Data Article

Assembly of genomic reads of elite indica rice cultivar onto 2101 reference bacterial genomes for identification of co-sequenced endophytic bacteria

Latha Battu, Mettu Madhavi Reddy, Burragoni Sravanthi Goud, Kayalvili Ulaganathan, Kandasamy Ulaganathan *

Centre for Plant Molecular Biology, Osmania University, Hyderabad 500007, Telangana, India

Article history:
Received 13 February 2017
Accepted 24 March 2017
Available online 31 March 2017

Abstract
Reference based assembly of genomic reads of the elite indica rice cultivar RP Bio-226 was carried out against 2101 reference bacterial genomes using Bowtie-2 genome assembly tool. Five types of data: Number of paired end reads concordantly aligned exactly only once, number of paired end reads concordantly aligned more than once, number of mates that make the pairs aligned exactly only once, number of mates that make the pairs aligned more than once and overall percentage of alignment were collected. Interpretation of the results and identification of endophytes based on these alignment statistics are described in detail in our research article “L.Battu, M.M. Reddy, B.S.Goud, K.Ulaganathan, K.Ulaganathan, Genome inside genome: NGS based identification and assembly of endophytic Sphingopyxis granuli and Pseudomonas aeruginosa genomes from rice genomic reads, Genomics (in press)”.© 2017 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 | Biology |
|--------------|---------|
| More specific subject area | Genomics |
| Type of data | Genome assembly data |
| How data was acquired | Computational reference based genome assembly |
| Data format | Analyzed |
| Experimental factors | Adapters were removed from the genomic reads |
| Experimental features | The pre-processing of raw reads was done with FastQC and the adapters were removed with Cutadapt tool. After pre-processing, the reads were aligned to the reference genome by using Bowtie2 (ver. 2.2.4). The reference genomes of 2101 bacterial species including *Pseudomonas aeruginosa* PAO1 and *Sphingopyxis granuli* were downloaded from NCBI. Reference based assembly of the reads against these reference genomes involved, indexing of the reference genomes and alignment of reads to the reference and creation of SAM files. |
| Data source location | Not applicable |
| Data accessibility | Data is with this article |

### Value of the data

- The data listed in tabular form provides information on assembly of reads of *Oryza sativa* indica cultivar RP Bio-226 onto 2101 reference bacterial genomes.
- This data can be compared to similar alignment of genomic reads from other rice cultivars onto reference bacterial genomes.
- The alignment data explains a new method of identification of endophytes from plant genomic reads and can be followed in other plants.

### 1. Data

Results of reference based assembly of genomic reads of elite indica rice cultivar RP Bio-226 in tabular form are the data mentioned in this paper. The tabulated data include five types of alignment statistics data: Number of paired end reads concordantly aligned exactly only once, number of paired end reads concordantly aligned more than once, number of mates that make the pairs aligned exactly only once, number of mates that make the pairs aligned more than once and overall percentage of alignment. (Supplementary Table-1) Additionally, genomes which showed maximum alignment with rice reads with respect to the above 5 different types of data are tabulated separately and enclosed (Tables 1–5).

### 2. Experimental design, materials and methods

Total DNA was isolated from the Leaves of in vitro grown *Oryza sativa* indica cultivar RP Bio-226 plants and sequencing library was prepared [1]. Whole genome sequencing was carried out with
Table 1
Reference based assembly of RP Bio-226 genomic reads on to bacterial genome: List of genomes to which more than 10,000 paired end reads aligned concordantly exactly only once.

| S.No | Name of the Bacterial Genomes used as reference | Reads aligned concordantly exactly 1 time | Reads aligned concordantly >1 time | Mates that make up the pairs aligned concordantly or discordantly exactly 1 time | Mates that make up the pairs aligned concordantly or discordantly >1 time | Overall Alignment rate % |
|------|-------------------------------------------------|----------------------------------------|-----------------------------------|-------------------------------------------------|-------------------------------------------------|------------------------|
| 1    | *Sphingopyxis granuli*                           | 189315                                 | 488                               | 17773                                           | 434                                             | 0.58                   |
| 2    | *Pseudomonas aeruginosa*                         | 33143                                  | 556                               | 730                                             | 1361                                           | 0.11                   |
| 3    | *Thermosynechococcus elongatus*                  | 22670                                  | 0                                 | 76053                                           | 0                                              | 0.18                   |
| 4    | *Thermosynechococcus sp.*                        | 18763                                  | 0                                 | 77881                                           | 0                                              | 0.17                   |
| 5    | *Sphingopyxis terrae*                            | 17082                                  | 156                               | 37916                                           | 165                                            | 0.11                   |
| 6    | *Sphingopyxis macrocollabida*                    | 16626                                  | 165                               | 39729                                           | 391                                            | 0.11                   |
| 7    | *Gloeobacter violaceus*                          | 16544                                  | 0                                 | 72171                                           | 0                                              | 0.15                   |
| 8    | *Gloeobacter klauweensis*                        | 15380                                  | 0                                 | 67349                                           | 0                                              | 0.14                   |
| 9    | *Sphingopyxis alaskensis*                        | 13147                                  | 23                                | 33753                                           | 84                                             | 0.09                   |
| 10   | *Sphingopyxis sp.*                               | 10580                                  | 139                               | 27461                                           | 177                                            | 0.07                   |
| 11   | *Sphingopyxis frigibergensis*                    | 10451                                  | 142                               | 33078                                           | 165                                            | 0.08                   |

* Cyanobacterial species are highlighted in grey;

Table 2
Reference based assembly of RP Bio-226 genomic reads on to bacterial genome: List of genomes to which more than 10,000 paired end reads aligned concordantly more than once.

| S. No | Name of the Bacterial Genomes used as reference | Reads aligned concordantly exactly 1 time | Reads aligned concordantly >1 time | Mates that make up the pairs aligned concordantly or discordantly exactly 1 time | Mates that make up the pairs aligned concordantly or discordantly >1 time | Overall Alignment rate % |
|-------|-------------------------------------------------|----------------------------------------|-----------------------------------|-------------------------------------------------|-------------------------------------------------|------------------------|
| 1     | *Cylindrospermum stagnale*                      | 2294                                   | 35317                             | 3106                                            | 77525                                           | 0.22                   |
| 2     | *Geitlerinema sp.*                              | 32249                                  | 1                                   | 84912                                           | 0                                                | 0.22                   |
| 3     | *Oscillatoria acuminata*                        | 132                                    | 31314                             | 180                                             | 83284                                           | 0.21                   |
| 4     | *Nostoc azollae*                                | 54                                     | 31289                             | 38                                              | 70590                                           | 0.19                   |
| 5     | *Microcystis panniformis*                       | 0                                      | 31104                             | 3                                               | 85117                                           | 0.21                   |
| 6     | *Microcystis aeruginosa*                        | 16                                     | 31094                             | 229                                             | 85131                                           | 0.21                   |
| 7     | *Cyanobacterium*                                | 0                                      | 30463                             | 8                                               | 82316                                           | 0.21                   |
| 8     | *Cyanobacterium endosymbiont*                   | 0                                      | 30463                             | 8                                               | 82316                                           | 0.21                   |
| 9     | *Nostoc punctiforme*                            | 1                                      | 29246                             | 1                                               | 88219                                           | 0.21                   |
| 10    | *Oscillatoriales cyanobacterium*                | 0                                      | 28541                             | 3                                               | 81824                                           | 0.20                   |
| 11    | *Oscillatoria nigroviridis*                     | 4                                      | 28176                             | 5                                               | 96119                                           | 0.22                   |
| 12    | *Fischerella sp.*                               | 431                                    | 28127                             | 173                                             | 73901                                           | 0.19                   |
| 13    | *Acaryochloris marina*                          | 10                                     | 27985                             | 8                                               | 75714                                           | 0.19                   |
| 14    | *Nodularia spumigena*                           | 4                                      | 27513                             | 100                                             | 80485                                           | 0.20                   |
| 15    | *Synechococcus elongatus*                       | 4                                      | 26252                             | 29                                              | 67922                                           | 0.17                   |
| 16    | *Cyanobium gracile*                             | 1                                      | 26016                             | 25                                              | 70894                                           | 0.18                   |
| 17    | *Anabaena variabilis*                           | 4                                      | 25773                             | 0                                               | 80169                                           | 0.19                   |
Illumina Nextseq. 500 system (Illumina, San Diego, CA). The raw data files in Fastq format were used for further analysis. The pre-processing of raw reads was done with FastQC and the adapters were removed with Cutadapt tool [2,3]. After pre-processing, the reads were aligned to the reference genome by using Bowtie2 (ver. 2.2.4) [4]. The reference genomes of 2101 bacterial species including Pseudomonas aeruginosa PAO1 and Sphingopyxis granuli were downloaded from NCBI. Reference based assembly of the reads against these reference genomes involved, indexing of the reference genomes and alignment of reads to the reference and creation of SAM files. Samtools (ver 0.1.18) was used for further analysis [5]. SAM files were converted into binary BAM files, sorted and indexed by using the 'view', 'sort' and 'index' functions of SAMtools. The consensus sequences were created with Samtools. Genome annotation was carried out with RAST and BaySys servers [6,7]. The tRNAs were identified with tRNAscan-SE software and rRNAs were identified with Rfam software [8,9].

Table 2 (continued)

| S. No | Name of the Bacterial Genomes used as reference | Reads aligned concordantly exactly 1 time | Reads aligned concordantly > 1 time | Mates that make up the pairs aligned concordantly or discordantly exactly 1 time | Mates that make up the pairs aligned concordantly or discordantly > 1 time | Overall Alignment rate % |
|------|--------------------------------------------------|----------------------------------------|----------------------------------|----------------------------------------|----------------------------------------|--------------------------|
| 18   | Trichodesmium erythraeum                        | 6                                      | 25746                            | 53                                      | 88351                                  | 0.20                     |
| 19   | Arthrospira platensis                           | 0                                      | 25699                            | 2                                       | 91181                                  | 0.21                     |
| 20   | Crinalium epipsammum                            | 5                                      | 24247                            | 17                                      | 88945                                  | 0.20                     |
| 21   | Chroococcidiopsis thermalis                     | 1                                      | 20908                            | 6                                       | 73861                                  | 0.17                     |
| 22   | Microcoleus sp.                                 | 4                                      | 20405                            | 25                                      | 89544                                  | 0.19                     |
| 23   | Dactylococcopsis salina                         | 52                                     | 20361                            | 175                                     | 72443                                  | 0.16                     |
| 24   | Candidatus Atelocyano-bacterimthalassa          | 0                                      | 19925                            | 4                                       | 69551                                  | 0.16                     |
| 25   | Pseudanabaena.sp.                               | 11                                     | 17760                            | 48                                      | 74103                                  | 0.16                     |
| 26   | Rivularia sp.                                   | 0                                      | 17695                            | 0                                       | 73210                                  | 0.16                     |
| 27   | Halothece sp.                                   | 209                                    | 17504                            | 1301                                    | 62653                                  | 0.14                     |
| 28   | Planktothrix agardhii                           | 268                                    | 15578                            | 1982                                    | 70470                                  | 0.15                     |
| 29   | Cyanobacterium aponinum                         | 0                                      | 15458                            | 0                                       | 48759                                  | 0.11                     |
| 30   | Chamaesiphon minutus                            | 1                                      | 14566                            | 4                                       | 68114                                  | 0.14                     |
| 31   | Cyanobacterium stanieri                         | 34                                     | 14415                            | 630                                     | 49301                                  | 0.11                     |
| 32   | Stanieria cyanosphaera                          | 2734                                   | 11866                            | 8319                                    | 72035                                  | 0.16                     |
| 33   | Geminocystis herdmanii                          | 2683                                   | 10717                            | 36938                                   | 16238                                  | 0.12                     |
Table 3
Reference based assembly of RP Bio-226 genomic reads on to bacterial genome: List of genomes to which more than 10,000 mates that make up the pairs aligned concordantly or discordantly exactly 1 time.

| S.No | Name of the Bacterial Genomes used as reference | Reads aligned concordantly exactly 1 time | Reads aligned concordantly >1 time | Mates that make up the pairs aligned concordantly or discordantly exactly 1 time | Mates that make up the pairs aligned concordantly or discordantly >1 time | Overall Alignment rate % |
|------|------------------------------------------------|----------------------------------------|----------------------------------|------------------------------------------|-------------------------------------------------|--------------------------|
| 1    | Thermosynechococcus sp.                         | 18763                                  | 0                                | 77881                                    | 0                                              | 0.17                     |
| 2    | Thermosynechococcus elongatus                  | 22670                                  | 0                                | 76053                                    | 0                                              | 0.18                     |
| 3    | Gloeobacter violaceus                          | 16544                                  | 0                                | 72171                                    | 0                                              | 0.15                     |
| 4    | Gloeobacter kilaeensis                         | 15380                                  | 0                                | 67349                                    | 0                                              | 0.14                     |
| 5    | Sphingopyxis macroglutabida                    | 16626                                  | 165                              | 39729                                    | 391                                            | 0.11                     |
| 6    | Sphingopyxis terrae                           | 17082                                  | 156                              | 37916                                    | 165                                            | 0.11                     |
| 7    | Geminicyclus hermandii                        | 2683                                   | 10717                            | 36938                                    | 16238                                          | 0.12                     |
| 8    | Sphingopyxis alaskensis                       | 13147                                  | 23                               | 33753                                    | 84                                             | 0.09                     |
| 9    | Sphingopyxis freibergensis                    | 10451                                  | 142                               | 33078                                    | 165                                            | 0.08                     |
| 10   | Sphingopyxis sp.                              | 10580                                  | 139                               | 27461                                    | 177                                            | 0.07                     |
| 11   | Candidatus Methylomirabilis oxyfera            | 1101                                   | 0                                | 20433                                    | 0                                              | 0.03                     |
| 12   | Enterococcus rotai                            | 3377                                   | 0                                | 19193                                    | 0                                              | 0.04                     |
| 13   | Sphingopyxis granuli                          | 189315                                 | 488                              | 17773                                    | 434                                            | 0.58                     |
| 14   | Enterococcus cecorum                         | 2012                                   | 156                              | 37916                                    | 165                                            | 0.11                     |
| 15   | Bacillus pseudonycoidei                       | 2580                                   | 147                              | 17473                                    | 0                                              | 0.03                     |
| 16   | Coprococcus sp.                               | 1111                                   | 0                                | 14869                                    | 0                                              | 0.02                     |
| 17   | Listeria marshi                               | 425                                    | 0                                | 14590                                    | 0                                              | 0.02                     |
| 18   | Spiroplasma eriocheiris                       | 2376                                   | 0                                | 14350                                    | 0                                              | 0.03                     |
| 19   | Spiroplasma atrichophaginis                   | 2155                                   | 0                                | 14299                                    | 0                                              | 0.03                     |
| 20   | Spiroplasma mirum                             | 2155                                   | 0                                | 14299                                    | 0                                              | 0.03                     |
| 21   | Spiroplasma helicoides                        | 2346                                   | 0                                | 13871                                    | 0                                              | 0.03                     |
| 22   | Spiroplasma cantharicola                      | 2620                                   | 0                                | 13837                                    | 0                                              | 0.03                     |
| 23   | Spiroplasma litorale                          | 1636                                   | 0                                | 13640                                    | 0                                              | 0.02                     |
| 24   | Spiroplasma turonicum                         | 1636                                   | 0                                | 13543                                    | 0                                              | 0.02                     |
| 25   | Spiroplasma poulsoni                          | 1290                                   | 0                                | 13542                                    | 0                                              | 0.02                     |
| 26   | Streptococcus pseudopneumoniae                | 842                                    | 0                                | 13478                                    | 0                                              | 0.02                     |
| 27   | Coprococcus catus                             | 406                                    | 0                                | 13325                                    | 0                                              | 0.02                     |
| 28   | Syntrophus aciditrophicus                     | 326                                    | 0                                | 13136                                    | 0                                              | 0.02                     |
| 29   | Moorella thermoacetica                        | 344                                    | 0                                | 12193                                    | 0                                              | 0.02                     |
| 30   | Gammaproteobacteria bacterium                 | 191                                    | 0                                | 11800                                    | 5                                              | 0.02                     |
| 31   | Eubacterium cylindroides                      | 930                                    | 0                                | 11475                                    | 0                                              | 0.02                     |
| 32   | Actinobacteria bacterium                      | 517                                    | 0                                | 11307                                    | 0                                              | 0.02                     |
| 33   | Ruminococcus champanellensis                 | 424                                    | 0                                | 10978                                    | 0                                              | 0.02                     |
| 34   | Acidobacteria bacterium                       | 396                                    | 0                                | 10904                                    | 0                                              | 0.02                     |
| 35   | Olsenella uli                                | 305                                    | 0                                | 10374                                    | 0                                              | 0.02                     |
| 36   | Atopobium parvulum                           | 288                                    | 0                                | 10311                                    | 0                                              | 0.02                     |
| 37   | Mycobacterium chelonae                        | 205                                    | 0                                | 10198                                    | 0                                              | 0.02                     |
| 38   | Jonquetella antroni                         | 491                                    | 0                                | 10155                                    | 0                                              | 0.02                     |
| 39   | Mycoplasma hyorhinis                          | 461                                    | 0                                | 10142                                    | 0                                              | 0.02                     |
| 40   | Mycoplasma flocculare                         | 460                                    | 0                                | 10139                                    | 0                                              | 0.02                     |
| 41   | Mycoplasma dispar                             | 461                                    | 0                                | 10138                                    | 0                                              | 0.02                     |
| 42   | Mycoplasma hyopneumonia                      | 461                                    | 0                                | 10138                                    | 0                                              | 0.02                     |
| 43   | Mycoplasma bovocati                           | 460                                    | 0                                | 10134                                    | 0                                              | 0.02                     |
| 44   | Mycoplasma conjunctivae                      | 460                                    | 0                                | 10134                                    | 0                                              | 0.02                     |
| 45   | Desulfobaca acetoxidans                      | 395                                    | 0                                | 10106                                    | 239                                            | 0.02                     |

*Cyanobacterial species are highlighted in grey*
Table 4
Reference based assembly of RP Bio-226 genomic reads on to bacterial genome: List of genomes to which more than 10,000 mates that make up the pairs aligned concordantly or discordantly more than once.

| S. No | Name of the Bacterial Genomes used as reference | Reads aligned concordantly exactly 1 time | Reads aligned concordantly > 1 time | Mates that make up the pairs aligned concordantly or discordantly exactly 1 time | Mates that make up the pairs aligned concordantly or discordantly > 1 time | Overall Alignment rate % |
|-------|-------------------------------------------------|----------------------------------------|---------------------------------|-----------------------------------------------|-------------------------------------------------|--------------------------|
| 1     | Oscillatoria nigroviridis                       | 4                                      | 28176                           | 5                                             | 96119                                           | 0.22                     |
| 2     | Arthrospira platensis                           | 0                                      | 25699                           | 2                                             | 91181                                           | 0.21                     |
| 3     | Microcoleus sp.                                 | 4                                      | 20405                           | 25                                            | 89544                                           | 0.19                     |
| 4     | Crinalium epipsammum                            | 5                                      | 24247                           | 17                                            | 88945                                           | 0.20                     |
| 5     | Trichodesmium erythraeum                        | 6                                      | 25746                           | 53                                            | 88351                                           | 0.20                     |
| 6     | Nostoc punctiforme                              | 1                                      | 29246                           | 1                                             | 88219                                           | 0.21                     |
| 7     | Microcystis aeruginosa                          | 16                                     | 31094                           | 229                                           | 85131                                           | 0.21                     |
| 8     | Microcystis panniformis                         | 0                                      | 31104                           | 3                                             | 85117                                           | 0.21                     |
| 9     | Getlerinema.sp.                                 | 1                                      | 32249                           | 3                                             | 84912                                           | 0.22                     |
| 10    | Oscillatoria acuminata                          | 132                                    | 31314                           | 180                                           | 83284                                           | 0.21                     |
| 11    | Cyanobacterium                                  | 0                                      | 30463                           | 8                                             | 82316                                           | 0.21                     |
| 12    | Cyanobacterium endosymbiont                     | 0                                      | 30463                           | 8                                             | 82316                                           | 0.21                     |
| 13    | Oscillatoriales cyanobacterium                  | 0                                      | 28541                           | 3                                             | 81824                                           | 0.20                     |
| 14    | Nodularia spumigena                             | 4                                      | 27513                           | 100                                           | 80485                                           | 0.20                     |
| 15    | Anabaena variabilis                             | 4                                      | 25773                           | 0                                             | 80169                                           | 0.19                     |
| 16    | Cylindropermum stagnale                         | 2294                                   | 35317                           | 3106                                          | 77525                                           | 0.22                     |
| 17    | Acaryochloris marina                            | 10                                     | 27985                           | 8                                             | 75714                                           | 0.19                     |
| 18    | Pseudanabaena.sp.                               | 11                                     | 17760                           | 48                                            | 74103                                           | 0.16                     |
| 19    | Fischerella.sp.                                 | 431                                    | 28127                           | 173                                           | 73901                                           | 0.19                     |
| 20    | Chroococcidiopsis thermalis                     | 1                                      | 20908                           | 6                                             | 73861                                           | 0.17                     |
| 21    | Rivularia sp.                                   | 0                                      | 17695                           | 0                                             | 73210                                           | 0.16                     |
| 22    | Dactylococcopsis salina                         | 52                                     | 20361                           | 175                                           | 72443                                           | 0.16                     |
| 23    | Stanieria cyanosphaera                          | 2734                                   | 11866                           | 8319                                          | 72035                                           | 0.16                     |
| 24    | Cyanobium gracile                               | 1                                      | 26016                           | 25                                            | 70894                                           | 0.18                     |
| 25    | Nostoc azollae                                  | 54                                     | 31289                           | 38                                            | 70590                                           | 0.19                     |
| 26    | Planktothrix agardhii                           | 268                                    | 15578                           | 1982                                          | 70470                                           | 0.15                     |
| 27    | Candidatus Atelocyanobacter thalassa            | 0                                      | 19925                           | 4                                             | 69551                                           | 0.16                     |
| 28    | Chamaesiphon minutus                            | 1                                      | 14566                           | 4                                             | 68114                                           | 0.14                     |
| 29    | Synechococcus elongatus                         | 4                                      | 26252                           | 29                                            | 67922                                           | 0.17                     |
| 30    | Halothece sp.                                   | 209                                    | 17504                           | 1301                                          | 62653                                           | 0.14                     |
| 31    | Cyanobacterium stanieri                         | 34                                     | 14415                           | 630                                           | 49301                                           | 0.11                     |
| 32    | Cyanobacterium aponinum                         | 0                                      | 15458                           | 0                                             | 48759                                           | 0.11                     |
Table 5
Reference based assembly of RP Bio-226 genomic reads on to bacterial genomes: List of genomes to which at least 0.1% paired end reads are aligned concordantly or discordantly.

| S.No | Name of the Bacterial Genomes used as reference | Reads aligned concordantly exactly 1 time | Reads aligned concordantly >1 time | Mates that make up the pairs aligned concordantly or discordantly exactly 1 time | Mates that make up the pairs aligned concordantly or discordantly >1 time | Overall Alignment rate % |
|------|-------------------------------------------------|-------------------------------------------|-----------------------------------|-------------------------------------------------|-------------------------------------------------|--------------------------|
| 1    | Sphingopyxis granulifera                        | 189315                                    | 488                               | 17773                                       | 434                                            | 0.58                     |
| 2    | Cylindrospermum stagnale                       | 2294                                       | 35317                             | 3106                                         | 77525                                         | 0.22                     |
| 3    | Oscillatoria nigr-viridis                      | 4                                           | 28176                             | 5                                            | 96119                                         | 0.22                     |
| 4    | Geitlerinema sp.                                | 1                                           | 32249                             | 3                                            | 84912                                         | 0.22                     |
| 5    | Microcystis aeruginosa                         | 16                                          | 31094                             | 229                                          | 85131                                         | 0.21                     |
| 6    | Oscillatoria acuminata                         | 132                                         | 31314                             | 180                                          | 83284                                         | 0.21                     |
| 7    | Cyanobacterium                                 | 0                                           | 30463                             | 8                                            | 82316                                         | 0.21                     |
| 8    | Cyanobacterium endosymbiont of Epithemia turgida isolate ESB Lake Yunoko | 0                                           | 30463                             | 8                                            | 82316                                         | 0.21                     |
| 9    | Microcystis panniformis                        | 0                                           | 31104                             | 3                                            | 85117                                         | 0.21                     |
| 10   | Arthrospira platensis                          | 0                                           | 25699                             | 2                                            | 91181                                         | 0.21                     |
| 11   | Nostoc punctiforme                             | 1                                           | 29246                             | 1                                            | 88219                                         | 0.21                     |
| 12   | Nodularia spumigena                            | 4                                           | 27513                             | 100                                          | 80485                                         | 0.20                     |
| 13   | Trichodesmium erythraemum                      | 6                                           | 25746                             | 53                                           | 88351                                         | 0.20                     |
| 14   | Ornithium epipsammum                          | 5                                           | 24247                             | 17                                           | 88945                                         | 0.20                     |
| 15   | Oscillatoriales cyanobacterium                 | 0                                           | 28551                             | 3                                            | 81824                                         | 0.20                     |
| 16   | Fischerella sp.                                | 431                                         | 28172                             | 173                                          | 73901                                         | 0.19                     |
| 17   | Nostoc azollae                                 | 54                                          | 31289                             | 38                                           | 70590                                         | 0.19                     |
| 18   | Microcoleus sp.                                | 4                                           | 20405                             | 25                                           | 89544                                         | 0.19                     |
| 19   | Acaryochloris marina                           | 10                                          | 27985                             | 8                                            | 75714                                         | 0.19                     |
| 20   | Anabaena variabilis                            | 4                                           | 25773                             | 0                                            | 80169                                         | 0.19                     |
| 21   | Thermosynechococcus elongatus                  | 22670                                       | 0                                 | 76053                                        | 0                                             | 0.18                     |
| 22   | Cyanobium gracile                              | 1                                           | 26016                             | 25                                           | 70894                                         | 0.18                     |
| 23   | Thermosynechococcus sp.                        | 18763                                       | 0                                 | 77881                                        | 0                                             | 0.17                     |
| 24   | Synechococcus elongatus                        | 4                                           | 26252                             | 29                                           | 67922                                         | 0.17                     |
| 25   | Chroococcidiopsis thermalis                    | 1                                           | 20908                             | 6                                            | 73861                                         | 0.17                     |
| 26   | Stanieria cyanospaera                          | 2734                                         | 11866                             | 8319                                         | 72035                                         | 0.16                     |
| 27   | Dactylococcuspsalis salina                    | 52                                          | 20361                             | 175                                          | 72443                                         | 0.16                     |
| 28   | Pseudanabaena sp.                              | 11                                          | 17760                             | 48                                           | 74103                                         | 0.16                     |
| 29   | Candidatus Atelocyanotheater thalassa isolate ALOHA | 0                                           | 19925                             | 4                                            | 69551                                         | 0.16                     |
| 30   | Rivularia sp.                                  | 0                                           | 17695                             | 0                                            | 73210                                         | 0.16                     |
| 31   | Golenobacter violaceus                         | 16544                                       | 0                                 | 72171                                        | 0                                             | 0.15                     |
| 32   | Planckothrix agarbii                           | 268                                         | 15578                             | 1982                                         | 70470                                         | 0.15                     |
| 33   | Gloeobacter klauensis                          | 15380                                       | 0                                 | 67349                                        | 0                                             | 0.14                     |
| 34   | Halotectec sp.                                 | 209                                         | 17504                             | 1301                                         | 62653                                         | 0.14                     |
| 35   | Chamaesiphon minutus                           | 1                                           | 14566                             | 4                                            | 68114                                         | 0.14                     |
| 36   | Geminicystis hermanii                         | 2683                                         | 10717                             | 36938                                        | 16238                                         | 0.12                     |
| 37   | Sphingopyxis macroglostabida                   | 16626                                       | 165                               | 39729                                        | 391                                           | 0.11                     |
| 38   | Sphingopyxis terrae                           | 17082                                       | 156                               | 37916                                        | 165                                           | 0.11                     |
| 39   | Pseudomonas aeruginosa                         | 33143                                       | 556                               | 730                                          | 1361                                          | 0.11                     |
| 40   | Cyanobacterium stanieri                        | 34                                          | 14415                             | 630                                          | 49301                                         | 0.11                     |
| 41   | Cyanobacterium aponinum                        | 0                                           | 15458                             | 0                                            | 48759                                         | 0.11                     |

*Non cyanobacterial species are highlighted in grey
Acknowledgements

Financial assistance from the Department of Science and Technology, Government of India through the DST-PURSE programme to Osmania University and the University Grants Commission through the UPE programme of Osmania University is acknowledged. Mettu Madhavi Reddy is supported by a Junior Research Fellowship from the Department of Biotechnology, Government of India under DBT-OU-ISLARE programme. Battu Latha is supported by the UGC-BSR-RFSMS Junior Research Fellowship. Burragoni Sravanthi Goud is supported by the UGC-BSR-RFSNS Senior Research Fellowship.

Transparency document. Supporting information

Transparency document. Supporting information Transparency data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.dib.2017.03.036.

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.2017.03.036.

References

[1] M.M. Reddy, K. Ulaganathan, Mitochondrial genome sequence of Oryza sativa indica cultivar RP Bio-226, Curr. Sci. 111 (2016) 258–260.
[2] S. Andrews, FastQC, A Quality Control Tool for High Throughput Sequence Data Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/, 2010.
[3] M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads, EMBO J. 17 (2011) 10.
[4] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with bowtie 2, Nat. Methods 9 (2012) 357–359.
[5] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, The sequence alignment/map format and SAMtools, Bioinformatics 25 (2009) 2078–2079.
[6] R.K. Aziz, D. Bartels, A.A. Best, M. DeJongh, T. Disz, R.A. Edwards, K. Formsma, S. Gerdes, E.M. Glass, M. Kubal, F. Meyer, The RAST Server: rapid annotations using subsystems technology, BMC Genom. 9 (2008) 75.
[7] G.H. Van Domselaar, P. Stotthard, S. Shrivastava, J.A. Cruz, A. Guo, X. Dong, P. Lu, D. Szafron, R. Greiner, D.S. Wishart, BASys: a web server for automated bacterial genome annotation, Nucleic Acids Res. 33 (2005) W455–W459.
[8] P. Schattner, A.N. Brooks, T.M. Lowe, The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs, Nucleic Acids Res. 33 (2005) W686–W689.
[9] K. Lagesen, P. Hallin, E.A. Redland, H.H. Staerfeldt, T. Rognes, D.W. Ussery, RNAmer: consistent and rapid annotation of ribosomal RNA genes, Nucleic Acids Res 35 (2007) 3100–3108.