Data Article

Genomics dataset of unidentified disclosed isolates

Bhagwan N. Rekadwad *

School of Life Sciences, Swami Ramanand Teerth Marathwada University, Nanded, Maharashtra 431606, India

ARTICLE INFO

ABSTRACT

Analysis of DNA sequences is necessary for higher hierarchical classi-fication of the organisms. It gives clues about the characteristics of organisms and their taxonomic position. This dataset is chosen to find complexities in the unidentified DNA in the disclosed patents. A total of 17 unidentified DNA sequences were thoroughly analyzed. The quick response codes were generated. AT/GC content of the DNA sequences analysis was carried out. The QR is helpful for quick iden-tification of isolates. AT/GC content is helpful for studying their stabi-lity at different temperatures. Additionally, a dataset on cleavage code and enzyme code studied under the restriction digestion study, which helpful for performing studies using short DNA sequences was reported. The dataset disclosed here is the new revelatory data for exploration of unique DNA sequences for evaluation, identification, comparison and analysis.

Specifications Table

Subject area	*Life Sciences*
More specific sub-ject area	*Microbiology, Genomics, Bioinformatics, Bacterial Systematics*
Type of data	*Table, Figures* *Through NCBI BioSample database*

* Corresponding author.
 E-mail address: rekadwad@gmail.com

How data was acquired	
Data format	*Raw and Analyzed*
Experimental factors	*Dataset obtained through bioinformatics tool*
Experimental features	*Only disclosed genome sequences were used*
Data source location	School of Life Sciences, S. R. T. M. University, Nanded, India
Data accessibility	Data available within article and via the *NCBI repository* http://www.ncbi.nlm.nih.gov/nuccore.

Value of the data

- Data provides information of the AT and GC percentage of unidentified isolates.
- This data would be valuable for qualitative and quantitative analysis newly isolated and unidentified strains.
- This data provides exact position of restriction sites to create blunt and sticky ends and gives an idea about cleavage affected by methylation.

1. Data

This paper contains data on data for QR codes, GC percentage and DNA sequence analysis of 17 unidentified strains. Genome sequences of unidentified bacterial strains which were disclosed from the patents US 6596510 and WO 9906567 were retrieved in FASTA format via NCBI nuccore database. These downloaded sequences were used to create quick response (QR) codes and digitized using ENDMEMO GC calculating and GC plotting tool. The AT and GC percentage, number of cleavage code (blunt end, 5′ and 3′ sticky ends) and number of enzyme code (cleavage affected methylation) were determined using BioLabs NEB cutter tool (NEW ENGLAND BioLabs. Inc. https://www.neb.com/).

2. Experimental design, materials and methods

A total of 17 genome sequences of disclosed unidentified bacteria (*AR360580, AR360581, AR360582, AR360583, AR360584, AR360585, AR360586, AR360587, AR360588, AR360589, AR360590, AX000218, AX000220, AX000221, AX000222, AX000224 and AX000225*) were saved in FASTA format via NCBI BioSample DNA database. DNABarID tool was used for creation of QR codes (Fig. 1). ENDMEMO GC calculating and GC plotting tool was used to determine percentage of nucleotides in the genome. Pattern of GC distribution in complete DNA sequence showed through graphical representations in Fig. 2. Upper and lower red line indicate maximum and minimum percentage of GC content distribution in complete DNA sequence, while middle blue line indicates average GC percentage [1–6]. NEB cutter tool was used analysis of DNA sequence of unidentified isolates. The number of cleavage to