Public Databases of 16s rRNA: Antibiotic Resistance

Vandana Singh[1], Kunal Maniar[2], Rajasri Bhattacharayya[1]*and Dibyajyoti Banerjee[1]

[1]Department of Experimental Medicine and Biotechnology, Post Graduate Institute of Medical Education and Research, Chandigarh, India

[2]Department of Pharmacology, Post Graduate Institute of Medical Education and Research, Chandigarh, India

*Corresponding author: Rajasri Bhattacharayya, Department of Experimental Medicine and Biotechnology, Post Graduate Institute of Medical Education and Research, Chandigarh, India; Tel: +91-9780134620; Fax: +91-172-2744401; E-mail: bdr.rajasri@yahoo.in .

Abstract

Next generation sequencing has resulted in generation of metagenomic databases of environmental samples containing partial or full length 16s rRNA gene sequences available in the public domain. 16s rRNA sequence has known antibiotic binding sites and therefore databases containing 16s rRNA gene sequence has the potential to be used as an excellent tool for understanding antibiotic resistance patterns in the environmental samples. However, currently some problems are existing in the various data repositories which are highlighted here for the purpose of addressing this issue and integrating the available databases in future so that the shortcomings can be minimized.

Keywords: NGS; 16s rRNA; 16s rRNA gene sequencing; Databases; Metagenomic; Antibiotic resistance; Antibiotic binding

Introduction

The sequencing experience of mankind in Human Genome Project (HGP) using Sanger method has revolutionized the application of knowledge base of genetic sequences [1]. The discovery of next generation sequencing (NGS) has made the process of identifying gene sequences rapid and user-friendly. Nowadays, it is believed as a parallel sequencing method as it overcome the limitations of traditional Sanger based sequencing systems by attaching the DNA molecules to be sequenced to a solid surface or beads; thus allowing millions of sequencing reactions to happen in parallel in a cost effective manner [2]. NGS can be performed from single DNA molecules fragmented into shorter reads as compared to previous sequencing methods that required DNA amplification and long reads fragments. Moreover, it can identify huge genomic sequencing data including species genotypic classification without culturing bacteria [2].

The NGS technique has wide application in whole genome sequencing and assembly, exome, transcriptome, metagenome, complete mitochondria sequence, ChiP, etc. analysis [3]. Thus, the sequences generated by NGS are assembled and resulted in various types of database formation. These databases are very much helpful in determining genetic variants, complex disease understanding, etc. [3].

Aims and Scope

16s rRNA is highly conserved and required for protein synthesis [4]. For bacterial identification and taxonomic classification 16s rRNA sequencing is routinely done [5]. Recently it is argued that it can also use to predict antibiotic resistance. There are some antibiotics which binds to 16S r RNA at specific region. For example, streptomycin binds at AAACUCAA sequence. This is present 915 loops of the 16s rRNA. Therefore, in its corresponding DNA level there will be AACTCAA. It is plausible to reason that if that specific sequence is mutated then that will cause a structural change to the 16s rRNA structure leading to antibiotic resistance [6-8].

There are numbers of databases available in public that all are resultant of sequencing through next generation sequencing like Green gene database (http://greengenes.secondgenome.com/downloads/database/13_5), EBI Metagenomics (https://www.ebi.ac.uk/metagenomics/), Silva Database (https://www.arb-silva.de/) and EzTaxon-e Database (https://www.ezbiocloud.net/taxonomy). In all these databases the DNA sequence of corresponding 16s ribosomal RNA are deposited. BLAST tools may be used for confirmation of the sequence at RNA level.

In GreenGene and EBI metagenomic databases the source of bacterial 16s rRNA gene is environment samples. However, the amplification length of 16s rRNA gene is not same in all the databases. EBI metagenomic 16s rRNA databases is prepared by NGS technique where samples were collected from forest, soil, water, human, etc. and reposited separately in different biomes. Although same NGS method has been used for amplification, but different regions and length of 16s rRNA gene has been amplified in different biomes like forest and soil (Table 1). Different length of 16s rRNA gene sequences in forest biome are observed as given in Table 1. Antibiotics like streptomycin binds with 16s rRNA and stops bacterial protein synthesis [4]. So, 16s rRNA gene mutation is a way for generation of antibiotic resistance. If streptomycin binding site of 16s rRNA is mutated, then streptomycin binding will be nonfunctional and that will generate streptomycin resistant strain of bacteria [8]. Since EBI metagenomic databases are constituted by DNA sequences collected from different environment, it is a good resource for comparative study of drug resistance among various biomes [9]. On the other hand, green gene database is curated database which addresses limitations of public repositories by providing chimera screening, standard alignment, and taxonomic classification using multiple published taxonomies [10].

Results and Discussion

Since different regions of 16s rRNA gene has been amplified in different biome (Table 1), hence only those antibiotic resistance or sensitivity can be predicted which binds in that regions. Streptomycin binds at 915 loop regions of 16s ribosomal RNA [8,9]. But in Forest biome of EBI metagenomic database this streptomycin binding region is amplified whereas in soil biome it is not amplified.

Database	Biome	Project	ID	Sequence	Classification
EBI metagenomics database	Forest	Bacterial community composition on leaf litter of F. Sylvatica Targeted Locus (Loci)	SRR1197858.5084-HWSOZGE01CFZDZ-2	**AAACTCAA**AGGAATTGACGGGGGCCCGCACAAGCGGTGGAGTATGTGGTTTAATTCGATGCAACGCGAAGAACCTTACCTGGCCTTGACATGTCCGGAATCCTGCAGAGATGCGGGAGTGCCTTCGGGAATCGGAACACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGGTTAAGT	k Bacteria; p Proteobacteria; c Alphaproteobacteria; o Rhizobiales;f Phyllobacteriaceae; g Mesorhizobium;
EBI metagenomics database	Forest	Bacterial community composition on leaf litter of F. Sylvatica Targeted Locus (Loci)	SRR1197858.5102-HWSOZGE01DZN3Q-2	**AAACTCAA**AGGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGAAGCGACGCGCAGAACCTTACCACCTTTTGACATGCCTGGACCGCCGGAGAGATCC	Not classified
EBI metagenomics database	Soil	16s r RNA soil metagenom	SRR1793463.5-5-1	TGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCAGTAGGGAAGAAGCGAAAGTGACGGTACCTGCAGAAGAAGCGCCGGCTAACTAAGTGCCAGCAGCCGCGGTAATACG	Not classified
Green gene database	Environment sample	-	179180	GATGAACGCTGGCGGCTGTGCTTAACACATGCAAGTCGAACGAACCTTCGGGTTAGTGGCGGACGGGTGAGTAACGCGTGAGAATCTGCCCTCAGGAGGGGGATAACGGTTGGAAACGACCGCTAATACCCCATATGCCGAGAGGTGAAGTGAAATTCGCCTGAGGATGAGCTCGCGTCTGATTAGCTAGTTGGTGTGGTAATGGCGCACCAAGGCTTCGATCAGTAGCTGGTCTGAGAGGATGATCAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATTTTCCGCAATGGGCGAAAGCCTGACGGAGCAACGCCGCGTGAGGGATGAAGGCCTCTGGGCTGTAAACCTCTTTTCTCAAGGAAGAAGATCTGACGGTACTTGAGGAATAAGCCACGGCTAATTCCGTGCCAGCAGCCCCGGTAATACGGGAGTGGCAAGCGTTATCCGGAATTATTGGGCGTAAAGCGTCCGCAGGCGGCCTCGAAAGTCTGTCGTTAAAGCGTGGAGCTTAACTCCATTTAAGCGGTGGAAACTACGAGGCTAGAGTGTGGTAGGGGCAGAGGGAATTCCCGGTGTAGCGGTGAAATGCGTAGATATCGGGAAGAACACCAGTGGCGAAGGCGCTCTGCTGGGCCATAACTGACGCTCATGGACGAAAGCCAGGGGAGCGAAAGGGATTAGATACCCCTGTAGTCCTGGCCGTAAACGATGAACACTAGGTGTCGGGGGAATCGACCCCCTCGGTGTCGTAGCTAACGCGTTAAGTGTTCCGCCTGGGGAGTACGCACGCAAGTGTG**AAACTCAA**AGGAATTGACGGGGGCCCGCACAAGCGGTGGAGTATGTGGTTTAATTCGATGCAACGCGAAGAACCTTACCAGGGCTTGACATCCTGCGAATCCCTTGGAAACTAGGGAGTGCCTTCGGGAGCGCAGAGACAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCACGTCTTTAGTTGCCAGCATTAAGTTGGGCACTCTAGAGAGACCGCCGGTGATAAACCGGAGGAAGGTGTGGATGACGTCAAGTCATCATGCCCCTTACGTCCTGGGCTACACACGTACTACAATGCTACGGACAAAGGGCAGCAAACTCGCGAGAGCTAGCAAATCCCATAAACCGTGGCTCAGTTCAGATCGTAGGCTGCAACTCGCCTACGTGAAGGAGGAATCGCTAGTAATCGCAGGTCAGCATACTGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGGAAGTTGGCCACGCCCGAAGTCGTTACTCCAACCCGTAAGGGAGGAGGATGCCGAAGGTGGGGCTGATGACTGGGGTGAAGTCGTA	k Bacteria; p Cyanobacteria; c Synechococcophycideae; o Synechococcales; f SynechococcaceaeSynechococcaceae; g Synechococcus;

Table 1: Showing classification of 16S r RNA and Streptomycin binding site (in bold) present in different databases.

Analyzing forest soil metagenomic database presence of streptomycin sensitive and resistance bacteria can be understood. However, from soil biome such type of study is not possible. In this metagenomic databases some 16s rRNA sequences are classified taxonomically. However, some sequences are reposited which remained as unclassified at the present moment (Table 1).

Conclusion

Therefore, in order to study these databases methodology has to be designed by the researchers for interpreting the data reposited in different databases. In our opinion to serve that purpose they should be interlinked with each other. We feel that uniformity of sequencing ensuring that either fully sequenced or same region of 16s rRNA gene sequence are reposited in different databases should be a prime concern while NGS sequencing. This will go a long way for understanding antibiotic resistance in the environment in a non-culture dependent method. Keeping in mind the widespread emergence of antibiotic resistance strains of pathogenic bacteria throughout the globe we encourage research in this direction.

References

1. Goodwin S, McPherson JD, McCombie WR (2016) Coming of age: Ten years of next-generation sequencing technologies. Nat Rev Genet 17: 333-351.

2. Hodkinson BP, Grice EA (2015) Next-generation sequencing: A review of technologies and tools for wound microbiome research. Adv Wound Care 4: 50-58.

Voelkerding KV, Dames S, Durtschi JD (2010) Next generation sequencing for clinical diagnostics-principles and application to targeted resequencing for hypertrophic cardiomyopathy: A paper from the 2009 William Beaumont Hospital Symposium on Molecular Pathology. J Mol Dia 12: 539-551.

3. Vázquez D (1979) Protein synthesis and translation inhibitors. In: Inhibitors of Protein Biosynthesis 1-14.

4. Kim OS, Cho YJ, Lee K, Yoon SH, Kim M, et al. (2012) Introducing EzTaxon-e: A prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species. Int J Syst Evol Microbiol 62: 716-721.

5. Hong W, Zeng J, Xie J (2014) Antibiotic drugs targeting bacterial RNAs. Acta Pharm Sin B 4: 258-265.

6. Richardson C (2012) Antibiotic resistant characteristics from 16S rRNA. arXiv preprint arXiv:1209.5801.

7. Singh V, Maniar K, Bhattacharayya R, Banerjee D (2017) A proposed method for identification of streptomycin resistance from 16s rRNA sequence by co-localization analysis of fluorescent signals: A step towards detection of streptomycin resistant Mycobacterium species in culture free and gene amplification independent technique. Ind J Clin Biochem 1-2.

8. Leclerc D, Melancon P, Brakier-Gingras L (1991) Mutations in the 915 region of Escherichia coli 16S ribosomal RNA reduce the binding of streptomycin to the ribosome. Nucleic Acids Res 19: 3973-3977.

9. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, et al. (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 72: 5069-5072.