Using the Endogenous CRISPR-Cas System of *Heliobacterium modesticaldum* To Delete the Photochemical Reaction Center Core Subunit Gene

Patricia L. Baker,^{a,b} Gregory S. Orf,^{a,b} Kimberly Kevershan,^{a*} Michael E. Pyne,^c Taner Bicer,^{a*} Kevin E. Redding^{a,b}

^aSchool of Molecular Sciences, Arizona State University, Tempe, Arizona, USA

^bCenter for Bioenergy and Photosynthesis, Arizona State University, Tempe, Arizona, USA

^cDepartment of Biology, Centre for Applied Synthetic Biology, Concordia University, Montréal, Québec, Canada

ABSTRACT In *Heliobacterium modesticaldum*, as in many *Firmicutes*, deleting genes by homologous recombination using standard techniques has been extremely difficult. The cells tend to integrate the introduced plasmid into the chromosome by a single recombination event rather than perform the double recombination required to replace the targeted locus. Transformation with a vector containing only a homologous recombination template for replacement of the photochemical reaction center gene *pshA* produced colonies with multiple genotypes, rather than a clean gene replacement. To address this issue, we required an additional means of selection to force a clean gene replacement. In this study, we report the genetic structure of the type I-A and I-E CRISPR-Cas systems from *H. modesticaldum*, as well as methods to leverage the type I-A system for genome editing. *In silico* analysis of the CRISPR spacers revealed a potential consensus protospacer adjacent motif (PAM) required for Cas3 recognition, which was then tested using an *in vivo* interference assay. Introduction of a homologous recombination plasmid that carried a miniature CRISPR array targeting sequences in *pshA* (downstream of a naturally occurring PAM sequence) produced nonphototrophic transformants with clean replacements of the *pshA* gene with ~80% efficiency. Mutants were confirmed by PCR, sequencing, optical spectroscopy, and growth characteristics. This methodology should be applicable to any genetic locus in the *H. modesticaldum* genome.

IMPORTANCE The heliobacteria are the only phototrophic members of the largely Gram-positive phylum *Firmicutes*, which contains medically and industrially important members, such as *Clostridium difficile* and *Clostridium acetobutylicum*. Heliobacteria are of interest in the study of photosynthesis because their photosynthetic system is unique and the simplest known. Since their discovery in the early 1980s, work on the heliobacteria has been hindered by the lack of a genetic transformation system. The problem of introducing foreign DNA into these bacteria has been recently rectified by our group; however, issues still remained for efficient genome editing. The significance of this work is that we have characterized the endogenous type I CRISPR-Cas system in the heliobacteria and leveraged it to assist in genome editing. Using the CRISPR-Cas system allowed us to isolate transformants with precise replacement of the *pshA* gene encoding the main subunit of the photochemical reaction center.

KEYWORDS CRISPR-Cas, heliobacteria, gene deletion, homologous recombination, photochemical reaction center

Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) proteins comprise a major system in bacteria and archaea for adaptive immunity against mobile invasive genetic elements. CRISPR-Cas systems have

Citation Baker PL, Orf GS, Kevershan K, Pyne ME, Bicer T, Redding KE. 2019. Using the endogenous CRISPR-Cas system of *Heliobacterium modesticaldum* to delete the photochemical reaction center core subunit gene. *Appl Environ Microbiol* 85:e01644-19. <https://doi.org/10.1128/AEM.01644-19>.

Editor M. Julia Pettinari, University of Buenos Aires

Copyright © 2019 American Society for Microbiology. All Rights Reserved.

Address correspondence to Kevin E. Redding, Kevin.Redding@asu.edu.

* Present address: Kimberly Kevershan, Marana High School, Tucson, Arizona, USA; Taner Bicer, Heliae Development, LLC, Gilbert, Arizona, USA.

Received 19 July 2019

Accepted 14 September 2019

Accepted manuscript posted online 20 September 2019

Published 14 November 2019

become a major new biotechnological tool (1). These systems are programmable and, when used in conjunction with a recombination system and a DNA template for recombination, can be utilized to produce precise genetic changes at the single base level, as well as large deletions and insertions. Variants of this system are thought to be present in over 40% of known bacteria and over 80% of known archaea (2); this high coverage is driven by extensive lateral gene transfer that has produced diverse CRISPR-Cas classes, types, and subtypes (3, 4).

Three processes are involved in producing CRISPR-Cas-based immunity: adaptation, expression/processing, and interference. During the adaptation phase, the host organism's Cas1 and Cas2 proteins acquire short sequences of DNA, called protospacers, from the foreign invaders, which are then incorporated into an area of the genome called the CRISPR array (5, 6). Once these short sequences are incorporated into the genome, they are termed spacers. Spacers are separated by nearly identical <50-bp palindromic sequences called direct repeats (7). During the expression/processing step, RNA is first transcribed from the CRISPR array. The transcribed array is processed by either Cas6 (8) or RNase III (9), which recognizes the direct repeats between the spacers and cuts each one into a separate RNA (henceforth called a CRISPR RNA or crRNA). Finally, during the interference step, a crRNA complexed with the multisubunit effector complex (formed from various Cas proteins) base pairs with the corresponding protospacer on the foreign DNA element. Successful base pairing recruits a Cas nuclease to the foreign DNA (10). Depending on the specific type of CRISPR-Cas system, the Cas nuclease can produce a double-strand DNA break or a single-strand DNA nick. Upon a double-strand break or single-strand nick, nucleases progressively degrade the foreign DNA (11, 12).

In type I and II CRISPR-Cas systems, the Cas nuclease requires recognition of a specific sequence located adjacent to the invading protospacer, which prevents self-attack of the CRISPR-Cas system on the spacers present in the CRISPR array on the genome. This sequence, called the protospacer-adjacent motif (PAM), is generally organism specific and must be discovered empirically.

The double-strand breaks or single-strand nicks caused by the Cas nucleases can be repaired by homology-directed repair (13). Researchers performing genome editing have taken advantage of this: if the CRISPR-Cas system is directed to target a host organism's gene by introduction of a synthetic crRNA, a break or nick will be executed on that gene, which can then be repaired by recombination using a homology-directed repair template that is simultaneously supplied. Changes ranging from site-directed mutagenesis, to single-gene replacements, to large-scale operonic reorganization, have been performed with high efficiency in other clostridia using this method (14–16).

Popular methods for CRISPR-Cas-mediated genome editing usually involve a heterologously expressed CRISPR-Cas9 system from *Streptococcus pyogenes*. However, Cas9 protein has been shown to be toxic to *Clostridium pasteurianum* and in other bacteria, even in the absence of a crRNA (17). The gene encoding this protein is also large (4,107 bp) and can increase plasmid size significantly. An alternative strategy for genome editing in prokaryotes is to use their endogenous CRISPR-Cas system. For example, Pyne et al. (17) showed that the genome of *C. pasteurianum* could be edited with high efficiency using its native type I-B CRISPR-Cas system. These authors speculated that their methodology could be applied to other members of the *Clostridium* genus. Zhang et al. (18) also reported high-efficiency genome editing in *C. tyrobutyricum* using its native type I-B system.

The *Heliobacteriaceae* family belongs to the class *Clostridia* (order *Clostridiales*) and contains the only photosynthetic members of the phylum *Firmicutes* (19–22). These bacteria are of interest in the study of photosynthesis due to, among other reasons, their unique photosynthetic architecture (22). In these bacteria, a single symmetrical reaction center (RC) composed of only two copies each of two subunits, PshA and PshX, is responsible for phototrophy. They are also the only known phototrophic organisms that use bacteriochlorophyll (BChl) *g* as the main pigment and that lack peripheral light-harvesting antenna proteins. Our laboratory recently succeeded in developing a

FIG 1 PCR-based probe of the *pshA* locus to determine genotype after initial homologous recombination attempts. The *pshA* locus of *H. modesticaldum* genome in the WT (A) and putative $\Delta pshA::cbp2p-aph3$ (B) strains. Numbering in each case reflects the nucleotide number in the genome when the origin is set at the same position as it is in *H. modesticaldum* genome in the NCBI database (accession number CP000930.2). The location of PCR primers outside the upstream and downstream flanking regions present in pPB1145 are indicated by arrows. Expected PCR products are shown as brackets, corresponding to the WT (3,163 bp) and the $\Delta pshA::cbp2p-aph3$ mutant (2,772 bp) strains. US, 574 bp cloned from the region upstream of *pshA* in the heliobacterial genome; DS, 527 bp cloned from the region downstream of *pshA*. (C) PCR genotyping of transconjugants. The primers “*pshA* probe for” and “*pshA* probe rev” were utilized in PCR with genomic DNA preparations from the individual tcg. Lane 1, linear DNA marker; lanes 2 to 4, tcg of plasmid pMTL86251 obtained from Em plates; lanes 5 to 13, tcg of plasmid pPB1145 obtained from Em plates (lanes 5 to 7) or Km plates (lanes 8 to 10) and pink in color or from Km plates that were green-brown in color (lanes 11 to 13).

transformation system for *Heliobacterium modesticaldum* (23). Here, we follow that work by performing genome editing to delete the *pshA* gene in *H. modesticaldum* strain Ice1 (referred to here as *H. modesticaldum*).

RESULTS

Gene editing: replacement of the *pshA* gene is hampered by a low rate of complete homologous recombination. In order to replace the *pshA* gene with a kanamycin resistance (Km^r) gene, we began by constructing the conjugative, nonreplicating vector pPB1199, which confers erythromycin resistance (Em^r). We inserted into this vector a template for homologous recombination designed to replace *pshA* with the thermostable Km^r gene *aph3* (derived from plasmid pIKM1) (24) driven by a 621-bp version of the *Clostridium thermocellum* cellobiose phosphorylase promoter (*cbp2p*) (25). The Km^r gene is flanked by sequences (~500 bp) found upstream and downstream of *pshA* (Fig. 1B). The resulting plasmid is called pPB1145 (see Fig. S1 in the supplemental material for plasmid maps).

Plasmids were introduced into *H. modesticaldum* using our recently developed method of premethylating plasmids by heliobacterial DNA methyltransferases and conjugal transfer from *Escherichia coli* (23). The results of the conjugations are reported in Table 1. As expected, Em^r colonies resulted from introducing the replicating plasmid pMTL86251, whereas no Em^r colonies resulted from introducing the nonreplicating plasmid pPB1199. Km^r and Em^r transconjugant (tcg) colonies were obtained by introducing pPB1145, indicating that homologous recombination had occurred, integrating the plasmid into the bacterial chromosome. Surprisingly, significantly more colonies

TABLE 1 Conjugation efficiency of plasmid designed to recombine with the *H. modesticaldum* genome at the *pshA* locus (with controls)^a

Plasmid	Ab ^r	Mean conjugation efficiency (tcg/donor cell) ± the SEM			
		Em ^r		Km ^r	
		Pink	Green-brown	Pink	Green-brown
pMTL86251	Em ^r	0	$2.8 \times 10^{-6} \pm 5.0 \times 10^{-8}$	0	0
pPB1199	Em ^r	0	0	0	0
pPB1145	Em ^r Km ^r	$1.6 \times 10^{-5} \pm 2.6 \times 10^{-6}$	0	$7.2 \times 10^{-5} \pm 2.6 \times 10^{-6}$	$7.1 \times 10^{-7} \pm 1.0 \times 10^{-7}$

^a*H. modesticaldum* cells were plated postconjugation on mPYE-agar supplemented with either Em or Km (antibiotic resistance [Ab^r]). Transconjugants were separated into categories based on the color of the colonies. The conjugation efficiency is expressed as the number of transconjugant heliobacterial colonies (tcg) normalized to the number of plasmid donor *E. coli* cells in the conjugation mixture. Em^r, erythromycin resistance; Km^r, kanamycin resistance.

were obtained than for the replicating vector, which suggests that the integrant chromosome may be more stable than the replicating plasmid. The vast majority of these colonies were pink in appearance; approximately 1% were green-brown (i.e., similar to the color of unmodified heliobacteria).

Three antibiotic-resistant colonies of each type of color were selected from each antibiotic plate (with the exception of Em^r colonies of pPB1145, for which there were no green-brown colonies) and screened by probe PCR and by replica plating on Km or Em plates. Genomic DNA was isolated and PCR was performed using primers designed to amplify the *pshA* gene region of the *H. modesticaldum* genome without annealing to any sequence present in the plasmid (Fig. 1A). Of the colonies that were pink (i.e., the vast majority), both the wild-type (WT) and the *pshA*-replacement genotypes persisted (Fig. 1C), even after ten serial colony isolations. Note that the chromosome with the integrated plasmid would be too large to be amplified by this PCR. We know that such chromosomes exist in the population, since the chromosome/plasmid junctions can be PCR amplified (data not shown). We speculate that these chromosomes are dominant in the population, explaining the weakness of the bands corresponding to the two chromosome types that can be amplified. Moreover, all of the pink colonies remained resistant to both Em and Km, indicating the presence of both resistance genes in the chromosome, consistent with the possession of integrant chromosomes. Of the three colonies that were green-brown in color, two had the WT genotype and one had the knockout genotype, although all were Em sensitive (Fig. 1). At this time, we cannot explain why the greenish clones with a wild-type *pshA* locus are Km resistant; the *aph3* gene may have inserted elsewhere on the chromosome.

These results demonstrate that the homologous recombination required for the plasmid to integrate into the chromosome occurs readily and with high frequency. If the recombination frequency at the second locus (to generate the *pshA* replacement) were similar to that of the first, it should have happened relatively quickly. The fact that it did not do so in the vast majority of cases indicates that the situation is not that simple. One potential complication hinted at by the observation of multiple genotypes in a clonal culture is that heliobacteria may be functionally polyploid, with individual cells harboring variant chromosomes. By using this strategy, the desired replacement was achieved at very low frequency (approximately 1 in 300 colonies) and would not have been detected at all were it not for the color difference between tcg colonies that were integrants and those that either were WT or had a replacement of the *pshA* gene. This was not very surprising, since similar difficulties had been reported in various *Clostridium* species and additional negative selection techniques were required to achieve homoplasmic gene deletions (26). Thus, although we were able to generate a homoplasmic $\Delta pshA::cbp2p-aph3$ strain using this rather standard technique, we knew that use of a visual color selection would not be generally applicable and that a gene replacement frequency of <1% would not be very useful for genome engineering of this species. We therefore turned to another methodology.

Analysis of the heliobacterial CRISPR arrays. Endogenous type I-B CRISPR-Cas systems have been used successfully as a tool for genome editing in *C. pasteurianum* (17) and *C. tyrobutyricum* (18). We decided to take a similar approach in *H. modestical-*

FIG 2 Maps of the genomic CRISPR loci from *H. modesticaldum*. (A) Overall architecture of each locus, including all Cas genes, arrays, and transposases. (B) Architecture of the individual CRISPR arrays. Hypothetical promoters (based on transcriptomic data and consensus sequences) are shown as bent black arrows. Light green, helicase domain of *cas3*; dark green, nuclease domain of *cas3*; light blue, components of the effector complex; red, Cas genes associated with protospacer acquisition; purple, IS4-family transposon; pink, IS1182-type transposon.

dum. The complete genome sequence of *H. modesticaldum* was published and deposited in the NCBI database in 2008 (20). We analyzed the two previously identified (27, 28) type I CRISPR-Cas loci of *H. modesticaldum* (Fig. 2): a putative type I-A system and a putative type I-E system.

The type I-A system does not conform straightforwardly to any of the subtype maps set forth in recent reviews (29–31). A recent review by Jackson et al. (32) suggests that phylogenetic analysis of the *cas3* helicase core domain can be used as a simple proxy for determining the CRISPR-Cas subtype when it is otherwise unclear. We performed a phylogenetic identification using the method of Jackson et al. (32), except that we added the helicase core domains from the two *cas3* genes from *H. modesticaldum* (see Fig. S2 in the supplemental material). The *cas3* gene in the putative type I-E locus clustered well with *cas3* genes from other type I-E systems, as expected. The *cas3* gene from the other locus clustered into the type I-A group, confirming its classification.

Comparing the putative *H. modesticaldum* type I-A system with the generalized map of type I-A systems (29, 30) reveals some differences, both in gene presence and organization. The *H. modesticaldum* type I-A locus contains the following genes (in order): *cas3*, a gene encoding a conserved hypothetical protein; *cas7*, *cas5*, *cas6*, *cas1*, *cas2*, two copies of a *cas4*; and a spacer array containing 19 spacers. The *cas7*, *cas5*, *cas6*, *cas1*, and *cas2* genes and both *cas4* genes seem to be typical of type I-A systems. Conspicuously missing, however, are the genes for the effector small subunit SS (sometimes called *cas8a1*) and large subunit *cas8a1* (sometimes called *casX*). Without the SS or *cas8a1* genes, it is unclear how the effector complex is formed. In many type I-A systems, the *cas3* helicase core domain (*cas3'*) and histidine-aspartate nuclease domain (*cas3''*) are separate proteins (33). However, the *cas3* gene in this locus is an example where both domains are fused into a single gene; this arrangement does not appear to be uncommon (30).

The gene encoding the hypothetical protein just downstream of the type I-A *cas3* is highly conserved among many prokaryotes, but no conserved domains or function are

annotated in the NCBI database. An appraisal of the top nucleotide BLAST (34) hits of this gene reveals that they are found in CRISPR loci of prokaryotes that also lack the large and small effector subunits (*Methanosaeta concilii* CP-6, *Brevibacillus parabrevis*, *Bacillus azotoformans*, *Deinococcus aerius*, and *Heliophilum fasciatum*; note that the last species in that list is another member of the family *Heliobacteriaceae*). Therefore, this gene likely plays a role in this variant type I-A locus. Transcriptomic analysis shows that this gene is expressed at an equal level to that of the immediately preceding *cas3*, indicating that they may be cotranscribed (see Table S1 in the supplemental material). We hypothesize that this gene may indeed encode a functional replacement for the missing canonical large and small effector subunits. We believe that this subject is worthy of further exploration because it may indicate that the *H. modesticaldum* putative type I-A system could actually be better classified as belonging to a hitherto-undescribed type I subtype.

The assignment of the other system as type I-E stems from the phylogenetic placement of the *cas3* gene (Fig. S2) and an effector complex operon that contains *cse1*, *cse2*, *cas7*, *cas5*, and *cas6* genes, followed by additional *cas1* and *cas2* genes responsible for protospacer acquisition. In this locus, the CRISPR array contains 17 spacers and lies upstream of the effector operon. We found that the previously annotated start site of the type I-E *cas3* gene (a GTG start codon at position bp 598258) is incorrect; this start site cuts off a portion of the nuclease domain (*cas3''*). We believe the correct start site to be an ATG start codon 192 bp upstream (at position bp 598450), resulting in a *cas3* gene predicted to contain a full nuclease domain fused to the 5' end of the helicase core domain (*cas3'*). This is consistent with the observed structure in canonical type I-E *cas3* genes. Based on the rest of the operon organization and recently published transcriptomics data from *H. modesticaldum* (35) (Table S1), the type I-E system appears to be transcribed.

Another interesting finding is the presence of transposons within the two CRISPR-Cas loci. In the type I-A locus, one transposon lies between *cas6* and *cas1*, while two others interrupt the CRISPR array between the 17th and 18th spacers and following the 19th spacer. The location of these transposons may interrupt transcription. It is possible that the *cas1*, *cas2*, and both *cas4* genes are expressed using a different promoter than the upstream genes. Similarly, it is likely that spacers 17 to 19 of the array are not expressed or processed consistently with the first 16. In the type I-E locus, two transposons lie between the CRISPR array and the effector operon. We do not expect the presence of these transposons to affect the expression of the locus, since the arrays and effector operons are probably controlled by separate promoters.

In silico determination of a possible consensus PAM sequence. In order to use the *H. modesticaldum* CRISPR-Cas system for genome editing, it was necessary to identify the PAM sequences that direct the effector complex to cut foreign DNA matching spacer sequences. To identify possible PAM sequences, we first searched for potential protospacer matches for the spacers in the *H. modesticaldum* CRISPR arrays. We used BLAST (34) and PHASTER (36) to search for matches in phages, prophage regions of bacterial genomes, plasmids, or transposons for each of the 19 type I-A and 17 type I-E spacer sequences. We omitted results in which there was a mismatch at the first nucleotide of the spacer, allowed no more than two mismatches within the first 7 nucleotides and no more than 35% total mismatch to the spacer sequence. Using these criteria, we did not identify any matches to mobile genetic elements. We then broadened the search to include all prokaryotic sequences, identifying six putative matches to each set of spacers (Table 2). The matches to the type I-A spacers were associated with spacers 1, 15 and 17, with half of the matches aligning to spacer 1. Matches to the type I-E spacers were associated with spacers numbered 6, 7, 9 and 10, with half of the matches aligning to spacer 7. All of the matches were imperfect, ranging from 67 to 85% identity to the spacer sequence, with some having gaps or insertions.

Since PAM sequences are most commonly found within the six nucleotides directly preceding the 5' end of type I CRISPR protospacers, we aggregated the six nucleotides

TABLE 2 Putative protospacer matches identified through *in silico* analysis of heliobacterial CRISPR spacers^a

T	S	Protospacer sequence	M	% identity	Genome (accession no.) ^b	Putative PAM ^c
A	1	GGGCTC-CTCGGCGGCCACGtACACCGaGtAggggg	9 (1 gap)	75	<i>Rubrobacter xylanophilus</i> DSM 9941 (CP000386)	CAGCGG
	1	GGGCcAGCTCGGCGGCCAgGAACAgCGGGGtgggcca	8	78	<i>Kocuria flava</i> strain HO-9041 (CP013254.1)	GGGCGA
	1	GGGCTCGGcGCTCGGCGGCTtCGAAgACCGGcggtt	12 (2 inserts)	67	<i>Planctomyces</i> sp. SH-PL62 (CP011273.1)	CAGCAG
	15	CCcACaGcAGCAACaGGcGGCTTACCGGGCaGcgctg	9	76	<i>Deinococcus proteolyticus</i> MRP (CP002536.1)	CCGCGA
	17	GCTCCCGG---GcGCCACCCCTTCTTctccagCtc	11 (3 gaps)	69	<i>Corynebacterium atypicum</i> strain R2070 (CP008944.1)	GTACGC
	17	GCTCCaGGTcGaGGCCCCACCCCTTcTCCcaaCgA	9	75	<i>Actinomyces meyeri</i> strain W712 (CP012072.1)	TCACGT
E	6	TAGACTtAAACCCCGCCTCgGCCaGGTccaggGC	8	77	<i>Desulfotomaculum kuznetsovii</i> DSM 6115 (CP002770.1)	ACCACG
	7	ACAaCCgGcACCGCaGATTACCGGCCcag	7	78	<i>Dickeya fangzhongdai</i> strain ND14b (CP009460.1)	TCGACC
	7	ACATCCAGCACcGcCaccTTAgCGGCCCGGC	6	81	<i>Pseudomonas chlororaphis</i> strain PCL1606 (CP011110.1)	CCGCGC
	7	ACATCCAGCAaGaCCGgTTCATCCGGCCCGgc	5 (1 insert)	84	<i>Octadecabacter arcticus</i> 238 (CP003742.1)	CCACCG
	9	GAAGcGCTCGCCAACCTTGTcCaAACGcAtC	5	85	<i>Ralstonia eutropha</i> JMP134 (CP000090.1)	CTGCGC
	10	CAAGTAGGATTGACCGAAAGTgGcGATGcGc	7	78	<i>Hyphomonas neptunium</i> ATCC 15444 (CP000158.1)	TTACTC

^aTable column headings: T, CRISPR-Cas type (A, type I-A; E, type I-E); S, spacer number. The protospacer sequence column indicates the spacer-protospacer mismatches in lowercase, while inserted bases not present in the spacer are highlighted by boldfacing. Gaps are indicated by dashed lines. M, number of mismatched nucleotides. The % identity column indicates the percent identity of nucleotides between spacer and protospacer sequences.

^bThe name of the organism in which the protospacer sequence was found is indicated, along with the NCBI GenBank accession number in parentheses.

^cThe putative PAM column indicates the six nucleotides found immediately 5' of the protospacer sequence. Potential conserved residues are indicated in boldface.

5' to the putative protospacers in their host organism as putative PAM sequences (Table 2; see also Fig. S3 in the supplemental material). Of the six putative PAM sequences identified for the type I-A locus, four had a G at the third position, all six had a C at the fourth position, and five had a G at the fifth position. Thus, the exercise yielded a consensus sequence of 5'-NNRCRN-3' (where N is any nucleotide and R is a purine) for the type I-A PAM. Of the potential PAM sequences identified for the type I-E locus, four of six had a C at positions 2, 4, and 6 (Table 2 and Fig. S3), yielding a consensus sequence of 5'-NCNCNC-3'.

***In vivo* interference assays to confirm the consensus PAM sequence.** We focused on empirically determining the operational PAM sequence for the type I-A CRISPR-Cas system for two reasons: there was more consensus among the type I-A putative PAM sequences than among the type I-E, and the type I-A system is more highly transcribed (Table S1), with the possible exception of the *cas2* gene, whose product functions in spacer acquisition (35, 37). To achieve this, we performed an *in vivo* interference assay (17). Spacer 1 from the type I-A array was cloned into the shuttle vector pMTL86251 with different putative PAM sequences placed immediately 5' of the spacer. By testing specific combinations of the pattern 5'-NNRCRN-3', we hoped to find some that would target the shuttle vector for destruction by the native type I-A system, which would be manifested as a reduction in the number of Em-resistant transconjugants.

Of the candidate PAM sequences tested, only 5'-CTGCGC-3' yielded zero colonies in the interference assay (Fig. 3); however, many other sequences produced significantly lower conjugation efficiencies than the control. Most of the candidate sequences that provided interference have a C at the fourth position, verifying its importance for effector complex recognition. There seemed to be no discernible specificity at the first, second, or sixth positions. As predicted, there was a strong preference for a purine at position 3. However, at position 5, every base except for A was compatible with interference. Thus, we tentatively identify the consensus PAM sequence as 5'-NNRCB N-3' (where R is G/C, B is C/G/T, and N is A/C/G/T) immediately preceding the spacer sequence. The one exception to this general rule was the sequence 5'-AGATCC-3', which was originally included as one of the negative controls but exhibited interference. This suggests that there may be more than one PAM consensus sequence, or that it follows a more general formula than we can discern.

Gene editing: leveraging the endogenous type I-A CRISPR-Cas system to produce a *pshA* replacement mutant. Using the PAM consensus sequence (NNRCBN), we identified 341 potential PAM sequences in both strands of the *pshA* gene. The short sequence immediately following any one of these PAM sequences could be targeted as if it were a protospacer sequence by making a synthetic CRISPR spacer matching it. This

FIG 3 *In vivo* interference assay to determine the consensus PAM sequence for the type I-A CRISPR array in *H. modesticaldum*. Putative PAM sequences were cloned directly 5' of spacer 1 from the type I-A CRISPR locus of *H. modesticaldum* and placed into the shuttle vector pMTL86251. These were then conjugated into *H. modesticaldum*. The numbers of tcg per donor cell are shown as averages of triplicate assays; error bars represent standard deviations of triplicates.

would be done by making a construct in which two type I-A CRISPR repeats are separated by a 37-bp sequence exactly matching the targeted protospacer sequence, and the construct is preceded by a promoter. The resulting transcript produced in heliobacterial cells should be processed into crRNA and the synthetic spacer would target the effector complex to cut the matching protospacer in *pshA*. This should be followed by degradation of the new 3' end, as Cas3 in other type I-A systems has been shown to nick DNA of the displaced strand inside the protospacer region (38) and to possess 3'-to-5' exonuclease activity (39, 40).

Reasoning that DNA nicks should be near both ends of a longer gene like *pshA* (1827 bp), so as not to be too far from the regions of homology used to replace the gene, we made synthetic mini-CRISPR arrays to target a protospacer near each end of the gene. We made two such arrays, in which the directions of 3'-to-5' exonucleolytic degradation were either divergent or convergent (see Fig. 4 for locations and sequences of spacers). Each array was constructed using three type I-A CRISPR repeats alternating with two spacer sequences derived from *pshA*, each one consisting of 37 bases immediately downstream of a PAM consensus sequence. In every case the PAM sequence used was the palindrome 5'-GCGC-3', allowing the same PAM sequence (on opposite strands) to be used for both synthetic CRISPR arrays. The arrays were placed under the transcriptional control of the constitutive enolase promoter (*enop*) from *C. thermocellum* (25). We also constructed plasmids that lacked or contained the recombination template previously used to replace *pshA* with *cbp2p-aph3* (Fig. 1; see Fig. S4 in the supplemental material for plasmid maps).

These plasmids were introduced into *H. modesticaldum* via conjugation as before. As expected, no Em^r or Km^r colonies resulted from introducing pPB1347 or pPB1349, which are nonreplicating plasmids and contain only the synthetic CRISPR array. When

FIG 4 Genome editing in *H. modesticaldum* using the endogenous CRISPR-Cas system. (A) Linear maps of the region of the *pshA* gene of *H. modesticaldum* show the location of spacers chosen for inclusion in the mini-synthetic CRISPR arrays. Plasmid pPB1145 has no mini-synthetic CRISPR array. Plasmids pPB1347 and pPB1348 and pPB1349 and pPB1350 contain mini-synthetic CRISPR arrays each containing two spacers targeting *pshA*. Blue arrow, *pshA*. The PAM sequence, 5'-GCGC-3', is found immediately 5' of the spacer in each case. Small scissors indicate the direction and strand of Cas3 exonuclease activity after a single-strand DNA nick. Dark gray boxes, protospacer 1 (CCGGGGGTAAATGCCATTTCGGCAGTGGCCATCAAA) and protospacer 2 (TGGCCACCTTCAACTCCCAACCCGCTCCATCAGCT); green boxes, protospacer 3 (AGGTCTTTGAGTACTTCAAAGACAAAGTGCCCGCTAC) and protospacer 4 (AGCCTGGTAGATGTCATTGTAGGAGCGCAACGGCT). (B) Conjugation efficiency of plasmids pPB1348 and pPB1350, with pPB1145 shown for comparison. Pink bars, colonies that initially appeared pink on plates; green bars, colonies that initially appeared greenish brown on plates. The data are based on three independent replicates. (C) Gene editing efficiency, expressed as the number of pure $\Delta pshA::cbp2p-aph3$ organisms produced as a percentage of the total number of kanamycin-resistant colonies obtained.

both the template for homologous recombination and the synthetic CRISPR array were present on the vector (pPB1348 and pPB1350; Fig. S4), no colonies were seen on Em plates. Taken together, these results indicate that the synthetic CRISPR arrays targeting the chromosome are not tolerated long term by the organism.

The tcg colonies of pPB1348 and pPB1350 that were selected on Km plates were initially of two different colors. Some appeared pink, while others were greenish brown. After two more days, the initially pink tcg colonies of pPB1348 had changed in appearance to look slightly green. In contrast, the pink tcg colonies of pPB1350 did not change appearance. Six colonies of each initial color for each plasmid were streaked three consecutive times on plates containing Km, and individual colonies of each were chosen for characterization. All colony-purified transconjugant strains were Km^r and Em^s, indicating that they no longer harbored all of the sequences of the vector that had been initially transferred via conjugation.

Genomic DNA of each transconjugant strain was screened by PCR of the *pshA* locus, as before (Fig. 1A). PCR products from 10 of 12 pPB1348 transconjugants (Fig. 5, lanes

FIG 5 PCR-based probe of the *pshA* locus in transconjugants. The primers "pshA probe for" and "pshA probe rev" were utilized in PCR using genomic DNA preparations from individual transconjugants. Lanes 1 to 6, transconjugants of pPB1348 that were initially pink in color; lanes 7 to 12 transconjugants of pPB1348 that were initially green in color; lanes 13 to 18, transconjugants of pPB1350 that were pink in color; lanes 19 to 24, transconjugants of pPB1350 that were green in color; lane 25, linear DNA markers. The expected size of the PCR product from *pshA* is 3,163 bp; the expected size of the PCR product from $\Delta pshA::cbp2p-aph3$ is 2,772 bp. See Fig. 1 for locus maps and primer locations.

FIG 6 Phenotypes of the $\Delta pshA::cbp2p-aph3$ mutant as measured by optical spectroscopy. (A) Absorbance spectra of mPYE cultures of WT (black) and $\Delta pshA::cbp2p-aph3$ (red) strains grown in light or the WT strain grown in dark (dashed). (B) Absorbance spectra of mPYE(2xP) dark-grown cultures. (C) Laser flash-induced photobleaching of P₈₀₀ in WT and $\Delta pshA::cbp2p-aph3$ cultures. The inset is plotted on a logarithmic time scale to visualize early time points.

1 to 12), but only 2 of 12 pPB1350 (Fig. 5, lanes 13 to 24) transconjugants, corresponded to the expected size for complete replacement of the *pshA* gene with the *cbp2p-aph3* construct. Sequencing of the PCR products confirmed the *pshA::cbp2p-aph3* genotype: the sequence was exactly as designed in all cases where the replacement had occurred (see Fig. S5 in the supplemental material). All of the pPB1350 transconjugants that were pink gave both PCR products corresponding to the expected sizes of WT and $\Delta pshA::cbp2p-aph3$. Of the six that looked somewhat green, two had only the product corresponding to the mutant genotype, while the other four retained the WT band at some level (Fig. 5). Thus, although more transconjugant colonies were obtained using plasmid pPB1350, the efficiency of genome editing to obtain the desired gene replacement was much higher with pPB1348, which used synthetic spacers that directed exonuclease activity toward the interior of *pshA* (Fig. 4). Of the two pPB1348 transconjugants that did not have a $\Delta pshA::cbp2p-aph3$ genotype, both appeared to be phenotypically wild type; while transconjugant 9 may possess an unaltered *pshA* locus, transconjugant 6 clearly has undergone some kind of insertion event.

Biochemical and biophysical characterization of the $\Delta pshA::cbp2p-aph3$ mutant. Helio bacteria are photoheterotrophs, requiring a reduced carbon source. They are capable of growing in the dark by fermenting simple organic acids, such as pyruvate, but attain much higher cell densities when grown in the light (Fig. 6A). The $\Delta pshA::cbp2p-aph3$ culture grown in the light reached roughly the same density as the WT culture grown in the dark, as evidenced by the similar optical density due to light scattering in regions of the visible and near-infrared (NIR) spectrum outside the main BChl *g* absorbance peak at 787 nm (Fig. 6A).

Pink cells obtained from conjugation with plasmids pPB1145 or pPB1350 lack the major absorbance peak at 787 nm (Fig. 6B). The lack of pigment absorption peaks and the resulting pink color of these cultures is likely due to a polar effect from integration of the plasmid adjacent to *pshA*, since this gene is embedded in a cluster of genes encoding enzymes for BChl *g* synthesis. The pink color may be due to the presence of cytochromes, whose synthesis would be unaffected. Cultures of the $\Delta pshA::cbp2p-aph3$ mutant are greenish and have a shift in the main peak from 787 to 765 nm. When heliobacterial cultures are grown in darkness, it is difficult to distinguish by eye between cells of the wild-type and those of the $\Delta pshA::cbp2p-aph3$ mutant, but the shift in the main NIR absorbance peak is readily visible (Fig. 6B).

To test for the presence of the heliobacterial RC (HbRC), a time-resolved spectroscopic experiment was performed using a Joliot-type spectrometer, in which a saturating laser flash induces photobleaching of P₈₀₀, the HbRC special BChl *g'* dimer. This bleaching and the subsequent recovery are monitored by 10- μ s LED flashes. It is a powerful and accurate measurement of the overall concentration and activity of the HbRC, even in whole cells. The P₈₀₀ photobleaching recovery time in WT cultures is

FIG 7 HPLC analysis of photosynthetic pigments in *H. modesticaldum* WT and $\Delta pshA::cbp2p-aph3$ strains. (A) HPLC chromatograms detected by absorbance at 400 nm. The molecules responsible for each peak were identified from their absorbance spectra (200 to 900 nm). (B) Absorbance spectra of peaks 3 and 4 (BChl *g* and its stereoisomer BChl *g'*) and peaks 6 and 7 (BPheo *g* and its stereoisomer BPheo *g'*). (B)Chlides, (bacterio)chlorophyllides; (B)Phs, (bacterio)pheophorbides; DANS, 4,4'-diaponeurosporene.

~1 ms, corresponding to reduction of P_{800}^{+} by cytochrome c_{553} at room temperature in this species (41). In the $\Delta pshA::cbp2p-aph3$ mutant, no photobleaching signal is observed whatsoever (Fig. 6C), indicating the absence of functional HbRC in the cells.

According to the X-ray crystal structure of the HbRC (22) (PDB ID [5V8K](#)), it noncovalently binds 54 BChl *g*, four BChl *g'* (the C13² epimer of BChl *g*), two 8¹-hydroxy Chl *a*, and two carotenes (4,4'-diaponeurosporene). There are no other pigment-binding proteins known in heliobacteria. Our initial hypothesis was that the greenish color of the $\Delta pshA::cbp2p-aph3$ mutant cells was due to the presence of free BChl *g* and/or Chl *a*-type pigments. We tested this by extracting the pigments from cultures with organic solvent and separating them via reversed-phase high-pressure liquid chromatography (HPLC) (Fig. 7). To our surprise, the main pigment that accumulated in $\Delta pshA::cbp2p-aph3$ cells did not comigrate with BChl *g*. Moreover, its absorbance spectrum was not the same as BChl *g*. The longer retention time and blue-shifted Q_y absorbance peak of the species were more consistent with that of bacteriopheophytin (BPheo) *g* (42), which is the product of the replacement of the central Mg^{2+} cation from BChl *g* with two protons. To confirm the assignment, a pigment extract from a WT culture was subjected to mild acid treatment, which demetallates (bacterio)chlorophylls, to produce molecular standards for BPheo *g* and the other demetallated pigments (Fig. 7A, blue trace). The retention time (Fig. 7A) and absorbance spectrum (Fig. 7B) of chemically prepared BPheo *g* is identical to the main pigment found in the $\Delta pshA::cbp2p-aph3$ strain.

DISCUSSION

While it is unknown why so many members of the *Clostridiales* perform plasmid integration via a single recombination event rather than gene replacement via a double recombination event, it is a factor hampering the genetic manipulation of many members of the order. One possibility is that the plasmid is able to reintegrate readily after any double crossovers that do occur. Since the plasmid we used is nonreplicating, this possibility seems unlikely, as the vector would be expected to be lost quickly, precluding the generation of stable clones. Another factor is that *H. modesticaldum* may be polyploid, which would further hamper selection of homoplasmic gene knockouts,

because multiple genetic arrangements (integrated plasmid, wild type, and gene knockout) could coexist in different chromosomal copies within the same cell.

The need for a way of forcing the gene replacement has led to our use of the native CRISPR-Cas system. Our results show that the gene editing efficiency is very low (<1%) after introduction of a construct for double recombination alone. Use of the endogenous CRISPR-Cas system as a further selection will be an essential tool for making mutations other than those that can be selected or screened for (e.g., antibiotic resistance or color change).

There are at least two possibilities for the manner in which the CRISPR-Cas system improves gene editing efficiency: (i) the nicking and exonuclease activity initiated by the CRISPR-Cas system in the target gene, and possibly into the homology arms, may accelerate the second recombination event, or (ii) the CRISPR-Cas system may act primarily by selecting against genomes in a polyploid organism that still contain the target protospacer sequence, eventually leading to degradation of those chromosomes. Our finding that the highest genome editing efficiency was observed when using two synthetic spacers that direct exonuclease activity in converging directions, perhaps leading to the creation of a double-strand break, could support either hypothesis. In fact, these hypotheses are not mutually exclusive. Further work will be required to address this question.

The number of transposons present in the CRISPR-Cas loci of this organism would seem remarkable. However, both of the transposons that interrupt these arrays are present numerous times throughout the *H. modesticaldum* chromosome. We assign the transposons interrupting the type I-A CRISPR array as members of the IS1182-type family. They are virtually identical to each other (1,909 of 1,914 bp) and one of them is entirely identical (1,914 of 1,914 bp) to one of the two located in the type I-E locus. This IS1182-type transposon is found in the chromosome in 31 locations, one of which is contained in a phage region identified by PHASTER. The other transposon located in the type I-E locus is identical to the one located between *cas6* and *cas1* in the type I-A locus. This transposon is found 13 times on the genome. We cannot assign its type beyond the broad class of IS4-family transposons. The large number of transposons annotated in the *H. modesticaldum* genome has been noted in the past (20), perhaps rendering their presence in the CRISPR-Cas loci somewhat less surprising.

Another possibility is that the transposon that we believe would interfere with the transcription and/or processing of spacers 17 to 19 of the type I-A CRISPR array in fact prevents targeting of the array against the *H. modesticaldum* genome. The first 14 bp of spacer 18 exactly match 14 bases in the *aroK* gene, which encodes Shikimate kinase, the product of which is involved in chorismate biosynthesis. Those 14 nucleotides in the *aroK* gene are immediately preceded by the sequence GCGC, which we have shown can act as a PAM sequence. None of the other spacers in the type I-A array have such matches in the genome preceded by a sequence fitting this PAM consensus sequence. In some of the CRISPR-Cas systems that have been studied (43, 44), a “seed sequence” of 6 to 12 bp was found to be sufficient to cause interference. If 14 bp are sufficient to act as a seed sequence for CRISPR-Cas cleavage in *H. modesticaldum*, then spacer 18 would direct self-targeting. Thus, it is possible that prevention of transcription of spacer 18 in the type I-A CRISPR array was of selective advantage to the organism.

In the search for spacer-protospacer matches, we had to lower the stringency of the matching criteria in order to achieve hits. This poor protospacer identification may be due to either (i) the lack of sequencing data from the natural habitat of heliobacteria or (ii) genetic drift of the spacers after their incorporation into the heliobacterial genome. Shmakov et al. have shown that using their higher stringency (95% sequence identity and 95% sequence coverage) analysis of all identified CRISPR spacers in bacterial and archaeal genomes, only 1 to 19% of spacers had identifiable protospacers (45). In addition, these investigators found that the majority of identified protospacers (80 to 90%) were from viral genomes and that the rest were mostly from genes incorporated in microbial genomes but involved in plasmid replication or conjugation. We were therefore surprised that all of our spacer-protospacer matches came from

seemingly nonmobile genetic elements of prokaryotic genomes. As more sequences from phages and other invasive genetic elements that may infect *helicobacteria* become available in the future, it is likely that better PAM consensus sequences can be devised from a purely bioinformatic approach. Nevertheless, testing of the predicted PAM by a functional assay, such as the interference assay used here, will still be necessary.

As for the $\Delta pshA::cbp2p-aph3$ mutant, it displays a nonphototrophic phenotype and relies solely on pyruvate fermentation, which explains why its growth is similar to that of the wild-type when it is grown in the dark. The accumulation of BPheo *g* in this strain has a few implications. First, the presence of any pigment absorbing light in the red/NIR range shows that the replacement of *pshA* has not disrupted any of the nearby BChl synthesis genes in the photosynthetic gene cluster. Second, since BPheo *g* is not an intermediate in the synthesis of BChl *g* (20, 46), it must therefore be an intermediate in the BChl *g* degradation process. This implies that BChl *g* is fully synthesized in the $\Delta pshA::cbp2p-aph3$ strain but is quickly degraded thereafter, probably due to its lack of incorporation into a HbRC complex.

The absorbance spectrum of the BPheo *g* in the cells is red shifted relative to its spectrum in organic solvent. Thus, it is possible that the BPheo *g* is bound by a protein or proteins in the cell. Although no antenna complexes binding any pigment have been identified in this organism, and no BPheo *g* is found associated with the HbRC, it is unlikely that BPheo *g* exists in a free form in the cellular membrane. Our tentative hypothesis is that proteins involved in BChl degradation and trafficking are binding the BPheo *g*. However, these pathways are largely unexplored in the *helicobacteria*. Future study of this organism will be required to unravel the regulation of BChl *g* biosynthesis, its pathway for degradation, and what proteins bind the degradation intermediates, including BPheo *g*.

MATERIALS AND METHODS

Growth conditions and media. *Helicobacter modesticaldum* strain lce1 was routinely grown in mPYE medium that had been bubbled with nitrogen to remove oxygen and autoclaved in Hungate bottles (23). For chemotrophic growth, the amount of sodium pyruvate was doubled to 4.4 g liter⁻¹ (2xP), and the pH was adjusted to 6.5 (instead of 7) prior to bubbling with nitrogen and autoclaving. The solid medium was mPYE or mPYE(2xP) with 1.5% (wt/vol) agar. *E. coli* was grown on Luria-Bertani (LB) 1.5% agar plates, which were supplemented with antibiotics to select for plasmids as needed (300 μ g ml⁻¹ Em, 50 μ g ml⁻¹ Km, 30 μ g ml⁻¹ chloramphenicol). When grown in liquid LB medium, antibiotic concentrations were lowered (100 μ g ml⁻¹ Em, 25 μ g ml⁻¹ Km, 15 μ g ml⁻¹ chloramphenicol).

Plasmid construction for interference assays. For interference assays, we used the vector pMTL86251, which we have previously shown replicates in *H. modesticaldum* (23). Pairs of oligonucleotides (see Table 3) were annealed to each other to create the PAM/spacer combination with 4-nucleotide 5' overhangs for ligation into compatible restriction sites of pMTL86251 without recreating the restriction sites. Ligations using T4 DNA ligase (New England Biolabs [NEB], Ipswich, MA) were performed with the restriction enzymes present in the ligation mix, using ten alternating 10-min cycles of 37 and 16°C, similar to a Golden Gate assembly. The ligations were then transformed into *E. coli* TOP10 cells with selection for Em resistance. Isolated plasmids were confirmed by Sanger sequencing.

Plasmid construction for gene replacement using the endogenous CRISPR-Cas system. Oligonucleotides and plasmids used are shown in Tables 3 and 4, respectively. Starting with plasmid pMTL86251, a plasmid with fewer restriction sites in the multiple-cloning site (MCS) and lacking the pIP404 replication module was constructed; this plasmid (pPB1199) possesses the *traJ* and *oriT* loci necessary for conjugal transfer but cannot replicate in *H. modesticaldum*. To make this plasmid, annealed oligonucleotides ("Custom sites for NotI" and "Custom sites rev Ascl" [see Table 3]) were cloned between the NotI/Ascl sites to remove part of the MCS and the annealed oligonucleotides "Ascl to FseI spacer for" and "Ascl to FseI spacer rev" were cloned between the Ascl/FseI restriction sites to remove the pIP404 replication module.

Using the primer pairs "pshA up for HindIII"/"pshA up rev XbaI" and "pshA dn for PacI"/"pshA dn rev HindIII," regions of DNA from upstream and downstream of *pshA* in the *H. modesticaldum* genome were amplified from genomic DNA. Using the "KI for XbaI" and "KI rev PacI" primers, the Km-resistance gene *aph3* was amplified from plasmid pIKM1, while the *cbp2* promoter was amplified from plasmid pDGO117 (25) using the primers "Pcbp_2 for NheI" and "Pcbp_2 rev NheI." All PCRs for cloning purposes were performed using the high-fidelity DNA-polymerase Q5 (NEB). These DNA elements were assembled and cloned into pPB1199 using standard restriction cloning methods to make plasmid pPB1145 (see Fig. S1 in the supplemental material).

In order to assemble the plasmids in which a synthetic mini CRISPR array would contain two spacers specifically targeting *pshA*, an array was designed to include two BsaI type IIS restriction enzyme sites, which result in dissimilar cohesive ends that overlap CRISPR repeats in the construct. The construct was synthesized (GenScript) with the following sequence (Sall, BsaI, and BglII recognition sites are underlined,

TABLE 3 Oligonucleotides used in this study

Oligonucleotide	Sequence (5'–3') ^a	Purpose ^b
pshA up for HindIII	TAAC TAAGCTT TCGGTCTCACGGATTATGAC	Amplification of <i>pshA</i> upstream region
pshA up rev XbaI	TAATAT CTAGAC AAACACGACCTCTTTTCAGA	
pshA dn for PacI	TAAC TTTAATTAAT CACACCAAGTCCATTTCGTTT	Amplification of <i>pshA</i> downstream region
pshA dn rev HindIII	TATTAC CAAGCTT CATCCATCGAAATGAGGATGTC	
Pcbp_2 for NheI	TAAT GCTAGC GAGTCGTGACTAAGAAGCTCAA	Amplification of <i>cbp_2</i> promoter
Pcbp_2 rev NheI	AATT GCTAGC TTTCCTGCTCCTTAAATTTTCG	
KI for XbaI	TAACCG CTAGA ATGGCTAAAATGAGAATATCACCG	Amplification of Km ^r gene <i>aph3</i>
KI rev PacI	AAATA TTAATTAAT CAAAAACAATTCATCCAGTAAAAATAAT	
Ascl to FseI spacer for	CGCGCCG ACTGTAATGCCTATGCGT CCGG	Removal of Gram-positive origin of replication
Ascl to FseI spacer rev	ACGCATAGGCATTACAGTCGG	
Custom sites for NotI	GGCC AGGTACCTCGAGAAGCTTACGCGTGG	Modifying MCS
Custom sites rev Ascl	CGCGCC ACGCGTAAGTTCGTGGCGCCGAGCGAGCCCT	
pshA probe for	AAACGGATGGTTCGGAAATACG	PCR to verify genomic constructs
pshA probe rev	TTTGAGGAAGCGAAACCATGC	
Spacer 1B 2PAM HindIII.S	AGCT ACAGCGGGGGCTCGCTCGGCGGCCACGAACACCGGGGATTACTT	Plasmid with putative PAM: CAGCGG
Spacer 1B 2PAM EcoRI.AS	AATTAAGTAAT CCCCGGTGTTCGTGGCCGCCGAGCGAGCCCCGCTGT	
Spacer 1B HindIII.S	AGCT AGGGCTCGCTCGGCGGCCACGAACACCGGGGATTACTT	Plasmid with putative PAM: AAGCTA
Spacer 1B EcoRI.AS	AATTAAGTAAT CCCCGGTGTTCGTGGCCGCCGAGCGAGCCCT	
Spacer 1B 1PAM HindIII.S	AGCT ACCGTGGGGGCTCGCTCGGCGGCCACGAACACCGGGGATTACTT	Plasmid with putative PAM: CCGTGG
Spacer 1B 1PAM EcoRI.AS	AATTAAGTAAT CCCCGGTGTTCGTGGCCGCCGAGCGAGCCCCACGTT	
Spacer 1B 3PAM HindIII.S	AGCT ACTGCGGGGGCTCGCTCGGCGGCCACGAACACCGGGGATTACTT	Plasmid with putative PAM: CTGCGG
spacer 1B 3PAM EcoRI.AS	AATTAAGTAAT CCCCGGTGTTCGTGGCCGCCGAGCGAGCCCCGCGAGT	
Spacer 1B 4PAM HindIII.S	AGCT ACTGAATGGGCTCGCTCGGCGGCCACGAACACCGGGGATTACTT	Plasmid with putative PAM: CTGAAT
Spacer 1B 4PAM EcoRI.AS	AATTAAGTAAT CCCCGGTGTTCGTGGCCGCCGAGCGAGCCCCATTAGT	
Spacer 1B 5PAM HindIII.S	AGCT ACTGAGCGGGCTCGCTCGGCGGCCACGAACACCGGGGATTACTT	Plasmid with putative PAM: CTGAGC
Spacer 1B 5PAM EcoRI.AS	AATTAAGTAAT CCCCGGTGTTCGTGGCCGCCGAGCGAGCCCCGCTCAGT	
Spacer 1B 6PAM HindIII.S	AGCT ACTGCGGGGCTCGCTCGGCGGCCACGAACACCGGGGATTACTT	Plasmid with putative PAM: CTGCGC
Spacer 1B 6PAM EcoRI.AS	AATTAAGTAAT CCCCGGTGTTCGTGGCCGCCGAGCGAGCCCCGCGAGT	
No PAM 1A1 NcoI for	CATG GGGCTCGCTCGGCGGCCACGAACACCGGGGATTACT	Plasmid with putative PAM: TCCATG
No PAM 1A1 BamHI rev	GATC AGTAATCCCCGGTGTTCGTGGCCGCCGAGCGAGCCCC	
CAGCAG NcoI for 1A1	CATG CAGCAGGGGCTCGCTCGGCGGCCACGAACACCGGGGATTACT	Plasmid with putative PAM: CAGCAG
CAGCAG BamHI rev 1A1	GATC AGTAATCCCCGGTGTTCGTGGCCGCCGAGCGAGCCCCGCTG	
CAACGG NcoI for 1A1	CATG CAACGGGGCTCGCTCGGCGGCCACGAACACCGGGGATTACT	Plasmid with putative PAM: CAACGG
CAACGG BamHI rev 1A1	GATC AGTAATCCCCGGTGTTCGTGGCCGCCGAGCGAGCCCCGCTG	
GAGCGG NcoI 1A1 for	CATG TGAGCGGGGCTCGCTCGGCGGCCACGAACACCGGGGATTACT	Plasmid with putative PAM: GAGCGG
GAGCGG BamHI 1A1 rev	GATC AGTAATCCCCGGTGTTCGTGGCCGCCGAGCGAGCCCCGCTCA	
GAGCAT NcoI 1A1 for	CATG AGCATGGGCTCGCTCGGCGGCCACGAACACCGGGGATTACT	Plasmid with putative PAM: GAGCAT
GAGCAT BamHI 1A1 rev	GATC AGTAATCCCCGGTGTTCGTGGCCGCCGAGCGAGCCCCATGCT	
GAACGT NcoI 1A1 for	CATG AACGTGGGCTCGCTCGGCGGCCACGAACACCGGGGATTACT	Plasmid with putative PAM: GAACGT
GAACGT BamHI 1A1 rev	GATC AGTAATCCCCGGTGTTCGTGGCCGCCGAGCGAGCCCCACGT	
GAACAT NcoI 1A1 for	CATG GAACATGGGCTCGCTCGGCGGCCACGAACACCGGGGATTACT	Plasmid with putative PAM: GAACAT
GAACAT BamHI 1A1 rev	GATC AGTAATCCCCGGTGTTCGTGGCCGCCGAGCGAGCCCCATGTT	
Spacer 1B for BglII	GATC CGGGCTCGCTCGGCGGCCACGAACACCGGGGATTACT	Plasmid with putative PAM: AGATCC
Spacer 1B rev EcoRI	AATT AGTAATCCCCGGTGTTCGTGGCCGCCGAGCGAGCCCC	
Spacer 1B PAM CCGCGG for HindIII	AGCT CCGCGGGGCTCGCTCGGCGGCCACGAACACCGGGGATTACT	Plasmid with putative PAM: CCGCGG
Spacer 1B PAM CCGCGG rev EcoRI	AATT AGTAATCCCCGGTGTTCGTGGCCGCCGAGCGAGCCCCGCGG	
Spacer 1B PAM CCGCGG for HindIII	AGCT CGGCGGGGCTCGCTCGGCGGCCACGAACACCGGGGATTACT	Plasmid with putative PAM: CCGCGG
Spacer 1B PAM CCGCGG rev EcoRI	AATT AGTAATCCCCGGTGTTCGTGGCCGCCGAGCGAGCCCCGCGG	
Spacer 1B PAM CCACGG for HindIII	AGCT CCACGGGGCTCGCTCGGCGGCCACGAACACCGGGGATTACT	Plasmid with putative PAM: CCACGG
Spacer 1B PAM CCACGG rev EcoRI	AATT AGTAATCCCCGGTGTTCGTGGCCGCCGAGCGAGCCCCGCTGG	
Target pshA spacer 1 for	aaag CGGGGGTTAAATGCCATTTCGGCAGTGGCCATCAAgt	Annealing oligonucleotides to target <i>pshA</i> spacer 1
Target pshA spacer 1 rev	tgaaac TTTGATGGCCACTGCCGAATGGCATTAAACCCCCG	
Target pshA spacer 2 for	ttgaaag TGGCCACCTTCAACTCCACAACCCGCTCCATCAGCT	Annealing oligonucleotides to target <i>pshA</i> spacer 2
Target pshA spacer 2 rev	aaac AGCTGATGAGCGGGTTGTGGGAGTTGAAGGTGGCCAtt	
Target pshA spacer 3 for (43-mer):	aaag AGGTCTTTGAGTACTTCAAAGACAAAGTGCCCGCTACgt	Annealing oligonucleotides to target <i>pshA</i> spacer 3
Target pshA spacer 3 rev (43-mer):	tgaaac GTAGCGGGCACTTTGTCTTTGAAGTACTCAAAGACCT	
Target pshA spacer 4 for (44-mer):	ttgaaag AGCCTGGTAGATGTCATTGTAGGGAGCGGCAACGGCT	Annealing oligonucleotides to target <i>pshA</i> spacer 4
Target pshA spacer 4 rev (44-mer):	aaac AGCCGTTGCCGCTCCTACAATGACATCTACAGGCTctt	
CRISPR repeat for	ttcact ggatgacccgctatgtagggga	5' phosphorylated oligonucleotides; annealing to form CRISPR repeat with 4-nt 5' overhangs
CRISPR repeat rev	tcaat ccctacatagcggtcatcag	

^aIn the nucleotide sequences, the restriction sites and 4-nt 5' overhangs are in boldface, and the spacer sequences are underlined. Nucleotides added to form a CRISPR repeat are in lowercase.

^bThat is, the reason the oligonucleotide was used in this study.

the CRISPR repeats are indicated in lowercase, and the locations of the 5' overhangs created by BsaI digestion are indicated in boldface, the first of which is created on the opposite strand): 5'-TAACGC GTCGACtttcaatccctacatagcggtcatcagtgaaacAGAGACCGGCGGTATAAGGTCTCA**acttt**caatccctacata gcggtgcatcagtgaaacTCTCATTAGATCTTAATCG-3'. This construct was digested with Sall and BglII

TABLE 4 Plasmids used in this study

Plasmid	Description	Source and/or reference
pKM1	Template for amplification of thermostable Km ^r gene <i>aph3</i>	J. Wiegel (24)
pDG0117	Template for amplification of <i>cbp2</i> promoter from <i>C. thermocellum</i>	D. Olson (25)
pMTL86251	Shuttle vector for clostridia with pIP404 replication module	23; Chain Biotech
pPB1073	pMTL86251 plus putative PAM CAGCGG 5' of CRISPR type I-A spacer 1	This study
pPB1089	pMTL86251 plus putative PAM AAGCTA 5' of CRISPR type I-A spacer 1	This study
pPB1091	pMTL86251 plus putative PAM CCGTGG 5' of CRISPR type I-A spacer 1	This study
pPB1093	pMTL86251 plus putative PAM CTGCGG 5' of CRISPR type I-A spacer 1	This study
pPB1095	pMTL86251 plus putative PAM CTGAAT 5' of CRISPR type I-A spacer 1	This study
pPB1097	pMTL86251 plus putative PAM CTGAGC 5' of CRISPR type I-A spacer 1	This study
pPB1099	pMTL86251 plus putative PAM CTGCGC 5' of CRISPR type I-A spacer 1	This study
pPB1332	pMTL86251 plus putative PAM TCCATG 5' of CRISPR type I-A spacer 1	This study
pPB1333	pMTL86251 plus putative PAM CAGCAG 5' of CRISPR type I-A spacer 1	This study
pPB1334	pMTL86251 plus putative PAM CAACGG 5' of CRISPR type I-A spacer 1	This study
pPB1335	pMTL86251 plus putative PAM GAGCGG 5' of CRISPR type I-A spacer 1	This study
pPB1336	pMTL86251 plus putative PAM GAGCAT 5' of CRISPR type I-A spacer 1	This study
pPB1337	pMTL86251 plus putative PAM GAACGT 5' of CRISPR type I-A spacer 1	This study
pPB1338	pMTL86251 plus putative PAM GAACAT 5' of CRISPR type I-A spacer 1	This study
pKRK3	pMTL86251 plus putative PAM AGATCC 5' of CRISPR type I-A spacer 1	This study
pKRK4	pMTL86251 plus putative PAM CCGCGG 5' of CRISPR type I-A spacer 1	This study
pKRK5	pMTL86251 plus putative PAM CGGCGG 5' of CRISPR type I-A spacer 1	This study
pKRK9-2	pMTL86251 plus putative PAM CCACGG 5' of CRISPR type I-A spacer 1	This study
pPB1199	Vector with modified MCS and no Gram-positive replication module	This study
pPB1145	pPB1199 with template for homologous recombination to replace <i>pshA</i> gene with the Km ^r cassette	This study
	<i>cbp2p-aph3</i>	
pPB1347	pPB1199 with mini-synthetic CRISPR array targeting <i>pshA</i> gene using artificial protospacers 1 and 2	This study
pPB1348	pPB1199 with mini-synthetic CRISPR array targeting <i>pshA</i> gene using artificial protospacers 3 and 4	This study
pPB1349	pPB1199 with both template for homologous recombination to replace <i>pshA</i> with <i>cbp2p-aph3</i> and with mini-synthetic CRISPR array targeting <i>pshA</i> using artificial protospacers 1 and 2	This study
pPB1350	pPB1199 with both template for homologous recombination to replace <i>pshA</i> with <i>cbp2p-aph3</i> and with mini-synthetic CRISPR array targeting <i>pshA</i> using artificial protospacers 3 and 4	This study

and cloned by restriction cloning into the KpnI and XhoI sites of plasmids pPB1199 and pPB1145 to create plasmids pPB1213 and pPB1344. Pairs of oligonucleotides were annealed to create the spacers with 4-nucleotide 5' overhangs: spacer 1 ("target *pshA* spacer 1 for"/"target *pshA* spacer 1 rev"), spacer 2 ("target *pshA* spacer 2 for"/"target *pshA* spacer 2 rev"), spacer 3 ("target *pshA* spacer 3 for"/"target *pshA* spacer 3 rev"), and spacer 4 ("target *pshA* spacer 4 for"/"target *pshA* spacer 4 rev"). The intervening CRISPR repeat was made the same way using prephosphorylated annealed oligonucleotides ("CRISPR repeat for"/"CRISPR repeat rev"). Spacer pairs 1/2 and 3/4 were cloned into plasmids pPB1213 and pPB1344 with the CRISPR repeat to form the mini-CRISPR arrays by Golden Gate assembly (47, 48), creating plasmids pPB1347, pPB1348, pPB1349, and pPB1350, respectively (see Table 4 and Fig. S4). Plasmid sequences were confirmed by Sanger sequencing.

When necessary, genomic DNA was isolated from *H. modesticaldum* cultures using a modification of the Wizard genomic DNA kit (Promega, Madison, WI) in which cells were exposed to lysozyme in the presence of 50 mM EDTA, and the supernatant was removed after centrifugation, before proceeding with the kit's included protocol (23).

All oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA).

Transformation by bacterial conjugation. Plasmids were introduced into *H. modesticaldum* by premethylation and conjugation with *E. coli* strain S17-1, as described previously (23). The conjugation was performed on solid mPYE(2xP) medium, followed by incubation for 2 h in the dark at 42°C. Resuspended mating mixtures were sonicated in an ultrasonic cleaner (VGT-2000; GTSONIC) for 30 s to separate clumped cells before plating them on mPYE(2xP) plates supplemented with appropriate antibiotics.

Spectroscopic measurements. Steady-state absorbance spectra of heliobacterial cells were collected using a fiber optic spectrometer (StellarNet, Inc., Tampa, FL) equipped inside an anaerobic glove box. The spectrometer was composed of a SL1 tungsten source lamp, an SL3 deuterium source lamp, and an SR-50 concave holographic grating detector unit. The outputs from the lamps were joined using a Y-shaped fiber optic cable to generate a broad spectrum, which was carried to a CUV1 cuvette holder by a fiber optic cable; transmitted light was carried to the detector unit by a second fiber optic cable. Fresh mPYE medium was used to blank the spectrometer. Alternatively, steady-state absorbance spectra were collected in an Epoch plate reader (BioTek, Winooski, VT).

Time-resolved absorbance kinetics were collected on a JTS-10 LED kinetic spectrometer (Bio-Logic, Inc., Claix, France) similarly to a previous study (35). Suspensions of living cells (optical density at 788 nm of ~0.5) were sealed in 10-mm screw-top glass cuvettes inside the anaerobic glove box before moving the cuvettes to the JTS-10. The P_{800} photobleaching recovery kinetics were measured using 10-μs measuring pulses from an 810-nm LED. The Si PIN photodiodes were covered by a 532-nm notch filter and a 780-nm long-pass filter (Omega Optical, Inc., Brattleboro, TN) to protect them from the laser light. Ten measuring pulses were delivered to establish a baseline before exciting the sample with a saturating,

single-turnover laser flash from a frequency-doubled Nd:YAG laser (532 nm, 20 mJ, 6-ns duration). After the laser flash, 25 measuring pulses were delivered over the μ s-to-s timescale to monitor the bleaching and recovery of P_{800} .

Pigment extraction and separation. Photosynthetic pigments were extracted from whole cells using the following method, based on published protocols (49, 50). A 2-ml sample of log-phase culture was harvested by centrifugation at $13,000 \times g$ in the anaerobic glove box in the dark. After removal of the supernatant, 500 μ l of a deoxygenated 7:2 acetone-methanol mixture was added, and the tube was vortexed thoroughly for 1 min. After centrifugation at $13,000 \times g$ for 1 min, the pigment-rich supernatant was removed by using a syringe and filtered through a 0.2- μ m-pore size solvent-resistant filter into a new microcentrifuge tube.

Pigment extracts were separated using an HPLC instrument described previously (51, 52) equipped with a PhenoSphere C_{18} analytical reverse-phase column (part 00G-3172-E0; Phenomenex Co., Torrance, CA) and analytical guard column cartridge (part KJ0-4282; Phenomenex Co.). A solvent gradient, based on a previously published method (53), was used to resolve (bacterio)chlorophylls, (bacterio)pheophytins, and carotenoids within a 40-min range. Briefly, two solvents were prepared for HPLC analysis. Solvent A was composed of 74:6:1 acetonitrile-methanol-0.2 M Tris-HCl (pH 8.0), and solvent B was composed of 7:1 methanol-hexane. The flow rate was set at 1.0 ml min^{-1} for the entire experiment, and the column was equilibrated with solvent A. The sample was removed from the anaerobic glove box, and a 100- μ l portion was immediately injected into the HPLC system. The following 40-min chromatography protocol was performed: isocratic flow of solvent A (8 min), a linear gradient from 100% solvent A to 100% solvent B (10 min), isocratic flow of solvent B (18 min), and linear gradient from 100% solvent B to 100% solvent A (4 min). The diode array detector was set to measure an absorbance spectrum from 200 to 900 nm (in 1-nm increments) once per second.

Chemical preparation of demetallated pigments. In darkness, 15 ml of a wild-type *H. modestical-dum* culture was pelleted in a 50-ml centrifuge tube at $5,000 \times g$ for 10 min. In the anaerobic glove box, the aqueous supernatant was removed by decanting, and 5 ml of a deoxygenated 7:2 acetone-methanol mixture was added. The tube was vortexed thoroughly for 1 min and then centrifuged at $5,000 \times g$ for 5 min. The supernatant was carefully decanted to a new tube and removed from the anaerobic glove box. From this point forward, the demetallation process had to be performed in the aerobic conditions of the laboratory hood due to the use of acid and volatile solvents. Therefore, great care was taken to perform the process in darkness, using only a dim green LED headlamp to aid vision, in order to minimize conversion of BChl *g* to hydroxy-Chl *a* (54).

The demetallation procedure was modified from previous protocols (53, 55). In the fume hood, the entire 5 ml of extracted pigments was added to a 50-ml glass separatory funnel. Hydrochloric acid (1 M, 5 ml) was slowly added over the course of 1 min with gentle swirling of the funnel. Diethyl ether (10 ml) was added to the funnel to extract pigments from the aqueous layer. The funnel was inverted and gently shaken for 20 s with periodic venting. The funnel was placed upright until the solvent layers separated. The lower aqueous layer was carefully drained, and the remaining ether layer (containing pigment) was washed twice with about 20 ml of a 4% (wt/vol) sodium bicarbonate solution in water (with the shaking/venting/separation/draining process repeated). The washed ether layer was drained into a 20-ml serum vial and dried under a gentle stream of N_2 gas. After drying, the pigment film was redissolved in 7:2 acetone-methanol, filtered through a 0.2- μ m-pore size solvent-resistant filter, and injected into the HPLC system. Chromatographic separation of the pigments was performed using the aforementioned protocol.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/AEM.01644-19>.

SUPPLEMENTAL FILE 1, PDF file, 1.3 MB.

ACKNOWLEDGMENT

This study was funded by the Division of Chemical Sciences, Geosciences, and Biosciences, Office of Basic Energy Sciences of the U.S. Department of Energy through grant DE-SC0010575 to K.E.R.

REFERENCES

- Adli M. 2018. The CRISPR tool kit for genome editing and beyond. *Nat Commun* 9:1911. <https://doi.org/10.1038/s41467-018-04252-2>.
- Makarova KS, Wolf YI, Koonin EV. 2013. Comparative genomics of defense systems in archaea and bacteria. *Nucleic Acids Res* 41:4360–4377. <https://doi.org/10.1093/nar/gkt157>.
- Godde JS, Bickerton A. 2006. The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. *J Mol Evol* 62:718–729. <https://doi.org/10.1007/s00239-005-0223-z>.
- Koonin EV, Makarova KS, Wolf YI. 2017. Evolutionary genomics of defense systems in archaea and bacteria. *Annu Rev Microbiol* 71:233–261. <https://doi.org/10.1146/annurev-micro-090816-093830>.
- Nuñez JK, Lee ASY, Engelman A, Doudna JA. 2015. Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature* 519:193–198. <https://doi.org/10.1038/nature14237>.
- Amitai G, Sorek R. 2016. CRISPR-Cas adaptation: insights into the mechanism of action. *Nat Rev Microbiol* 14:67–76. <https://doi.org/10.1038/nrmicro.2015.14>.
- Haft DH, Selengut J, Mongodin EF, Nelson KE. 2005. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput Biol* 1:474–483.
- Carte J, Wang R, Li H, Terns RM, Terns MP. 2008. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev* 22:3489–3496. <https://doi.org/10.1101/gad.1742908>.

9. Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, Eckert MR, Vogel J, Charpentier E. 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* 471:602–607. <https://doi.org/10.1038/nature09886>.
10. Brouns SJJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RHJ, Snijders APL, Dickman MJ, Makarova KS, Koonin EV, van der Oost J. 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321:960–964. <https://doi.org/10.1126/science.1159689>.
11. Gasiunas G, Barrangou R, Horvath P, Siksnys V. 2012. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A* 109:E2579–E2586. <https://doi.org/10.1073/pnas.1208507109>.
12. Westra ER, van Erp PBG, Künne T, Wong SP, Staals RHJ, Seegers CLC, Bollen S, Jore MM, Semenova E, Severinov K, de Vos WM, Dame RT, de Vries R, Brouns SJJ, van der Oost J. 2012. CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by cascade and Cas3. *Mol Cell* 46:595–605. <https://doi.org/10.1016/j.molcel.2012.03.018>.
13. Kuzminov A, Stahl FW. 2014. Overview of homologous recombination and repair machines, p 349–368. *In* The bacterial chromosome. American Society for Microbiology, Washington, DC.
14. Wasels F, Jean-Marie J, Collas F, López-Contreras AM, Lopes Ferreira N. 2017. A two-plasmid inducible CRISPR/Cas9 genome editing tool for *Clostridium acetobutylicum*. *J Microbiol Methods* 140:5–11. <https://doi.org/10.1016/j.mimet.2017.06.010>.
15. Wang Y, Zhang ZT, Seo SO, Lynn P, Lu T, Jin YS, Blaschek HP. 2016. Bacterial genome editing with CRISPR-Cas9: deletion, integration, single nucleotide modification, and desirable “clean” mutant selection in *Clostridium beijerinckii* as an example. *ACS Synth Biol* 5:721–732. <https://doi.org/10.1021/acssynbio.6b00060>.
16. Wang Y, Zhang ZT, Seo SO, Choi K, Lu T, Jin YS, Blaschek HP. 2015. Markerless chromosomal gene deletion in *Clostridium beijerinckii* using CRISPR/Cas9 system. *J Biotechnol* 200:1–5. <https://doi.org/10.1016/j.jbiotec.2015.02.005>.
17. Pyne ME, Bruder MR, Moo-Young M, Chung DA, Chou CP. 2016. Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in *Clostridium*. *Sci Rep* 6:25666. <https://doi.org/10.1038/srep25666>.
18. Zhang J, Zong W, Hong W, Zhang Z-T, Wang Y. 2018. Exploiting endogenous CRISPR-Cas system for multiplex genome editing in *Clostridium tyrobutyricum* and engineer the strain for high-level butanol production. *Metab Eng* 47:49–59. <https://doi.org/10.1016/j.ymben.2018.03.007>.
19. Sattley WM, Blankenship RE. 2010. Insights into heliobacterial photosynthesis and physiology from the genome of *Heliobacterium modesticaldum*. *Photosynth Res* 104:113–122. <https://doi.org/10.1007/s11120-010-9529-9>.
20. Sattley WM, Madigan MT, Swingle WD, Cheung PC, Clocksin KM, Conrad AL, Dejesa LC, Honchak BM, Jung DO, Karbach LE, Kurdoglu A, Lahiri S, Mastrian SD, Page LE, Taylor HL, Wang ZT, Raymond J, Chen M, Blankenship RE, Touchman JW. 2008. The genome of *Heliobacterium modesticaldum*, a phototrophic representative of the *Firmicutes* containing the simplest photosynthetic apparatus. *J Bacteriol* 190:4687–4696. <https://doi.org/10.1128/JB.00299-08>.
21. Tang K-H, Yue H, Blankenship RE. 2010. Energy metabolism of *Heliobacterium modesticaldum* during phototrophic and chemotrophic growth. *BMC Microbiol* 10:150. <https://doi.org/10.1186/1471-2180-10-150>.
22. Gisriel C, Sarrou I, Ferlez B, Golbeck JH, Redding KE, Fromme R. 2017. Structure of a symmetric photosynthetic reaction center–photosystem. *Science* 357:1021–1025. <https://doi.org/10.1126/science.aan5611>.
23. Baker P, Orf GS, Khan Z, Espinoza L, Leung S, Kevershan K, Redding KE. 2019. A molecular biology toolkit for the photosynthetic firmicute, *Heliobacterium modesticaldum* Ice1. *Appl Environ Microbiol* 85:e01287-19. <https://doi.org/10.1128/AEM.01287-19>.
24. Mai V, Lorenz WW, Wiegand J. 2006. Transformation of *Thermoanaerobacterium* sp. strain JW/SL-YS485 with plasmid pK1M1 conferring kanamycin resistance. *FEMS Microbiol Lett* 148:163–167. [https://doi.org/10.1016/S0378-1097\(97\)00028-1](https://doi.org/10.1016/S0378-1097(97)00028-1).
25. Olson DG, Maloney M, Lanahan AA, Hon S, Hauser LJ, Lynd LR. 2015. Identifying promoters for gene expression in *Clostridium thermocellum*. *Metab Eng Commun* 2:23–29. <https://doi.org/10.1016/j.meteno.2015.03.002>.
26. Pyne ME, Bruder M, Moo-Young M, Chung DA, Chou CP. 2014. Technical guide for genetic advancement of underdeveloped and intractable *Clostridium*. *Biotechnol Adv* 32:623–641. <https://doi.org/10.1016/j.biotechadv.2014.04.003>.
27. Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Néron B, Rocha EPC, Vergnaud G, Gautheret D, Pourcel C. 2018. CRISPRCasFinder, an update of CRISPRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res* 46:W246–W251. <https://doi.org/10.1093/nar/gky425>.
28. Grissa I, Vergnaud G, Pourcel C. 2007. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* 8:172. <https://doi.org/10.1186/1471-2105-8-172>.
29. Makarova KS, Koonin EV. 2015. Annotation and classification of CRISPR-Cas systems, p 47–75. *In* Lundgren M, Charpentier E, Fineran P (ed), *Methods in molecular biology*, vol 1311. Humana Press, New York, NY. https://doi.org/10.1007/978-1-4939-2687-9_4.
30. Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, Barrangou R, Brouns SJJ, Charpentier E, Haft DH, Horvath P, Moineau S, Mojica FJM, Terns RM, Terns MP, White MF, Yakunin AF, Garrett RA, van der Oost J, Backofen R, Koonin EV. 2015. An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol* 13:722–736. <https://doi.org/10.1038/nrmicro3569>.
31. Makarova KS, Haft DH, Barrangou R, Brouns SJJ, Charpentier E, Horvath P, Moineau S, Mojica FJM, Wolf YI, Yakunin AF, van der Oost J, Koonin EV. 2011. Evolution and classification of the CRISPR–Cas systems. *Nat Rev Microbiol* 9:467–477. <https://doi.org/10.1038/nrmicro2577>.
32. Jackson RN, Lavin M, Carter J, Wiedenheft B. 2014. Fitting CRISPR-associated Cas3 into the helicase family tree. *Curr Opin Struct Biol* 24:106–114. <https://doi.org/10.1016/j.sbi.2014.01.001>.
33. Koonin EV, Makarova KS, Zhang F. 2017. Diversity, classification, and evolution of CRISPR-Cas systems. *Curr Opin Microbiol* 37:67–78. <https://doi.org/10.1016/j.mib.2017.05.008>.
34. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
35. Sheehy D, Kuang Lu Y, Osman F, Alattar Z, Flores C, Sussman H, Zaare S, Dooling M, Meraban A, Baker P, Touchman JW, Redding KE. 2018. Genome-wide transcriptional response during the shift to N₂-fixing conditions in *Heliobacterium modesticaldum*. *J Proteomics Bioinform* 11:143–160. <https://doi.org/10.4172/jpb.1000481>.
36. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* 44:W16–W21. <https://doi.org/10.1093/nar/gkw387>.
37. Nuñez JK, Kranzusch PJ, Noeske J, Wright AV, Davies CW, Doudna JA. 2014. Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity. *Nat Struct Mol Biol* 21:528–534. <https://doi.org/10.1038/nsmb.2820>.
38. Sinkunas T, Gasiunas G, Fremaux C, Barrangou R, Horvath P, Siksnys V. 2011. Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *EMBO J* 30:1335–1342. <https://doi.org/10.1038/emboj.2011.41>.
39. Beloglazova N, Petit P, Flick R, Brown G, Savchenko A, Yakunin AF. 2011. Structure and activity of the Cas3 HD nuclease MJ0384, an effector enzyme of the CRISPR interference. *EMBO J* 30:4616–4627. <https://doi.org/10.1038/emboj.2011.377>.
40. Gasiunas G, Sinkunas T, Siksnys V. 2014. Molecular mechanisms of CRISPR-mediated microbial immunity. *Cell Mol Life Sci* 71:449–465. <https://doi.org/10.1007/s00018-013-1438-6>.
41. Redding KE, Sarrou I, Rappaport F, Santabarbara S, Lin S, Reifschneider KT. 2014. Modulation of the fluorescence yield in heliobacterial cells by induction of charge recombination in the photosynthetic reaction center. *Photosynth Res* 120:221–235. <https://doi.org/10.1007/s11120-013-9957-4>.
42. Michalski TJ, Hunt JE, Bowman M, Smith U, Bardeen K, Gest H, Norris JR, Katz JJ. 1987. Bacteriopheophytin g: properties and some speculations on a possible primary role for bacteriochlorophylls b and g in the biosynthesis of chlorophylls. *Proc Natl Acad Sci U S A* 84:2570–2574. <https://doi.org/10.1073/pnas.84.9.2570>.
43. Semenova E, Jore MM, Datsenko KA, Semenova A, Westra ER, Wanner B, van der Oost J, Brouns SJJ, Severinov K. 2011. Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc Natl Acad Sci* 108:10098–10103. <https://doi.org/10.1073/pnas.1104144108>.
44. Xue C, Seetharam AS, Musharova O, Severinov K, Brouns SJ, Severin AJ, Sashital DG. 2015. CRISPR interference and priming varies with individ-

- ual spacer sequences. *Nucleic Acids Res* 43:10831–10847. <https://doi.org/10.1093/nar/gkv1259>.
45. Shmakov SA, Sitnik V, Makarova KS, Wolf YI, Severinov KV, Koonin EV. 2017. The CRISPR spacer space is dominated by sequences from species-specific mobilomes. *mBio* 8:e01397–17. <https://doi.org/10.1128/mBio.01397-17>.
 46. Ortega-Ramos M, Canniffe DP, Radle MI, Neil Hunter C, Bryant DA, Golbeck JH. 2018. Engineered biosynthesis of bacteriochlorophyll *g* F in *Rhodobacter sphaeroides*. *Biochim Biophys Acta Bioenerg* 1859:501–509. <https://doi.org/10.1016/j.bbabi.2018.02.006>.
 47. Engler C, Kandzia R, Marillonnet S. 2008. A one-pot, one-step, precision cloning method with high-throughput capability. *PLoS One* 3:e3647. <https://doi.org/10.1371/journal.pone.0003647>.
 48. Weber E, Engler C, Gruetzner R, Werner S, Marillonnet S. 2011. A modular cloning system for standardized assembly of multigene constructs. *PLoS One* 6:e16765. <https://doi.org/10.1371/journal.pone.0016765>.
 49. Orf GS, Collins AM, Niedzwiedzki DM, Tank M, Thiel V, Kell A, Bryant DA, Montañó GA, Blankenship RE. 2017. Polymer–chlorosome nanocomposites consisting of non-native combinations of self-assembling bacteriochlorophylls. *Langmuir* 33:6427–6438. <https://doi.org/10.1021/acs.langmuir.7b01761>.
 50. Shah VB, Ferris C, Orf GS, Kavadiya S, Ray JR, Jun Y-S, Lee B, Blankenship RE, Biswas P. 2018. Supramolecular self-assembly of bacteriochlorophyll *c* molecules in aerosolized droplets to synthesize biomimetic chlorosomes. *J Photochem Photobiol B Biol* 185:161–168. <https://doi.org/10.1016/j.jphotobiol.2018.04.032>.
 51. Sarrou I, Khan Z, Cowgill JB, Lin S, Brune DC, Romberger SP, Golbeck JH, Redding KE. 2012. Purification of the photosynthetic reaction center from *Heliobacterium modesticaldum*. *Photosynth Res* 111:291–302. <https://doi.org/10.1007/s11120-012-9726-9>.
 52. Kashey TS, Luu DD, Cowgill JC, Baker PL, Redding KE. 2018. Light-driven quinone reduction in heliobacterial membranes. *Photosynth Res* 138: 1–9. <https://doi.org/10.1007/s11120-018-0496-x>.
 53. Siefermann-Harms D. 1988. High-performance liquid chromatography of chloroplast pigments. *J Chromatogr A* 448:411–416. [https://doi.org/10.1016/S0021-9673\(01\)84604-5](https://doi.org/10.1016/S0021-9673(01)84604-5).
 54. Ferlez B, Dong W, Siavashi R, Redding K, Hou HJM, Golbeck JH, Van Der Est A. 2015. The effect of bacteriochlorophyll *g* oxidation on energy and electron transfer in reaction centers from *Heliobacterium modesticaldum*. *J Phys Chem B* 119:13714–13725. <https://doi.org/10.1021/acs.jpcc.5b03339>.
 55. Schwartz SJ, Woo SL, von Elbe JH. 1981. High-performance liquid chromatography of chlorophylls and their derivatives in fresh and processed spinach. *J Agric Food Chem* 29:533–535. <https://doi.org/10.1021/jf00105a025>.