MGI's 16S rDNA Sequencing Solution based on the DNBSEQ-E25 Platform Enpowers the Study of Deep-sea Microbiome in the South China Sea

The extreme environment in the deep sea provides unique study conditions for life sciences, materials science and other fields, which can better understand the origin and diversity of life on Earth[1]. BGI Group, together with the Institute of Deep-sea Science and Engineering, CAS, Shanghai Jiao Tong University, Hong Kong University of Science and Technology and other institutions, went to the cold seep area of the South China Sea for deep-diving scientific research, and completed the microbial community analysis of deep-sea sediment samples utilizing the 16S rDNA sequencing product based on the MGI's DNBSEQ-E25 platform, detecting bacteria and archaea that have not been cultured by humans, and providing an important reference for in-depth understanding of the biogeochemical mechanisms of important deep-sea microbial groups.

Recommended application: Marine environment microbiome;
Recommended model: DNBSEQ-E25RS;

- **Fast detection**

  ATOPlex 16S V4 rDNA library preparation kit with PE150 sequencing and self-developed MetaSIS analysis software enable the report output on the same day of sampling.

- **Wide range of species monitoring**

  The solution enables the monitoring of species include bacteria, fungi, algae, plants, zooplankton, benthic animals and fish, fully meeting various application needs.

- **No fear of sea turbulence**

  DNBSEQ-E25's unique anti-turbulence design can provide reliable and accurate sequencing data at sea.

- **One-stop product solution**

  This combination product integrates the entire process of nucleic acid extraction, library preparation, sequencing and analysis, making it a truly one-stop product solution.

- **High degree of automation**

  With the automated nucleic acid extractor, automated sample preparation system and analysis software, automated sample preparation and data analysis can be realized, minimizing human intervention.

Background

Environmental DNA sequencing is to sequence biological DNA in environmental media (water, soil, sediment, etc.) to achieve qualitative and abundance analysis of biological communities, and use the response of biological individuals, populations or communities to environmental changes to assess environmental conditions. It is an important means to assess ecological environmental conditions and biodiversity, and is also an important tool in the field of environmental research. The sensitivity advantage of high-throughput sequencing can detect DNA with extremely low content in the environment[2].

The deep sea refers to the area deeper than 1,000 meters below sea level. It is a dark, cold and high-pressure environment that has long been considered unsuitable for biological survival, and is known as the "forbidden zone of life". However, in some areas, such as cold seeps or hydrothermal vents, fluids rich in hydrogen sulfide or methane are leaked into deep waters, and these chemicals provide a source of energy for symbiotic microorganisms. The deep sea is rich in resources, including mineral resources, biological resources, etc. With the increasing depletion of land resources, the deep sea has become a new frontier for resource development. Deep-sea ecosystems also have an important impact on the earth's environment. Understanding and protecting the deep-sea environment is of great significance for maintaining global ecological balance and responding to climate change[3].

The Haima Cold Seep is located in the southeastern waters of the South China Sea. It is one of the largest deep-sea cold seep ecosystems discovered in China. It provides an important field for studying scientific issues such as biodiversity and evolution in the deep sea of the South China Sea[4]. Recently, BGI Group, together with several scientific research institutions, went to the cold seep area of the South China Sea for deep-diving scientific research, completed a microbial community sequencing analysis of a deep-sea seabed mud sample utilizing the 16S rDNA sequencing product solution based on the MGI's DNBSEQ-E25RS platform. The study detected bacteria and archaea that have not been cultured by humans, such as *Chloroflexus*, *Acidobacteria*, *Proteobacteria*, etc., and carried out the functional gene annotation, providing an important reference for in-depth understanding of the biogeochemical mechanisms of important deep-sea microbial groups.

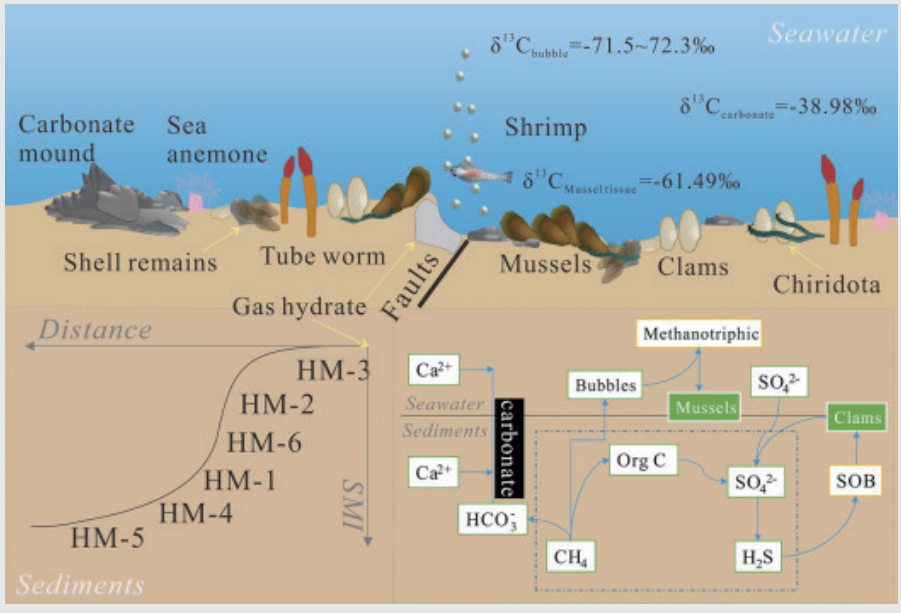

Figure 1. "Haima Cold Seep" ecosystem[4]

Study Description

To study what substances can be produced by the microbial community of deep-sea cold seeps in the process of geochemical coupling reactions, thus constituting the most primitive elements of the origin of life, the BGI's deep-sea scientific research team boarded the "Deep Sea Warrior"[5] manned submersible and dived to the seabed 1,000 meters below the sea level in the Haima Cold Seep and Ganquan Platform in the South China Sea to collect sediment samples, and completed the 16S V4 region sequencing based on the ATOPlex 16S V4 rDNA library preparation kit and DNBSEQ-E25RS sequencer. The sequencing data were analyzed by MetaSIS software. Bacteria and archaea that have not been cultured by humans, such as *Chloroflexus, Acidobacteria, Proteobacteria*, etc., were detected, and the functional gene annotation was performed, providing an important reference for in-depth understanding of the biogeochemical mechanisms of important deep-sea microbial groups.

Figure 2. The BGI's scientific research team boarded the "Deep Sea Warrior" submersible to operate in areas such as the Haima Cold Seep.

Figure 3. MGI sequencing platform installed on the research vessel.

Materials and Methods

Sample collection

The sediment columns were collected from the seabed using a customized Pushcore by the "Deep Sea Warrior" manned submersible. The sediment columns were fixed to a customized sediment sample splitter in the laboratory, manually dispensed into sterile containers, and stored in a refrigerator at -80°C.

Library preparation and sequencing

The nucleic acid was extracted from the sediment samples using the MGI's MGIEasy environmental DNA extraction kit and the MGISP-NE32 automated nucleic acid extractor. The extracted DNA samples and standard control were prepared into PCR libraries using the ATOPlex 16S V4 rDNA library preparation kit, followed by circularization and DNB preparation, and finally sequenced on the DNBSEQ-E25 platform with pair-ended 150 (PE150) sequencing recipe. It is worth noting that when the sample size is large, the process can be done with MGI's automated extraction and library construction platform, which can greatly save manpower and improve efficiency.

Sample ID	Original ID	Source
Sample-1	SPL240724174820-1	Standard control
Sample-2	SPL240724174820-2	Standard control
Sample-3	SPL240724174820-3	Deep-sea sediments
Sample-4	SPL240724174820-4	Deep-sea sediments

Table 1. Sample information.

Bioinformatics Analysis

The sequencing data were analyzed by MetaSIS software after splitting the Barcode. The software first processed the raw data to obtain Clean Reads, then performed OTU cluster and filter, and finally performed in-depth analysis. The analysis content includes OTU analysis, species composition analysis, functional gene prediction, Alpha diversity analysis, Beta diversity analysis, etc.

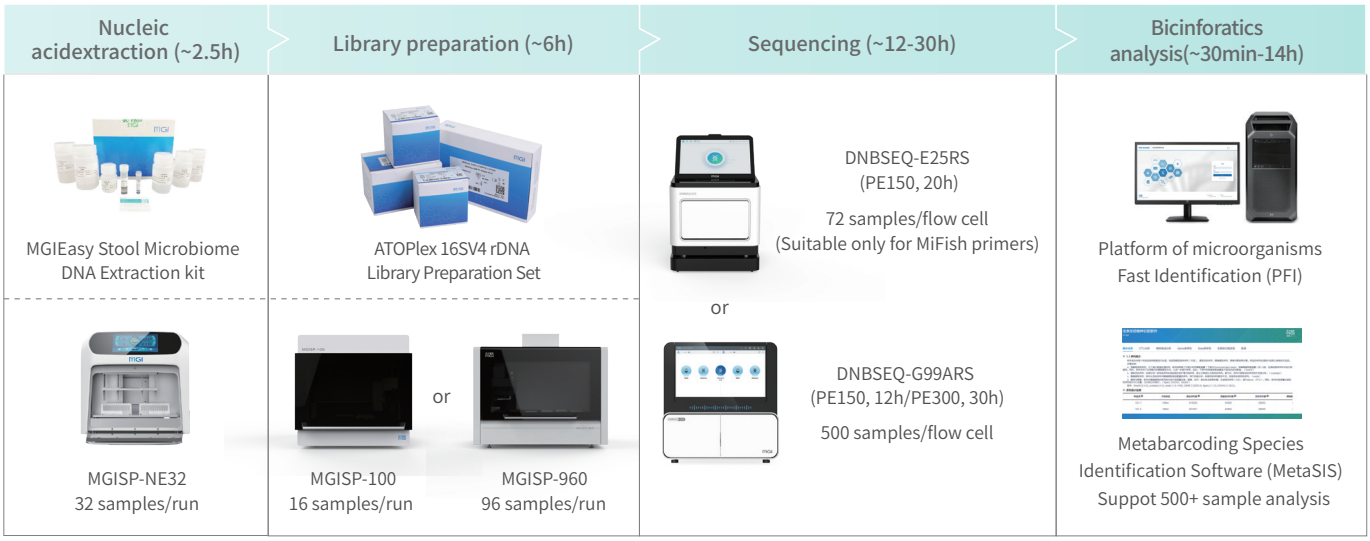

Figure 4. the MGI environmental DNA metabarcoding sequencing package.

Results

DNBSEQ-E25 provides high-quality sequencing data

Although sequencing was performed at sea, it can be seen from Table 2 and Figure 5 that the DNBSEQ-E25 platform produced high-quality sequencing data for both standard samples and deep-sea sediment samples. The Q30 > 96% and data output > 29M reads, which fully met the needs of downstream deep analysis, indicating that the DNBSEQ-E25 sequencing platform is completely unaffected by sea turbulence and can still complete high-quality sequencing stably.

QC indexes	Values
TotalReads(M)	29.43
ESR%	88.25
Q30%	96.89
SplitRate%	93.72
CycleNumber	320

Table 2. QC information of sequencing data.

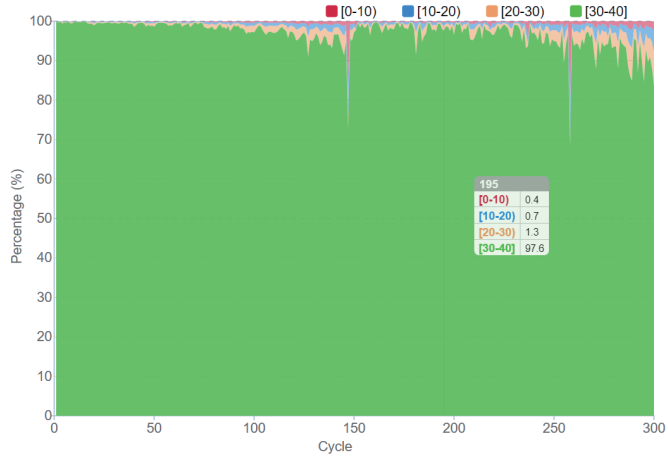

Figure 5. Base sequencing quality distribution diagram.

Sample OTU Analysis

OTU is a unified mark set for a certain taxon in phylogenetic or population genetics research for the convenience of analysis 6. The MetaSIS software clusters sequences with 100% similarity into one OTU. After obtaining the OTU, the QIIME 2 module is used to compare and annotate the OTU sequence with the annotation database to obtain the species identification results at different classification levels. By counting the species identification of OTUs, the number of biological community components of each sample at different classification levels (composed of kingdom, phylum, class, order, family, genus, and species) was obtained, as shown in Table 3.

Species Composition Analysis

Without considering the abundance, based on the species identification results of all samples, the GraphlAn module is used to generate a species composition visualization graph, as shown in Figure 6. This graph can intuitively visualize the species composition of each classification level of the sample, helping users discover dominant biological groups. From the inside to the outside, each circle represents a level, which is the level of phylum, class, order, family, genus, and species. Different phyla are marked with different colors. The outermost circle text shows the name of the previous level of the terminal classification level of the top 20 abundances, and the branches of the classification level are marked with darker colors.

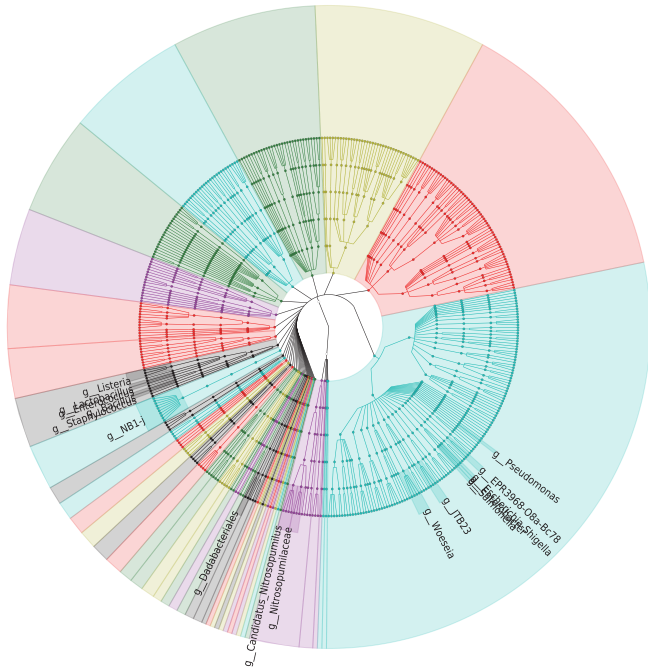

Figure 6. Species Composition Chart.

Sample ID	Number of OTU types	Number of OTU sequence	Number of identifiable OTU types	Number of identifiable OTU sequences	Kingdom	Phylum	Class	Order	Family	Genus	Species
Sample-1	11	41065	10	41030	1	2	2	5	7	9	9
Sample-2	13	40996	11	40956	1	2	2	5	7	9	11
Sample-3	843	32742	840	32709	2	36	86	159	182	190	54
Sample-4	840	31041	837	30980	2	40	88	158	177	189	52

Table 3. Statistical table of OTU.

Table 4 shows the relative abundance of species at different classification levels (phylum, class, order, family, genus, and species) of each sample. Due to space limitations, this article only shows the top ten species. It can be seen that bacteria and archaea that have not been cultured by humans have been detected, such as *Chloroflexus*, *Acidobacteria*, *Proteobacteria*, etc.

Select the top 18 species identification results of each sample at the species classification level to generate a stacked bar chart, which can intuitively show the composition of the biological community at different classification levels in the sample, and better understand the differences and similarities between samples. As shown in Figure 7, the abscissa stands for the sample, and the ordinate stands for the relative abundance of the annotated species. Each bar graph represents a sample, and each taxon is distinguished by color. The longer the bar, the higher the relative abundance of the taxon in the corresponding sample.

level	Latin name	Sample-1	Sample-2	Sample-3	Sample-4
Species	uncultured_bacterium	15.3449%	15.6973%	57.8373%	56.6946%
Species	Lactobacillus_fermentum	22.9442%	21.9089%	0.0000%	0.0000%
Species	Bacillus_subtilis	13.8557%	13.0481%	0.0000%	0.0000%
Species	Klebsiella_sp.	11.7865%	13.7098%	0.0000%	0.0000%
Species	Escherichia_coli	12.0741%	12.4060%	0.0000%	0.0000%
Species	uncultured_archaeon	0.0000%	0.0000%	13.0820%	14.3544%
Species	Listeria_monocytogenes	8.6717%	8.4823%	0.0000%	0.0000%
Species	Pseudomonas_aeruginosa	7.4994%	7.2029%	0.0000%	0.0000%
Species	Enterococcus_faecalis	6.3295%	5.8331%	0.0000%	0.0000%

Table 4. Statistical table of relative abundance of species.

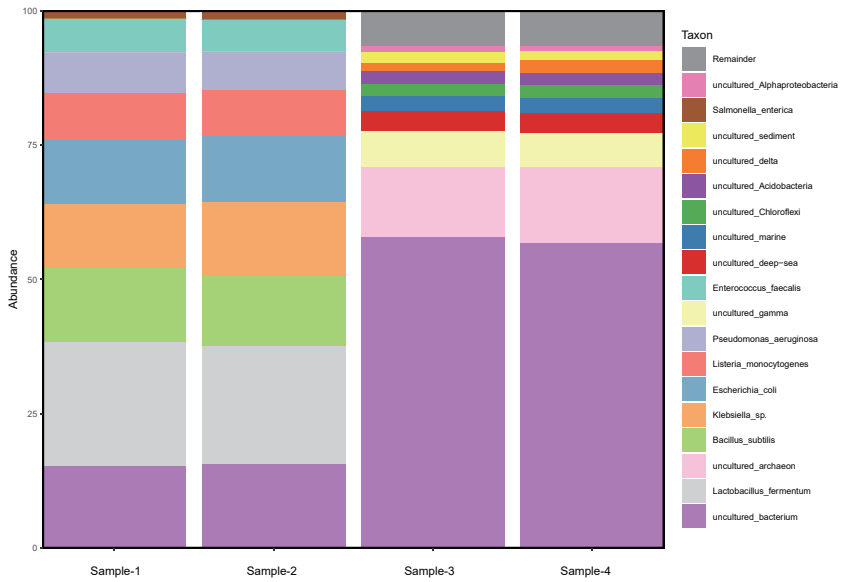

Figure 7. Species distribution stacking diagram.

Select the top 50 species identification results of each sample at the species level to generate a heat map, which can intuitively show the composition of the biological community at different classification levels in the sample, and better understand the differences and similarities between samples, as shown in Figure 8. The abscissa stands for the sample, and the ordinate stands for the annotated species name. The color reflects the abundance information of the annotated species in the sample. The numerical value can be intuitively represented by the depth of color, and the color gradient and similarity can be used to reflect the similarities and differences of species distribution of multiple samples at various classification levels.

To more intuitively display the species identification information of all samples at different classification levels (phylum, class, order, family, genus, and species), the software generates a Krona diagram. As shown in Figure 9, the circles represent different classification levels from inside to outside. The size of the fan represents the relative proportion of different identification results.

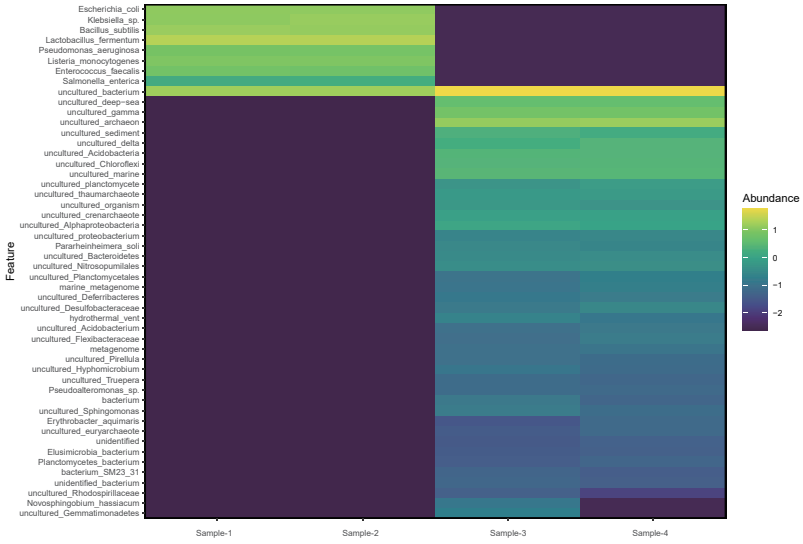

Figure 8. Species distribution heat map.

Figure 9. Species distribution Krona diagram.

Functional Gene Prediction

The research team also annotated the functional genes in the samples, using COG functional gene annotation, which is usually used to annotate new genome sequences. By comparing the new gene sequence with the protein sequence in the COG database, the possible function and structure of the new gene sequence can be determined 7. Through comparison and annotation, some high-frequency known and unknown functional gene clusters were discovered, such as the S gene cluster, the R gene cluster, and the J gene cluster related to transcription, ribosome structure and synthesis, as shown in Figure 10, which provides an important reference for the subsequent study of gene functions of deep-sea microorganisms.

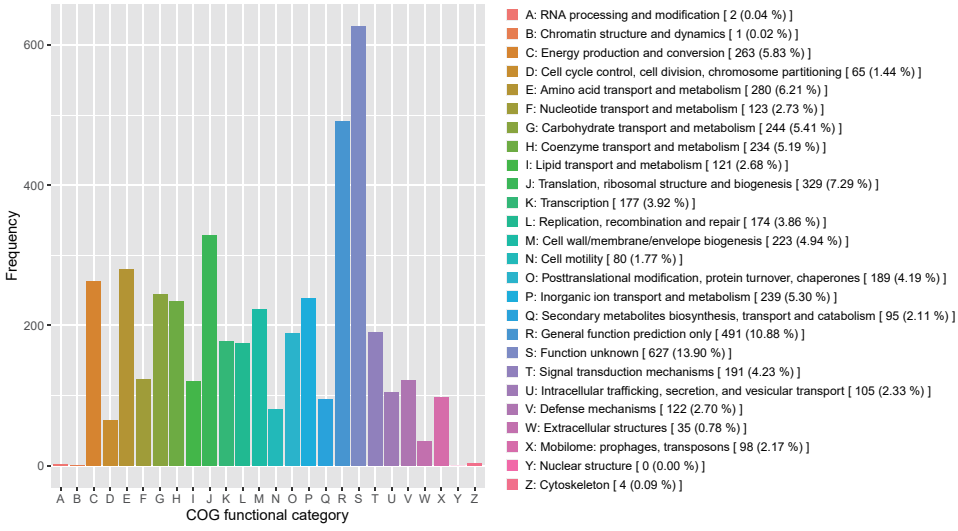

Figure 10. Functional gene prediction diagram.

Summary

MGI's 16S V4 sequencing product solution based on the DNBSEQ-E25 platform has a fast detection speed and can complete data analysis on the day of sampling at sea. With no fear of sea turbulence, the sequencing data quality is excellent and stable. The powerful software has discovered valuable bacterial species and functional genes. In short, the MGI's microbial sequencing product has provided reliable and efficient tool support for this deep-sea scientific research in the South China Sea.

Genetic Sequencer DNBSEQ-E25RS

DNA Sequencing Library Preparation System MGISP-100RS

References

1. Acinas, S. G. *et al.*, Deep ocean metagenomes provide insight into the metabolic architecture of bathypelagic microbial communities. *COMMUN BIOL* 4 604 (2021).
2. Ashish Sahu, N. K. C. P., Environmental DNA (eDNA): Powerful technique for biodiversity conservation. *J NAT CONSERV* (2023).
3. Michael A. Rex, R. J. E., Deep-sea Biodiversity: *Pattern and Scale*. (2010).
4. 梁永志 & 冯强强, 海洋日专题|“海马冷泉”：深蓝世界的生命交响曲. *http://www.gmgs.cgs.gov.cn/hykp/hyzs/202406/t20240617_764071.html* (2024).
5. 佚名, “深海勇士”号4500米载人潜水器正式交付 落户中科院. *https://www.cas.cn/jh/201801/t20180119_4632981.shtml* (2017).
6. Llado, F. S., Vetrovsky, T. & Baldrian, P., The concept of operational taxonomic units revisited: genomes of bacteria that are regarded as closely related are often highly dissimilar. *FOLIA MICROBIOL* 64 19 (2019).
7. Tatusov, R. L., Galperin, M. Y., Natale, D. A. & Koonin, E. V., The COG database: a tool for genome-scale analysis of protein functions and evolution. *NUCLEIC ACIDS RES* 28 33 (2000).

Recommended Ordering Information

Category	Product	Cat. NO.
Instruments	Genetic Sequencer DNBSEQ-E25RS	900-000537-00
	Genetic Sequencer DNBSEQ-G99ARS	900-000609-00
	Automated Nucleic Acid Extractor MGISP-NE32RS	950-000020-00
	DNA Sequencing Library Preparation System MGISP-100RS	900-000206-00
	MGISP-960RS Automated Sample Preparation System(config 9)	900-000154-00
Sample Extraction	MGIEasy Environmental DNA Extraction Kit(96 Preps)	940-002484-00
	MGIEasy Environmental DNA Extraction Kit (384 Preps)	940-002486-00
Library Prep	ATOPlex 16SV4 rDNA Library Preparation Set (576 RXN)	940-002559-00
	ATOPlex 16SV4 rDNA Library Preparation Set (96 RXN)	940-002560-00
	Standard Library Reagent (PCR Product)	1000027585
Sequencing Reagents	DNBSEQ-E25RS High-throughput Sequencing Set(FCL PE150)	940-000567-00
	High-throughput Sequencing Set（G99 SM FCL PE150）	940-000410-00
Bioinformatics	Metabarcoding Species Identification Software	970-000417-00
	Platform of microorganisms Fast Identification	900-000393-00

MGI Tech Co.,Ltd

Building 11, Beishan Industrial Zone, Yantian District, Shenzhen, CHINA, 518083

Version: December 2024

+86-4000-688-114

en.mgi-tech.com

MGI-service@mgi-tech.com

Authors: Jia Lu

Editor-in-Charge: Qiwei Wang

Reviewer: Yao Jiang