Data Article

Data on graphical representation (CGR and FCGR) of bacterial and archaeal species from two Soda Lakes

Bhagwan N. Rekadwad *, Chandrahasya N. Khobragade

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

A R T I C L E   I N F O

A B S T R A C T

In this paper, we presented the datasets generated using Chose Game representation (CGR) and Choase Game Representation of Frequencies (FCGR) of bacterial and archaeal 16S rRNA sequences. The data in the form of graphical representations was yielded with the help of ENDMEMO tool. The computational representation of these data datasets is useful for studies and interpretation of microbial sequences. Based on a technique from chaotic dynamics, the method produces a picture of any gene (DNA and RNA) sequence which displays both local and global patterns. Eukaryotes and prokaryotes can be identified merely based on their generated visual representation/DNA structures.

Specifications Table

Subject area	*Biology*
More specific subject area	*Microbiology; Bioinformatics*
Type of data	*Table, figure*
How data was acquired	*Bioinformatics tools*
Data format	*Raw, analyzed*
Experimental factors	*Standard and default*

---

* Corresponding author.
  E-mail address: rekadwad@gmail.com (B.N. Rekadwad).

Experimental features	*Graphical representations*
Data source location	*School of Life Sciences, SRTM University, Nanded*
Data accessibility	*Data is available with this article*

Value of the data

- Data generated in this study permits the representation and investigation of patterns in any type of sequences which visually revealed previously unknown pattern.
- The generated graphical data by means of sequences using a new tool derived from the "chaotic dynamical systems" which allowed the depiction of frequencies of oligonucleotides in the form of images.
- Data on CGR and FCGR are the main factors explaining the variability observed among sequences. The distance between images helpful for measurement of phylogenetic proximity.

1. Data

This paper describes data on 16S rRNA sequence of bacterial and archaeal species isolated from Soda Lakes such as Sambhar Lake and Chilka Lake (India). The data generated in the form of graphical representations contains information on their oligonucleotides distribution and numbers.

2. Experimental design, materials and methods

115 bacterial and archaeal 16S rRNA sequences (both short and long) were obtained in FASTA format from NCBI repository (Table 1). These sequences of bacteria and archaea were used for graphical representations. The generated graphical representations in the form of Chaose Game Representations (Fig. 1) and Chose Game Representations of Frequencies (Fig. 2) obtained in the form of visual images [1,2]. Graphical representations of oligonucleotides in the form of CGR and FCGR pictorial representations were created using ENDMEMO tool [3,4] for studies on primary sequence organization and representation of oligonucleotides frequency in the given sequence.

Table 1
Bacterial and archaeal species isolated from Soda Lakes.

Accession number	Species/Strain
AF472595	Sambhar Salt Lake archaeon HA1
AF472596	Sambhar Salt Lake archaeon HA6
AJ889020	*Marichromatium chilcum*
EU669822	Haloalkaliphilic bacterium EMB4
FJ984520	*Marinobacter alkaliphilus* strain NBSL05
FJ984521	*Marinobacter alkaliphilus* strain NBSL06
FJ984522	*Marinobacter hydrocarbonoclasticus* strain NBSL04
FJ984523	*Halomonas* sp. NBSL08
FJ984524	*Marinobacter alkaliphilus* strain NBSL03
FJ984525	*Halomonas* sp. NBSL10
FJ984526	*Halomonas* sp. NBSL14
FJ984527	*Tsukamurella* sp. NBSL21
FJ984528	*Ochrobactrum haemophilum* strain NBSL23
FJ984529	*Bacillus horikoshii* strain NBSL26
FJ984530	*Bacillus horikoshii* strain NBSL27
FJ984531	*Micrococcus luteus* strain NBSL29