Liuyun Du and Jeffrey Zheng, J Comput Sci Syst Biol 2019, 12:2
DOI: 10.4172/0974-7230.1000298

Journal of
Computer Science & Systems Biology

Research Article Open Access

16S rRNA Sequences of Soil Microbial Diversity on Variant Maps

Liuyun Du* and Jeffrey Zheng

School of Software, Yunnan University, Kunming, China

Abstract

Soil microorganisms affect global climate change, food security, soil ecosystem change, pollutant transformation, and the healthy development of human society. In this paper, the 16S rRNA gene sequence of soil microorganisms is visualized on the variation model. Using soil microorganisms as a two-dimensional statistical map list provides some clues for the study of soil microbial diversity.

Keywords: Variation model; 16S rRNA; Visualization; Soil microbial diversity

Introduction

Soil microbial diversity refers to the population and interspecific differences of soil microbial communities, which is of great significance in maintaining soil quality and ecosystem stability [1]. In recent years, soil microbial diversity has been seriously disturbed by population growth, overexploitation of natural resources, and aggravation of environmental pollution and invasion of alien species. As the main body of soil organisms, soil microbial diversity research can monitor the changes of soil environmental quality, and effectively respond to global climate change and various environmental pollution controls. At the same time, the study of soil microbial community and its function is of great significance to explore the natural life mechanism and its role in the ecosystem [2]. The research and development of soil microbial diversity depends on the breakthrough and improvement of related research methods.

At present, research methods can be roughly divided into traditional microbial isolation and culture technology, biochemical-based technology and modern molecular biology-based technology. In this paper, several methods of soil microbial diversity research are introduced, and the probabilistic model variable statistics method is elaborated in detail. The 16S rRNA gene sequences of bacteria, cyanobacteria and archaea in soil microorganisms are processed to show their distribution characteristics.

Research Techniques of Soil Microbial Diversity

Traditional microbial isolation and culture technology

Traditional methods of isolation and cultivation of soil microorganisms are intuitive, fast and easy to operate. At the same time, they can also provide living and heterotrophic population information. Soil microorganisms were isolated, cultured and counted by artificial nutrient substrates, and their morphological, physiological and biochemical characteristics were classified and identified by pure culture [3].

Biochemical methods

BIOLOG microanalysis and PLFA atlas analysis are commonly used biochemical methods in the study of soil microbial diversity. BIOLOG microanalysis means that microbial cells use 31 kinds of carbon sources to metabolize. By analyzing the fingerprint of soil microbial metabolism characteristics, the changes of soil microbial community under different environments can be reflected. PLFA is an important component of living cell membranes. Different groups of microorganisms can synthesize different PLFAs through different biochemical pathways. Therefore, the types, quantities and relative proportions of microorganisms can be analyzed by different kinds and contents of PLFA in PLFA spectra, mainly for bacterial and fungal community analysis [4].

Modern molecular biology technology

With the rapid development of molecular biology technology, many new techniques have been applied to the study of soil microbial diversity, such as DGGE, T-RFLP, cloning library, high throughput sequencing and so on.

Research and development of soil microbial diversity

Different methods have been widely used in the study of soil microbial diversity, such as the combination of PLFA and BIOLOG, the combination of PLFA and high-throughput sequencing, and the combination of DGGE and high-throughput sequencing. With the development of computer technology, the analysis of soil microbial diversity will make a breakthrough in the interdisciplinary study of soil microorganisms and statistics, bioinformatics and computer science. Meanwhile, single-molecule sequencing technology is also being explored and developed.

Although some technical methods, such as high-throughput sequencing and PCR-based sequencing, require higher DNA gene sequence requirements, for many years, the National Institute of Human Genome Research (NHGRI) has tracked the costs associated with DNA sequencing carried out by the Institute-funded sequencing centers. Figures 1 and 2 accurately reflect the nature of the reduced cost of DNA sequencing. Each chart also shows hypothetical data reflecting Moore's Law, which describes the long-term trends in the computer hardware industry. Therefore, the trend of using DNA research methods in the future is bound to rise in a straight line [5-9].

Variable Maps Theory and Experiments

Introduction of variable graphics theory

In this paper, 16S rRNA gene sequence analysis was used to evaluate

***Corresponding author:** Du L, School of Software, Yunnan University, Kunming, China, Tel: +8615827133602; E-mail: 617279723@qq.com

Received April 08, 2019; **Accepted** April 16, 2019; **Published** April 23, 2019

Citation: Liuyun Du, Jeffrey Zheng (2019) 16S rRNA Sequences of Soil Microbial Diversity on Variant Maps. J Comput Sci Syst Biol 12: 42-46. doi:10.4172/0974-7230.1000298

Copyright: © 2019 Liuyun Du, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

soil microbial diversity [10]. In the gene sequence, base pairing obeys strict complementary symmetry. From the 15 parameters of probability measure (A, T, G, C, A+T, A+G, A+C, T+G, T+C, G+C, A+T+G, A+T+C, A+G+C, A+G+C, A+G+C, T+G+C, A+T+G+C), various one-dimensional or multi-dimensional visualization modes can be formed. In order to get better graphical results, this paper mainly chooses a two-dimensional visualization framework, focusing on statistics of the probability of A+T and A+G parameters. Then, in the visualization model, a 2D map is generated according to the number of parameters [11].

Architecture

The data in this paper is handled by Benli Chai, Ph.D., Center for Microbial Ecology, University of Michigan. First, we screen and tailor the initial data (20183 KB). Finally, we select the size (69 KB) which is easy to study and compare. After we get the format we need, we process the data in the same segment and method. Each data can get a 2D map output [12]. The whole frame structure of 16S rDNA gene sequence is shown in Figure 3.

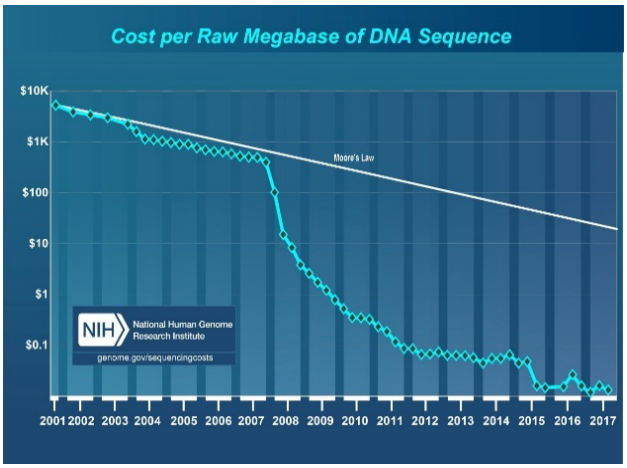

Figure 1: Circos diagram corresponding to the chromosome sequence of Homo sapiens and chimpanzees. (a), (b) images are derived from: http://bioinfo.konkuk.ac.kr/synteny_portal/htdocs/synteny_circos.php.

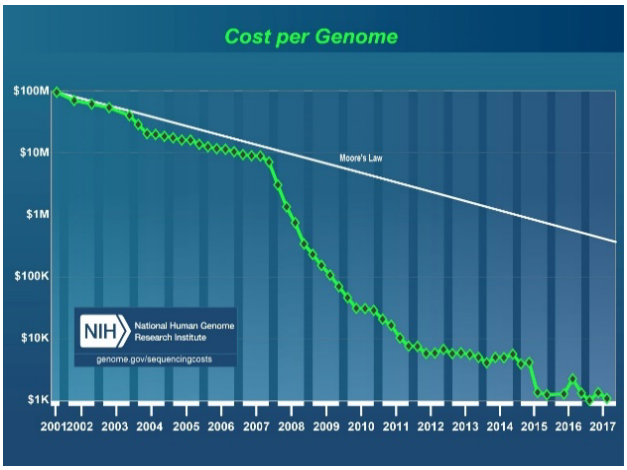

Figure 2: "Cost per genome" - the cost of sequencing a human-sized genome.

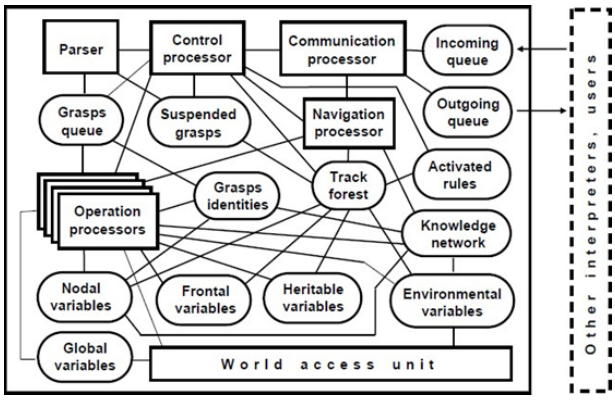

Figure 3: Architecture.

Experiment

The 16S rDNA gene sequence was divided into several equal-length sub-sequences (Figure 4), and the number of data in each sub-sequence was calculated according to the measured parameters [12].

Parameter settings: M: 16S rDNA Sequence Segment Length. Number (A+T)=Number (A)+Number (T); Number (A+G)=Number (A)+Number (G); Mapping to Figure X, Y settings: X=Number (A+T); Y=Number (A+G).

The 16S rDNA sequence was read, N genes were selected as the selected DNA sequence, and the input data were m segments. After quantitative statistics, the normalized measure is obtained. Through Number (A+T) and Number (A+G) of each segment, the position of each point in Cartesian coordinates X and Y can be determined. Each coordinate point is used as the input value of image projection part, and all selected 16S rDNA sequences are collected for image projection. Finally, the characteristic distribution map of 16S rDNA sequence of soil microorganisms can be obtained.

Experimental results

Under the controllable parameter of m, a two-dimensional graph can be formed. By using the method of controlling variables, we can get a better visualization effect when the value of segment length m is taken (Figures 5a-5c).

As shown in the Figure 5, Mycobacterium under Actinobacteria door was selected to compare the visualization effect when m=20, 30, 60.

By counting the number of A+T and A+G in each paragraph, we can see that there are certain distribution characteristics in the three graphs of a, b and c. Segment length is distinct, although the aggregation area is different, but the aggregation shape has some similarities, which shows that A+T, A+G have some regularity in 16S rDNA gene sequence. Observing the clustering effect of this group of graphs, when m is 20, the visualization of Mycobacterium sequence under this model is relatively blurred. When m is 60, the distribution result of Mycobacterium sequence under this model is smaller. When m is 30, the graphics projection effect is better. In the model, the heat of Matplotlib Library in Python is used to realize the visualization of the statistical number, and the two-dimensional image formed by mapping is obtained. The darker the medium color, the denser the distribution of AT and AG. We chose m=30 to observe the visualization results of three species of bacteria belonging to the Firmicutes.

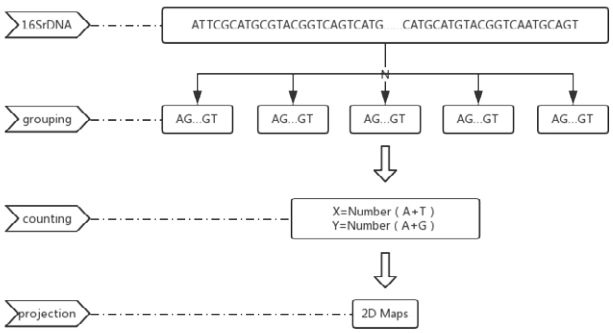

Figure 4: Measurement and projection module.

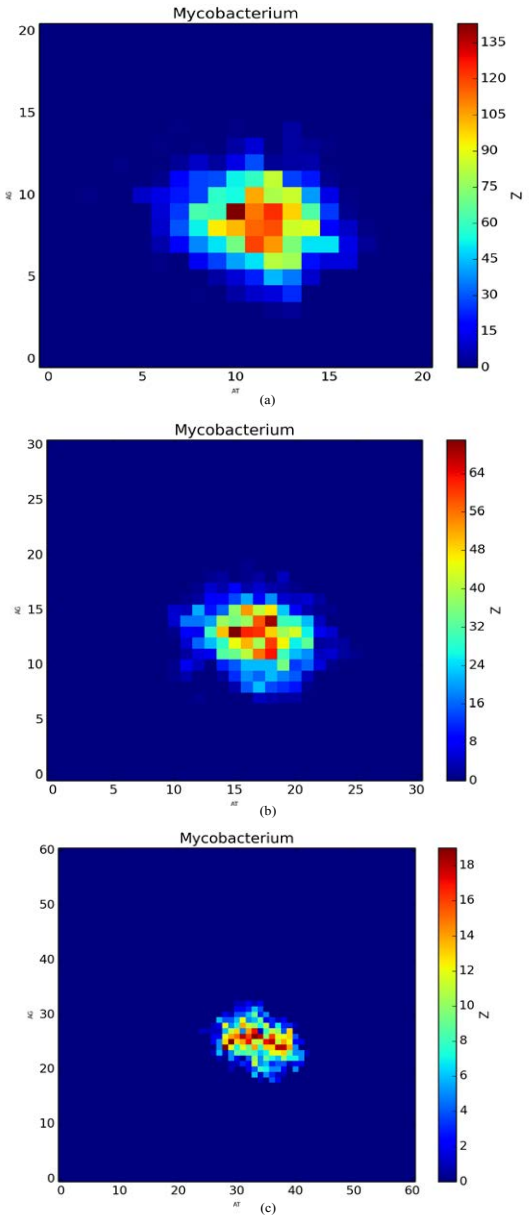

Figure 5: Variant maps of mycobacterium on various segmented lengths (a-c) m={20, 30, 60} (a) m=20; (b)=30; (c)=60.

The following phenomena and characteristics can be observed through the above visualization diagram:

It shows that there is a certain rule in the number of A+T and A+G of bacterial 16S rDNA sequence, that is, they gather in a fixed region and have a central point, and the depth of color expands outward with the central point.

Through the observation of the Figure 6, a kind of Archaea under the door of Euryarchaeota and two kinds of bacteria under the door of Tenericutes were selected for visualization.

From the comparison of the three groups of maps, the similarity of the three microorganisms in gene sequence is that all the results of their maps are projected into (7,27) (7,27) the square region. The difference is that the centers of their visualization results are different, but the centers of the bacteria under the same phylum are almost the same. For example, the central points of the three bacteria under Firmicutes are (17.5,14). Euryarchaeota belongs to archaea. The central points of the pictures of the bacteria under Firmicutes are (17.5,12.5). The central points of the two bacteria under Tenericutes are (17.5,15).

From Figures 6-8, we can see that the three groups of microorganisms are clustered into different groups according to the 16S rRNA gene homology. The results of visualization of 16S rRNA sequences of all microorganisms are similar in shape, aggregation point and so on, which indicates that they do have certain characteristics in gene sequences. It is daring to speculate that there are some other information among microbial gene sequences besides the principle of base complementary pairing.

Discussion

The 16S rRNA located on the small ribosome subunit of prokaryotic cells is about 1540 BP in length. Its structure and base arrangement are moderately complex and easy to sequence and analyze. Moreover, through a large number of studies, 16S rRNA sequence libraries of bacteria have been obtained. Zhao et al. [13] of the Chinese Academy of Sciences developed a microarray based on 16S rRNA gene using ARB software. Through a series of experiments, it was shown that 16S rRNA can be used as a convenient tool for monitoring planktonic bacterial communities in the marine environment.

In recent years, the rapid development of marker gene, genome and macrogenome analysis has greatly expanded our ability to analyze the characteristics of soil microbial community and identify the main factors driving soil microbial community at different spatial and temporal scales. Although most soil microorganisms are still unknown, we can classify soil microorganisms based on their ecological strategies. This should be a proven and effective method for predicting the functional properties of a single microbial group using genomic information. Developments in this area will help to clarify how we can improve soil fertility, increase crop yields, and raise awareness of how terrestrial ecosystems respond to environmental changes through the operation and management of soil microorganisms [14].

A new method for observing soil microbial community is proposed in this paper. This method is not very precise in identifying microbial species, and it is impossible to get which microorganism a gene sequence belongs to by this method. But it can be concluded that they belong to some kind of microorganism. At present, there are some online software (such as Blastn) which can compare sequence homology in database [15]. Now people have recognized that 16S rRNA/rDNA gene sequence can be used to evaluate genetic polymorphism and phylogenetic relationship of organisms, and it can be used as a scientific

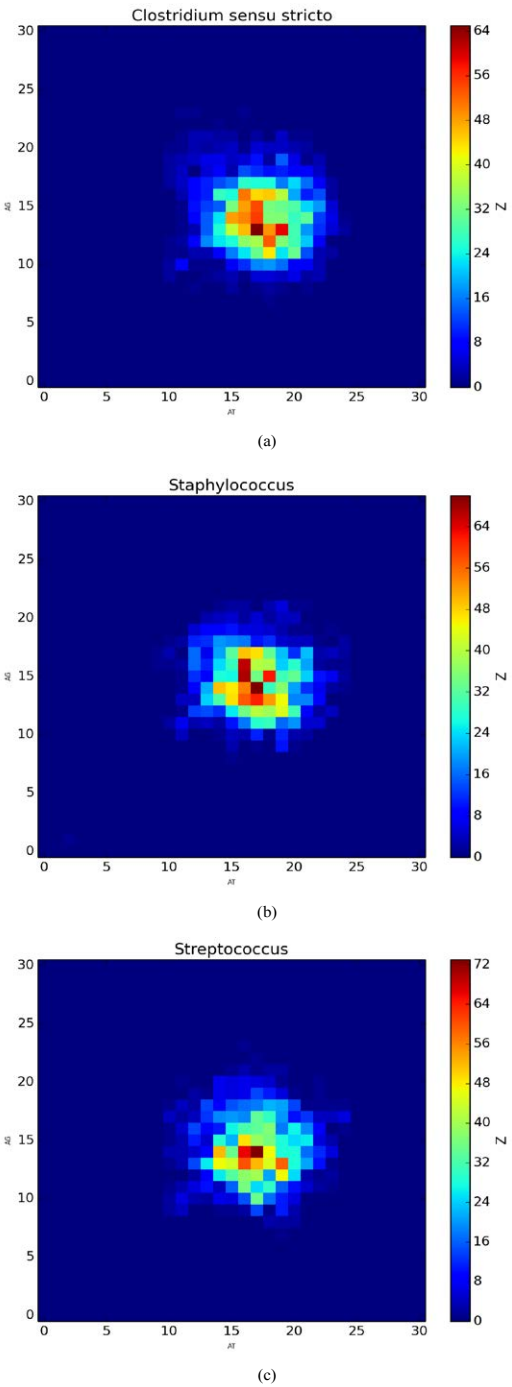

Figure 6: Variant maps of three species of bacteria under the firmicutes on m=30 (a) Clostridium sensu stricto; (b) Staphylococcus; (c) Streptococcus.

and reliable index in bacteriological taxonomy. When classifying and identifying microorganisms in some special environments, it is difficult to obtain their pure culture for physiological and biochemical analysis due to the limitation of isolation and culture technology. The advantages of 16S rRNA sequence analysis are fully reflected.

Soil microbial taxonomy and ecology have revolutionized through the application of molecular methods, and they are increasingly closely related to genomics and functional biology [16,17]. Many research

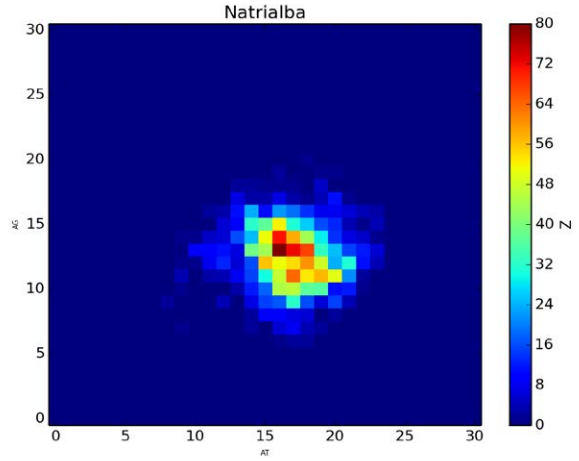

Figure 7: Variant maps of a species of archaea under the Euryarchaeota on m=30 (a) Natrialba.

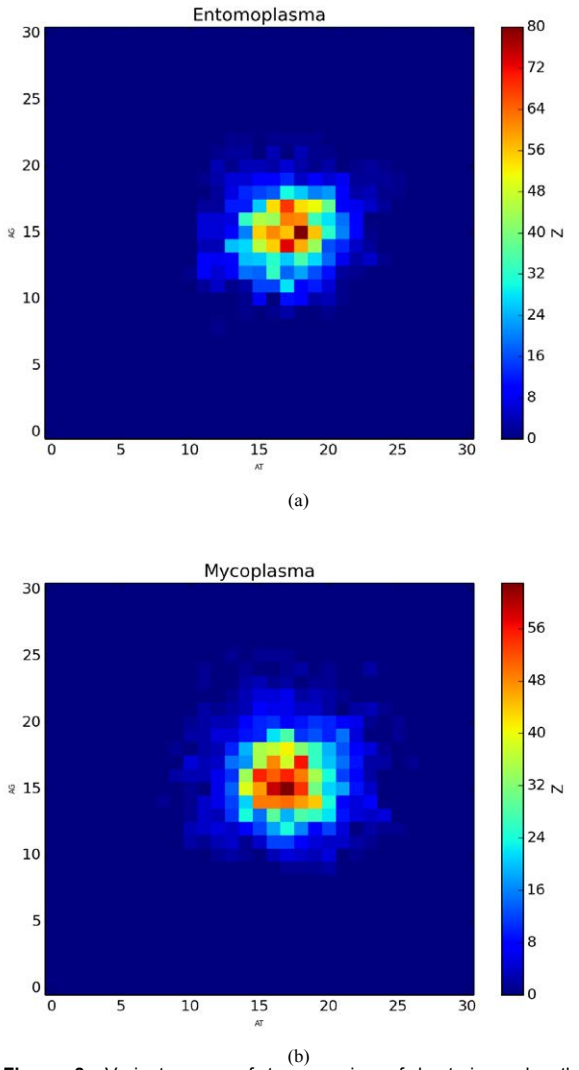

Figure 8: Variant maps of two species of bacteria under the Tenericuteson m=30 (a) Entomoplasma; (b) Mycoplasma.

centers around the world, such as the Center for Microbial Ecology at the University of Michigan, have been working on improving tools for alignment and analysis of DNA and rDNA sequences [18,19]. Not only that, with the rapid development of the Internet, there are many open source data on the website, which can be downloaded for free, and even some websites can provide you with online comparison data [20].

In the future, we can subdivide microbial species by obtaining open source data, optimize the algorithm, and make the visualization results show the fine characteristics of 16S rRNA sequence as much as possible. With specific applications as the starting point, such as focusing on the study of microbial communities in specific locations (hot springs, etc.), we can use the established model to find the distribution and characteristics and solve practical problems.

Conclusion

Soil microbiome is undoubtedly a key component of natural and managed ecosystems. Despite the challenges in soil, a gram of soil can contain thousands of microbial groups, including viruses and members of all three-domain organisms. The structure and composition of microbial communities are relatively complex. There are still some microbial communities whose biological significance and diversity cannot be detected and explained by modern technology and need to be further developed by gene sequencing technology.

Although systematic data from traditional soil microbial diversity research techniques have not been fully integrated with data streams from macrogenomics and roundness studies. But the way of data processing through computer technology will grow rapidly, which is also the value of our research. From the experimental, we can know that there is a certain clustering relationship between soil microbial diversity and gene sequence. The distribution characteristics of 16S rRNA sequence can be described stably, quickly and conveniently by using the gene sequence measurement and visualization model in this paper. This measurement and variation model is easy to understand and operate, and greatly shortens the research process. We will further explore the different combinations of ATGC from the addition, subtraction, multiplication and division of mathematical relationships to find more characteristics and rules, and then combine them with more practical problems related to soil microbial diversity to obtain more meaningful information.

References

1. Yimin T, Yuanqu H, Wenping G (2014) Advances in soil microbial diversity based on molecular technology. Journal of Central South Forestry University 10: 1-9.
2. Huiqing C, Xiaochen L, Xuefeng Y (2018) Advances in soil ecosystem microbial diversity. Earth and Environment.
3. Zhenjiang J, Hongwei Z, Qiang L (2016) Comparison of microbial biomass, biomass and soil enzyme activities between paddy fields originating from karst cave wetland and upland soils. Environmental Science 37: 335-341.
4. Pengxian T, Jingying Y, Zhiying X (2016) BIOLOG in the study of functional diversity of soil microbial communities. Southern Agriculture 10: 242-243.
5. Wetterstrand KA (2018) DNA sequencing costs: Data from the NHGRI Genome Sequencing Program.
6. Mardis ER (2011) A decade's perspective on DNA sequencing technology. Nature 470: 198-203.
7. Metzker ML (2010) Sequencing technologies-the next generation. Nature Genetics 11: 31-46.
8. Stein DL (2010) The case for cloud computing in genome informatics. Genome Biology 11: 207-213.
9. Nature (2010) Human genome at ten: the sequence explosion. Nature 464: 670-671.
10. Dunbar J, Ticknor LO, Kuske CR (2000) Assessment of microbial diversity in four southwestern United States soils by 16S rRNA gene terminal restriction fragment analysis. Applied & Environmental Microbiology 66: 2943-2950.
11. Wanzhu W, Zhijie Z (2013) Visualization of one-dimensional segmented measurement distribution of DNA sequence. Journal of Yunnan University (Natural Science Edition) 35: 1-6.
12. Yuyuan M, Zheng J, Liu W (2017) Mapping whole DNA sequence on variant maps. ACM, pp: 1037-1040.
13. Zhao W, Wang J, Liang Y, Huang Z (2017) Development of a 16S rRNA gene-based microarray for the detection of marine bacterioplankton community. Acta Oceanologica Sinica 36: 106-114.
14. Fierer N (2017) Embracing the unknown: disentangling the complexities of the soil microbiome. Nature Reviews Microbiology 15: 579-590.
15. Feizhou Z, Liyu C, Hanchun C (2013) 16S rRNA gene sequencing for identification of pathogenic bacteria. Journal of Central South University: Medical Edition 38: 1035-1041.
16. Shun W, Ji W, Huo-Wang C (2006) Study on visual representation of fractal-based DNA sequences. p: 7.
17. Xikui L, Yan L, Jin X (2004) Two-dimensional representation and related analysis of DNA sequences. Progress in Natural Science 14: 1032-1038.
18. Cole R, Farris RJ, Wang Q, Chai B, Wang Q, et al. (2019) The ribosomal database project: sequences and tools for high-throughput rRNA. Analysis 1: D294-D296.
19. Kitts CL (2001) Terminal restriction fragment patterns: A tool for comparing microbial communities and assessing community dynamics. Current Issues in Intestinal Microbiology 2: 17-25.
20. Yuanke S, Xiaomin Z, Tonghui L, Yanxue L, Changyong C (2017) Analysis and research of environmental microorganisms in clean rooms based on 16S rRNA. Journal of Food Safety Quality Detection 8: 4161-4168.