficiency, we designed another five pairs of sgRNAs (*sfGFP*-sgRNAs 1–6) targeting five genome segments (50, 75, 100, 150 and 200 kb in length,

respectively) at different genome locations (2,677,407–2,828,959 in the *sfGFP*-knock-in *E. coli* strain *E. coli* DH5a str. WPN25-*sfGFP* (Fig. 3a) versus 166,441–366,735 in the *lacZ*-gene-containing strain *E. coli* str. K-12 substr. MG1655 (Fig. 2a)). The results suggested that the cleavage efficiency (Figs 2b and 3b) and cloning efficiency (see results below) were unaffected when different locations of the genome were targeted.

**Cloning various long genome segments.** Having successfully cleaved the long genomic sequences of interest from bacterial chromosomes, we attempted to ligate the target DNA into BAC vectors in a Gibson assembly mix<sup>3</sup>, taking advantage of the high sequence specificity of Gibson assembly, without the need for size selection on PFGE and gel purification from the background genomic DNA. We first prepared BAC vectors that shared 30-bp terminal sequence overlaps with the target DNA (Fig. 1; see Methods). The remaining of the Cas9-digested plugs (one- and two-thirds of gel plugs, after 1/3 of a plug has been used for PFGE from a total of two starting plugs) were pooled, melted and digested using



**Figure 3 | CATCH cleavage and cloning of long genomic sequences of various lengths from different genomic locations.** (a) A total of five sgRNA pairs were designed to target segments of different lengths (50, 75, 100, 150 and 200 kb, respectively) in the *E. coli* *DH5a* str. WPN25-sfGFP genome, all containing an *sfGFP* gene. The genome location, target sequences, PAM sequences and cleavage sites are detailed and highlighted. (b) *E. coli* chromosomes in agarose gel plugs were digested by Cas9 with the corresponding sgRNA pairs and analysed by PFGE. M, marker.

agarase, after which the DNA was purified using ethanol precipitation and resuspension in nuclease-free water (see Methods). We then mixed the recovered DNA and vector in a Gibson assembly mix containing T5 5'–3' exonuclease, *Taq* DNA ligase and high-fidelity polymerase. Finally, the ligation mix was electrotransformed into electrocompetent *E. coli* cells. Depending on the length of the target DNA to be cloned, we obtained 50–100 colonies on selective LB plates containing chloramphenicol, isopropyl- $\beta$ -D-thiogalactoside (IPTG) and X-gal, among which 20–65% appeared blue (for detailed data, see Fig. 4a, Table 1 and Supplementary Table 2). All blue clones were selected and validated using PCR and sequencing at one of the two junction sites opposite to the *lacZ* gene (see Methods, Supplementary Fig. 2). The cloned BAC plasmids were purified, linearized and analysed with PFGE (Fig. 4b). All of the blue colonies appeared to be positive clones with correct insert sizes ranging from 50 to 150 kb; however, only one positive clone was obtained with the 150-kb insert and none with 200 kb (Table 1 and Supplementary Table 2).

We next tested whether the orientation of sgRNA targeting and the location of target genome segments could affect the cloning efficiency. We showed that the cloning efficiency was unaffected by different sgRNA-targeting orientations (Fig. 4a '*E. coli-lacZ* 50 kb (R)', Table 1 and Supplementary Table 2). We attribute these results in part to the robustness of the Gibson assembly that ensures the ligation with the target sequence. Furthermore, the cloning efficiency of genome segments of the same lengths but at different genome locations (166,441–366,735 in *E. coli* str. K-12 substr. MG1655 (Fig. 2a) versus 2,677,407–2,828,959 in *E. coli* *DH5a* str. WPN25-sfGFP (Fig. 3a)) were comparable (Fig. 4a, Table 1, Supplementary Fig. 2 and Supplementary Table 2). Taken together, our results suggested that the positive rates were size-dependent (where the cloning of DNA segments of longer than 100 kb was less efficient), but not genome-location- or sgRNA-orientation-dependent.

**Cloning large gene clusters from different bacteria.** Encouraged by the successful targeted isolation and cloning of genomic sequences from *E. coli* at various lengths, we turned to the cloning of large gene clusters from other bacterial genomes. Here we tested our method on cloning the 78-kb bacillaene-producing *psk* gene cluster (the largest gene cluster in *Bacillus subtilis*)<sup>24</sup>, a gene cluster of particular interest to chemical biologists for producing polyketides<sup>5,6</sup>, into BAC vector. Using the same method

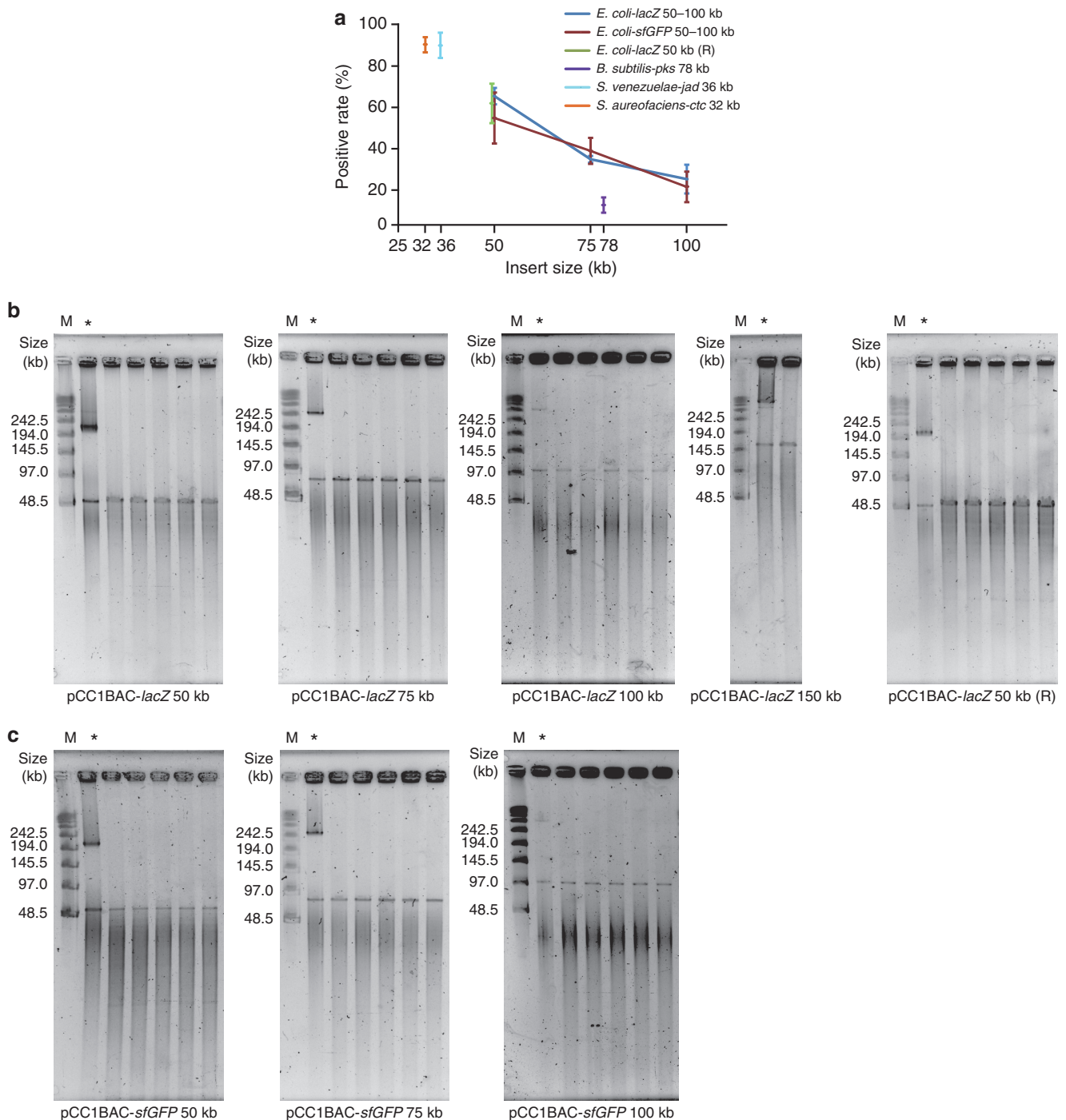
mentioned above, we obtained a total of 12 positive colonies in three trials with an ~12% positive rate (Fig. 4a '*B. subtilis-psk* 78 kb', Fig. 5a, Table 1, Supplementary Fig. 2 and Supplementary Table 2). In addition, we cloned the 36-kb jadomycin-producing *jad* gene cluster from *Streptomyces venezuelae*<sup>25</sup> and the 32-kb chlortetracycline-producing *ctc* gene cluster from *S. aureofaciens*<sup>26</sup> into the p15A vector (Fig. 4a '*S. venezuelae-jad* 36 kb' and '*S. aureofaciens-ctc* 32 kb', Fig. 5b,c). Overall, we obtained ~60 positive colonies in each experiment, with positive rates at ~90% (Fig. 4a, Table 1, Supplementary Fig. 2 and Supplementary Table 2), demonstrating the versatility of CATCH on cloning various bacterial genomic sequences into different cloning vectors.

## Discussion

A long-standing challenge in traditional cloning is how to isolate and clone specific genomic sequences that are difficult to obtain using PCR or restriction-enzyme-based methods. Taking advantage of the RNA-guided CRISPR-Cas9 system *in vitro* to cleave and clone long DNA sequences, CATCH can synergize with gene synthesis, synthetic biology and chemical biology efforts for purposes such as producing biofuels, pharmaceuticals or other high-added-value biomolecules<sup>1–6</sup>.

So far, we have been able to efficiently clone bacterial genomic sequences of up to 100 kb, a sufficient length for most known bacterial gene clusters. We showed that CATCH cloning of DNA segments longer than 100 kb was less efficient, despite that each step (Cas9 digestion and DNA segment separation, Gibson assembly and BAC maintenance) alone could handle longer DNA fragments, suggesting that ~150 kb could be a size limit to the CATCH cloning method itself with the current experimental set-up. The reduced cloning efficiency with longer inserts may in part result from mechanical shearing of long DNA fragments during manipulation, the lower ligation efficiency of Gibson assembly with longer DNA and the lower electroporation efficiency of larger plasmids. It is worth noting that the amount of DNA available from the *in vitro* Cas9 digestion of bacterial chromosomes was significantly less than what would be available for typical Gibson assembly:  $\sim 1.5 \times 10^7$  cells are embedded in an agarose gel plug, resulting in only a few nanograms of DNA of the desired length. Furthermore, the efficiency of CATCH cloning can be affected by other factors related to the cleavage efficiency of Cas9, for example, the design of the sgRNA sequences<sup>27</sup>.



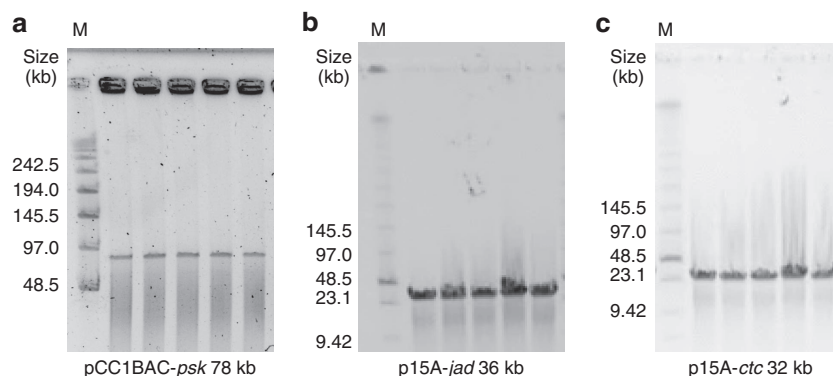


**Figure 4 | Efficiency of CATCH cloning of different genomic sequences.** (a) The positive rates of CATCH cloning with different insert sizes and target sequences from different genome locations, as well as from different bacteria are analysed (error bar: s.d.;  $n = 3$ ). *E. coli-lacZ* 50 kb (R) is the cloning result of a 50-kb genome segment targeted by *lacZ*-sgRNAs 7 and 8, with the PAM sequences falling on the flanking sequences instead of the target genome segment. (b) Plasmids carrying the target sequences of different lengths cloned from *E. coli* str. K-12 substr. MG1655 (ligated to an 8-kb BAC vector) were purified from the blue-white-screening- and PCR-positive clones, linearized and analysed by PFGE. (c) Plasmids carrying the target sequences of different lengths cloned from *E. coli* *DH5a* str. WPN25-*sfGFP* were purified, linearized and analysed by PFGE. M, marker; asterisk,  $\lambda$ -terminase-free control.

Recent studies showed that the use of truncated or extended versions of guide RNA can reduce the off-target effects of Cas9 nucleases<sup>28,29</sup>. In the current study, we have used a variety of PAM sequences (Supplementary Table 3), and the cleavage and cloning results suggested that the selection of PAM sequences (including their orientations with respect to the target genome segment) did not appear to affect the in-gel cleavage and cloning

efficiency of CATCH (Fig. 4a, Table 1 and Supplementary Table 2). The off-target effect of Cas9 is less concerning in the cloning of large gene clusters from bacteria genomes; however, this could be a concern when CATCH is applied to other complex genome systems in the future.

To understand the mechanisms responsible for the generation of negative clones, we analysed the plasmids purified from the



**Figure 5 | Cloning large gene clusters from different bacteria into different vectors.** Plasmids carrying the target large gene clusters cloned from *B. subtilis* (a), *S. venezuelae* (b) or *S. aureofaciens* (c), respectively (see Table 1 and Supplementary Table 2), were purified from the PCR-positive clones, linearized and analysed using PFGE. M, marker.

negative clones by Sanger sequencing, which suggested that most negative clones arose from self-ligation of vectors at sites that share identical sequence of  $\geq 6$  bp (Supplementary Fig. 3). We reasoned that vectors nicked at the sites with identical sequences were chewed by the T5 5'–3' exonuclease, annealed and self-ligated in the Gibson assembly mix; these small, self-ligated plasmids, containing the origin of replication and antibiotic resistance gene but without the target DNA insert, have higher electrotransformation efficiency and faster replication rate than those with long inserts. Thus, we avoided long identical sequences ( $\geq 6$  bp) on the vector when designing target sites and purified the vectors from PCR products with gentle handling (by pipetting slowly with wide-bore tips).

The ability of CATCH to isolate near-arbitrary DNA segments from the known, flanking sequences may greatly accelerate efforts to fill the gaps in sequenced genomes. Adaptations of CATCH to larger genomes would make it a powerful tool, enabling numerous applications such as targeted cloning of large genes from plants and mammals, targeted sequencing of disease-specific genes in humans or identification of copy-number variations. Efforts to quantify and optimize the specificity and efficiency, and reduce the off-target effects of the CRISPR-Cas9 system<sup>30–32</sup>, as well as discoveries of other targeted nuclease systems with improved specificity should be made to facilitate the application of CATCH to other areas.

While this paper was being reviewed, an alternative method for cloning long genomic DNA segments on the basis of in-solution Cas9 digestion of purified human genomic DNA and yeast TAR was published<sup>33</sup>. While the in-solution digestion by Cas9 is more convenient to carry out, the CATCH in-gel cleavage protects chromosomal DNA from shearing, permitting the cloning of longer DNA sequences of up to 150 kb. Furthermore, another apparent advantage of using the in-gel Cas9 cleavage instead of in-solution digestion of purified genomic DNA is that, because the rest of the genomic DNA is protected by the agarose gel matrix, the background DNA fragments that could be falsely ligated and cloned, leading the generation of negative clones, can be greatly reduced. In addition, using in-gel cleavage as opposed to in-solution digestion allows for immediate PFGE assessment of the cleavage efficiency (Supplementary Fig. 1), and thus makes the troubleshooting and optimization of the Cas9 cleavage feasible. Although TAR uses long hooks of up to a few hundred bp for homologous recombination, the cloning has to be carried out *in vivo* in yeast, and an additional step would be required to clone the inserts into BAC for maintenance and further genetic manipulation. The use of *in vitro* Gibson assembly in CATCH, on the other hand, permits one-step ligation and cloning into BAC to

be accomplished. Future adaptations of both methods, for example, combining the in-gel digestion method of CATCH and TAR, may facilitate efforts to clone large gene clusters from bacteria, as well as long genome segments from other organisms.

## Methods

**Strains and plasmids.** *E. coli* str. K-12 substr. MG1655 was obtained from the laboratory of Professor Guoqiang Chen (Tsinghua University). *E. coli* DH5a str. WPN25-sfGFP was obtained by integrating an *sfGFP* expression cassette into the corresponding chromosomal integration site (2,875,434–2,875,653) using the One-Step Integration Plasmid<sup>34</sup>. The sequence of the integration site was verified using PCR and sequence analysis. *B. subtilis* str. 168 was obtained from the laboratory of Professor Tingyi Wen (Institute of Microbiology, Chinese Academy of Sciences). *S. venezuelae* ISP52030 and *S. aureofaciens* ATCC 10762 were obtained from the laboratory of Professor Keqian Yang (Institute of Microbiology, Chinese Academy of Sciences). Electrocompetent *E. coli* cells TransforMax EPI300 and BAC vector pCC1BAC were purchased from Epicentre. The p15A vector was provided by Professor Keqian Yang (Institute of Microbiology, Chinese Academy of Sciences).

**Expression and purification of Cas9 and T7 RNA polymerase.** The plasmid for *S. pyogenes* Cas9 protein expression was provided by Professor Zhen Xie (Tsinghua University). The histidine-tagged fusion protein Cas9 was purified using a Ni-NTA column and the AKTA system with a Mono S column<sup>18</sup>. In total,  $\sim 10$  mg of Cas9 protein was purified and stored in 20 mM HEPES, 150 mM KCl, 1 mM dithiothreitol (DTT), 50% glycerol, pH 7.5 at  $-20^\circ\text{C}$ . The plasmid for T7 RNA polymerase expression was provided by Professor Nieng Yan (Tsinghua University) and the enzyme was purified using a Ni-NTA column and the AKTA system with a Mono Q column<sup>35</sup>.

**sgRNA design and preparation.** To design the sgRNAs used for in-gel digestion, we first searched for PAM sequence 'NGG' near the target genomic region, with 'G' being the starting nucleotide of a transcript because it is ideal for both Cas9 loading<sup>36</sup> and *in vitro* transcription (IVT) by T7 RNA polymerase. The sgRNA IVT templates were prepared by overlapping PCR of three primers: a primer (X-sgRNA-P) containing the T7 promoter and target sequence, and two others (sgRNA-F and sgRNA-R) containing crRNA–tracrRNA chimera sequence of the sgRNA. All the primers used in this study are listed in Supplementary Table 1. The PCR product was purified by phenol (pH > 7.8)/chloroform extraction and isopropanol precipitation, followed by resuspension in RNase-free water. The IVT was performed at  $37^\circ\text{C}$  for 2 h in 100 mM Tris-HCl, pH 8.0, 10 mM MgCl<sub>2</sub>, 30 mM DTT, 2 mM Spermidine, 2.5 mM (each) nucleoside triphosphate (NTP), 10% dimethylsulphoxide, 100 ng  $\mu\text{l}^{-1}$  PCR product and 50  $\mu\text{g ml}^{-1}$  T7 RNA polymerase. After incubation, the IVT product was purified by phenol (pH < 5.2)/chloroform extraction and isopropanol precipitation. Finally, the sgRNA was resuspended in RNase-free water at a concentration of 300 ng  $\mu\text{l}^{-1}$  and stored at  $-80^\circ\text{C}$ .

**Genome digestion and recovery.** *E. coli* or *B. subtilis* cells were embedded in agarose gel plugs at a concentration of  $5 \times 10^8$  cells  $\text{ml}^{-1}$ . These plugs were treated with lysozyme, proteinase K and washed with buffer successively according to instructions of the CHEF Bacterial Genomic DNA Plug Kit (Bio-Rad). In the second wash, 1 mM of phenylmethyl sulphonyl fluoride was added to inactivate the residual proteinase K and 0.1  $\times$  wash buffer was used for the last wash. The

*Streptomyces* mycelia were harvested and used for making gel plugs, as described in ref. 37. The well-washed plugs can be stored in  $1 \times$  wash buffer at  $4^\circ\text{C}$  for 2 months, and another round of wash should be performed using  $0.1 \times$  wash buffer immediately before Cas9 digestion. For the cleavage reaction, two plugs were first equilibrated at room temperature for 30 min in 1 ml RNase-free cleavage buffer containing 20 mM HEPES, 150 mM KCl, 10 mM  $\text{MgCl}_2$ , 0.5 mM DTT and 0.1 mM EDTA at pH 7.5, and then transferred into a new batch of cleavage buffer that contains Cas9 protein ( $0.1 \text{ mg ml}^{-1}$ ) and the corresponding sgRNA pair (each at  $30 \text{ ng } \mu\text{l}^{-1}$ ), and incubated at  $37^\circ\text{C}$  for 2 h. After the reaction, the plugs were washed with  $0.1 \times$  wash buffer, and 1/3 of a gel plug was cut out and used to assess the cleavage efficiency by PFGE. The PFGE was performed with 1% agarose gel in  $0.5 \times$  TBE using the CHEF Mapper XA System (Bio-Rad) set to auto-algorithm programme with 5–250 kb parameters ( $6 \text{ V cm}^{-1}$ , 0.22–21.79 s, 15 h 16 min,  $120^\circ$ ) and with circulation at  $14^\circ\text{C}$ . After PFGE, the gel was stained with SYBR Gold (Life Technologies) and the DNA bands were visualized using a ChemiDoc XRS + Imaging System (Bio-Rad). The rest of the plugs were melted and digested by agarase according to instructions of the GELase Agarose Gel-Digesting Preparation Kit (Epicentre). The digested DNA was precipitated by ethanol and resuspended gently in  $20 \mu\text{l}$  DNase-free water with wide-bore tips. The obtained DNA can be stored at  $4^\circ\text{C}$  for several days, although immediate ligation is recommended.

**Vector ligation and cloning.** The vectors were prepared by PCR amplification using pCC1BAC (Epicentre) or p15A vectors as templates, followed by DpnI (NEB) treatment<sup>38</sup>. Each PCR primer consisted of an  $\sim 20$ -nucleotide (nt) sequence that annealed to the vector template and an  $\sim 30$ -nt overhang that overlapped with one end of the target DNA. The Gibson assembly mix was prepared as described in the one-step isothermal DNA assembly protocol<sup>9</sup>. A volume of  $1 \mu\text{l}$  vector ( $\sim 50 \text{ ng}$ ) and  $4 \mu\text{l}$  of previously prepared target DNA (with a viscous appearance) were added to  $15 \mu\text{l}$  of Gibson assembly mix by gentle pipetting and incubated at  $50^\circ\text{C}$  for 1 h. After ligation,  $2 \mu\text{l}$  of the mix was transformed into  $50 \mu\text{l}$  TransforMax EPI300 (Epicentre) electrocompetent *E. coli* cells in a 1-mm cuvette (BTX) at 1,300 V using the ECM 399 Electroporation system (BTX). The cells were recovered at  $37^\circ\text{C}$  for 2 h in 1 ml LB medium without antibiotics and then plated on LB medium containing  $12.5 \mu\text{g ml}^{-1}$  chloramphenicol, IPTG and X-gal. After incubation at  $37^\circ\text{C}$  for 12 h, the blue colonies were selected for PCR and sequencing at one of the two junction sites opposite to the *lacZ* gene (with one primer on pCC1BAC (BAC-vF) and the other on the insert (*lacZ*-X kb-vR)). The positive clones were grown in 5 ml LB medium overnight at  $37^\circ\text{C}$  containing  $12.5 \mu\text{g ml}^{-1}$  chloramphenicol and  $30 \mu\text{l}$  CopyControl BAC Autoinduction Solution (Epicentre). Plasmids were extracted from these cells using the QIAprep Miniprep Kit using the QIAcube system (Qiagen) according to the manufacturer's instructions. The purified plasmids were linearized with  $\lambda$ -terminase (Epicentre) and analysed using PFGE as described above. For cloning the *sfGFP*-containing genome segments from *E. coli*, no IPTG or X-gal was added to the LB medium, and the green colonies were selected for PCR and sequencing at one of the two junction sites opposite to the *sfGFP* gene (with one primer on pCC1BAC (BAC-vF) and the other on the insert (*sfGFP*-X kb-vR)). For cloning the *pks* gene cluster from *B. subtilis*, all colonies were PCR- and sequence-verified at both junction sites (Supplementary Fig. 2). The *jad* gene cluster from *S. venezuelae* and *ctc* gene cluster from *S. aureofaciens* were cloned into p15A vector using a similar method, except that the colonies were grown on LB medium with  $50 \mu\text{g ml}^{-1}$  ampicillin and the plasmids were linearized by XbaI.

## References

- Endy, D. Foundations for engineering biology. *Nature* **438**, 449–453 (2005).
- Keasling, J. D. Synthetic biology and the development of tools for metabolic engineering. *Metab. Eng.* **14**, 189–195 (2012).
- Cobb, R. E. & Zhao, H. Direct cloning of large genomic sequences. *Nat. Biotechnol.* **30**, 405–406 (2012).
- Gronenberg, L. S., Marcheschi, R. J. & Liao, J. C. Next generation biofuel engineering in prokaryotes. *Curr. Opin. Chem. Biol.* **17**, 462–471 (2013).
- Pfeifer, B. A., Admiraal, S. J., Gramajo, H., Cane, D. E. & Khosla, C. Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. *Science* **291**, 1790–1792 (2001).
- Wong, F. T. & Khosla, C. Combinatorial biosynthesis of polyketides—a perspective. *Curr. Opin. Chem. Biol.* **16**, 117–123 (2012).
- Barnes, W. M. PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. *Proc. Natl Acad. Sci. USA* **91**, 2216–2220 (1994).
- Li, M. Z. & Elledge, S. J. SLIC: a method for sequence- and ligation-independent cloning. *Methods Mol. Biol.* **852**, 51–59 (2012).
- Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
- Zhang, H. B. *et al.* Construction of BIBAC and BAC libraries from a variety of organisms for advanced genomics research. *Nat. Protoc.* **7**, 479–499 (2012).
- Fu, J. *et al.* Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nat. Biotechnol.* **30**, 440–446 (2012).
- Yamanaka, K. *et al.* Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *Proc. Natl Acad. Sci. USA* **111**, 1957–1962 (2014).

- Wang, R. Y., Shi, Z. Y., Chen, J. C. & Chen, G. Q. Cloning large gene clusters from *E. coli* using *in vitro* single-strand overlapping annealing. *ACS Synth. Biol.* **1**, 291–295 (2012).
- Garneau, J. E. *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67–71 (2010).
- Sapranaukas, R. *et al.* The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res.* **39**, 9275–9282 (2011).
- Hale, C. R. *et al.* RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* **139**, 945–956 (2009).
- Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007).
- Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
- Doudna, J. A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096 (2014).
- Chen, B. *et al.* Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**, 1479–1491 (2013).
- Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).
- Karvelis, T., Gasiunas, G. & Siksnys, V. Programmable DNA cleavage *in vitro* by Cas9. *Biochem. Soc. Trans.* **41**, 1401–1406 (2013).
- Kim, J. M., Kim, D., Kim, S. & Kim, J. S. Genotyping with CRISPR-Cas-derived RNA-guided endonucleases. *Nat. Commun.* **5**, 3157 (2014).
- Vargas-Bautista, C., Rahlwes, K. & Straight, P. Bacterial competition reveals differential regulation of the *pks* genes by *Bacillus subtilis*. *J. Bacteriol.* **196**, 717–728 (2014).
- Han, L., Yang, K., Ramalingam, E., Mosher, R. H. & Vining, L. C. Cloning and characterization of polyketide synthase genes for jadomycin B biosynthesis in *Streptomyces venezuelae* ISP5230. *Microbiology* **140**, 3379–3389 (1994).
- Vanek, Z., Cudlin, J., Blumauerova, M. & Hostalek, Z. How many genes are required for the synthesis of chlortetracycline? *Folia Microbiol. (Praha)*. **16**, 225–240 (1971).
- Doench, J. G. *et al.* Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat. Biotechnol.* **32**, 1262–1267 (2014).
- Cho, S. W. *et al.* Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res.* **24**, 132–141 (2014).
- Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M. & Joung, J. K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat. Biotechnol.* **32**, 279–284 (2014).
- Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827–832 (2013).
- Ran, F. A. *et al.* Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154**, 1380–1389 (2013).
- Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62–67 (2014).
- Lee, N. C., Larionov, V. & Kouprina, N. Highly efficient CRISPR/Cas9-mediated TAR cloning of genes and chromosomal loci from complex genomes in yeast. *Nucleic Acids Res.* **43**, e55 (2015).
- St-Pierre, F. *et al.* One-step cloning and chromosomal integration of DNA. *ACS Synth. Biol.* **2**, 537–541 (2013).
- Zawadzki, V. & Gross, H. J. Rapid and simple purification of T7 RNA-polymerase. *Nucleic Acids Res.* **19**, 1948–1948 (1991).
- Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80–84 (2014).
- Keijsers, B. J. F., van Wezel, G. P., Canters, G. W., Kieser, T. & Vijgenboom, E. The ram-dependence of *Streptomyces lividans* differentiation is bypassed by copper. *J. Mol. Microbiol. Biotechnol.* **2**, 565–574 (2000).
- Gibson, D. G. *et al.* Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* **319**, 1215–1220 (2008).

## Acknowledgements

We thank Professors Guoqiang Chen, Keqian Yang, Tingyi Wen, Zhen Xie and Nieng Yan for kindly providing the bacteria strains or plasmids. This work was supported in part by funding from the Ministry of Science and Technology of China (973 grant No. 2015CB553402, No. 2011CBA00805 and No. 2013CB734001), the National Natural Science Foundation of China (grant No. 31470532 and No. 31470818), the Tsinghua University-Peking University Center for Life Sciences (CLS), the Tsinghua University XIN Center, the Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, the i-Core programme of the Israel Science Foundation (grant No. 1902/12), the Marie Curie Career Integration Grant and the European Research Council Starter Grant (grant No. 337830).

## Author contributions

T.F.Z., Y.E. and C.L. conceived and supervised the study. W.J., X.Z. and T.G. performed experiments and analysed data. T.F.Z., W.J., Y.E. and C.L. wrote the manuscript.

**Additional information**

**Supplementary information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing financial interests:** T.F.Z., Y.E., C.L., W.J., X.Z. and T.G. have filed a provisional patent application related to this work.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**How to cite this article:** Jiang, W. *et al.* Cas9-Assisted Targeting of CHromosome segments CATCH enables one-step targeted cloning of large gene clusters. *Nat. Commun.* 6:8101 doi: 10.1038/ncomms9101 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>