Data in Brief 53 (2024) 110237

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Dataset on CRISPR/Cas9 system targeting hydrogenase genes in *Rhodobacter johrii* MAY2 isolate

Leo Agustin F. Barcelo[a], Nacita B. Lantican[b], Aprill P. Manalang[c], Jey-R S. Ventura[a,*]

[a] *Biomaterials and Environmental Engineering Laboratory, Department of Engineering Science, College of Engineering and Agro-Industrial Technology, University of the Philippines Los Baños, College, Laguna 4031 Philippines*
[b] *Microbiology Division, Institute of Biological Sciences, College of Arts and Sciences, University of the Philippines Los Baños, College, Laguna 4031 Philippines*
[c] *Genetics and Molecular Biology Division, Institute of Biological Sciences, College of Arts and Sciences, University of the Philippines Los Baños, College, Laguna 4031 Philippines*

ARTICLE INFO

Article history:
Received 28 December 2023
Accepted 19 February 2024
Available online 28 February 2024

Dataset link: Gene Targets and Pre-validation results for CRISPR Cas-9 editng of hydrogenase genes in Rhodobacter johrii isolate MAY2 (Original data)

Keywords:
Biohydrogen
Gene-editing
Guide RNA
Purple non-sulfur bacteria

ABSTRACT

This dataset contains the gene sequences of the small and large sub-unit of the hydrogenase enzyme obtained from the annotated genome of *Rhodobacter johrii* MAY2. The whole genome sequence of the isolate was performed using SEED genome viewer on the Rapid Annotation using the Subsystem Technology (RAST) platform. Concurrently, guide RNA sequences and primers were meticulously crafted using the CHOPCHOP v.3.0 web tool, specifically designed for the precise editing and amplification of the target genes. The primers were optimized via gradient PCR to determine appropriate amplification conditions. Furthermore, the guide RNA was tested via in-vitro cleavage assay, gauging its efficacy in cleaving the intended target genes. The dataset, including the optimization and the cleavage assay, was deposited in Mendeley Data with DOI no: 10.17632/rcx3mcssnx.2.

* Corresponding author.
E-mail address: jsventura@up.edu.ph (J.S. Ventura).
Social media: @leobacillus (L.A.F. Barcelo), @jeyr1123 (J.S. Ventura)

https://doi.org/10.1016/j.dib.2024.110237

Specifications Table

Subject	Biological Sciences
Specific subject area	Gene-editing
Data format	Raw, Analyzed, Filtered
Type of data	Table, Image, Chart, Graph, Figure
Data collection	For the design of guide RNA, the initial step involved obtaining the gene sequences of the target genes — the small sub-unit and large sub-unit of the uptake hydrogenase enzyme. These sequences were derived from a de novo whole genome sequence, accomplished through next-generation sequencing (NGS) on the Illumina® HiSeq 2500 platform. Subsequently, the genome was thoroughly annotated using Rapid Annotation using Subsystem Technology (RAST). The annotated sequences were then inputted into the CHOPCHOP web tool (https://chopchop.cbu.uib.no/), which, in turn, yielded CRISPR targets accompanied by corresponding guide RNA (gRNA) and primer designs. The sequences with the highest predicted efficiency were selected for synthesis. The primer designs were tested and optimized via gradient PCR and then gRNA designs were pre-validated for cleavage ability using in vitro cleavage assay.
Data source location	Biomaterials and Engineering Laboratory, Department of Engineering Science, College of Engineering and Agro-Industrial Technology, University of the Philippines Los Banos, Laguna, Philippines
Data accessibility	Repository name: Mendeley data Data identification number: 10.17632/rcx3mcssnx.2 Direct URL to data: https://data.mendeley.com/datasets/rcx3mcssnx/2 Instructions for accessing these data: The data can be accessed and downloaded by creating an account of logging-in in Mendeley Data (https://data.mendeley.com/) and entering the DOI no. 10.17632/rcx3mcssnx.2 in the “Find Research Data” field or via accessing the hyperlink above.

1. Value of the Data

- In the Philippines, the exploration of purple non-sulfur bacteria (PNSB) remains limited, with existing studies predominantly confined to isolation and characterization. The present dataset, focused on PNSB genes, aims to explore the potential applications of PNSB.
- CRISPR/Cas9 systems are gaining global popularity, but local studies are limited. Most of these local studies are focused on editing eukaryotic systems. This dataset describes a CRISPR/Cas9 intended for a prokaryotic system. The methods, gene sequences, design of primers, and guide RNAs are made available to serve as a reference for researchers conducting experiments on CRISPR/Cas9 with intended application in PNSB or other types of bacteria.
- Hydrogenase is a general enzyme found in hydrogen-producing microorganisms. The design of primers and gRNA could be utilized for other PNSB species or other hydrogen-producing microorganisms.

2. Background

This dataset originated from a project aimed at generating hydrogenase-deficient purple non-sulfur bacteria (PNSB). The approach was to design guide RNAs targeting the genes for the small and large sub-units of the enzyme and attempt to directly introduce the gRNA and Cas9 protein in the PNSB cells. The method employed was markerless, relying on the high transformation

efficiency of the CRISPR/Cas system, leading to the selection of potential transformants through a random process. The genome of each of the chosen transformants was extracted and the genes of interest were amplified using the designed primers. The amplicons were then sequenced to confirm if there were mutations in the target genes. Also, the hydrogen production of the chosen transformants was assessed to correlate if the confirmed mutation resulted in a change in its hydrogen production capability.

3. Data Description

The dataset is available on Mendeley data (https://data.mendeley.com/datasets/rcx3mcssnx/2). The data contains a folder consisting of three files, two of which are fasta files containing the sequences for the small (MAY2_RAST_HyaA small subunit.fasta) and the large (MAY2_RAST_HyaB large subunit.fasta) sub-unit of the hydrogenase enzyme. These sequences were obtained via the annotated whole genome sequence data of *Rhodobacter johrii* MAY2 isolate in RAST.

The other file is a PDF containing the two sequences of the guide RNA and four primer pairs designed for the target genes. The file provides comprehensive details on the results of amplification experiments conducted with the primer pairs and in-vitro cleavage assays using the guide RNAs. These experiments were performed on *Rhodobacter johrii* MAY2, a purple non-sulfur bacteria isolated from Laguna Lake in Los Banos, Laguna, Philippines.

4. Experimental Design, Materials and Methods

4.1. Primers and guide RNA design

Using the SeedViewer genome browser tool of RAST v2.0, the sequences coding for genes linked to the uptake hydrogenase enzyme in *R. johrii* MAY2 were obtained. This curated assembly unveiled that the uptake hydrogenase comprises two sub-units, namely the small (coded by *hyaA*, also known as *hupS*) and the large (coded by *hyaB*, also known as *hupL*) sub-units. Four primer pairs and two sequence guide RNAs (Table 1) targeting *hyaA* and large *hyaB* sub-units of the uptake hydrogenase enzyme in MAY2 were designed using the CHOPCHOP v.3.0 web tool (https://chopchop.cbu.uib.no/). The selection of the guide RNAs was based on the predicted transformation efficiency provided by the design tool. The coding sequence of the two genes was uploaded to the web tool and it was set to find target sites for CRISPR/Cas9 gene knock-out application. After multiple tool runs, guide RNA target sequences were suggested, and the designs with the highest predicted efficiency, along with two primer sets flanking each guide RNA, were chosen.

To assess potential off-target sites of the designed guide RNAs against the MAY2 genome, Geneious Prime software free trial version (https://www.geneious.com) was employed. The

Table 1
Guide RNAs and primer pairs designed targeting the small sub-unit and large sub-unit of the hydrogenase gene in *R. johrii* MAY2.

Gene Target	Guide RNA	Primer pairs
hyaA (small sub-unit)	CAGCATGTAGGTGATGACG CCGG (*gRNA-hA*)	1. GACAAGCCGATCATCAAGGT (*hA1-F*) CCATCTTGTAGAGGCAGTAGCC (*hA1-R*) 2. GACAAGCCGATCATCAAGGT (*hA2-F*) CATCTTGTAGAGGCAGTAGCCC (*hA2-R*)
hyaB (large sub-unit)	GAACATCCTCGAGCTCGAC GAGG (*gRNA-hB*)	1. ATCACTCCTGGTATGCCTATGG (*hB1-F*) CTCGACCTGGTTCTTGATGTC (*hB1-R*) 2. ATTCGTCGATCACTCCTGGTAT (*hB2-F*) CTCGACCTGGTTCTTGATGTC (*hB2-R*)

whole genome assembly was imported into the software, and the “Find CRISPR site” tool was utilized to identify sites matching the guide RNAs. The sequence of the guide RNA and PAM region was entered, using the Doench et al. (2016) activity scoring with a set maximum mismatch allowance of “3” against off-targets and "0" for mismatches leading to indels.

4.2. Amplification of the hypA and hypB gene

For the amplification of target genes (*hyaA* and *hyaB*), gradient polymerase reaction was performed using the extracted genomic DNA of MAY2 as a template. The Thermo Fisher Scientific Tm Calculator was utilized to determine optimal theoretical annealing temperatures ranging from 51 to 60 °C, with the selected annealing temperatures for the gradient PCR being 51.5 °C, 54.0 °C, 57.2 °C, 59.2 °C, and 60.0 °C.

The PCR mixture was prepared as follows: 1X PCR buffer, 0.2 mM DNTPs, 1.5 mM $MgCl_2$, 0.2 µL of primers (left and right), 2 units of Taq polymerase and approximately 500 ng of genomic DNA of MAY2 and diluted up to 30 µL using nuclease-free water. Duplicates of the PCR mixture were prepared for each primer pair and each annealing temperature. The PCR mixtures were then quick spun for 10 s and loaded to the SeeAmp™ thermocycler (SeeGene Inc., South Korea). The PCR profile was set as follows: initial denaturation at 94 °C (2 min), then 30 cycles of denaturation at 94 °C (1 min), annealing at 51.5 °C, 54.0 °C, 57.2 °C, 59.2 °C, and 60.0 °C (1 min), extension 72 °C (50 s), and final extension at 72 °C (45 s).

To determine if the gradient PCR protocol produced the target PCR products, a gel electrophoresis set-up was run using 1.0% agarose. The agarose was prepared by heating 0.1 g of agarose for every 10 mL of Tris-acetate-EDTA buffer used. Gel red was added at 1.0 µl per 10 mL of the heated solution. The gel was then allowed to cool and solidify in a gel cast. Five microliters of the PCR product were added with 3.0 µL 6× loading dye and loaded in the sample wells. Five microliters of Vivantis 1kb DNA ladder were also loaded in one of the wells to serve as molecular weight markers. The electrophoresis was run for 25 min at 100 volts using Mupid® exU (Mupid Co., Japan), after which the gel was then visualized using auto imaging in Azure Biosystems c200 system (Azure Biosystems Inc., USA). set to detect ethidium bromide dye. Furthermore, the purity and amount of the amplicons were determined using the Thermo Fisher Scientific Multiskan™ Sky microplate spectrophotometer.

4.3. In vitro cleavage assay and checking for off-target sites

This protocol was done to pre-validate if the designed guide RNAs and Cas-9 system could recognize and cleave the target site before introducing the gene editing system in vivo [1].

The reaction mixture for the assay was prepared in a 0.2 microcentrifuge tube (PCR tube) by combining 1 µL of ∼100 ng of the target gene amplicons, 1 µL of 400 ng of the guide RNA corresponding to the target gene, 10× reaction buffer (20 mM HEPES, 100 mM NaCl, 5 mM $MgCl_2$, 0.1 mM EDTA adjusted to pH 6.5), and 3.0 µL of 0.5 µg/µL Cas9 nuclease. The reaction mixtures were then quickly spun for 10 s, after which were incubated using a thermocycler set at 37 °C for one hour. Immediately after the incubation, 4 µl of RNAse A was added to each tube and re-incubated for another 20 min using the same temperature setting.

The cleavage products were subjected to gel electrophoresis using a 1.5% agarose gel to enhance gel resolution. The framework of the process, starting from the guide RNA design to the confirmation of their cleavage ability, is presented in (Fig. 1).

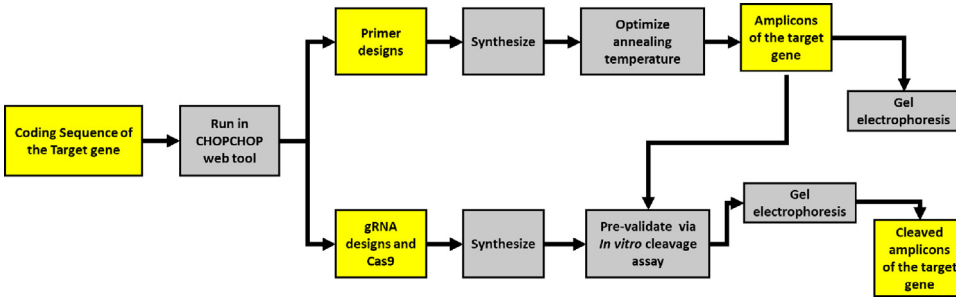

Fig. 1. Framework on the design of gRNA and primers.

Limitations

The data described here is limited to the local isolate of interest *Rhodobacter johrii* MAY2. This was isolated and characterized by the same institution which provided this dataset.

Ethics Statement

The authors have read and followed the ethical requirements for publication in Data in Brief and confirmed that the current work does not involve human subjects, animal experiments, or any data collected from social media platforms.

Data Availability

Gene Targets and Pre-validation results for CRISPR Cas-9 editng of hydrogenase genes in Rhodobacter johrii isolate MAY2 (Original data) (Mendeley Data)

CRediT Author Statement

Leo Agustin F. Barcelo: Validation, Conceptualization, Methodology, Software, Writing – original draft, Data curation; **Nacita B. Lantican:** Supervision, Resources, Methodology, Writing – review & editing; **Aprill P. Manalang:** Methodology, Software, Writing – review & editing; **Jey-R S. Ventura:** Conceptualization, Writing – review & editing, Supervision, Funding acquisition.

Acknowledgments

The authors express sincere appreciation to the National Research Council of the Philippines under the Department of Science and Technology for their generous funding support. Additionally, gratitude is extended to our collaborative partners, the Institute of Biological Sciences and the National Institute of Molecular Biology and Biotechnology (UPLB-BIOTECH), for their invaluable contributions to this endeavor.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Reference

[1] A.K. Challa, D. Stanford, A. Allen, L. Rasmussen, F.K. Amanor, S. Raju, Validation of gene editing efficiency with CRISPR-Cas9 system directly in rat zygotes using electroporation mediated delivery and embryo culture, MethodsX 8 (2021) 101419, doi:10.1016/j.mex.2021.101419.