



## Article

# The Restriction–Modification Systems of *Clostridium carboxidivorans* P7

Patrick Kottenhahn <sup>1,2,†</sup>, Gabriele Philipps <sup>1,\*</sup>, Boyke Bunk <sup>3</sup> , Cathrin Spröer <sup>3</sup> and Stefan Jennewein <sup>1,\*</sup>

<sup>1</sup> Department of Industrial Biotechnology, Fraunhofer Institute for Molecular Biology and Applied Ecology IME, 52074 Aachen, Germany

<sup>2</sup> Department of Biology, RWTH Aachen University, 52074 Aachen, Germany

<sup>3</sup> Department Bioinformatics and Databases, Leibniz Institute DSMZ-German Culture Collection for Microorganisms and Cell Cultures, 38124 Braunschweig, Germany

\* Correspondence: gabriele.philipps@ime.fraunhofer.de (G.P.); stefan.jennewein@ime.fraunhofer.de (S.J.)

† Current address: Nutrition & Care, Evonik Operations GmbH, 33790 Halle (Westfalen), Germany.

**Abstract:** *Clostridium carboxidivorans* P7 (DSM 15243) is a bacterium that converts syngas (a mixture of CO, H<sub>2</sub>, and CO<sub>2</sub>) into hexanol. An optimized and scaled-up industrial process could therefore provide a renewable source of fuels and chemicals while consuming industry waste gases. However, the genetic engineering of this bacterium is hindered by its multiple restriction–modification (RM) systems: the genome of *C. carboxidivorans* encodes at least ten restriction enzymes and eight methyltransferases (MTases). To gain insight into the complex RM systems of *C. carboxidivorans*, we analyzed genomic methylation patterns using single-molecule real-time (SMRT) sequencing and bisulfite sequencing. We identified six methylated sequence motifs. To match the methylation sites to the predicted MTases of *C. carboxidivorans*, we expressed them individually in *Escherichia coli* for functional characterization. Recognition motifs were identified for all three Type I MTases (CAYNNNNNCTGC/GCAGNNNNNRTG, CCANNNNNNNTTCG/CGANNNNNNNTGG and GCANNNNNNNTNCG/CGNNANNNNNNNTGC), two Type II MTases (GATAAT and CRAAAAR), and a single Type III MTase (GAAAT). However, no methylated recognition motif was found for one of the three Type II enzymes. One recognition motif that was methylated in *C. carboxidivorans* but not in *E. coli* (AGAAGC) was matched to the remaining Type III MTase through a process of elimination. Understanding these enzymes and the corresponding recognition sites will facilitate the development of genetic tools for *C. carboxidivorans* that can accelerate the industrial exploitation of this strain.

**Keywords:** DNA methylation; restriction enzyme; restriction–modification system; *Clostridium*; genetic transformation



**Citation:** Kottenhahn, P.; Philipps, G.; Bunk, B.; Spröer, C.; Jennewein, S.

The Restriction–Modification Systems of *Clostridium carboxidivorans* P7. *Microorganisms* **2023**, *11*, 2962.

<https://doi.org/10.3390/microorganisms11122962>

<https://doi.org/10.3390/microorganisms11122962>

Academic Editor: Yingang Feng

Received: 7 November 2023

Revised: 30 November 2023

Accepted: 1 December 2023

Published: 12 December 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Synthesis gas (syngas) is a promising feedstock for the production of platform chemicals from sources such as municipal waste, industrial process gases, or gasified biomass [1–5]. This could reduce our dependence on fossil resources while limiting greenhouse gas emissions by capturing and using the carbon contained in these gases. So-called acetogenic bacteria can use syngas as their sole carbon and energy source and can convert it into various organic acids and alcohols. However, *Clostridium carboxidivorans* P7 can synthesize medium-chain alcohols such as butanol and hexanol from syngas, which can be used as fuels or as raw materials for the synthesis of plastics and chemicals. Recent developments combining medium and process optimization have already resulted in yields in the range of several grams of hexanol/L [6–8].

One barrier hindering the optimization of medium-chain alcohol production is the genetic inaccessibility of *C. carboxidivorans*. Although one study reported four transgenic *C. carboxidivorans* strains with improved ethanol and butanol yields, gene transfer was achieved through conjugation rather than standard electro-transformation procedures because the latter appeared to be inhibited by endogenous restriction–modification (RM)

systems [9]. RM systems are ubiquitous in bacteria and are often considered as barriers to genetic engineering and transformation, with a prominent example being bacteria from the genus *Clostridium* [10–14]. They act as an innate immune system for bacteria, restricting phage infection of the host cell [15,16]. Typical RM systems consist of restriction enzymes (restriction endonucleases) that recognize and/or cleave DNA at a particular site and methyltransferases (MTases) that protect the DNA from cleavage by modifying the bases. Type I systems involve ATP-dependent enzyme complexes consisting of two MTase units (M), two restriction endonuclease units (R), and one specificity peptide (S), which selectively target asymmetric, non-palindromic sites consisting of six to seven specific bases interrupted by five to eight variable bases [17]. The actual cleavage occurs up to several kb away from the recognition site, and cleavage results in the complete degradation of DNA containing an unmethylated recognition sequence [18,19]. Type II systems are the most widely used in molecular biology because the R and M components are separate enzymes that target the same site, which is specific and palindromic [20]. The typical modifications are m<sup>6</sup>A, m<sup>5</sup>C, or m<sup>4</sup>C [17]. In Type III systems, the R and M components form a tetrameric complex that recognizes an asymmetric target site and cleavage occurs downstream from the unmethylated site, although digestion is usually incomplete [21,22]. In contrast to Types I, II, and III, Type IV systems consist of restriction enzymes without a corresponding MTase and only cut methylated DNA motifs. Some Type IV enzymes have specific and precise cleavage sites whereas others are non-specific and variable [17]. In summary, Type I–III systems restrict DNA lacking the host's native methylation pattern, whereas Type IV systems restrict foreign methylation patterns [23].

The genus *Clostridium* (and even individual species within it) features diverse strains ranging from those without RM systems, which are easy to manipulate, to strains with many active RM systems spanning several different types, which are genetically inaccessible [12,14,24,25]. Inconvenient RM systems can be circumvented by ensuring the absence of recognition sites in plasmids used for transformation [26] or by using donor strains with compatible methylation patterns [14,27]. However, neither method is suitable if the methylation patterns of the recipient strain are unknown. Genome sequencing can be used to predict the presence of MTase genes, allowing the creation of a donor strain expressing the recipient's native methyltransferases, but this approach becomes increasingly complex if the host possesses multiple RM systems.

Herein, we characterized the RM systems of *C. carboxidivorans* by screening the genome for putative RM genes and analyzing the methylation status of *C. carboxidivorans* genomic DNA using PacBio and bisulfite sequencing. The genes encoding each MTase (and specificity unit, where applicable) were cloned and expressed in *Escherichia coli* to determine the induced methylation patterns. This enabled us to match each enzyme to the corresponding recognition motif.

## 2. Materials and Methods

### 2.1. Strains and Cultivation Conditions

*Clostridium carboxidivorans* P7 (DSM 15243) was obtained from the DSMZ (Braunschweig, Germany). The cells were grown heterotrophically in modified minimal medium ATCC 1754 [7] with an oxygen-free atmosphere (5% H<sub>2</sub>, 10% CO<sub>2</sub>, and 85% N<sub>2</sub>) at 30 °C without agitation in a Whitley A55 Anaerobic Workstation (Don Whitley Scientific, Herzlake, Germany). *Escherichia coli* NEB 5- $\alpha$  cells (C2987H; New England Biolabs, Ipswich, MA, USA) were used for plasmid propagation and were incubated at 30 °C in a rotary shaker at 150 rpm. For the cultivation of pMT\_Solo strains based on pCDFDuet-1, we added spectinomycin (100 mg mL<sup>-1</sup> dissolved in water) to the LB medium, with a final concentration of 100  $\mu$ g mL<sup>-1</sup>. The *E. coli* strains used in this study are listed in Table A1.

### 2.2. Plasmid Construction, Sequencing and Transformation

MTase genes and specificity subunit genes (where present) were amplified from *C. carboxidivorans* genomic DNA using Q5 High-Fidelity DNA Polymerase (New England

Biolabs) and the specific primers listed in Table A2. The amplicons were inserted into vector pCDFDuet-1 under the control of the inducible T7 promoter. For the analysis of single MTases, Gibson assembly sites were selected so that only one expression site remained in the pCDFDuet-1 vector, enabling the construction of pMT-Solo plasmids. Plasmids assembled through Gibson assembly were introduced into chemically competent *E. coli* NEB 5-alpha cells, which were spread on LB agar plates (Carl Roth, Karlsruhe, Germany) with the appropriate antibiotics and incubated at 30 °C. Single colonies were then transferred to liquid LB medium (Carl Roth) supplemented with 100 µg mL<sup>-1</sup> spectinomycin, and the cultures were used for plasmid isolation and the preparation of glycerol stocks for storage at -80 °C. Plasmids were isolated using the NucleoSpin Plasmid Mini kit (Macherey-Nagel, Düren, Germany). Plasmid integrity was verified through restriction digestion and in-house Sanger sequencing and/or sequencing using Microsynth Seqlab (Göttingen, Germany). Plasmid sequencing reads were assembled and confirmed using the CLC workbench.

### 2.3. Methylation Analysis

Methylation sites were identified using PacBio SMRT sequencing. *C. carboxidivorans* cells were grown as described above and harvested in the late exponential to early stationary growth phase. *E. coli* NEB 5-alpha cells were grown overnight in LB medium at 30 °C while being shaken at 150 rpm. The medium contained 100 µg mL<sup>-1</sup> spectinomycin for the selection of pCDFDuet-1 plasmids and 2 mM isopropyl-β-D-thiogalactopyranoside (IPTG) for induction. *E. coli* NEB 5-alpha does not encode a T7 polymerase gene in its genome. However, the T7 promoter region is expected to be accessible to native polymerases. Additionally, pCDFDuet-1 contains cryptic -35 and cryptic -10 promoter boxes adjacent to the T7 promoter, resulting in low-level expression of the MTases. Cells were harvested in the mid-to-late exponential growth phase. Genomic DNA was isolated from *E. coli* using the standard protocol of the NucleoSpin gDNA Mini kit (Macherey-Nagel). For *C. carboxidivorans*, we used the isolation protocol for hard-to-lyse bacteria according to the manufacturer's recommendations. Genomic DNA was stored at 4 °C and sent to the DSMZ in a cooled overnight parcel for PacBio single molecule real-time (SMRT) sequencing, which can detect the presence of m<sup>6</sup>A and m<sup>4</sup>C [28,29].

SMRTbell template libraries were prepared by following the Procedure & Checklist—Preparing Multiplexed Microbial Libraries Using SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences, Menlo Park, CA, USA). Briefly, 10 kb libraries were prepared by shearing 1 µg of genomic DNA in g-tubes (Diagenode, Denville, NJ, USA) according to the manufacturer's instructions. The DNA was end-repaired and ligated to barcoded adapters using the SMRTbell Express Template Prep Kit 2.0. Samples were pooled as recommended by the Microbial Multiplexing Calculator. Conditions for primer annealing and the binding of polymerase to the purified SMRTbell template were assessed using the SMRT-Link calculator. Three genomic libraries were sequenced on a Sequel IIe device (Pacific Biosciences) taking one 15 h movie per SMRT cell. One SMRT cell was used for *C. carboxidivorans* and one was used for *E. coli*.

For bioinformatic analysis, all datasets were processed using the Base Modification Analysis Protocol in SMRT-Link 10.0.0.108728. Essentially, the detection of base modifications is based on a (statistical) increase in the inter-pulse duration (IPD) values. The details are described by Feng et al. 2013 [30]. The genomes of *C. carboxidivorans* P7 (Ref-Seq NZ\_CP011803.1) and *E. coli* K-12 NEB 5-alpha (GenBank CP017100.1) were used as references. We applied a modification threshold (Qmod) score of 50 if not stated otherwise.

To verify the absence of m<sup>5</sup>C methylation in *C. carboxidivorans*, genomic DNA was sent to CD Genomics (CD Biosciences, New York, NY, USA) for bisulfite conversion using the Bisulfite-Seq Library Prep Kit (Acegen, Shenzhen, China) followed by sequencing on an Illumina NovaSeq device in PE150 mode. Briefly, 1 µg of genomic DNA was fragmented through sonication (200–400 bp mean size), followed by end-repair, 5' phosphorylation, 3'-dA-tailing, and ligation to methylated adapters. The methylated adapter-ligated DNAs were purified using 0.8× Agencourt AMPure XP magnetic beads before bisulfite conversion

using the ZYMO EZ DNA Methylation-Gold Kit (Zymo Research, Irvine, CA, USA). The converted DNA was amplified with 25  $\mu$ L KAPA HiFi HotStart Uracil+ ReadyMix (2X) and 8-bp index primers (final concentration 1  $\mu$ M each). The library quality was confirmed using an Agilent 2100 Bioanalyzer and quantified using a Qubit fluorometer with the Quant-iT dsDNA HS Assay Kit (Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA). The library was then sequenced using an Illumina HiSeq X ten platform in PE150 mode.

#### 2.4. RNA Isolation and Semi Quantitative RT-PCR

Cells were grown in 50 mL of modified ATCC 1754 medium with fructose in a 500 mL bottle to OD = 0.8 at 30 °C in a Don Whitley anaerobic workstation. The cells were harvested through centrifugation at 4 °C and the pellet was stored at  $-80$  °C. RNA was isolated using the Macherey–Nagel RNA Isolation Kit (with DNase digest) and was reverse transcribed to cDNA before semi-quantitative PCR using the Qiagen QuantiTect Reverse Transcription Kit (with a second DNase digest prior to reverse transcription), Phusion<sup>®</sup> High-Fidelity DNA Polymerase, and primers specific to each RM component gene (Table A3). The annealing temperature was set to 55 °C, with elongation for 30 s over 35 amplification cycles. Genomic DNA was used as the positive control, total RNA after DNA digestion was used as the negative control, and cDNA was used as the test sample.

### 3. Results and Discussion

#### 3.1. Analysis of Methylation Activity in *C. carboxidivorans*

Undiscovered RM systems can be predicted based on genome sequences by screening for homology to genes representing known restriction enzymes and MTases, which are collected in the online resource REBASE (<http://rebase.neb.com>; [31]). Using this approach, ten restriction enzyme genes were predicted in the *C. carboxidivorans* P7 genome, representing three Type I RM systems, two Type II systems, two Type III systems, and three Type IV systems (Table 1). The *C. carboxidivorans* genome encodes an unusually large number of restriction enzymes compared to other *Clostridium* species (Table 2). Furthermore, eight MTase genes correlating with the Type I, II, and III restriction enzymes, as well as one additional orphan Type II MTase were predicted for *C. carboxidivorans*. Two of the Type II restriction enzymes and MTases were present as fusion proteins.

To determine whether the predicted RM genes in *C. carboxidivorans* P7 were expressed, the total RNA was isolated, followed by two DNase treatments to destroy any genomic DNA contamination that could lead to false positive results. The RNA was then reverse transcribed into cDNA and amplified using primers specific to each predicted RM component (Table A3). We used genomic DNA as a positive control for the RM genes and RNA, following the DNase step before cDNA synthesis, as a control for DNA contamination. After 35 amplification cycles, the RNA negative control lane contained only weak bands, confirming that the samples contained negligible amounts of residual genomic DNA (Table A1). In contrast, all of the cDNA lanes produced strong bands, indicating that all of the RM components in the *C. carboxidivorans* genome were expressed under our selected experimental conditions (Figure A1). Prior to this study, methylation patterns in the *C. carboxidivorans* genome were partially predicted but not experimentally verified. To identify the comprehensive set of native *C. carboxidivorans* RM motifs, we isolated genomic DNA from heterotrophic cultures for PacBio sequencing and methylation analysis.

During SMRT sequencing, labeled nucleotides are polymerized on the complementary DNA strand. Delays in nucleotide incorporation caused by base modifications such as methylation increase the gap between fluorescence pulses, known as the inter-pulse duration. Different modifications affect polymerase kinetics in specific ways, enabling a simultaneous readout of the primary nucleotide sequence and certain base modifications [29]. It is therefore possible to detect m<sup>6</sup>A (N<sup>6</sup>-methyladenosine) and m<sup>4</sup>C (N<sup>4</sup>-methylcytosine) even with a low sequencing coverage [33].

**Table 1.** Restriction modification (RM) systems predicted in the genome of *C. carboxidivorans* P7. The table shows the RM type, gene names representing individual subunits, predicted recognition motifs (if available), genome coordinates, and the short names used for individual enzymes in this study. The genome sequence was obtained from GenBank (CP011803, 5,732,880 bp). Table modified from [http://tools.neb.com/genomes/report.php?genome\\_id=20798](http://tools.neb.com/genomes/report.php?genome_id=20798) (accessed on 2 September 2021) based on the published *C. carboxidivorans* genome sequence [32].

Type	Subunit	Gene Name	Predicted Recognition Site	Coordinates	ORF (bp)	Short Name
I	S	S.CcaP7I	CAYNNNNNCTGC	2727394–2728617 c	1224	Type I.1
	M	M.CcaP7I	CAYNNNNNCTGC	2728619–2730115 c	1497	
	R	CcaP7IP	CAYNNNNNCTGC	2731721–2735029 c	3309	
I	R	CcaP7IIP	CCANNNNNNNNTCG	827150–830422	3273	Type I.2
	M	M.CcaP7II	CCANNNNNNNNTCG	830425–831846	1422	
	S	S.CcaP7II	CCANNNNNNNNTCG	831846–833189	1344	
I	R	CcaP7IIIP	GCANNNNNNNTNCG	5411396–5414605	3210	Type I.3
	S	S.CcaP7III	GCANNNNNNNTNCG	5414864–5416072	1209	
	M	M.CcaP7III	GCANNNNNNNTNCG	5416106–5417695	1590	
II	M	M.CcaP7ORF16660P	CTSAG	3757440–3758204 c	765	Type II.1
II	RM	CcaP7IV	GATAAT	5299217–5300845	1629	Type II RM.1
II	RM	CcaP7ORF7120P	-	1647209–1650934	3726	Type II RM.2
III	R	CcaP7ORF17830P	-	4003447–4006389 c	2943	Type III.1
	M	M.CcaP7ORF17830P	-	4006403–4008355 c	1953	
III	R	CcaP7ORF8705P	-	2023100–2025358 c	2259	Type III.2
	M	M.CcaP7ORF8705P	-	2025352–2027229 c	1878	
IV	R	CcaP7ORF3495P	-	805395–808574	3180	Type IV.1
IV	R	CcaP7McrB2P	-	815415–816887	1473	Type IV.2
IV	R	CcaP7McrBP	-	816865–818568	1704	Type IV.3

c—complement.

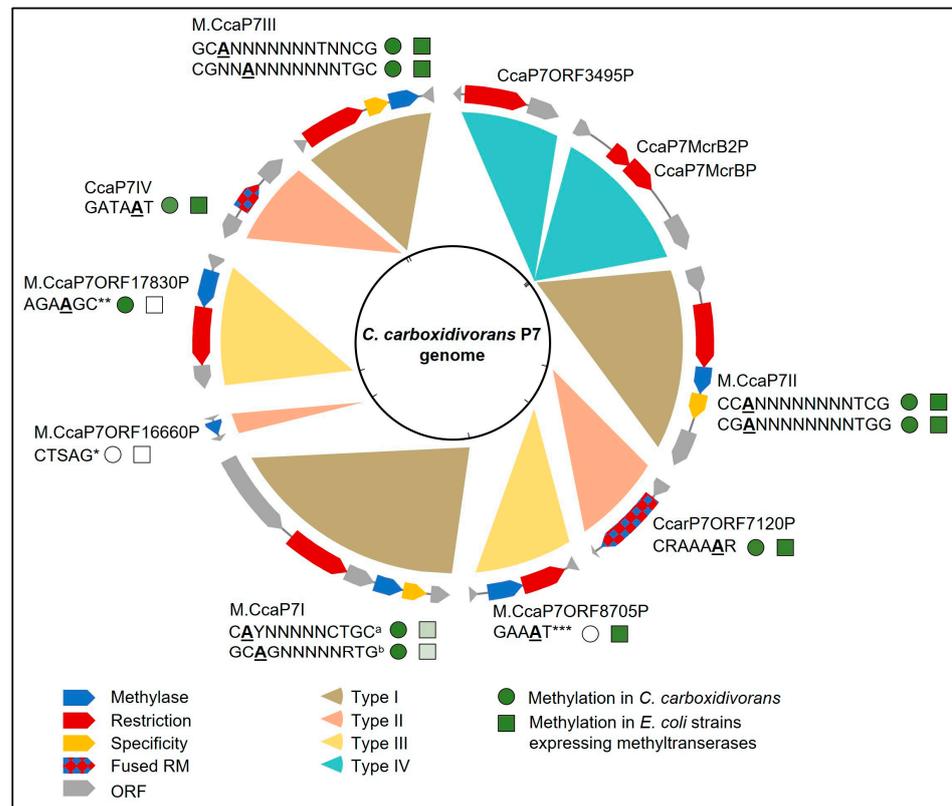
**Table 2.** Comparison of genome sizes and the number of putative restriction enzymes (REs) and methyltransferase (MTases) encoded in the genomes of different *Clostridium* species. Data were retrieved from REBASE (<http://rebase.neb.com>, accessed 24 February 2023).

Organism	Genome Size (bp)	Accession Number	PacBio	Putative REs	Putative MTases
<i>Clostridium acetobutylicum</i> ATCC 824	3,940,880	AE001437 (NC_003030)	No	3	6
<i>Clostridium autoethanogenum</i> DSM 10061	4,352,446	CP012395	Yes	6	4
<i>Clostridium beijerinckii</i> NCIMB 8052	6,000,632	CP000721 (NC_009617)	No	2	2
<i>Clostridium botulinum</i> A ATCC 19397	3,863,450	CP000726 (NC_009697)	Yes	2	5
<i>Clostridium carboxidivorans</i> P7	5,732,880	CP011803	Yes	10	8
<i>Clostridium cellulolyticum</i> H10	4,068,724	CP001348 (NC_011898)	No	4	10
<i>Clostridium cellulovorans</i> 743B	5,262,222	CP002160 (NC_014393)	No	6	13
<i>Clostridium difficile</i> 630	4,290,252	AM180355 (NC_009089)	Yes	2	5
<i>Clostridium diolis</i> DSM 15410	5,940,808	CP043998	Yes	2	1
<i>Clostridium kluyveri</i> DSM 555	3,964,618	CP000673 (NC_009706)	No	5	13 (2 <sup>a</sup> )
<i>Clostridium ljungdahlii</i> DSM 13528	4,630,065	CP001666 (NC_014328)	Yes	5	7
<i>Clostridium pasteurianum</i> BC1	4,990,707	CP003261	No	4	5
<i>Clostridium pasteurianum</i> DSM 525 = ATCC 6013	4,352,852	CP013018	Yes	4	8
<i>Clostridium perfringens</i> ATCC 13124	3,256,683	CP000246 (NC_008261)	Yes	6	7
<i>Clostridium sporogenes</i> DSM 795	4,142,990	CP011663	Yes	2	4
<i>Clostridium thermocellum</i> ATCC 27405	3,843,301	CP000568 (NC_009012)	Yes	5	11

<sup>a</sup> The number in brackets shows the additional number of predicted putative MTases located on *C. kluyveri* DSM 555 plasmid pCKL555A [CP000674]. No additional MTases or restriction enzymes were predicted on *C. carboxidivorans* plasmid p19 [CP011804].

The coverage diagram for the circular *C. carboxidivorans* genome revealed high coverage between 4 and 5 Mbp of the reference sequence and low coverage between 1 and 2 Mbp (Figure A2a). This matches the Z-shaped curve obtained for actively growing cultures and the predicted position of the *oriC* region between positions 4,610,600 and 4,611,172 in the reference sequence [34]. All motifs detected by PacBio sequencing contained m<sup>6</sup>A residues, with no evidence of m<sup>4</sup>C cytosine methylation. We also screened for putative m<sup>5</sup>C (5-methylcytosine) sites using bisulfite treatment followed by sequencing on an Illumina NovaSeq instrument in PE150 mode.

Bisulfite treatment converts unmethylated cytosine to thymidine, whereas methylated cytosine and other bases remain unaffected, allowing the position of m<sup>5</sup>C sites to be confirmed by comparing the bisulfite sequence to the reference sequence. We found no evidence of m<sup>5</sup>C cytosine methylation, suggesting that all RM systems in *C. carboxidivorans* introduce m<sup>6</sup>A modifications. These data have been deposited in the NCBI sequence read archive (BioProject-No. PRJNA994489). The results are summarized in Figure 1 and are set out in more detail in Table A4. The PacBio analysis of *C. carboxidivorans* genomic DNA revealed nine methylation motifs (Figure A2b). Three of them were pairs of non-palindromic complementary sequences, which are methylated by one enzyme each (grouped in Table A4).



**Figure 1.** Schematic map of the *Clostridium carboxidivorans* P7 genome showing genes encoding RM system components and the results of methylome analysis in *C. carboxidivorans* and *E. coli* strains expressing the corresponding methyltransferases. Genes representing RM systems are color coded according to their assumed functional unit and flanking genes are shown in gray. The type of the RM system is shown as a colored segment (image adapted and modified from REBASE). Each methyltransferase gene is accompanied by an assigned recognition motif. Methylated bases are underlined and in bold (**A**). The percentage of methylated motifs in *C. carboxidivorans* and *E. coli* strains expressing the corresponding methyltransferases is visualized by the intensity of the colored

symbols (○ and □, respectively). Genomic DNA isolated from cells at the late exponential/early stationary growth phase was analyzed using PacBio sequencing and sequencing after bisulfite conversion. More details are provided in Table A4. \* The predicted motif CTSAG for MTaseII.1 (Table 1) was not detected in *C. carboxidivorans* or in the *E. coli* expression strains we investigated, but the gene was transcribed in *C. carboxidivorans* (Figure A1). \*\* The motif AGAAGC was detected in *C. carboxidivorans* but not in the *E. coli* expression strains. Because all other MTases were accounted for, this motif is probably recognized by Mtase III.1. \*\*\* The motif GAAAT was detected in *E. coli* but not in *C. carboxidivorans*. In the *E. coli* strain expressing the MTase and the specificity unit of M.CcaP7I, the precise motifs were not found but very similar motifs (<sup>a</sup> tCAYbNNNNCTGC and <sup>b</sup> GCAGNNNNNRTGnnnh, with differing bases in lower case) were detected using a lower modification threshold (Qmod) score of 25.

However, we did not find the predicted motif CTSAG for the Type II.1 MTase. Given that the corresponding native gene was expressed (Figure A1), we assume that the quantity of the transcript was insufficient to produce enough enzyme, or that the enzyme was inactive. Furthermore, the sequences GATAAT, CRAAAR, and AGAAGC were methylated at high frequencies of 84.8%, 97.3%, and 99.6%, respectively. Methylome analysis in *C. carboxidivorans* therefore confirmed the activity of six MTases, but it was still necessary to determine which of the eight predicted MTases recognize which of the six identified motifs and which enzymes are inactive.

### 3.2. Expression of *C. carboxidivorans* MTases for In Vivo Methylation in *E. coli*

In order to pair the eight *C. carboxidivorans* MTases with their specific target sites, we inserted each methyltransferase gene into the expression vector pCDFDuet-1 under the control of the lactose-inducible T7 promoter for the transformation of *E. coli* NEB 5-alpha cells. To ensure sequence recognition by the Type I MTases, these genes were co-expressed with the corresponding specificity subunit but not the restriction subunit. A schematic representation of one of the resulting pMT\_solo plasmids, named according to the specific MTase, is shown in Figure A3. Following the overnight induction of expression with 2 mM IPTG, total DNA was isolated from all eight *E. coli* strains for PacBio sequencing to annotate the putative recognition motifs associated with each MTase (Figure 1 and Table A4). Almost all of the detected motifs were methylated at a frequency of nearly 100%. The exception was Type I.1 M.CcarP7I, where we detected two complementary motifs only when using a lower modification threshold (tCAYbNNNNCTGC and GCAGNNNNNRTGnnnh). In *E. coli*, these motifs were modified at frequencies of 32.0% and 20.6%, respectively, indicating low enzyme activity (or low expression). In contrast, both motifs were modified at a frequency of nearly 100% by the native enzyme in *C. carboxidivorans* (Table A4). These results confirmed that most *C. carboxidivorans* MTases were active in the *E. coli* expression strains. Recognition motifs could be identified for all MTases encoded in the *C. carboxidivorans* genome by combining the predicted motifs from REBASE with the methylome data from *C. carboxidivorans* and the *E. coli* strains expressing single MTases (Table A4).

REBASE allows a comprehensive search of organisms with similar motifs when a particular recognition sequence is used as a search query. The motifs detected by methylome analysis were therefore used to search REBASE for matches in other organisms (Table 3).

**Table 3.** MTase recognition motifs in *C. carboxidivorans* and their conservation in other bacterial species. Data were retrieved from REBASE (accessed 24 February 2023).

Methylation Motif	Type	Name	Homology
CAYNNNNNCTGC GCAGNNNNNRTG	Type I.1	M.CcaP7I	None
CCANNNNNNNTCG CGANNNNNNNTGG	Type I.2	M.CcaP7II	Motif included in CCANNNNNNNTCGT/ ACGANNNNNNNTGG found in <i>Vibrio harveyi</i> NCTC12970
GCANNNNNNNTNCG CGNNANNNNNNNTGC	Type I.3	M.CcaP7III	Motif included in GGCANNNNNNNTNCG/ CGNNANNNNNNNTGC found in <i>Klebsiella pneumoniae</i> AR_0139
CTSAG *	Type II.1	M.CcaP7ORF16660P	Predicted for many strains (including Clostridia) and several confirmed by PacBio, but also other species such as <i>Klebsiella pneumoniae</i> NCTC9151, <i>Lactobacillus acidophilus</i> , and <i>Bacillus stearothermophilus</i> Isl 15-111 (with gold standard enzyme)
GATAAT	Type II RM.1	CcaP7IV	Several strains from <i>Clostridium botulinum</i> and <i>Clostridium sporogenes</i>
CRAAAAR	Type II RM.2	CcaP7ORF7120P	Motif CRAAAAR: <i>Clostridium cadaveris</i> IFB3C5 <sup>a</sup> Motif CAAAAAR: several strains from <i>Clostridium botulinum</i> and <i>Clostridium sporogenes</i> ; other species such as <i>Clostridium pasteurianum</i> , <i>Clostridium tetani</i> , <i>Clostridium autoethanogenum</i> , <i>Clostridium ljungdahlii</i> <sup>b</sup> , <i>Eubacterium limosum</i> B2 <sup>c</sup> and <i>Acetobacterium woodii</i> DSM 1030 Motif CAAAAA: <i>Clostridium difficile</i> 630 (with gold standard enzyme)
AGAAGC **	Type III.1	M.CcaP7ORF17830P	<i>Lactococcus lactis</i> subsp. <i>lactis</i> strain UC073
GAAAT ***	Type III.2	M.CcaP7ORF8705P	<i>Pseudopropionibacterium propionicum</i> NCTC11666 <i>Corynebacterium diphtheriae</i> NCTC10838 <i>Helicobacter fennelliae</i> NCTC13102 <i>Arachnia propionica</i> F0231

\* The methylation of motif CTSAG for MTaseII.1 (Table 1) was not detected in *C. carboxidivorans* or in the *E. coli* expression strains in this study, but the native gene was expressed in *C. carboxidivorans* (Table A1) \*\* The motif AGAAGC was detected in *C. carboxidivorans* but not in the *E. coli* expression strains. However, because all the other MTases were accounted for, this motif is probably recognized by MTase III.1. \*\*\* The motif GAAAT was detected in *E. coli* but not in *C. carboxidivorans*. Other organisms with similar recognition motifs were found by screening REBASE. Further data taken from the literature: <sup>a</sup> [35], <sup>b</sup> [36], <sup>c</sup> [37].

The complementary pair CAYNNNNNCTGC/GCAGNNNNNNRTG did not match any other organisms, whereas the pair CCANNNNNNNTCG/CGANNNNNNNTGG is also found in *Vibrio harveyi* NCTC12970 (as part of the slightly longer motif CCANNNNNNNNTCGT/ACGANNNNNNNTGG) and the pair GCANNNNNNNTNCG/CGNNANNNNNNNTGC is also found in *Klebsiella pneumoniae* AR\_0139 (as part of the longer motif GGCANNNNNNNTNCG/CGNNANNNNNNNTGCC). The Type II MTase motif GATAAT matched several strains of *C. botulinum* and *C. sporogenes*. Although we found no matches for the motif CRAAAAR in REBASE, a recently published article discussing the complete genome sequence of *Clostridium cadaveris* IFB3C5 reported the same motif [35]. However, the non-generalized (more specific) recognition sequence CAAAAAR was detected multiple times in the genus *Clostridium*, including in *C. botulinum* and *C. sporogenes*; in others such as *C. pasteurianum*, *C. tetani*, *C. autoethanogenum* [38,39], and *C. ljungdahlii* [36,39]; and even in *Eubacterium limosum* B2 [37] and *Acetobacterium woodii* DSM1030 [39]. A closely related motif (CAAAA) was found in *Clostridium difficile* 630. In a study of 36 different *C. difficile* strains, the motif CAAAA was ubiquitously methylated and was proposed to have a conserved function influencing sporulation [25]. Due to their high similarity, the motifs CRAAAAR and CAAAAAR might have the same function.

The predicted Type II MTase target site CTSAG was not modified in *C. carboxidivorans* or in the *E. coli* strain expressing the corresponding enzyme. However, when we screened REBASE with this sequence, we recovered 263 entries representing many different species, including *C. botulinum* and *C. perfringens*. Furthermore, *K. pneumoniae* NCTC9151 carries a Type III restriction enzyme with this recognition sequence, and the motif is also found in *Lactobacillus acidophilus* and *Bacillus stearothermophilus* Isl 15-111. In the latter case, the corresponding MTase gene was cloned and its enzyme activity was verified [40]. Those characterized enzymes are classified in REBASE as gold standard enzymes. Given that no corresponding restriction enzyme was predicted for this MTase in *C. carboxidivorans*, it may have evolved to fulfil a different, perhaps regulatory function. Alternatively, it might be active only under specific growth conditions that were not tested in this study. Finally, it may have lost its function after the corresponding restriction enzyme was lost from the genome, thus removing the selection pressure for CTSAG methylation.

The Type III MTase motif AGAAGC was the only motif that was methylated in *C. carboxidivorans* but not in *E. coli*. The only other organism to match this motif in REBASE was *Lactococcus lactis* ssp. *lactis* strain UC073. Interestingly, the second Type III MTase motif GAAAT was only methylated in *E. coli* but not in *C. carboxidivorans*. PacBio data confirmed that the motif is also methylated in *Arachnia propionica* F0231, *Corynebacterium diphtheriae* NCTC10838, *Helicobacter fennelliae* NCTC13102, and *Pseudopropionibacterium propionicum* NCTC11666, but the corresponding enzymes have not yet been assigned.

The diverse results arising from the comparison of methylation in different bacteria raises some interesting questions. Whereas some motifs, such as the one recognized by Type II.1, were shared by a very heterogeneous group of organisms, others, such as the one recognized by Type II RM.1, were limited to selected *Clostridium* species and the Type I.1 motif was not listed in the database at all. Species in the same genus as *C. carboxidivorans* were present, but also Gram-negative species such as *K. pneumoniae* and *V. harveyi*. This phylogenetically and environmentally diverse group sharing methylation motifs with *C. carboxidivorans* suggests that MTase genes may have been acquired by horizontal gene transfer or convergent evolution. Further studies are required to determine which of these scenarios is most likely.

#### 4. Conclusions

The *C. carboxidivorans* genome encodes an unusually large number of active RM enzymes that are likely to hinder the genetic modification of this species. By defining the restriction motifs and corresponding MTases, it should be possible to develop reliable transformation protocols for *C. carboxidivorans*, as already shown for other *Clostridium* species. These methods include restriction alleviation, in which plasmid sequences are

designed to avoid recognition motifs or plasmids are specifically methylated [12,26,41,42]. The analysis of RM systems in *C. carboxidivorans* provides important knowledge that will allow this species to be tailored for the industrial utilization of syngas. We observed interesting similarities between recognition motifs in *C. carboxidivorans* and distantly related bacteria from different habitats. Future research should consider how these RM traits were acquired, including convergent evolution and horizontal gene transfer, to provide insight into the evolution of RM systems in diverse ecosystems.

**Author Contributions:** Conceptualization, P.K., G.P. and S.J.; methodology and investigation, P.K. (microbiology) B.B. and C.S. (methylation study); software, B.B.; data discussion and analysis, P.K., G.P. and B.B.; writing—original draft preparation, P.K.; writing—review and editing, P.K., G.P. and S.J.; visualization, G.P.; resources, B.B. and C.S.; supervision, project administration and funding acquisition, S.J. and G.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the German Federal Ministry of Education and Research (BMBF) as part of the BioCOnversion project (grant number 03INT513B).

**Data Availability Statement:** Data is contained within the article or in Appendix A. The *C. carboxidivorans* methylome data were deposited at NCBI SRA (BioProject-No. PRJNA994489) and will be made available after publication. Further data will be made available on reasonable request.

**Acknowledgments:** The authors thank Nicole Heyer for technical assistance with SMRT sequencing and Richard M Twyman for manuscript editing.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results. The first author conducted this research while affiliated with the Fraunhofer Institute for Molecular Biology and Applied Ecology IME and RWTH Aachen University. Currently, he is employed at Evonik Operations GmbH.

## Appendix A

**Table A1.** List of *Escherichia coli* strains used in this study.

Strain Name	Properties
<i>E. coli</i>	NEB 5-alpha <sup>a</sup>
<i>E. coli</i>	soloI.1
<i>E. coli</i>	soloI.2
<i>E. coli</i>	soloI.3
<i>E. coli</i>	soloII.1
<i>E. coli</i>	soloIIRM1
<i>E. coli</i>	soloIIRM2
<i>E. coli</i>	soloIII.1
<i>E. coli</i>	soloIII.2

<sup>a</sup> Provided by New England Biolabs (Ipswich, MA, USA).

**Table A2.** List of primers used for the cloning of CMTase genes in pMT\_solo constructs.

Solo MTs	Primer Sequence (5'–3')	Anneals
pMT1-3 fw (pCDF1-3 fw)	TTAACCTAGGCTGCTGCCACCGCTGAGCAATAACTAGC	pCDFDuet backbone
pCDF Duet rv neu	CATGGTATATCTCCTTATTAAAGTTAAACAAAATTATTTCTAC	pCDFDuet backbone
TypeIII MT1_fwd	TAATAAGGAGATATACCATGGCTAACTTAATTGAAAAC	<i>C. carboxidivorans</i> Type III M1
TypeIII MT1_rev	GTGGCAGCAGCCTAGGTTAATATCTATATACTCATAATATCTT TTATATCATTTT	<i>C. carboxidivorans</i> Type III M1
TypeIII MT2_fwd	TGTTTAACTTTAATAAGGAGATATACCATGGAAAAAGTATATG CATTGG	<i>C. carboxidivorans</i> Type III M2
TypeIII MT2_rev	TTGCTCAGCGGTGGCAGCAGCCTAGGTTAAATGCCTACCATG ACTTTAG	<i>C. carboxidivorans</i> Type III M2
TypeII MT1_fwd	TGTTTAACTTTAATAAGGAGATATACCATTAGAAATAAAGTAA TAAATAAAGAGTGC	<i>C. carboxidivorans</i> Type II M1

Table A2. Cont.

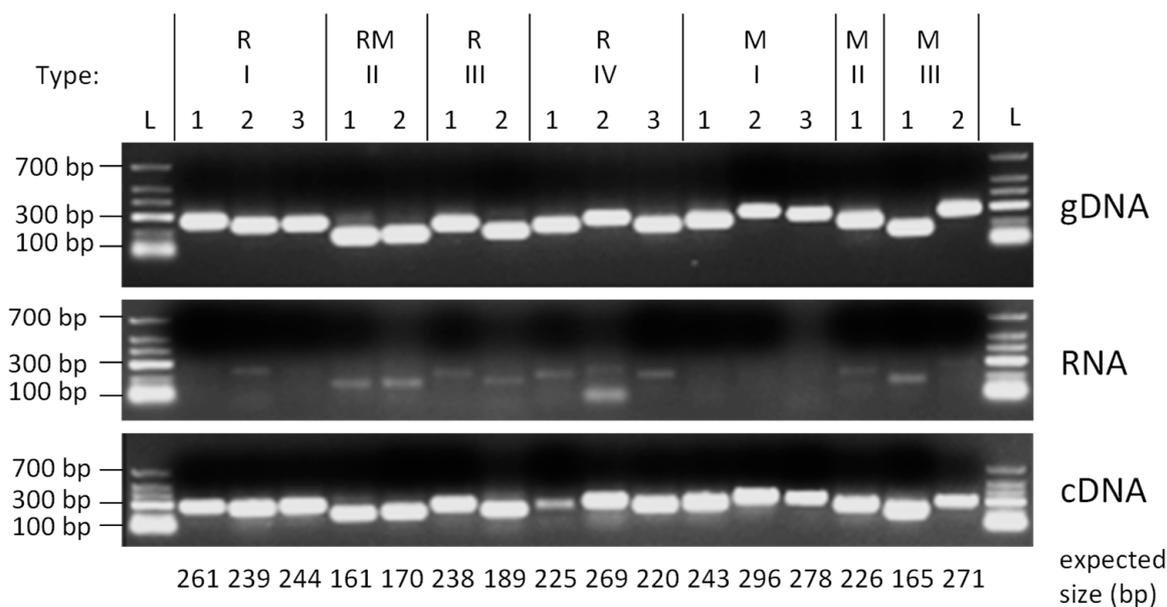
Solo MTs	Primer Sequence (5'–3')	Anneals
TypeII MT1_rev	TTGCTCAGCGGTGGCAGCAGCCTAGGTTAAGCTCTAAACAAA TTTAAGCTG	<i>C. carboxidivorans</i> Type II M1
TypeII RM1_fwd	TGTTTAACTTTAATAAGGAGATATACCATGTATAAAGACGTTA AATTAGAAAAAAG	<i>C. carboxidivorans</i> Type II R+M 1
TypeII RM1_rev	TTGCTCAGCGGTGGCAGCAGCCTAGGTTAATTCTTACAGCAGT TTATCTC	<i>C. carboxidivorans</i> Type II R+M 1
TypeII RM2_fwd	TGTTTAACTTTAATAAGGAGATATACCATGGATAAGACTAAAG TAAAAATCC	<i>C. carboxidivorans</i> Type II R+M 2
TypeII RM2_rev	TTGCTCAGCGGTGGCAGCAGCCTAGGTTAAGTTAATTATATCT TAAAAAATAAATCCTTTTAAAC	<i>C. carboxidivorans</i> Type II R+M 2
TypeI.1_fwd	TGTTTAACTTTAATAAGGAGATATACCATGTAAACACGCGAGA CAAAAAG	<i>C. carboxidivorans</i> Type I M+S 1
TypeI.1_rev	TTGCTCAGCGGTGGCAGCAGCCTAGGTTAATTAATTAATAGT TCTCCTTTGAAAAG	<i>C. carboxidivorans</i> Type I M+S 1
TypeI.2_fwd	TGTTTAACTTTAATAAGGAGATATACCATGAATACACAAGAGA TAGTAAG	<i>C. carboxidivorans</i> Type I M+S 2
TypeI.2_rev	TTGCTCAGCGGTGGCAGCAGCCTAGGTTAACTATATATCTTTAC TCAATATTTCCC	<i>C. carboxidivorans</i> Type I M+S 2
TypeI.3_fwd	TGTTTAACTTTAATAAGGAGATATACCATGGAAAAAACA AAA AATAAACCC	<i>C. carboxidivorans</i> Type I M+S 3
TypeI.3_rev	TTGCTCAGCGGTGGCAGCAGCCTAGGTTAACTAAATTTTACA CCTAAAATCTTCAATTG	<i>C. carboxidivorans</i> Type I M+S 3

Table A3. List of RT-PCR primers used in this study.

Primer	Primer Sequence (5'–3')	Product Size (bp)	Anneals
RI.1_fw	AGTGAGCCTAGACAGGTTTG	261	<i>C. carboxidivorans</i> Type I R 1
RI.1_rv	CCAGCTTGCCCTCCATTAATC		<i>C. carboxidivorans</i> Type I R 1
RI.2_fw	AGCAACAGTACAGGCTATGG	239	<i>C. carboxidivorans</i> Type I R 2
RI.2_rv	GGCTGGTGTAGCTGTAAGTG		<i>C. carboxidivorans</i> Type I R 2
RI.3_fw	AAGTGGCAGTTACGTTTAGC	244	<i>C. carboxidivorans</i> Type I R 3
RI.3_rv	AGTTCTGGTGCATCAAATCC		<i>C. carboxidivorans</i> Type I R 3
RMII.1_fw	AGAAAGAATAAGCAGTGCAAAG	161	<i>C. carboxidivorans</i> Type II R+M 1
RMII.1_rv	AGTCTATCAGGCAGTACAAATC		<i>C. carboxidivorans</i> Type II R+M 1
RMII.2_fw	GACATAGGAGCAAGGTATTGTC	170	<i>C. carboxidivorans</i> Type II R+M 2
RMII.2_rv	TCGCCTACCTGGATATTGTAAG		<i>C. carboxidivorans</i> Type II R+M 2
RIII.1_fw	TCGGCCTTAAGAGAAGGTTG	238	<i>C. carboxidivorans</i> Type III R 1
RIII.1_rv	TTACGCCACCATCTTCTTCG		<i>C. carboxidivorans</i> Type III R 1
RIII.2_fw	TGGATGGGATTGTCCGAGAG	189	<i>C. carboxidivorans</i> Type III R 2
RIII.2_rv	GAAAGGCTGCCGACTTTAAC		<i>C. carboxidivorans</i> Type III R 2
RIV.1_fw	AAGTGCTGGATAGAGCAAATAC	225	<i>C. carboxidivorans</i> Type IV R 1
RIV.1_rv	CAAAGTGCATGTCACATTGTTT		<i>C. carboxidivorans</i> Type IV R 1
RIV.2_fw	TACACAGCTATTCGCAATGATG	269	<i>C. carboxidivorans</i> Type IV R 2
RIV.2_rv	TCCAGCCACTTTATTGTTTAC		<i>C. carboxidivorans</i> Type IV R 2

Table A3. Cont.

Primer	Primer Sequence (5'–3')	Product Size (bp)	Anneals
RIV.3_fw	ATGGACTAGAGGCGGATATG	220	<i>C. carboxidivorans</i> Type IV R 3
RIV.3_rv	GCTGCTTCTCCAAGTACTG		<i>C. carboxidivorans</i> Type IV R 3
MI.1_fw	AACCCATGTGCTGAGGATAAG	243	<i>C. carboxidivorans</i> Type I M1
MI.1_rv	GTCATGGAGGCTATTCTAACC		<i>C. carboxidivorans</i> Type I M1
MI.2_fw	TGGCTCTTATGAATGCTATGC	296	<i>C. carboxidivorans</i> Type I M2
MI.2_rv	TATCTTTGTTCCGTCACCTTCC		<i>C. carboxidivorans</i> Type I M2
MI.3_fw	AGTAGTTACGAGCGGTGTAG	278	<i>C. carboxidivorans</i> Type I M3
MI.3_rv	GTGGATGGTCAATCCCTTTC		<i>C. carboxidivorans</i> Type I M3
MII.1_fw	TGCCTATGAAAGCACATGAAG	226	<i>C. carboxidivorans</i> Type II M1
MII.1_rv	TGTGTTTGGTGCAGAGGATAAG		<i>C. carboxidivorans</i> Type II M1
MIII.1_fw	GCTGCAGGGTATGAAAGTTG	165	<i>C. carboxidivorans</i> Type III M1
MIII.1_rv	AAGAGGCGATGGTCGTTTAG		<i>C. carboxidivorans</i> Type III M1
MIII.2_fw	GCAGCAGTACCAATTCTCAATC	271	<i>C. carboxidivorans</i> Type III M2
MIII.2_rv	AAAGTAACTCCTCCCTCAGAAG		<i>C. carboxidivorans</i> Type III M2

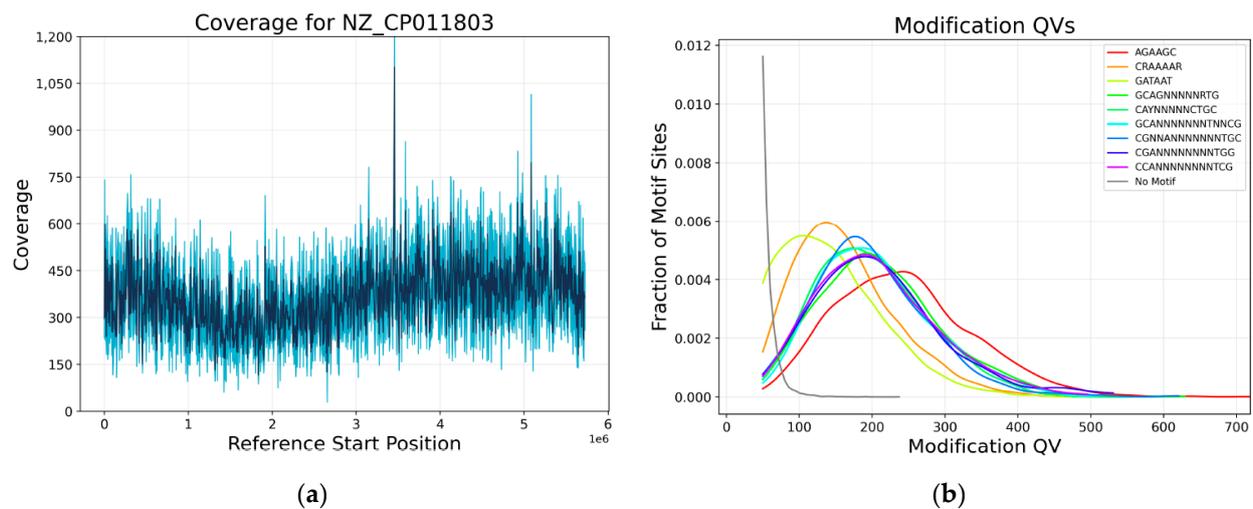


**Figure A1.** Verification of the transcription of predicted RM genes in *C. carboxidivorans*. The predicted RM genes are shown in Table 1. Total RNA was extracted from exponentially growing *C. carboxidivorans* cells. After two DNase treatments, gene-specific primers were used to amplify all RM-related genes in the *C. carboxidivorans* genome. Lanes: L—GeneRuler Low Range DNA Ladder, R—restriction enzyme (subunit); M—methyltransferase; RM—fused restriction enzyme and methyltransferase. Type indicates the designation of the RM system as Type I, II, III, or IV. Number refers to the short name used in Table 1. For example, the first sample lane on the left is a PCR product for the primers specific to Type I restriction enzyme 1.

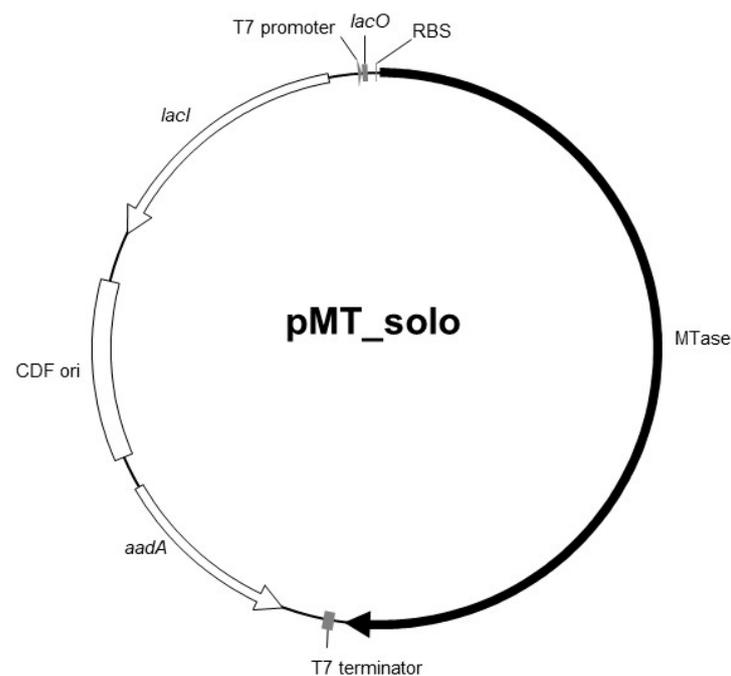
**Table A4.** Methylome analysis of *C. carboxidivorans* P7 and *E. coli* strains expressing corresponding MTases. Methylated bases are underlined and in bold (**A**). Gray rows indicate complementary motifs, where the same enzyme acts on complementary DNA strands. Genomic DNA was isolated from cells at the late exponential to early stationary phase and was analyzed using PacBio sequencing and bisulfite sequencing.

Methylation Motif	Type	Name	<i>C. carboxidivorans</i>				<i>E. coli</i>					
			% of Motifs Detected	# of Motifs Detected	# of Motifs in Genome	Mean QV	Mean Coverage	% of Motifs Detected	# of Motifs Detected	# of Motifs in Genome	Mean QV	Mean Coverage
CAYNNNNNCTGC <sup>a</sup>	I.1	M.CcaP7I	99.5	1047	1052	201.2	177.0	32.0 <sup>a</sup>	91 <sup>a</sup>	284 <sup>a</sup>	62.1 <sup>a</sup>	209.6 <sup>a</sup>
GCAGNNNNNRTG <sup>b</sup>			99.5	1047	1052	212.9	177.0	20.6 <sup>b</sup>	219 <sup>b</sup>	1060 <sup>b</sup>	59.4 <sup>b</sup>	210.1 <sup>b</sup>
CCANNNNNNNTCG	I.2	M.CcaP7II	99.1	241	243	207.1	182.1	99.6	3126	3137	188.0	142.4
CGANNNNNNNTGG			98.7	240	243	211.3	182.9	98.7	3097	3137	177.6	143.1
GCANNNNNNNTNCG	I.3	M.CcaP7III	99.5	451	453	208.0	183.4	99.3	4075	4100	146.2	108.3
CGNNANNNNNNNTGC			99.5	451	453	201.3	183.3	99.4	4078	4100	144.9	108.0
CTSAG *	II.1	M.CcaP7ORF 16660P	-	-	-	-	-	-	-	-	-	-
GATAAT	II RM.1	CcaP7IV	84.8	6251	7367	152.9	179.7	98.2	3609	3672	248.7	211.1
CRAAAAR	II RM.2	CcaP7ORF 7120P	97.3	6321	6490	166.7	177.8	99.3	4368	4395	247.5	257.0
AGAAGC **	III.1	M.CcaP7ORF 17830P	99.6	4629	4647	245.0	180.4	-	-	-	-	-
GAAAT ***	III.2	M.CcaP7ORF 8705P	-	-	-	-	-	99.2	12,981	13,076	167.1	133.6

%—percent. #—number. \* The predicted motif CTSAG for MTaseII.1 (Table 1) was not detected in *C. carboxidivorans* or in the *E. coli* expression strains, but the corresponding native gene was expressed in *C. carboxidivorans* (Figure A1). \*\* The motif AGAAGC was detected in *C. carboxidivorans* but not in the *E. coli* expression strains. Because all other MTases were accounted for, this motif is probably recognized by MTase III.1. \*\*\* The motif GAAAT was detected in *E. coli* but not in *C. carboxidivorans*. In the *E. coli* strain expressing M.CcaP7I MTase and its corresponding specificity unit, the exact motifs could not be found, but very similar motifs (<sup>a</sup> tCAYbNNNNCTGC and <sup>b</sup> GCAGNNNNNRTGnnnh, differing bases are in lower case) were detected at a lower modification threshold (Qmod) score of 25.



**Figure A2.** Coverage display and modification quality values (modQVs) from *C. carboxidivorans* P7 PacBio sequencing data. (a) PacBio sequencing depth obtained by read mapping onto the *C. carboxidivorans* reference genome NZ\_CP011803 (b) Overview of the modification QVs of the different motif sites detected.



**Figure A3.** Representative plasmid map of MTase expression vectors for the investigation of *C. carboxidivorans* MTase recognition motifs. The plasmid is based on pCDFDuet-1, with a lactose/IPTG-inducible T7 promoter. The second open reading frame, including its promoter and ribosome binding site (RBS), was removed from this construct. The black arrow indicates the location of the MTase gene. Abbreviations: *lacI*—lactose inhibitor, controlling the activity of the T7 promoter (by binding the *lacO* sequence); *lacO*—lac operator; CDF ori—CloDF13 origin of replication; *aadA* (*smR*)—streptomycin/spectinomycin resistance gene.

## References

1. Phillips, J.R.; Clausen, E.C.; Gaddy, J.L. Synthesis Gas as Substrate for the Biological Production of Fuels and Chemicals. *Appl. Biochem. Biotechnol.* **1994**, *45/46*, 145–157. [\[CrossRef\]](#)
2. Köpke, M.; Held, C.; Hujer, S.; Liesegang, H.; Wiezer, A.; Wollherr, A.; Ehrenreich, A.; Liebl, W.; Gottschalk, G.; Durre, P. *Clostridium ljungdahlii* represents a microbial production platform based on syngas. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 13087–13092. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Liew, M.F.; Köpke, M.; Simpson, D.S. Gas Fermentation for Commercial Biofuels Production. In *Liquid, Gaseous and Solid Biofuels*; Fang, Z., Ed.; IntechOpen: London, UK, 2013. [\[CrossRef\]](#)
4. Sun, X.; Atiyeh, H.K.; Huhnke, R.L.; Tanner, R.S. Syngas fermentation process development for production of biofuels and chemicals: A review. *Bioresour. Technol. Rep.* **2019**, *7*, 100279. [\[CrossRef\]](#)
5. Fackler, N.; Heijstra, B.D.; Rasor, B.J.; Brown, H.; Martin, J.; Ni, Z.; Shebek, K.M.; Rosin, R.R.; Simpson, S.D.; Tyo, K.E.; et al. Stepping on the Gas to a Circular Economy: Accelerating Development of Carbon-Negative Chemical Production from Gas Fermentation. *Annu. Rev. Chem. Biomol. Eng.* **2021**, *12*, 439–470. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Shen, S.; Gu, Y.; Chai, C.; Jiang, W.; Zhuang, Y.; Wang, Y. Enhanced alcohol titre and ratio in carbon monoxide-rich off-gas fermentation of *Clostridium carboxidivorans* through combination of trace metals optimization with variable-temperature cultivation. *Bioresour. Technol.* **2017**, *239*, 236–243. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Kottenhahn, P.; Philipps, G.; Jennewein, S. Hexanol biosynthesis from syngas by *Clostridium carboxidivorans* P7—Product toxicity, temperature dependence and in situ extraction. *Heliyon* **2021**, *7*, e07732. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Oh, H.J.; Gong, G.; Ahn, J.H.; Ko, J.K.; Lee, S.-M.; Um, Y. Effective hexanol production from carbon monoxide using extractive fermentation with *Clostridium carboxidivorans* P7. *Bioresour. Technol.* **2023**, *367*, 128201. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Cheng, C.; Li, W.; Lin, M.; Yang, S.T. Metabolic engineering of *Clostridium carboxidivorans* for enhanced ethanol and butanol production from syngas and glucose. *Bioresour. Technol.* **2019**, *284*, 415–423. [\[CrossRef\]](#)
10. Raleigh, E.A.; Murray, N.E.; Revel, H.; Blumenthal, R.M.; Westaway, D.; Reith, A.D.; Rigby, P.W.; Elhai, J.; Hanahan, D. McrA and McrB restriction phenotypes of some *E. coli* strains and implications for gene cloning. *Nucleic Acids Res.* **1988**, *16*, 1563–1575. [\[CrossRef\]](#)
11. Pyne, M.E.; Moo-Young, M.; Chung, D.A.; Chou, C.P. Development of an electrotransformation protocol for genetic manipulation of *Clostridium pasteurianum*. *Biotechnol. Biofuels* **2013**, *6*, 50. [\[CrossRef\]](#)
12. Pyne, M.E.; Bruder, M.; Moo-Young, M.; Chung, D.A.; Chou, C.P. Technical guide for genetic advancement of underdeveloped and intractable *Clostridium*. *Biotechnol. Adv.* **2014**, *32*, 623–641. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Huang, C.-N.; Liebl, W.; Ehrenreich, A. Restriction-deficient mutants and marker-less genomic modification for metabolic engineering of the solvent producer *Clostridium saccharobutylicum*. *Biotechnol. Biofuels* **2018**, *11*, 264. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Woods, C.; Humphreys, C.M.; Rodrigues, R.M.; Ingle, P.; Rowe, P.; Henstra, A.M.; Köpke, M.; Simpson, S.D.; Winzer, K.; Minton, N.P. A Novel Conjugal Donor Strain for Improved DNA transfer into *Clostridium* spp. *Anaerobe* **2019**, *59*, 184–191. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Bernheim, A.; Sorek, R. The pan-immune system of bacteria: Antiviral defence as a community resource. *Nat. Rev. Microbiol.* **2020**, *18*, 113–119. [\[CrossRef\]](#)
16. Seong, H.J.; Han, S.-W.; Sul, W.J. Prokaryotic DNA methylation and its functional roles. *J. Microbiol.* **2021**, *59*, 242–248. [\[CrossRef\]](#)
17. Suzuki, H. *Host-Mimicking Strategies in DNA Methylation for Improved Bacterial Transformation*; INTECH Open Access Publisher: London, UK, 2012; ISBN 9535108816. [\[CrossRef\]](#)
18. Murray, N.E. Type I restriction systems: Sophisticated molecular machines (a legacy of Bertani and Weigle). *Microbiol. Mol. Biol. Rev.* **2000**, *64*, 412–434. [\[CrossRef\]](#)
19. Loenen, W.A.; Dryden, D.T.; Raleigh, E.A.; Wilson, G.G. Type I restriction enzymes and their relatives. *Nucleic Acids Res.* **2014**, *42*, 20–44. [\[CrossRef\]](#)
20. Pingoud, A.; Wilson, G.G.; Wende, W. Type II restriction endonucleases—a historical perspective and more. *Nucleic Acids Res.* **2014**, *42*, 7489–7527. [\[CrossRef\]](#)
21. Wilson, G.G. Organization of restriction-modification systems. *Nucleic Acids Res.* **1991**, *19*, 2539–2566. [\[CrossRef\]](#)
22. Wilson, G.G.; Murray, N.E. Restriction and modification systems. *Annu. Rev. Genet.* **1991**, *25*, 585–627. [\[CrossRef\]](#)
23. Loenen, W.A.M.; Raleigh, E.A. The other face of restriction: Modification-dependent enzymes. *Nucleic Acids Res.* **2014**, *42*, 56–69. [\[CrossRef\]](#)
24. Kirk, J.A.; Fagan, R.P. Heat shock increases conjugation efficiency in *Clostridium difficile*. *Anaerobe* **2016**, *42*, 1–5. [\[CrossRef\]](#)
25. Oliveira, P.H.; Ribis, J.W.; Garrett, E.M.; Trzilova, D.; Kim, A.; Sekulovic, O.; Mead, E.A.; Pak, T.; Zhu, S.; Deikus, G.; et al. Epigenomic characterization of *Clostridioides difficile* finds a conserved DNA methyltransferase that mediates sporulation and pathogenesis. *Nat. Microbiol.* **2020**, *5*, 166–180. [\[CrossRef\]](#)
26. Johnston, C.D.; Cotton, S.L.; Rittling, S.R.; Starr, J.R.; Borisy, G.G.; Dewhurst, F.E.; Lemon, K.P. Systematic evasion of the restriction-modification barrier in bacteria. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 11454–11459. [\[CrossRef\]](#)
27. Riley, L.A.; Ji, L.; Schmitz, R.J.; Westpheling, J.; Guss, A.M. Rational development of transformation in *Clostridium thermocellum* ATCC 27405 via complete methylome analysis and evasion of native restriction-modification systems. *J. Ind. Microbiol. Biotechnol.* **2019**, *46*, 1435–1443. [\[CrossRef\]](#)

28. Eid, J.; Fehr, A.; Gray, J.; Luong, K.; Lyle, J.; Otto, G.; Peluso, P.; Rank, D.; Baybayan, P.; Bettman, B.; et al. Real-time DNA sequencing from single polymerase molecules. *Science* **2009**, *323*, 133–138. [[CrossRef](#)]
29. Flusberg, B.A.; Webster, D.R.; Lee, J.H.; Travers, K.J.; Olivares, E.C.; Clark, T.A.; Korlach, J.; Turner, S.W. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* **2010**, *7*, 461–465. [[CrossRef](#)]
30. Feng, Z.; Fang, G.; Korlach, J.; Clark, T.; Luong, K.; Zhang, X.; Wong, W.; Schadt, E. Detecting DNA Modifications from SMRT Sequencing Data by Modeling Sequence Context Dependence of Polymerase Kinetic. *PLoS Comput. Biol.* **2013**, *9*, e1002935. [[CrossRef](#)]
31. Roberts, R.J.; Vincze, T.; Posfai, J.; Macelis, D. REBASE—A database for DNA restriction and modification: Enzymes, genes and genomes. *Nucleic Acids Res.* **2015**, *43*, D298–D299. [[CrossRef](#)] [[PubMed](#)]
32. Li, N.; Yang, J.; Chai, C.; Yang, S.; Jiang, W.; Gu, Y. Complete genome sequence of *Clostridium carboxidivorans* P7(T), a syngas-fermenting bacterium capable of producing long-chain alcohols. *J. Biotechnol.* **2015**, *211*, 44–45. [[CrossRef](#)] [[PubMed](#)]
33. Biosciences, P. Detecting DNA base modifications using single molecule, real-time sequencing. *White Pap. Base Modif.* 2015. Available online: [https://www.pacb.com/wp-content/uploads/2015/09/WP\\_Detecting\\_DNA\\_Base\\_Modifications\\_Using\\_SMRT\\_Sequencing.pdf](https://www.pacb.com/wp-content/uploads/2015/09/WP_Detecting_DNA_Base_Modifications_Using_SMRT_Sequencing.pdf) (accessed on 24 February 2023).
34. Dong, M.-J.; Luo, H.; Gao, F. Ori-Finder 2022: A Comprehensive Web Server for Prediction and Analysis of Bacterial Replication Origins. *Genom. Proteom. Bioinform.* **2022**, *20*, 1207–1213. [[CrossRef](#)]
35. McGlinchey, A.S.; Zepeda-Rivera, M.A.; Stepanovica, M.; Baryames, A.A.; Jones, D.S.; LaCourse, K.D.; Bullman, S.; Johnston, C.D. Complete Genome Sequence of *Clostridium cadaveris* IFB3C5, Isolated from a Human Colonic Adenocarcinoma. *Microbiol. Resour. Announc.* **2022**, *11*, e0113521. [[CrossRef](#)]
36. Philipps, G.; de Vries, S.; Jennewein, S. Development of a metabolic pathway transfer and genomic integration system for the syngas-fermenting bacterium *Clostridium ljungdahlii*. *Biotechnol. Biofuels* **2019**, *12*, 112. [[CrossRef](#)]
37. Pregon, G.; Minton, N.P.; Soucaille, P. Genome Sequence of *Eubacterium limosum* B2 and Evolution for Growth on a Mineral Medium with Methanol and CO<sub>2</sub> as Sole Carbon Sources. *Microorganisms* **2022**, *10*, 1790. [[CrossRef](#)]
38. Utturkar, S.M.; Klingeman, D.M.; Bruno-Barcena, J.M.; Chinn, M.S.; Grunden, A.M.; Köpke, M.; Brown, S.D. Sequence data for *Clostridium autoethanogenum* using three generations of sequencing technologies. *Sci. Data* **2015**, *2*, 150014. [[CrossRef](#)]
39. REBASE. Available online: <http://rebase.neb.com> (accessed on 24 February 2023).
40. Jurenaite-Urbanaviciene, S.; Kazlauskienė, R.; Urbelyte, V.; Maneliene, Z.; Petruyte, M.; Lubys, A.; Janulaitis, A. Characterization of BseMII, a new type IV restriction–modification system, which recognizes the pentanucleotide sequence 5'-CTCAG(N)10/8↓. *Nucleic Acids Res.* **2001**, *29*, 895–903. [[CrossRef](#)]
41. Minton, N.P.; Ehsaan, M.; Humphreys, C.M.; Little, G.T.; Baker, J.; Henstra, A.M.; Liew, F.; Kelly, M.L.; Sheng, L.; Schwarz, K.; et al. A roadmap for gene system development in *Clostridium*. *Anaerobe* **2016**, *41*, 104–112. [[CrossRef](#)] [[PubMed](#)]
42. Yang, X.; Xu, M.; Yang, S.-T. Restriction modification system analysis and development of *in vivo* methylation for the transformation of *Clostridium cellulovorans*. *Appl. Microbiol. Biotechnol.* **2016**, *100*, 2289–2299. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.