Environmental DNA Metabarcoding Sequencing Package

——Accurate and powerful, providing an ideal tool for environmental monitoring and scientific research

Features

- **One-stop product solution**
  Integrated sample treatment, library preparation, sequencing, and analysis, a true one-stop product solution.
- **Wide range of species covered**
  The monitored species cover bacteria, fungi, algae, plants, planktonic animals, benthic animals, and fish, meeting various application needs.
- **Excellent data quality**
  The unique DNBSEQ technology provides high-quality sequencing data.
- **Automation friendly**
  Combined with automated nucleic acid extraction instruments, automated sample preparation systems and analysis software, to achieve automated sample preparation and data analysis, save labor.
- **Powerful analysis function**
  Self-developed software provides various analysis functions, such as OTU analysis, species composition analysis, alpha diversity analysis, beta diversity analysis, biomarker analysis, and more.
- **Support large-scale parallel sequencing**
  Up to 4608 barcodes can be provided for library preparation, meeting the demands of ultra-high-throughput sequencing.

Introduction

Strengthening environmental monitoring is crucial for environmental protection. Environmental biomonitoring utilizes the responses of biological individuals, populations, or communities to assess the environmental conditions. It can be applied to monitor atmospheric, aquatic, and soil environments, serving as an important approach to evaluate natural environmental conditions and biodiversity, and is also a significant aspect of environmental research. Environmental DNA sequencing technology is one of the most revolutionary technologies in the field of ecological and environmental sciences since the 20th century. It involves extracting DNA from environmental media such as water, soil, and sediments, followed by PCR amplification and high-throughput sequencing of specific DNA fragments in the genome, enabling qualitative and quantitative analysis of the biological community.

MGI Environmental DNA metabarcoding sequencing package is a combination of self-developed reagents for nucleic acid extraction, ATOPlex multiplex PCR library preparation, MGISP-960 automated sample preparation system, DNBSEQ-G99 and DNBSEQ-E25 sequencers, and MetaSIS analysis software. This comprehensive solution covers the entire process from sample processing to analysis reporting. It enables rapid and accurate analysis of the biological community structure in environmental samples, providing tool for ecological environmental monitoring and environmental scientific research.

Table 1. Product parameters

Panel	Applicable species types	Recommended read length	Recommended data per sample	Sequencer	Recommended sample number /FC
MiFish	fish	PE150	≥65K reads	DNBSEQ-G99/E25	500/72
Ac12S	fish	PE300		DNBSEQ-G99	500
16S V3V4	bacteria				
COI	benthic animals, zooplankton				
18S V4	algae, invertebrates				
ITS1	Fungus				

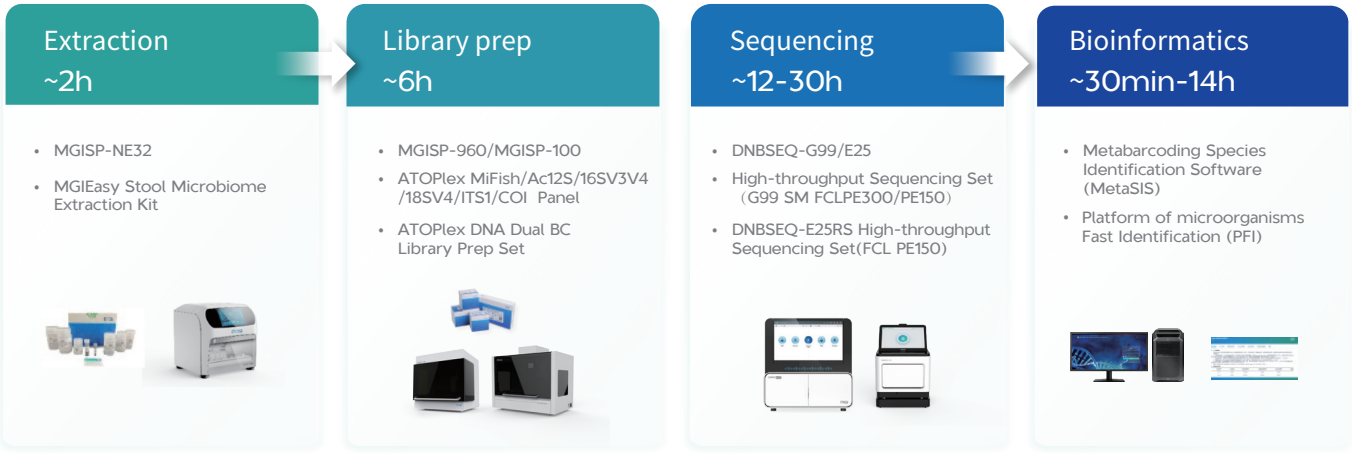


Figure1 Workflow

Performance

The following three sets of data were to evaluate the performance of the solution, such as repeatability, species detection rate, and quantitative assessment of species abundance, using simulated samples with known species composition and abundance.

Samples S1-S3 are three replicates of mixed samples containing 8 bacterial DNA standards. The samples were prepared to 16S V3V4 library, sequenced and analyzed on the DNBSEQ-G99. The results are shown in Table 2 and Figure 2. all expected species were detected, and the measured abundance similar with the theoretical abundance. The three replicates showed minimal deviation.

Table 2. Species abundance analysis results

Species	Theoretical abundance	S1 Measured abundance	S2 Measured abundance	S3 Measured abundance
Pseudomonas	4.20%	6.83%	7.86%	6.94%
Escherichia	10.10%	12.62%	9.15%	13.29%
Salmonella	10.40%	11.48%	8.12%	9.97%
Lactobacillus	18.40%	16.24%	20.45%	16.58%
Enterococcus	9.90%	10.63%	11.14%	13.02%
Staphyloccus	15.50%	12.44%	18.34%	12.58%
Listeria	14.10%	15.40%	12.56%	13.97%
Bacillus	17.40%	14.36%	12.37%	13.64%

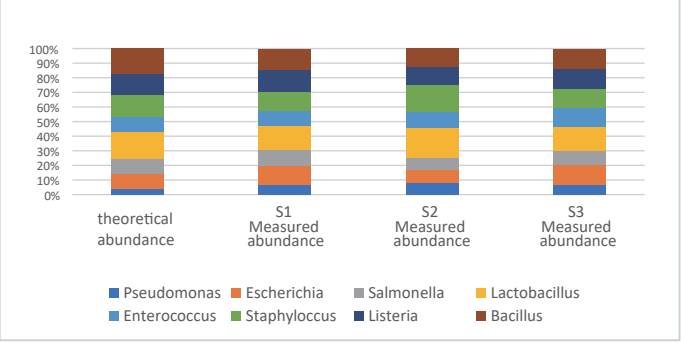


Figure 2. Species composition and abundance stacked histogram

Samples S4-S6 are three replicates of mixed samples containing 10 species fish tissue. The samples were prepared to MiFish library, sequenced on the DNBSEQ-E25 platform and analysis. The results are shown in Table 3 and Figure 3. From the results, we can see all expected species were detected, and the measured abundance similar with the theoretical abundance. Additionally, the three replicates showed excellent reproducibility.

Table 3. Species abundance analysis results

Species	theoretical abundance	S4 measured abundance	S5 measured abundance	S6 measured abundance
Acanthopagrus	15%	21%	21%	21%
Carassius cuvieri	3%	2%	2%	2%
Channa maculata	6%	8%	8%	8%
Ctenopharyngodon idella	10%	12%	12%	13%
*Hypophthalmichthys nobilis /Hypophthalmichthys molitrix**	5%	4%	4%	4%
Larimichthys crocea	16%	14%	15%	15%
*Lateolabrax japonicus /Laterolabrax maculatus***	17%	15%	15%	14%
Oreochromis niloticus	5%	3%	3%	3%
Sciaenops ocellatus	12%	9%	9%	9%
Sebastiscus marmoratus	12%	12%	12%	12%

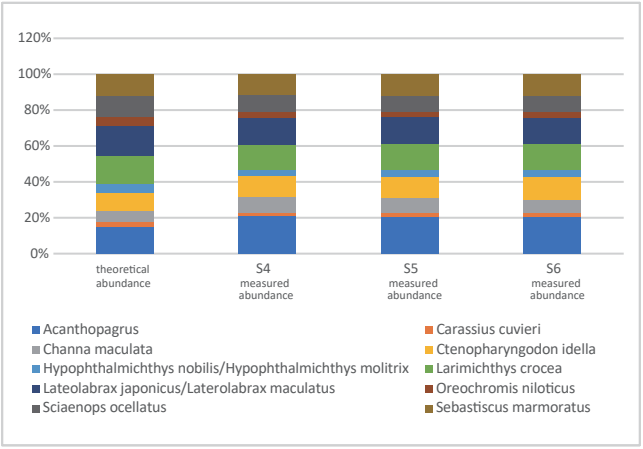


Figure 3. Species composition and abundance stacked histogram

Samples S7 and S8 are mixed samples containing 10 fish tissue species but with different proportions. The samples were prepared to AC12S library, sequenced and analyzed on the DNBSEQ-G99. The results are shown in Table 4 and Figure 4. we can see all expected species were detected, and the measured abundance similar with the theoretical abundance.

Table 4. Species abundance analysis results

Species	S7 theoretical abundance	S7 measured abundance	S8 theoretical abundance	S8 measured abundance
Acanthopagrus	3.79%	6.81%	1.94%	1.76%
Carassius cuvieri	21.00%	21.97%	17.19%	19.01%
Channa maculata	14.73%	19.70%	15.92%	23.68%
Ctenopharyngodon idella	10.32%	8.65%	13.03%	13.35%
*Hypophthalmichthys nobilis /Hypophthalmichthys molitrix**	11.47%	14.86%	9.88%	14.36%
Larimichthys crocea	4.01%	4.30%	2.01%	1.31%
Lateolabrax maculatus	4.24%	3.48%	6.13%	5.41%
Oreochromis niloticus	18.49%	11.52%	15.98%	8.49%
Sciaenops ocellatus	3.97%	2.32%	5.95%	3.39%
Sebastiscus marmoratus	7.99%	6.39%	11.97%	9.25%

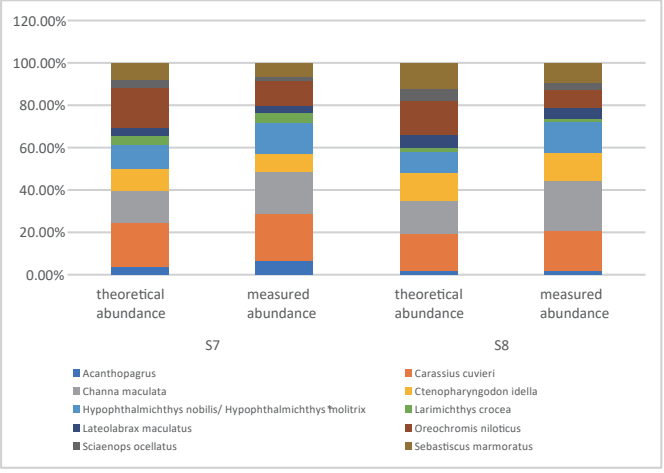


Figure 4. Species composition and abundance stacked histogram

Real environmental samples mainly include water, sediment, and soil. Water are suitable for monitoring fish, zooplankton, phytoplankton, and benthic organisms. Sediment are suitable for monitoring benthic organisms and fish. Soil contain a large number of microbial communities. In the following tests, we detected fish in water, benthic organisms in sediment, and fungi in soil.

Samples D3-D8 are water samples collected from 6 different locations in Donghu Lake, Wuhan. Each location was sampled three times. After filtration, the filter membrane were processed using MGI's solution for nucleic acid extraction, MiFish library preparation, DNBSEQ-G99 sequencing and analysis. The results of quality control and data processing are shown in the table 5. We can see the Q30 > 94% for all samples, indicating excellent sequencing quality. Each sample was trimmed to 50,000 reads and conducted to low-quality filtering, denoising, merging, and chimeric removal to generate feature sequences, also known as OTUs.

Table 5. Quality control and data processing results

Sample ID	Raw reads	Target reads	Down sampled reads	Filtered reads	Q30(%)	Denoised reads	Merged reads	Chimeric reads	Feature reads
D3-1	815929	58832	50000	47434	96.17	47220	43499	1961	41538
D3-2	831507	58949	50000	47327	96.09	47104	42609	1860	40749
D3-3	845240	59063	50000	47385	95.71	47106	44854	3223	41631
D4-1	856348	58529	50000	47643	96.77	47333	41628	1471	40157
D4-2	734924	58997	50000	47638	96.71	47486	45974	2068	43906
D4-3	856431	58840	50000	47626	96.99	47352	44918	2304	42614
D5-1	749180	59064	50000	47522	96.03	47258	42392	2887	39505
D5-2	724429	58958	50000	47553	96.46	47352	42476	2329	40147
D5-3	740403	58817	50000	47454	96.34	47225	41429	2633	38796
D6-1	711722	58710	50000	46795	94.86	46558	42725	4450	38275
D6-2	683555	58926	50000	47461	96.3	47213	43249	5111	38138
D6-3	667056	58724	50000	47513	96.43	47189	42167	5503	36664
D7-1	1136152	58868	50000	47186	95.62	46946	43338	1567	41771
D7-2	673020	59033	50000	47415	96.62	47231	44439	2473	41966
D7-3	989925	58825	50000	47087	95.2	46687	44323	2607	41716
D8-1	895021	58825	50000	47236	95.67	47036	45033	4283	40750
D8-2	862580	58705	50000	47422	95.94	47122	44556	2717	41839
D8-3	631941	58872	50000	47487	96.13	47208	44753	2410	42343

After obtaining the OTUs, the system aligns the OTU sequences with the annotation database to get the identification results at different taxonomic levels, including kingdom, phylum, class, order, family, genus, and species. The results are shown in Table 6.

Table 6. Species abundance analysis results

Sample ID	OTU number	OTU tag number	Kingdom	Phylum	Class	Order	Family	Genus	Species
D3-1	37	41538	1	1	1	3	7	17	26
D3-2	43	40749	1	1	1	4	8	18	27
D3-3	43	41631	1	1	1	3	7	17	25
D4-1	75	40157	1	1	1	3	6	12	16
D4-2	95	43906	1	1	1	5	9	16	19
D4-3	98	42614	1	1	1	4	7	14	18
D5-1	60	39505	1	1	1	4	7	18	25
D5-2	59	40147	1	1	1	4	9	25	33
D5-3	60	38796	1	1	1	5	8	22	28
D6-1	40	38275	1	1	1	5	9	16	20
D6-2	48	38138	1	1	1	5	8	14	18
D6-3	49	36663	1	1	1	4	7	14	18
D7-1	40	41771	1	1	1	5	8	14	19
D7-2	46	41966	1	1	1	4	8	13	15
D7-3	46	41716	1	1	1	5	8	15	19
D8-1	49	40750	1	1	1	4	7	17	24
D8-2	47	41839	1	1	1	4	7	17	24
D8-3	50	42343	1	1	1	5	8	18	25

To visually display the composition and abundance of the biological communities at species level in the samples, the system selects the top 18 species in terms of abundance at species level for all samples and generates a stacked bar chart. The chart, shown in Figure 5, illustrates that the species composition and abundance are consistent among the three replicates from the same location. Additionally, a total of 59 fish species were identified across the six locations.

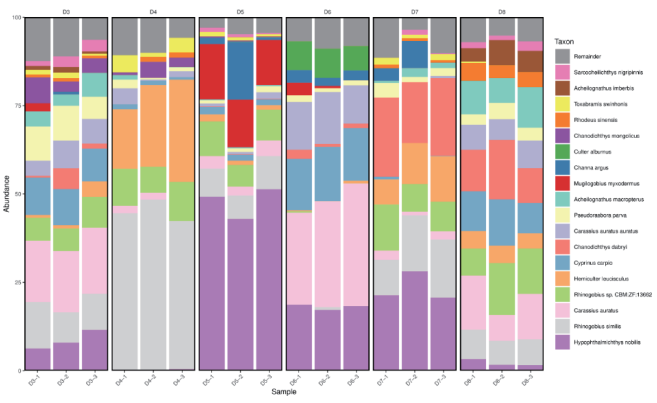


Figure 5. Species composition and abundance stacked histogram

The rank-abundance curve is another function of the system analyzing diversity. It reflects the diversity and evenness of species in the sample by arranging species in descending order of abundance and calculating the relative abundance and ranking of each species. In the horizontal direction, the width of the curve reflects the abundance of species, with a wider range on the x-axis indicating higher species abundance. The shape (smoothness) of the curve reflects the evenness of species distribution in the sample, with a smoother curve indicating a more even distribution of species. The rank-abundance curves for all samples are shown in Figure 6, where the x-axis represents the number of OTUs after ranking, and the y-axis represents the relative abundance of each OTU.

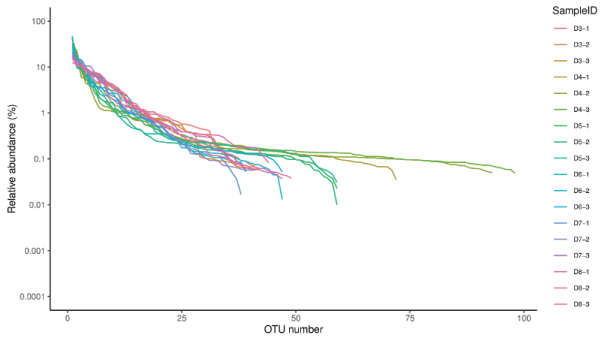


Figure 6. The rank-abundance curve for samples

To display species identification and abundance information at different taxonomic levels (phylum, class, order, family, genus, species) for individual samples more intuitively, the system drawn Krona charts. Taking sample D3 as an example, the Krona chart result is shown below in Figure 7. The circles represent different taxonomic levels from the inner to the outer, and the size of the sectors represents the relative proportion of different identification results.

Figure 7. Krona chart for sample D3

To evaluate whether the sequencing depth can detect most of species, the system drawn Shannon-Wiener curves using the sequencing depth of each sample at different sequencing depths. These curves reflect the biodiversity of each sample at different sequencing depths. When the curve becomes flat, indicates the sequencing data is enough to detect most of the species information in the sample. As shown in Figure 8, we can see the samples curves tend to flatten at 30,000 reads. This suggests that trimming to 50,000 reads is sufficient to detect the majority of species in the samples.

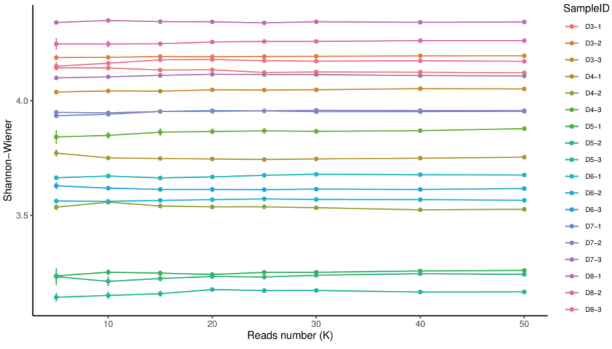


Figure 8. Shannon-Wiener curves of samples

To analyze the differences in species diversity between different locations, the system computed the dissimilarity index of species composition between different samples to assess the degree of species composition differences. The compositional distance analysis plot for all samples is shown in Figure 9, where both the x-axis and y-axis represent the samples, and the color reflects the distance between samples horizontally and vertically. We can see samples from locations D3 and D8 have some similarity in fish species composition, while there is a significant difference in species composition between samples from other locations.

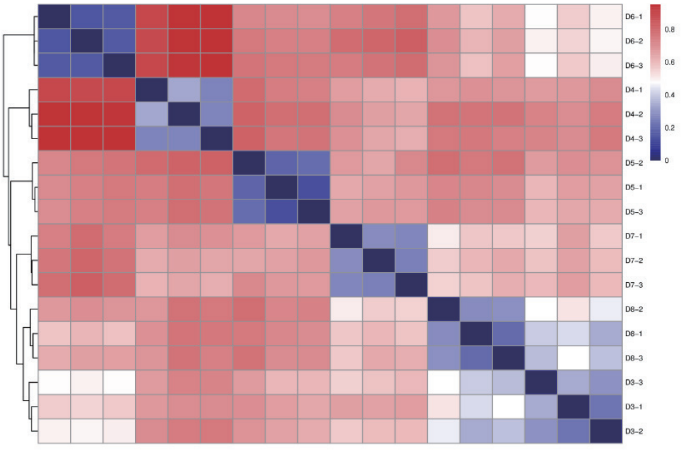


Figure 9. Inter-group distance analysis

Principal Component Analysis (PCA) is another function available in the system for analyzing the differences in species diversity between different locations. The analysis results are shown in Figure 10, where each point represents a sample, and points of the same colour belong to the same group. The closer the distance between two points, the smaller the difference between them. We can see the samples from locations D3 and D8 are closer in distance, indicating a certain similarity in fish species composition between these two locations. This is consistent with the conclusion from the inter-group distance analysis mentioned above.

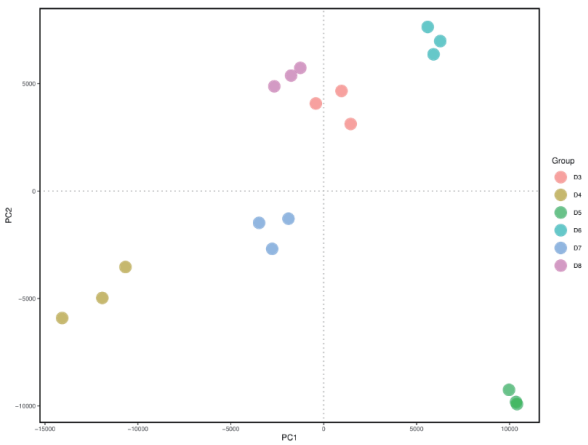


Figure 10. PCA chart

The OTU sequences were aligned and annotated according to the database to obtain species identification results at different taxonomic levels. The results are shown in Table 8.

Table 8. OTU classification statistics table

Sample ID	OTU number	OTU tag number	Kingdom	Phylum	Class	Order	Family	Genus	Species
DJ1-1	278	34884	2	12	20	35	67	102	117
DJ1-2	316	33828	2	13	20	34	72	108	123
DJ1-3	314	33731	2	13	22	34	75	111	128
S1-1	157	36142	2	9	13	23	42	67	81
S1-2	78	35914	2	8	12	20	32	46	51
S1-3	129	34254	2	10	12	20	39	61	72
JX1-1	1087	33282	2	17	27	52	114	206	246
JX1-2	987	36449	2	14	23	46	110	187	230
JX1-3	1197	34670	2	15	25	41	104	195	245

Sample DJ1, S1 and JX1 are sediment samples collected from Diaojiang, Jiuxu River, and Jiuzhou River in Guangxi. Each sampling location was replicated 3 times. After DNA extraction, COI libraries were prepared using the MGI's solution. The libraries were then subjected to DNBSEQ-G99 sequencing and analysis. The results of the quality control and data processing are shown in Table 7. all samples' Q30 > 96%. Each sample was trimmed to 50,000 reads and conducted to low-quality filtering, denoising, merging, and chimera removal to generate feature sequences.

Table 7. Quality control and data processing results

Sample ID	Raw reads	Target reads	Down sampled reads	Filtered reads	Q30(%)	Denoised reads	Merged reads	Chimeric reads	Feature reads
DJ1-1	873230	61516	50000	38417	97.34	37127	35922	1038	34884
DJ1-2	937803	61632	50000	36734	96.66	35879	35112	1284	33828
DJ1-3	824197	61491	50000	36386	96.65	34886	33869	138	33731
S1-1	797266	61790	50000	39331	97.56	38459	37514	1372	36142
S1-2	779515	61741	50000	38541	97.24	38110	37634	1720	35914
S1-3	275573	61804	50000	38650	97.57	38024	36518	2264	34254
JX1-1	253806	61694	50000	38895	97.38	36448	33462	180	33282
JX1-2	352053	61795	50000	40172	97.54	38325	36595	145	36450
JX1-3	205548	61814	50000	38918	97.44	36917	34859	189	34670

Figure 11 shows the stacked bar chart of species composition and abundance at the species level for the three groups of sediment samples. It shows that the species composition and abundance of benthic or planktonic organisms detected in the three replicates of the same sampling location are similar and have similar abundances. However, there are significant differences in species composition and abundance between the three sampling locations.

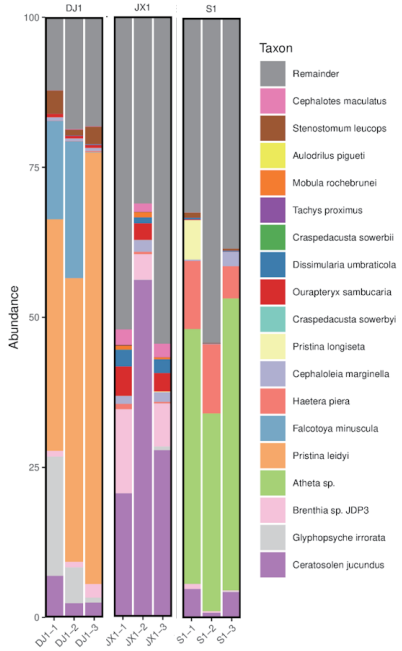


Figure 11. Species composition and abundance stacked histogram

According to the species identification results of OTUs at different taxonomic levels, Krona plot drawn, as shown in Figure 12, there are significant differences in species composition between the three sampling locations, which is consistent with the above conclusion.

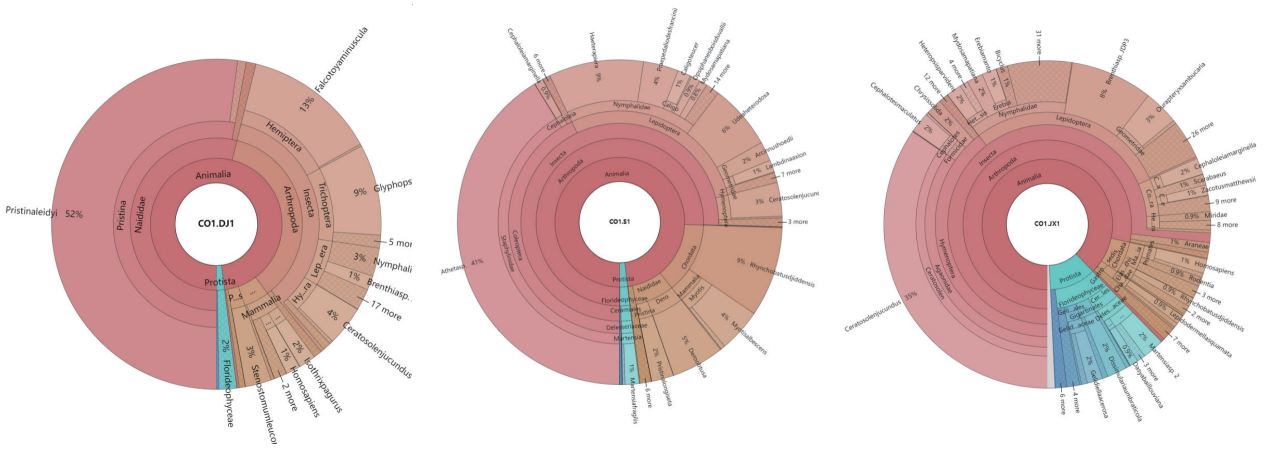


Figure 12. Krona chart for 3 set samples

Three soil samples (T1-T3) collected from the same location were prepared into 18S V4 and ITS1 libraries respectively . The libraries were then sequenced and analyzed on the DNBSEQ-G99. The results of quality control and data processing are shown in Table 9. all samples Q30 > 93%. After trimming to 50,000 reads per sample, the data conducted to low-quality filtering, denoising, merging, and chimera removal. The resulting OTU sequences were then aligned and annotated, as shown in Tables 9 and 10.

Table 9. Quality control and data processing results

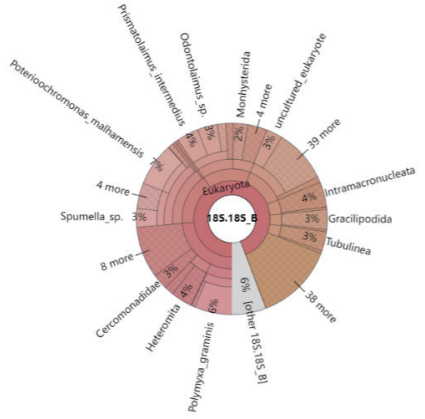
Sample ID	Panel	Raw reads	Target reads	Down sampled reads	Filtered reads	Q30 (%)	Denoised reads	Merged reads	Chimeric reads	Feature reads
T1	18S V4	131907	64259	50000	31730	96.79	30345	24598	339	24259
T2	18S V4	128451	64147	50000	32647	96.91	31313	24752	705	24047
T3	18S V4	93458	63723	50000	33346	97.03	32071	26233	404	25829
T1	ITS1	188746	64919	50000	47416	93.88	45077	42088	507	41581
T2	ITS1	227863	64846	50000	47488	93.84	45245	41634	165	41469
T3	ITS1	218857	64773	50000	47269	93.94	45132	42317	503	41814

Table 10. OTU classification statistics table

Sample ID	Panel	OTU number	OTU tag number	Kingdom	Phylum	Class	Order	Family	Genus	Species
T1	18S V4	439	24258	1	37	58	83	90	130	171
T2	18S V4	417	24046	1	37	53	74	81	115	158
T3	18S V4	452	25828	1	38	58	83	90	134	175
T1	ITS1	172	41581	1	6	16	32	58	73	79
T2	ITS1	158	41469	1	7	17	33	59	70	75
T3	ITS1	164	41814	1	6	16	31	53	65	72

Based on the species identification results of OTUs at different taxonomic levels, Krona plot drawn, as shown in Figure 13.

In summary, MGI environmental DNA metabarcoding sequencing package has excellent powerful performance, various bioinformatics functions, and compatibility with main types of environmental samples. It is an ideal tool for ecological environmental monitoring and environmental science research.



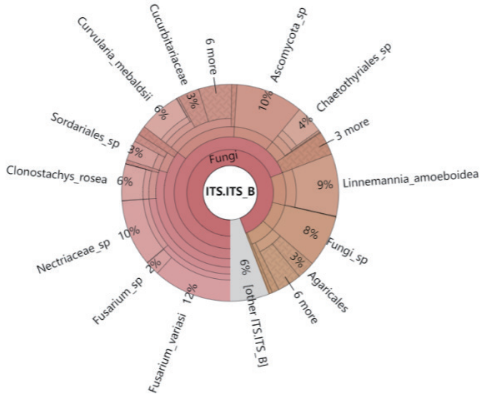


Figure 13. Krona chart for 2 set soil samples

Ordering information

Name	Specification	PN
Instrument		
DNBSEQ-G99ARS	Include server	900-000609-00
DNBSEQ-E25RS	Standard Config	900-000537-00
Automated Nucleic Acid Extractor MGISP-NE32RS	/	950-000020-00
MGISP-960RS High-throughput Automated Sample Preparation System	Custom Configuration 9-V7	900-000154-00
Reagent		
MGIEasy Stool Microbiome Extraction Kit	48 Preps/Kit	940-000122-00
MGIEasy Stool Microbiome Extraction Kit	192 Preps/Kit	940-000123-00
MGIEasy Tissue Grinding Beads	96 Preps/ Bottle	940-000136-00
ATOPlex ITS1 rDNA Library Prep Set	96 RXN	Group010-000013-00
ATOPlex ITS1 rDNA Library Prep Set	576 RXN	Group010-000014-00
ATOPlex 18S V4 rDNA Library Prep Set	96 RXN	Group010-000015-00
ATOPlex 18S V4 rDNA Library Prep Set	576 RXN	Group010-000016-00
ATOPlex COI mtDNA Library Prep Set	96 RXN	Group010-000017-00
ATOPlex COI mtDNA Library Prep Set	576 RXN	Group010-000018-00
ATOPlex Ac12S mtDNA Library Prep Set	96 RXN	Group010-000019-00
ATOPlex Ac12S mtDNA Library Prep Set	576 RXN	Group010-000020-00
ATOPlex MiFish Library Prep Set	96 RXN	Group010-000021-00
ATOPlex MiFish Library Prep Set	576 RXN	Group010-000022-00
ATOPlex 16SV3V4 rDNA Library Preparation Set	96 RXN	940-001261-00
ATOPlex 16SV3V4 rDNA Library Preparation Set	576 RXN	940-000725-00
ATOPlex MiFish Panel	96RXN/Kit	940-001539-00
ATOPlex MiFish Panel	576RXN/Kit	940-001538-00
ATOPlex 16SV3V4 rDNA panel	96RXN/Kit	940-001260-00
ATOPlex 16SV3V4 rDNA panel	576RXN/Kit	940-000724-00
ATOPlex Ac12S mtDNA Panel	96RXN/Kit	940-001535-00
ATOPlex Ac12S mtDNA Panel	576RXN/Kit	940-001540-00
ATOPlex COI mtDNA Panel	96RXN/Kit	940-001534-00
ATOPlex COI mtDNA Panel	576RXN/Kit	940-001537-00

Name	Specification	PN
Reagent		
ATOPlex 18SV4 rDNA Panel	96RXN/Kit	940-001533-00
ATOPlex 18SV4 rDNA Panel	576RXN/Kit	940-001541-00
ATOPlex ITS1 rDNA Panel	96RXN/Kit	940-001536-00
ATOPlex ITS1 rDNA Panel	576RXN/Kit	940-001532-00
ATOPlex DNA Dual BC Library Prep Set	96RXN	940-001191-00
ATOPlex DNA Dual BC Library Prep Set	576RXN	940-001190-00
MGIEasy Dual Barcode Circularization Kit V1.0	16RXN/Kit	1000020570
DNBSEQ OneStep DNB Make Reagent Kit	4 RXN	1000026466
ATOPlex E450 Dual Barcode Balanced Library Reagent	26ng/Tube	940-000637-00
Standard Library Reagent (PCR Product)	1500ng/Tube	1000027585
High-throughput Sequencing Set	G99 SM FCL PE150	940-000410-00
High-throughput Sequencing Set	G99 SM FCL PE300	940-000415-00
DNBSEQ-E25RS High-throughput Sequencing Set	FCL PE150	940-000567-00
Software		
Metabarcoding Species Identification Software	/	970-000417-00
Metabarcoding Species Identification Software Package	96 reports	970-000456-00
Platform of microorganisms Fast Identification	/	900-000393-00
Platform of microorganisms Fast Identification and assembly evolution	/	900-000399-00

MGI Tech Co.,Ltd.

Building 11, Beishan Industrial Zone,Yantian District,Shenzhen,China

+86-4000-688-114

en.mgi-tech.com

MGI-service@mgi-tech.com