



ELSEVIER

Contents lists available at ScienceDirect

Data in Brief

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



Data Article

Digital data of quality control strains under general deposit at Microbial Culture Collection (MCC), NCCS, Pune, India: A bioinformatics approach

Bhagwan N. Rekadwad*, Chandrahasya N. Khobragade

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

ARTICLE INFO

Article history:

Received 4 April 2016

Received in revised form

16 April 2016

Accepted 19 April 2016

Available online 26 April 2016

Keywords:

Chaos Game Representation

GC content

Microbial Culture Collection

QR codes

Standard type strains

ABSTRACT

A total of 13 short DNA sequences of quality control strains (MCC 2052, MCC 2077, MCC 2078, MCC 2080, MCC 2309, MCC 2322, MCC 2408, MCC 2409, MCC 2412, MCC 2413, MCC 2415, MCC 2483 and MCC 2515) were retrieved from NCBI BioSample database and generated quick response (QR) codes for sequences. 16S rRNA was used for creation of Chaos Game representation (CGR), Chaos Game Representation of Frequencies (FCGR) and measurement of GC percentage. Digital data in the form of QR codes, CGR, FCGR and GC plot would be useful for identification, visual comparison and evaluation of newly isolated strains with quality control strains. The digital data of QR codes, CGR, FCGR and GC content all the quality control strains are made available to users through this paper. This generated digital data helps to evaluate and compare newly isolated strains, less laborious and avoid misinterpretation of newly isolated species.

© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Specifications Table

Subject area *Microbiology*

* Corresponding author.

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

<http://dx.doi.org/10.1016/j.dib.2016.04.048>

2352-3409/© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

More specific subject area	<i>Bioinformatics</i>
Type of data	<i>Table, figure and graphs</i>
How data was acquired	<i>Through NCBI BioSample database</i>
Data format	<i>Analyzed</i>
Experimental factors	<i>Bioinformatics tools were used for creation of digital information</i>
Experimental features	<i>Analysis and digitization of DNA sequences were carried out using bioinformatics approach</i>
Data source location	<i>Microbial Culture Collection, National Centre for Cell Sciences, Pune, Maharashtra, India (18° 32' 14.00" N; 73° 47' 29.79" E)</i>
Data accessibility	<i>Data available within this paper.</i>

Value of the data

- Digital information generated provides most comprehensive and analysed data on quality control strains present under general deposit facility in Microbial Culture Collection, National Centre for Cell Sciences, Pune, Maharashtra, India.
- Data would be valuable for identification, evaluation and comparison of other species with quality control strains.
- Data would be valuable for analysis and interpretation correlation between GC content and thermophilic nature of bacteria.
- These experiments with quality control strains were carried out first time by us and made available to users.

1. Data

A total of 13 short DNA sequences of quality control strains (MCC 2052, MCC 2077, MCC 2078, MCC 2080, MCC 2309, MCC 2322, MCC 2408, MCC 2409, MCC 2412, MCC 2413, MCC 2415, MCC 2483 and

Table 1

Details of digitised short DNA sequences of quality control strains under general deposit at MCC, NCCS, Pune, Maharashtra, India.

MCC accession no.	Name	NCBI BioSample ID (Accession no.)	Maximum GC content	Average GC content
2052	<i>A. hydrophila</i> (ATCC 7966 T)	X74677	67.5	55.3
2077	<i>A. calcoaceticus</i>	X81661	67.5	52.8
2078	<i>Citrobacter freundii</i> 16 S rRNA gene (strain DSM 30039)	AJ233408	65	52.2
2080	<i>Pseudomonas aeruginosa</i> strain ATCC 27853	AY268175	65	54.3
2309	<i>A. caviae</i> (ATCC 15468 T)	X74674	67.5	55.1
2322	<i>Aeromonas caviae</i>	X60409	65	54.9
2408	<i>Staphylococcus aureus</i> subsp. <i>aureus</i>	AB594753	62.5	51.2
2409	<i>Citrobacter freundii</i>	AJ233408	65	55.2
2412	<i>Escherichia coli</i> (strain ATCC 25922)	DQ360844	67.5	54.7
2413	<i>Escherichia coli</i>	AM980865	67.5	55
2415	<i>Streptococcus pneumoniae</i>	AY281082	65	52.7
2483	<i>Stenotrophomonas maltophilia</i>	AB008509	70.4	55.3
2515	<i>M. phlei</i>	M29566	70	57.5

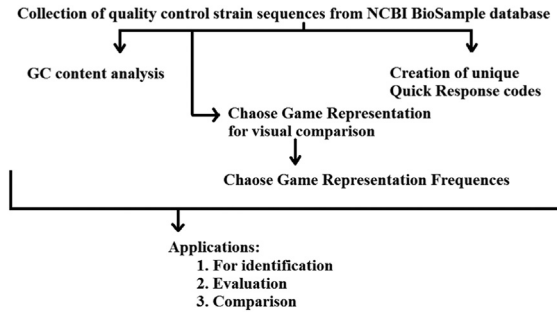


Fig. 1. Flow diagram for digitisation of quality control strains and their application.

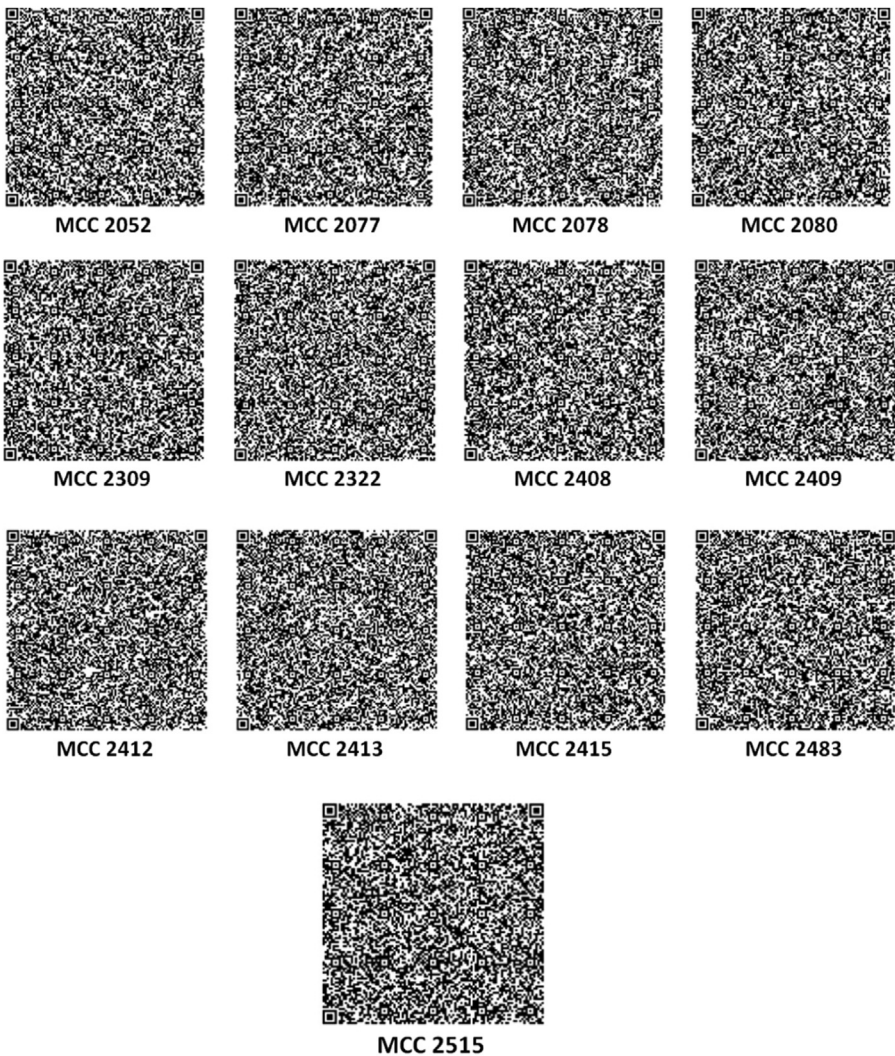


Fig. 2. QR codes of quality control strains.

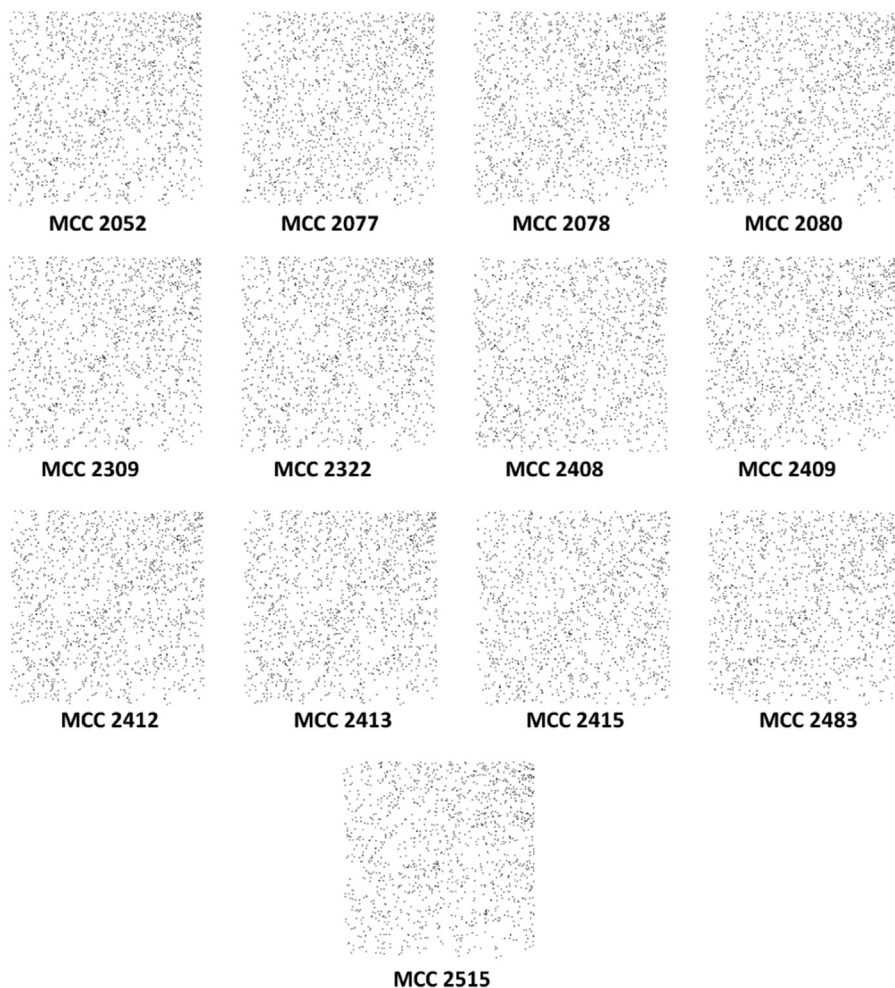


Fig. 3. Chaose Game Representation of quality control strains.

MCC 2515) were obtained through NCBI's BioSample database. QR codes for these species were generated through DNABarID tool. Chaose Game Representation (CGR) and CGR of Frequencies were graphically represented using BioPHP tool [1]. Data on GC content in percentage were generated using ENDMEMO GC calculating tool (Table 1 and Figs. 1–5) [2]. See also NCBI repository <http://www.ncbi.nlm.nih.gov/nuccore> and MCC – list of bacteria under general deposit <http://www.nccs.res.in/mcc/Bacteria.html>.

2. Experimental design, materials and methods

Short DNA sequences in FASTA format were retrieved from NCBI BioSample database. The data in the form of QR codes, CGR, FCGR and GC plot was generated using DNABarID, BioPHP and ENDMEMO GC calculating and plotting tools respectively. DNABarID tool was used for creation of unique QR codes. BioPHP tool was used for graphical representation of nucleotide in quality control strains. ENDMEMO tool was used for calculation of GC content and plotting GC distribution graph.

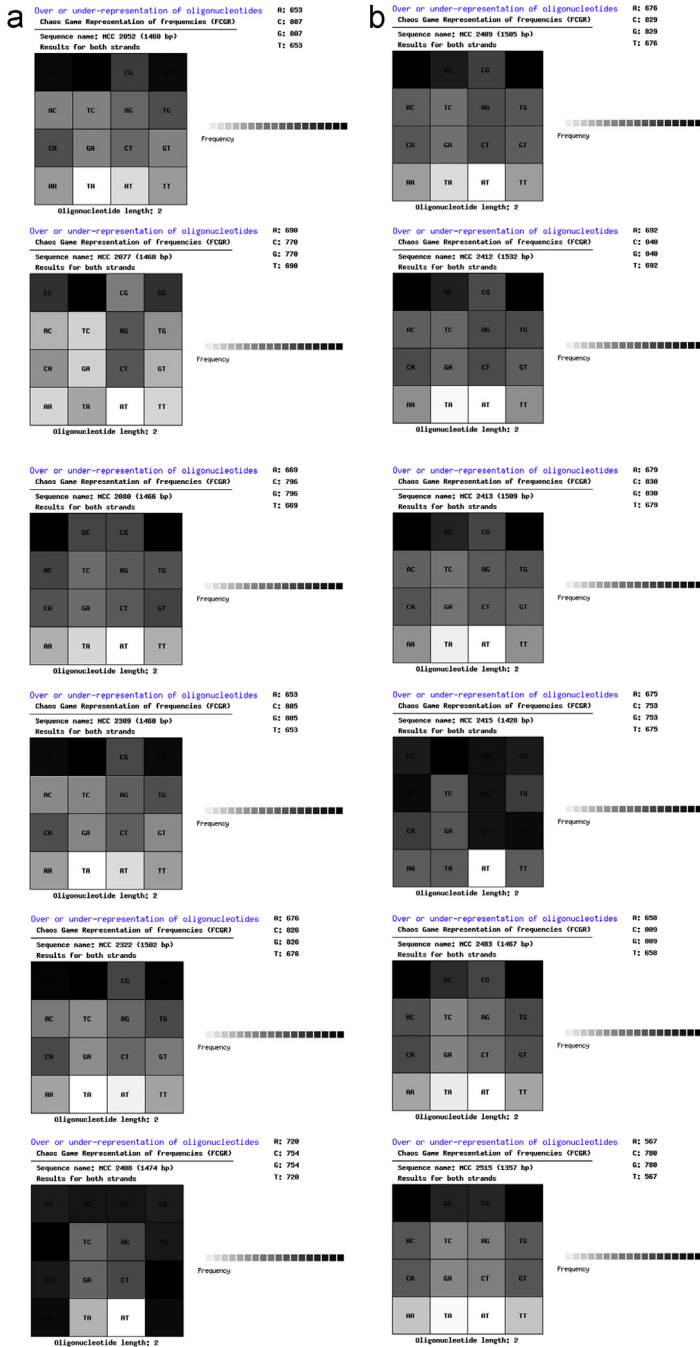


Fig. 4. Chaos Game Representation of Frequencies: quality control strains.

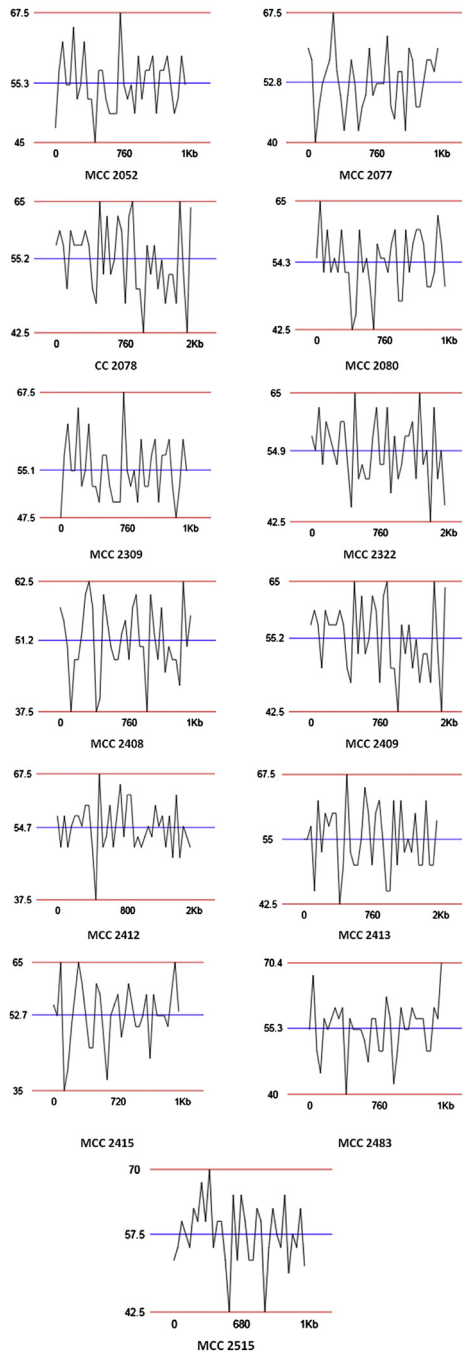


Fig. 5. GC percentage in short DNA sequences quality control strains under general deposit at MCC, NCCS, Pune, Maharashtra, India.

3. Interpretation of generated digital data

The generated QR codes, CGR, FCGR and GC plot using 16 S rRNA sequences of quality control strains would be useful for identification, evaluation and comparison of newly isolated species. The data displayed in QR codes is exactly similar to the data available on NCBI web portal. The number of dots appeared in a CGR image is directly proportional to number of base pairs in a field. Like CGR, FCGR images also represent the concentration of base pairs in a quantitative manner. FCGR comprises twelve squares which mimics Ramanujan's Magic Square but different than it. Dark colour of square indicates that higher frequency of base in 16 S rRNA. Frequency tape given in each figure is helpful to understand pattern of distribution of base pairs in DNA. Darkness of the square is directly proportional to the number of base pairs in the square. GC plot shows maximum (upper; red), average (middle; blue) and minimum (lower; red) GC percent lines

- I) Upper red line indicate maximum % of GC.
- II) Middle blue line indicate average % of GC.
- III) Lower red line indicate minimum % of GC in short DNA sequence.

Acknowledgements

BNR is thankful to University Grants Commission (UGC), New Delhi, India for financial support (F. No. PDFSS-2013-14-ST-MAH-4350_website).

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.dib.2016.04.048>.

References

- [1] B.N. Rekadwad, C.N. Khobragade, *Data in Brief*, 6 (2016) 53–57.
- [2] B.N. Rekadwad, C.N. Khobragade, *Data in Brief*, (2016), in press, <http://dx.doi.org/10.1016/j.dib.2016.04.038>.