SIP Protocol Series No.8

Functional assessment of microbiota in various environments using MAPLE

Cross-ministerial Strategic Innovation Promotion Program

The Strategic Innovation Promotion Program (SIP) was launched by the Council for Science, Technology, and Innovation (CSTI), which oversees projects that target scientific and technological innovation in line with Japanese government directions as stated in the Comprehensive Strategy on Science Technology and Innovation and the Japan Revitalization Strategy. This interdisciplinary program among government agencies, academic institutes and private sectors addresses eleven issues. One of these issues is Next-Generation Technology for Ocean Resources Exploration.

Zipangu in the Ocean Program and Protocols for Environmental Survey Technologies

Zipangu in the Ocean Program is a technical study of the development of submarine mineral deposits that takes into consideration the wise use of these resources.

One research area is the ecological survey of organisms and their long-term monitoring. However, an ecosystem consists of various interrelated factors; thus, in addition to a comprehensive understanding of the system, observation and analysis of each component to its most elemental level are unavoidable. Recently, increased environmental awareness and the necessity of forming a consensus have become key issues in conducting development activities. Growing concern for the environment by the public and the diversification of the use of maritime areas have complicated the interests of stakeholders. To facilitate the formation of a consensus under these conditions, it is important for standardized methods to be implemented. This will ensure that research processes are transparent and that the collection of survey data is objective.

This protocol series aims to introduce more accurate, user-friendly, objective and effective underlying technologies required to understand the environmental impact of submarine mineral resource development. We believe that creating such a technology tool-kit will allow us to develop these resources in a sustainable manner.

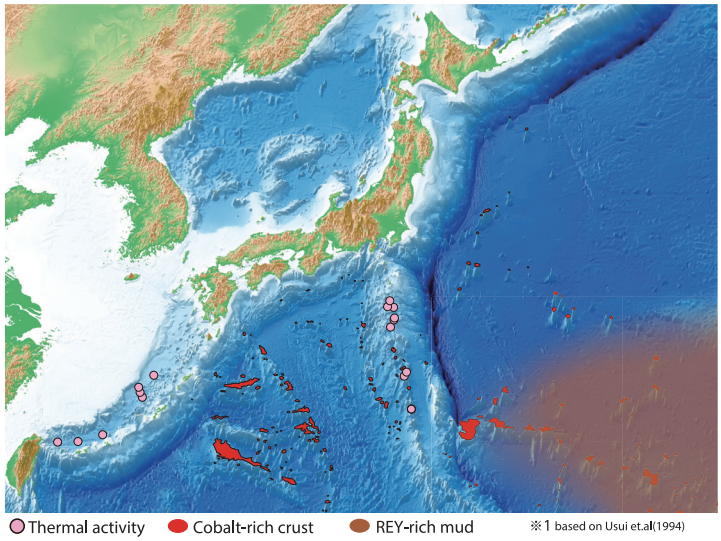


Table of Contents

chapter 1

Introduction

Considering that culturable microbes constitute less than 1–10% of all microbes thriving in natural environments, metagenomic analysis is one of the most powerful ways to understand species and functional diversity in whole communities (WCs) of microbes. Since the first report of a comprehensive metagenomic analysis of a marine environment was published [1], many metagenomic analyses focused on microbial communities ranging from natural environments to human gut microbiota have been reported. However, the diversity estimates based on 16S rDNA and some other key genes, which enable the characterization of the habitats have just been discussed in most metagenomic papers. However, a main goal of metagenomics is the actual deduction of not only community structure but also potential comprehensive functions (the functionome) harbored by entire communities across various environments. We define the functionome as the comprehensive functions occurring through combinations of individual functions, such as carbon fixation, nitrogen fixation, nitrification, denitrification, and amino acid metabolism, encoded by multiple genes [2]. This goal remains poorly addressed because the evaluation of the potential functionomes still remains difficult compared with the functional annotation of individual genes or proteins, mainly because there is no established standard methodology for extracting functional category information such as data pertaining to metabolism, energy generation, and membrane transport systems.

To resolve this problem, a new method for evaluating the potential functionome was developed based on calculating the completion ratio of four types of Kyoto Encyclopedia of Genes and Genomes (KEGG) modules (see Note 1): pathways, molecular complexes, functional sets, and signatures [3]. This is represented as the percentage of a modulecomponent filled with the input KEGG Orthology (KO)-assigned genes by KAAS [4]. A prototype system of MAPLE (Metabolic And Physiological potentiaL Evaluator) for automating the newly developed method was then launched in December 2013. MAPLE first assigns KO to the query gene, maps the KO-assigned genes to the corresponding KEGG functional modules, and then calculates the module completion ratio (MCR) of each functional module to characterize the potential functionome of the user' s own genomic and metagenomic data. Afterwards, two new useful functions for calculating module abundance and Q-value, which indicate the functional abundance and statistical significance of the MCR results, respectively, were added to the new MAPLE ver. 2.1.0 to enable more detailed comparative genomic and metagenomic analyses [5].

Although this system was very useful at the time, especially for metagenomic data analysis, the high computational time associated with analyzing the often massive amino acid sequence datasets (with 1–3 million sequences typically

employed in metagenomic research) reduced its utility. Thus, the MAPLE system was further developed to reduce calculation time by adapting KAAS to use GHOSTX [6], a much faster homology search program, instead of BLAST. The latest MAPLE system 2.3.1 is now available through a web interface (https://maple.jamstec.go.jp/maple/maple-2.3.1/).

Although next-generation sequencing (NGS) can easily produce massive sequence datasets, these raw data cannot be directly applied to the system because the data submitted to MAPLE must be a multi-FASTA file of amino acids. Unfortunately, it is difficult for researchers who are unfamiliar with bioinformatic tools to process such massive raw datasets properly. To increase user convenience, we developed MAPLE Submission Data Maker (MSDM), which can convert raw NGS data into multi-FASTA files of amino acid sequences. This useful software can be installed on personal computers that run MacOS X or Windows OS.

MAPLE results, such as MCR, Q-value, and module abundance, can be easily downloaded as an Excel file. However, MAPLE results are difficult to judge visually as they consist of rows of numerical values. Thus, we developed two kinds of downloadable program that is available through the website to draw histograms based on the MAPLE results (MAPLE Graph Maker (MGM). With the program, users can easily visualize MAPLE results by importing resulting Excel files. This protocol presents the methodology for functional metagenomics using MAPLE system to reveal the functional diversity of individual microbes and WCs of microbes in the actual environments.

chapter 2

MAPLE system

2-1. Overview of functional potential evaluation

MAPLE is an automatic system that can perform a series of steps used in the evaluation of potential comprehensive functions (i.e., functionomes) harbored by genomes and metagenomes. From April (2016) through March (2017), MAPLE was accessed 2.5 million times. However, beginners still have difficulty in processing such massive raw datasets produced by NGS prior to data submission to MAPLE and in interpreting MAPLE results, which contain many rows of numerical values. Thus, we now provide a complete system to support every step from initial data processing to final visualization of the MAPLE results (Fig. 1).

MAPLE first assigns a KO ID to the query gene using KAAS (Fig. 2B), then maps the KO-assigned genes to the KEGG functional modules (Fig. 2C), and finally calculates the MCR of each functional module and its abundance when the module is complete (Fig. 2D). There are two methods for KO assignment by KAAS: bidirectional best hit (BBH) and single-directional best hit (SBH). The BBH method is suitable for complete gene sets identified from complete genomes, while the SBH method is mainly appropriate for short-read sequences of metagenomes or incomplete genomes. When the query sequences are submitted to MAPLE (Fig. 2A), the MCR and abundance of each KEGG module as well as the taxonomic information for the KO-assigned genes mapped to the module are displayed along with a mapping pattern. When a KO mapped to a module is shared by two or more modules, the module IDs sharing the same KO are listed (Fig. 3). The MCR calculation is performed based on a Boolean algebra-like equation defined by KEGG for each module. The results of KO assignment produced by KAAS, taxonomic information for the genes mapped to the KEGG modules, and calculated MCRs are downloadable in an Excel spreadsheet format. MAPLE can display the results of comparative analyses of mapping patterns, MCR results, and the abundance of complete modules between the different metagenomic samples. To evaluate the working probability of the physiological function in the incomplete modules, we proposed the Q-value as a more appropriate way to interpret MCR results. The Q-value, which indicates the probability that a reaction module is identified by chance, is calculated based on the statistics of the sequence similarity score and KO abundance using the concept of multiple testing corrections according to Boolean algebra-like equations.

The previous version of MAPLE system (ver. 2.1.0) can accept 1 million amino acid sequences (<160 Mbytes) for each job; however, the 80-hour running time to finish all calculation steps for 1 million sequences strongly motivated efforts to reduce the calculation time for many users. To resolve this urgent problem, we employed GHOSTX [6], a homology search program that is much faster than BLAST, which had been used in a previous version of MAPLE. In the current MAPLE version 2.3.1, approximately 18 hours are required to complete jobs when users select GHOSTX instead of BLAST as a homology search program, depending

on the size and content of the metagenomic sequences. For example, a job using the BBH method for a 4.4-Mb individual genome containing 4,035 protein sequences can be completed within 45 minutes. In the case of metagenomic sequences containing 3 million sequence reads, a job will take about 12–18 hours to complete when the SBH method is used. Fundamentally, the SBH method should be used for metagenomic analysis even when using contigs assembled from metagenomic sequences if contigs are a part of the genome of an individual organism (i.e., draft genome). An e-mail address can be specified for users to receive a message with a URL for the results from the system upon completion of the job.

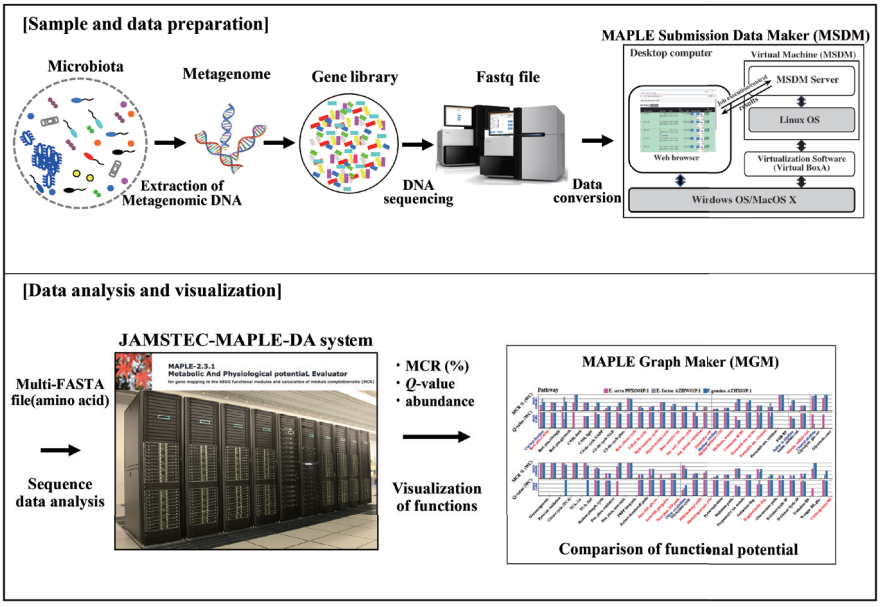


Fig. 1

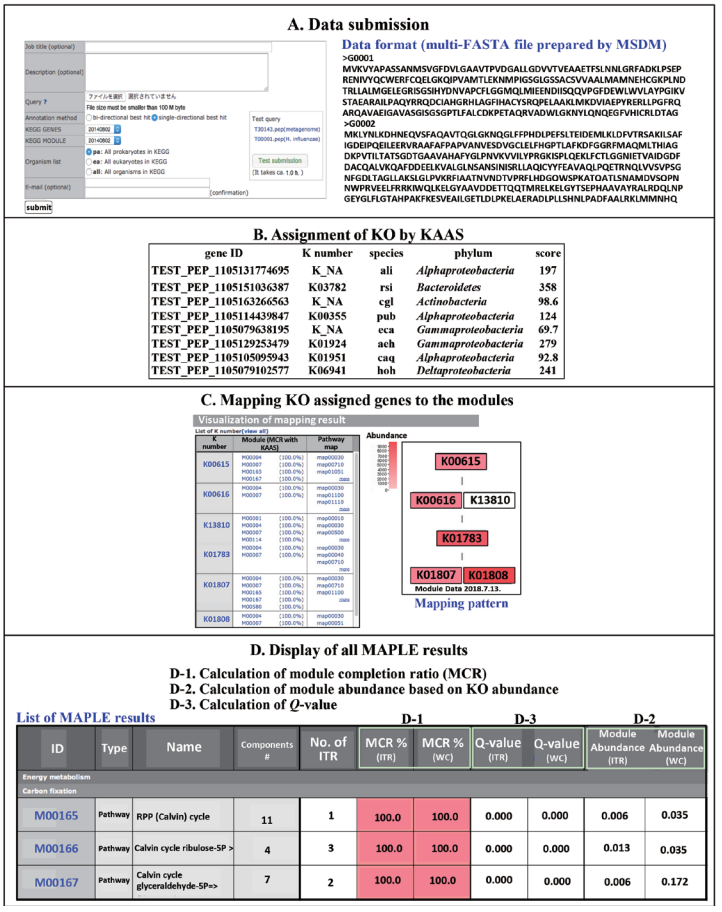


Fig.2

2-1-1. Calculation of the module completion ratio based on the Boolean algebra-like equation

The completion ratio of all KEGG functional modules in each organism was calculated on the basis of a Boolean algebra-like equation. For this analysis, 1 genome was selected from each of the 5,257 available prokaryotic and 457 eukaryotic species genomes, and a reference genome set was constructed to cover all completely sequenced organisms, excluding draft genomes as of July 13, 2018 (3,622 total genomes, 3,186 prokaryote genomes, and 436 eukaryotic genomes). For example, M00001 is a pathway module for glycolysis, comprising 10 reactions as shown in Fig. 3. In each KO number set depicted, the horizontally arranged rows of K numbers indicate alternatives, which are related to each other by "Or" or "," in the equation [4]. When genes are assigned to all KO IDs in each reaction according to the Boolean algebra-like equation, the module completion ratio becomes 100%. For example, when genes are not assigned to KO IDs in two reactions, the module completion ratio is calculated to be 80% (8/10 × 100 = 80).

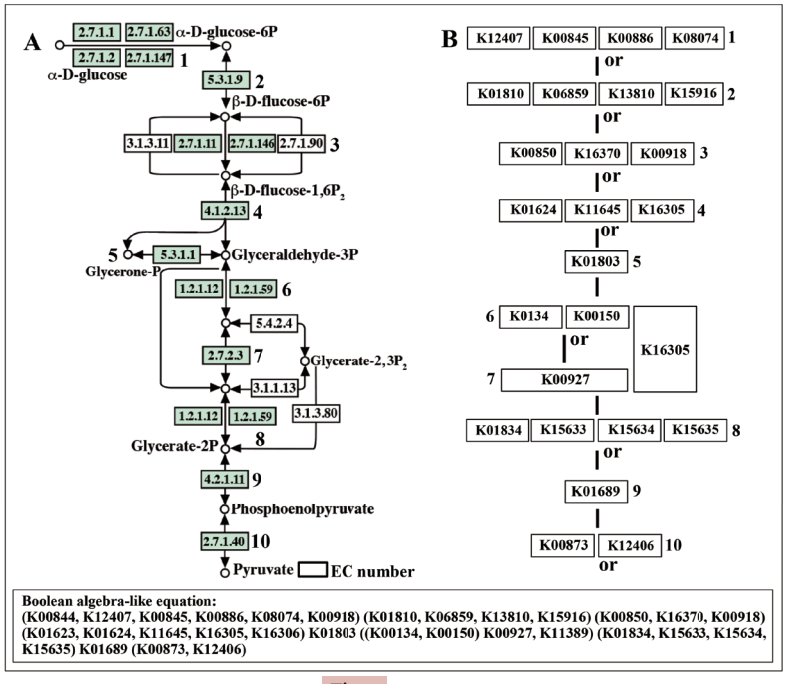


Fig. 3

2-1-2. Evaluation of the module completion ratio according to Q-values

Generally, it is expected that the MCR is linked to the likelihood that the organisms perform the physiological function corresponding to a particular module. However, when the KOs used for a module are shared with the other modules, the MCR does not necessarily reflect the working probability of each functional module (Fig. 4). Thus, the MCRs of the targeted module, module completion of other modules to which the same KOs are assigned, and the contribution of specific KOs of each module to module completion should be considered when a module is incomplete. That is, even if the same MCR was observed among different modules, the working probability of the physiological function is not necessarily equal among these MCRs. To avoid these problems, we propose the use of the Q-value for determining the significance of module completion. This measure, which represents the probability that a reaction module is identified by chance, is

calculated based on statistical sequence similarity scores (e.g., E-values) using the concept of multiple testing corrections, according to the definition of the KEGG reaction module (i.e., Boolean algebra-like equation) [5].

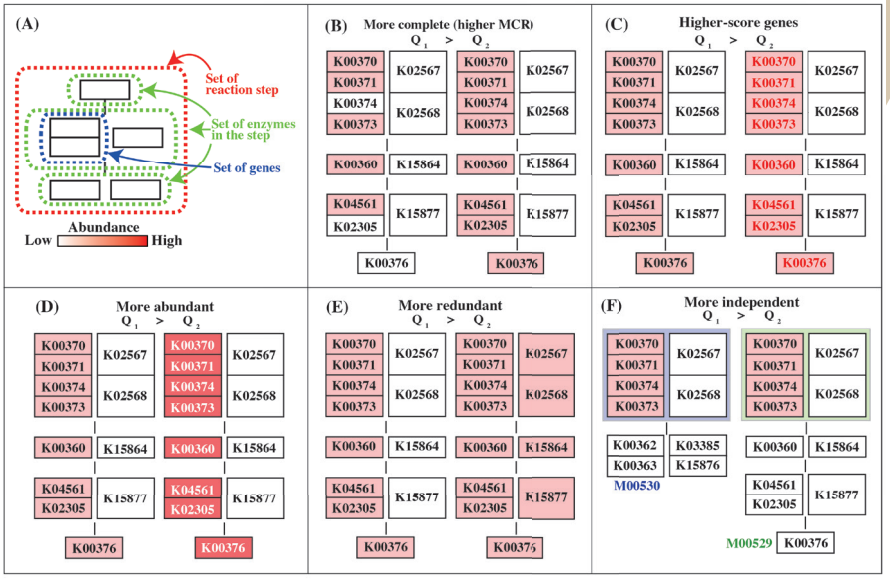


Fig.4

2-1-3. Calculation of module abundance

MAPLE highlights the difference in the potential functions of organisms and environmental samples if the MCRs of various modules differ. MAPLE can also determine diversity in individual taxonomic rank (ITR) for completed modules and abundance of the modules for every ITR to clarify the difference in the functional potential of modules commonly completed across multiple samples. ITR is a second taxonomic rank defined in the KEGG Organisms database such as phylum, class, and order (http://www.genome.jp/kegg/catalog/org_list.html).

The total number of sequence reads assigned to each KO in constructing a module was divided by the average length of each KO group to normalize KO abundance. This normalized KO abundance is described in the lowerright corner of each KO box when the mapping pattern of the query genes for the module are displayed (Fig. 5). The module abundance is calculated based on the normalized KO abundance. For example, the module M00529 (Figs. 4B–F and 5), defined as a denitrification reaction, is composed of four reaction steps. For each K number set, vertically arranged K numbers represent a complex whereas horizontally arranged K numbers represent alternatives [4, 5]. Because the enzyme responsible for the first reaction (nitrate reductase) is composed of four (Fig. 5, left side) or two KO complexes (Fig. 5, right side), the abundance of the first reaction step becomes 0 unless all KOs vertically connected are filled with the KO-assigned genes. When all vertically connected KOs are filled, the minimal value of KO abundance in the vertically connected boxes becomes the abundance at the first step. Thus, the abundance of the first step in module M00529 is 63. As the second step, when two horizontally located KOs are filled with the KO-assigned genes, the abundance at the second step becomes 1,129, which is the sum of both KOs. The abundance at the third step becomes 56 in a similar manner as the first step, and that of the last step is 46 (Fig. 5). Because the module abundance

becomes the minimum value among all steps,the abundance of module M00529 is calculated to be 46. To facilitate normalization by ribosomal proteins, we defined a new virtual module (M91000) for all ribosomes (i.e., prokaryote + eukaryote ribosomes) comprising 130 ribosomal proteins (excluding accessory proteins) because the 31 KOs corresponding to the ribosomal proteins are shared by Bacteria and Archaea and 26 KOs are shared between Archaea and Eukaryote (see 1.2). Thus, we calculated the module abundance per ribosomal protein to enable the comparison among different environmental sites.

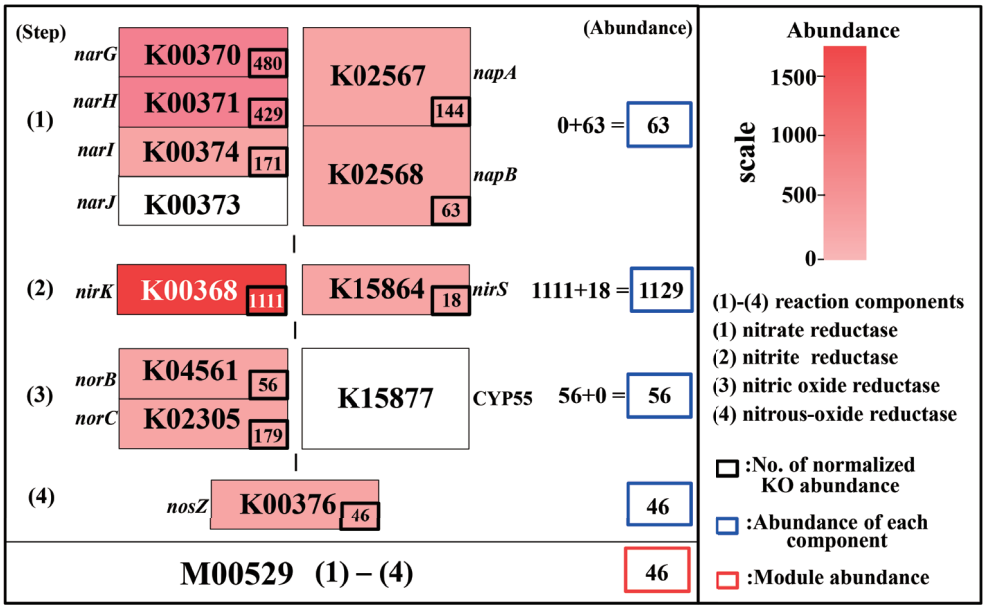


Fig. 5

2-2-1. Microbial community structure based on ribosomal proteins

In many metagenomic analyses, 16S rRNA gene sequences obtained by PCR amplification are used to compare microbial community structures among different environments. In recent studies, this PCR-based amplicon approach has been used to target the V4 region because different regions of the 16S rRNA gene yield varying degrees of accuracy in taxonomic assignments [8]. However, prokaryotic species exhibit variation in copy number of the 16S rRNA gene even within the same species (https://rrndb.umms.med.umich.edu/) [9], and it is impossible to determine the copy numbers of individual unculturable and unknown microbes present in actual microbial communities. Thus, because taxonomic compositions based on amplicon sequences are strongly influenced by copy number in addition to basic PCR bias, this approach is not useful for the analysis of microbial community structure.

Ribosomal proteins are well conserved among all organisms and possess sequences specific to each individual organism; therefore, ribosomal proteins can be used for the identification of organisms. Actually, it has been confirmed that MAPLE can be effectively used to identify organisms by constructing a metagenome based on ribosomal proteins [5]. To apply MAPLE to taxonomic analysis, we calculated the proportions of Bacteria, Archaea, and Eukaryote in the metagenome based on the mapping pattern of the virtual module M91000 for all ribosomes (Fig. 6) and the taxonomic

annotation of each ribosomal protein, whereas eukaryotic taxonomic information is limited in the KEGG database. As mentioned above, because archaeal and eukaryotic ribosomes, which contain 58 and 77 ribosomal proteins, respectively, have 6 and 15 more proteins than the bacterial ribosome, we normalized the total number of archaeal ribosomal proteins to the number of bacterial ribosomal proteins by multiplying the archaeal and eukaryotic ribosome count data by 52/58 and 52/77, respectively. We summed the number of bacterial ribosomes and normalized archaeal and eukaryotic ribosomes and then used this sum as a denominator for calculating the proportions of Archaea, Eukaryote, and Bacteria. We can also calculate the proportion for each taxonomic level defined by KEGG, such as phylum and class, in the metagenome using the same method. When a metagenome is composed of only prokaryotic sequences, module M90000 is useful for analysis of prokaryotic community structure, but module M91000 for ribosomes of all organisms is also useful for community structure analyses when the microbial community also contains eukaryotic species together with prokaryotes.

On the other hand, a new method based on universal single-copy genes, which provides prokaryotic species boundaries at a higher resolution than possible using the 16S rRNA gene, has been used to estimate the relative abundances of known and unknown microbial community members with metagenomic data at a species-level resolution [10]. However, community structure analysis at such a high resolution is not necessarily required in metagenomic analyses of natural environments, unlike that of the human gut microbiome, because many community members have not yet been cultivated or identified at the species level. Thus, community structure analyses at the phylum or class level based on ribosomal proteins using the latest version of the MAPLE system (version 2.3.1) are thought to be feasible for analyses of metagenomes from natural environments.

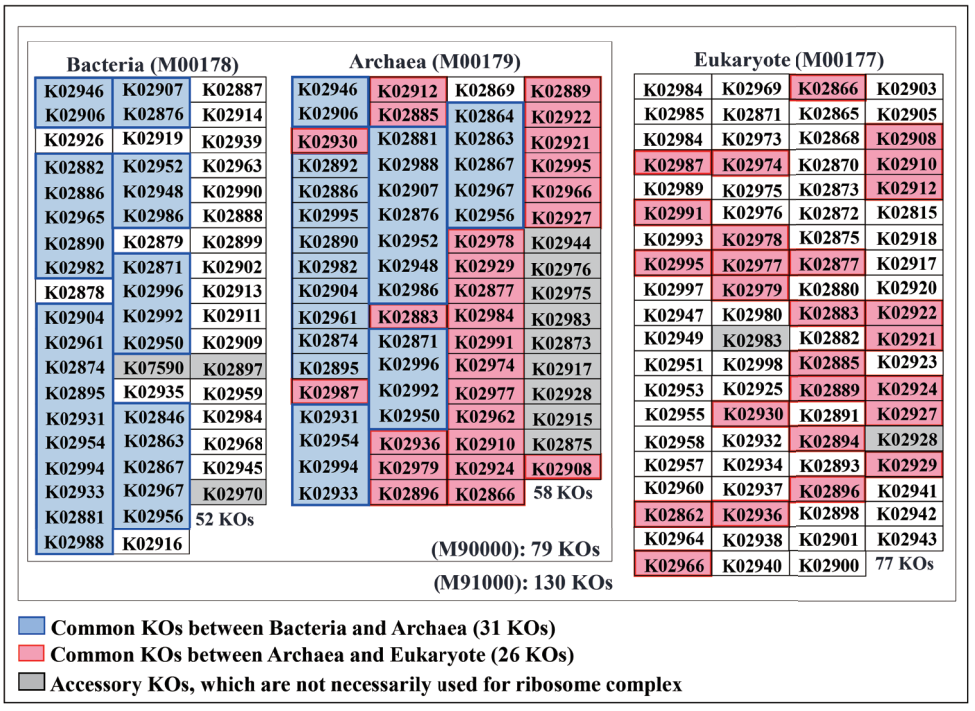


Fig. 6

chapter 3

User's guide for MAPLE version 2.3.1

3-1. Data submission

Massive NGS datasets consist of raw data that cannot be directly imported into the MAPLE system as a standard FASTQ file because data submitted to MAPLE must be in multi-FASTA files of amino acid (a.a.) sequences with unique IDs in the comment lines and without tab delimiters (Fig. 2A). This compatibility issue can be an impediment to researchers who would like to utilize MAPLE but are unfamiliar with the relevant bioinformatic tools. Accordingly, we developed MAPLE Submission Data Maker (MSDM), an application that automatically converts paired-end FASTQ files into multi-FASTA files. This software can be installed on personal computers with MacOS X and Windows OS, as detailed in the installation manual. Users can obtain the software from the MAPLE software download site (https://maple.jamstec.go.jp/maple/maple-2.3.1/softdownload/) after registration of their account. The query sequence does not necessarily need to be a complete gene, but a.a. sequences longer than 100 residues are generally recommended for accurate KO assignment, though some genes consist of fewer than 100 amino acids, such as ribosomal proteins. The number of query sequences may not exceed 3 million sequences owing to the limitations of computational resources; an error message is displayed when the file size exceeds this limit.

When more sequence data are required for an analysis, the user can submit several sub-datasets consisting of fewer than 3 million sequences derived from the same metagenomic sample and then merge the results from all sub-datasets into one dataset by clicking the "Merge" button on the job list page. When two jobs with 3 million sequences are merged as one job (i.e., 6 million total sequences), it will take 10 hours. Datasets containing 3 million sequences are ideal for the accurate evaluation of MCRs by considering the results of KO rarefaction curves to determine whether sufficient sequences have been included (https://maple.jamstec.go.jp/maple/maple-2.3.1/help.html),
particularly for determining the abundance of completed modules. However, we can sufficiently elucidate the overall trends of the functionome indicated by metagenomic data even when using fewer than 3 million sequences. Rarefaction curves are automatically drawn when the MAPLE analysis is complete and can be accessed from the results page. Indeed, we have successfully determined the metabolic potential of the human gut microbiome from 13 healthy individuals by comparative analysis with total sequences from each consisting of fewer than 100,000 amino acids [4]. After submission of a dataset, a URL address for accessing the results is displayed along with the job ID. The results are also

displayed on the current page upon completion of the job.

3-2. Results pages

To access the tables containing the results of completion ratios for all types of KEGG modules (i.e., pathway, structural complex, functional set, and signature modules), the user clicks on the URL address displayed on the submission page; alternatively, the user can be notified of job completion by e-mail and then click on the job ID in that email. The user can view detailed information of the mapping results for each KEGG module by clicking on the module ID in each table. In addition, the user can access an overview of the MCR results by clicking on the "histogram (PNG)" button on the results page and accessing the mapping results of the KEGG module by placing the cursor on each module name (Fig. 7).

Additional information for each module, such as taxonomy, class, and definition, are also displayed on the results page (Fig. 7). Taxonomy is defined as the biological classification based on the MCR patterns of reference organisms with the determined genomic sequences. For example, if a module contains more than four prokaryotic species (i.e., Bacteria or Archaea) belonging different phyla, the module is represented by a prokaryotic taxonomy. Similarly, if a module contains only species belonging to Proteobacteria, the taxonomy of the module is Proteobacteria. Class indicates the module type based on the MCR patterns of reference organisms as previously defined.

The distribution of MCRs among 3186 prokaryotic or 436 eukaryotic species (one genome per species) can be categorized into four patterns (i.e., universal, restricted, diversified, and nonprokaryotic/noneukaryotic) regardless of the module type (i.e., pathway, structural complex, signature, or functional set) (see Note 2). A Boolean algebra-like equation is defined by KEGG for each module, and MCRs are calculated based on this equation. Since KOs are often assigned to two or more modules, all module IDs that share each KO composing a particular module are listed together with the pathway IDs containing the KO. For example, K01623, a member of M00167 (reductive pentose phosphate cycle), is shared by M00001 (glycolysis) and M00003 (gluconeogenesis) and is used in seven pathway maps (Fig. 7). In the case of metagenomic sequences, taxonomic information for KO-assigned genes is displayed, and the details of the phylum or class level for every KO, which facilitate the classification of organisms contributing to the completion of each functional module, are listed. For example, the module for virtual prokaryotic ribosomes (M90000), comprising 21 bacterial, 27 archaeal, and 31 common ribosomal proteins between Bacteria and Archaea (Fig. 6), can be used to represent the taxonomic breakdown of prokaryotes in the metagenome instead of the 16S rRNA gene, whose copy number varies among prokaryotic species. Ribosomal proteins can be used effectively because most are encoded by single-copy genes in the genome, and there is only minor variation in length among orthologous groups. The results are removed from the server 14 days after the job is completed; however, the user can download the results by clicking the "MAPLE results" button (Fig. 7).

The user can browse all MAPLE data by re-uploading previously downloaded data from the first page. In addition, the user can also easily download the data in an Excel-readable format containing KO assignments by KAAS, MCRs, the

abundances of KO-assigned genes and completed modules, analyses of module significance results, and taxonomic information for the KO-assigned genes mapped to each module from the results page (Fig. 8).

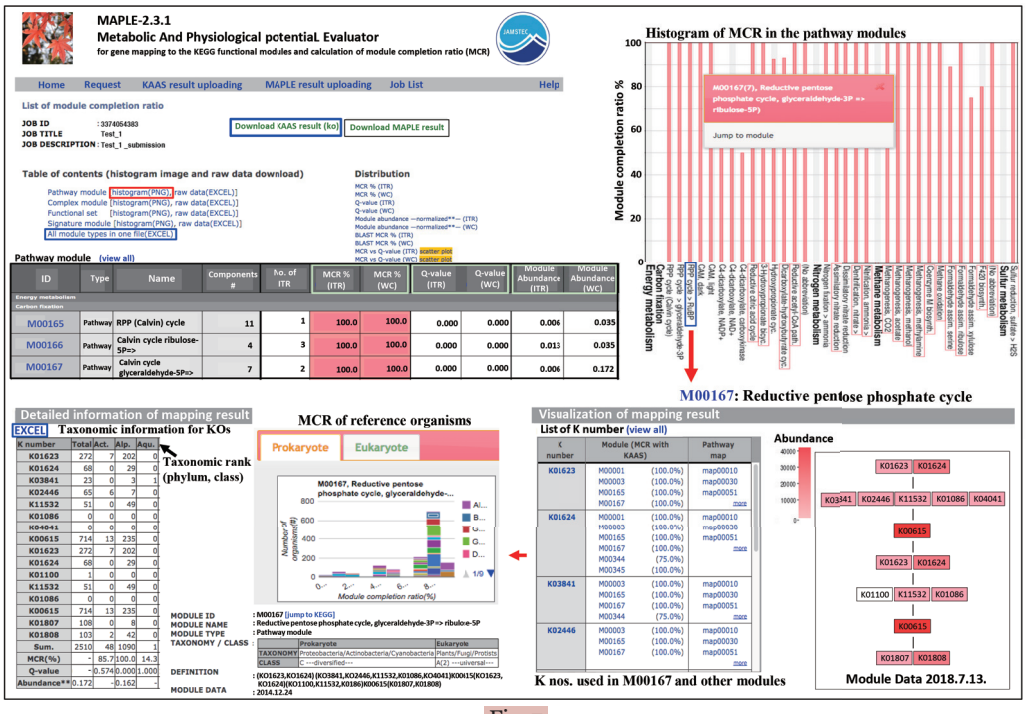
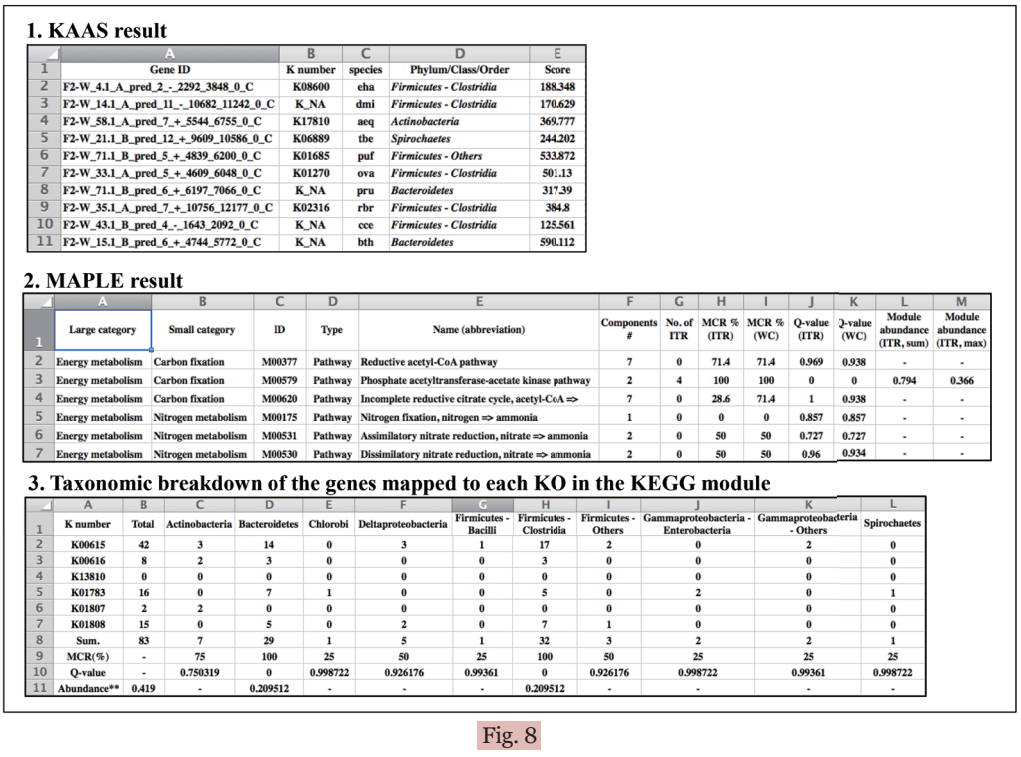


Fig. 7



1. KAAS result

	Gene ID	K number	species	Phylum/Class/Order	Score
2	F2-W_4.1_A_pred_2_-_2292_3848_0_C	K08600	cha	*Firmicutes - Clostridia*	188.348
3	F2-W_14.1_A_pred_11_-_10682_11242_0_C	K_NA	dmi	*Firmicutes - Clostridia*	170.629
4	F2-W_58.1_A_pred_7_+_5544_6755_0_C	K17810	aeq	*Actinobacteria*	369.777
5	F2-W_21.1_B_pred_12_+_9609_10586_0_C	K06889	tbe	*Spirochaetes*	244.202
6	F2-W_71.1_B_pred_5_+_4839_6200_0_C	K01685	puf	*Firmicutes - Others*	533.872
7	F2-W_33.1_A_pred_5_+_4609_6048_0_C	K01270	ova	*Firmicutes - Clostridia*	501.13
8	F2-W_71.1_B_pred_6_+_6197_7066_0_C	K_NA	pru	*Bacteroidetes*	317.39
9	F2-W_35.1_A_pred_7_+_10756_12177_0_C	K02316	rbr	*Firmicutes - Clostridia*	384.8
10	F2-W_43.1_B_pred_4_-_1643_2092_0_C	K_NA	cce	*Firmicutes - Clostridia*	125.561
11	F2-W_15.1_B_pred_6_+_4744_5772_0_C	K_NA	bth	*Bacteroidetes*	590.112

2. MAPLE result

	Large category	Small category	ID	Type	Name (abbreviation)	Components #	No. of ITR	MCR % (ITR)	MCR % (WC)	Q-value (ITR)	Q-value (WC)	Module abundance (ITR, sum)	Module abundance (ITR, max)
2	Energy metabolism	Carbon fixation	M00377	Pathway	Reductive acetyl-CoA pathway	7	0	71.4	71.4	0.969	0.938	-	-
3	Energy metabolism	Carbon fixation	M00579	Pathway	Phosphate acetyltransferase-acetate kinase pathway	2	4	100	100	0	0	0.794	0.366
4	Energy metabolism	Carbon fixation	M00620	Pathway	Incomplete reductive citrate cycle, acetyl-CoA =>	7	0	28.6	71.4	1	0.938	-	-
5	Energy metabolism	Nitrogen metabolism	M00175	Pathway	Nitrogen fixation, nitrogen => ammonia	1	0	0	0	0.857	0.857	-	-
6	Energy metabolism	Nitrogen metabolism	M00531	Pathway	Assimilatory nitrate reduction, nitrate => ammonia	2	0	50	50	0.727	0.727	-	-
7	Energy metabolism	Nitrogen metabolism	M00530	Pathway	Dissimilatory nitrate reduction, nitrate => ammonia	2	0	50	50	0.96	0.934	-	-

3. Taxonomic breakdown of the genes mapped to each KO in the KEGG module

	K number	Total	Actinobacteria	Bacteroidetes	Chlorobi	Deltaproteobacteria	Firmicutes - Bacilli	Firmicutes - Clostridia	Firmicutes - Others	Gammaproteobacteria - Enterobacteria	Gammaproteobacteria - Others	Spirochaetes
2	K00615	42	3	14	0	3	1	17	2	0	2	0
3	K00616	8	2	3	0	0	0	3	0	0	0	0
4	K13810	0	0	0	0	0	0	0	0	0	0	0
5	K01783	16	0	7	1	0	0	5	0	2	0	1
6	K01807	2	2	0	0	0	0	0	0	0	0	0
7	K01808	15	0	5	0	2	0	7	1	0	0	0
8	Sum.	83	7	29	1	5	1	32	3	2	2	1
9	MCR(%)	-	75	100	25	50	25	100	50	25	25	25
10	Q-value	-	0.750319	0	0.998722	0.926176	0.99361	0	0.926176	0.998722	0.99361	0.998722
11	Abundance**	0.419	-	0.209512	-	-	-	0.209512	-	-	-	-

Fig. 8

3-3. Comparison of results

Users can compare results not only between their own jobs but also between job (s) and KEGG-annotated genomes by clicking the “MAPLE job comparison” button on the top of the page and then inputting an e-mail address to access the job list. To conduct a comparative analysis, the “Comparison” button can be selected after checking the job IDs to be compared on the job list page. When several IDs of jobs to be compared are checked in the job list and the “Comparison” button is clicked, the job arrangement page is displayed. The user

can add pre-analyzed individual organisms from the organism list if necessary and change the display order. The comparison results of MCR values and mapping patterns for each KEGG module are displayed side by side in parallel (Fig. 9). Detailed information for each KEGG module and the taxonomy of the KO-assigned genes mapped to the module are displayed, and the MCR and taxonomy results of the KO-assigned genes can be downloaded, as previously mentioned. Comparisons between KEGG-annotated genomes, excluding the user' s jobs, are also possible. The user can directly access the comparison page without submitting an e-mail address using the "MAPLE genome comparison" button at the top of the page.

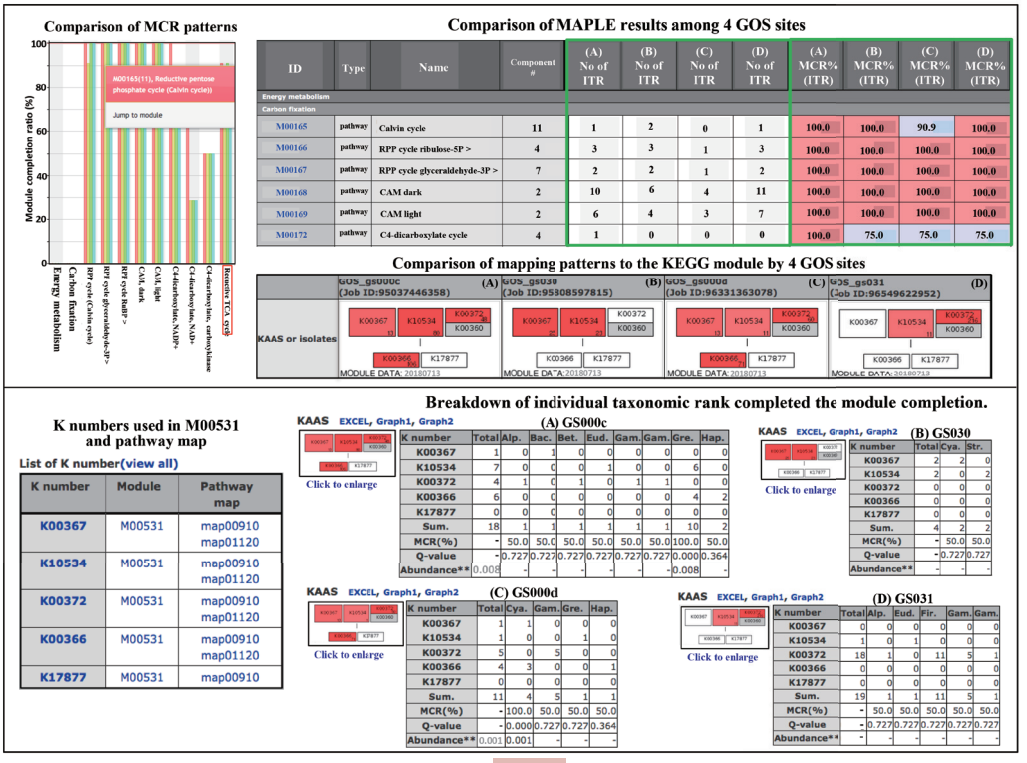


Fig. 9

3-4. Visualization of MAPLE results using MAPLE Graph Maker

MAPLE results, such as MCR, Q-value, and module abundance, can be easily downloaded as an Excel file. Drawing histograms using these files is laborious because there are 800 functional modules. To address this difficulty, we developed MAPLE Graph Maker (MGM) to automatically draw histograms of MAPLE results (Fig. 1). Users can easily create histograms by importing Excel files containing their MAPLE results. After a MAPLE job has been completed, the user can download an Excel file summarizing all the MAPLE results (MCR, Q-value, and module abundance) from the first results page (Fig. 7). When MGM is started, the initial menu is shown.

When the user selects an Excel file and clicks the "Read Excel" button, a new menu is displayed after 20 seconds, from which the user can select data (i.e., MCR%, Q-value, and module abundance) and drawing parameters (combination and order of the data and the elimination of the histogram showing that MCR% is zero in all samples used for comparison). When the user clicks the "Show Graph" button after making the appropriate selections, the histogram is automatically displayed. This histogram can be saved as a PDF file.

chapter 4

Notes

4-1. KEGG module

KEGG MODULE [8] is a collection of pathway modules and other functional units designed for automatic functional annotation and pathway enrichment analysis. As of July 13, 2018, a total of 305 pathway modules have been defined for energy, carbohydrate, lipid, nucleotide, and amino acid metabolisms, including genetic and environmental information processing pathways. A total of 796 KEGG modules (305 pathways, 294 structure complexes, 157 functional sets, and 40 signatures) can be accessed through the website (http://www.genome.jp/kegg-bin/get_htext?ko00002.keg). Pathway modules in KEGG MODULE correspond to smaller portions of subpathways (Fig. 3A), manually defined as consecutive reaction steps, operons, or other regulatory units, and phylogenetic units obtained by genome comparisons (Fig. 3B). Other functional units include (1) structural complexes representing sets of protein subunits for molecular machineries such as ATPase, (2) functional sets representing other types of essential sets such as aminoacyl-tRNA synthases, and (3) signature modules representing markers of phenotypes such as the enterohemorrhagic Escherichia coli pathogenicity signature for Shiga toxin. Each module is defined by a combination of KO identifiers (IDs) such that it can be used for annotation and interpretation purposes in individual genomes or metagenomes. Notations of the Boolean algebra-like equation (Fig. 3) for this definition include space-delimited items indicating pathway elements, comma-separated items in parentheses indicating alternatives, plus signs indicating complexes, and minus signs indicating optional items.

4-2. Distribution patterns of the module completion ratio for 3,186 prokaryotes

Each KEGG module is designed for automatic functional annotation by a Boolean algebra-like equation of KEGG Orthology IDs. However, it remains unclear which species possess common modules or whether certain modules demonstrate universality or rareness among specific taxa. Specific information regarding the phylogenetic profiles of each module holder would be especially useful for annotating metagenomes [4]. Thus, we first examined distribution patterns of the MCRs of the KEGG modules for the 3,186 prokaryotic species whose genomic sequences have been completed. Although the distribution of the MCRs for the 3,186 species varied greatly depending on the kind of module (Fig. 10), we found that there were essentially four patterns (i.e., universal, restricted, diversified, and non-prokaryotic) regardless of the module type (i.e., pathway, structural complex, signature, or functional set) when considering 70% of all species to represent a majority measurement for the patterns.

Pattern A (defined as “universal”) comprised modules completed for more

than 70% of the 3,186 species (Fig. 10A-1), and more than 70% of the 3,186 species possessed MCRs of >80%, referred to as pattern A-2 (Fig. 10A-2). Only 9.0% of the pathway modules were grouped into pattern A, and they mainly belonged to the categories of central carbohydrate metabolism and cofactor and vitamin biosynthesis. Although there are many species, more than 70% of the 3,186 prokaryotes possessed MCRs of 80%. Species with a 100% completion ratio were very limited within the pattern A-2 group. M00096, a representative of pattern A-2 (Fig. 10), is a pathway module for C5 isoprenoid biosynthesis, a non-mevalonate pathway comprising eight reaction steps. Pattern B (defined as "restricted") is composed of modules completed by less than 30% of the species (Fig. 10B) and accounted for 24.7% of all the pathway modules, and 66 modules were rare modules, completed for less than 10% of the 3,186 species.

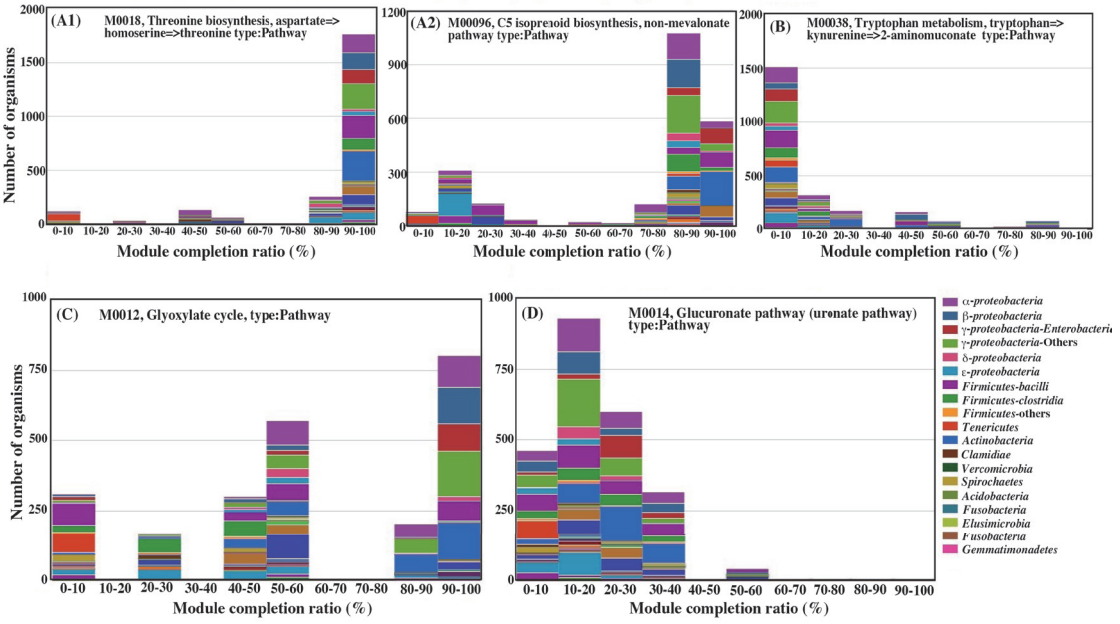
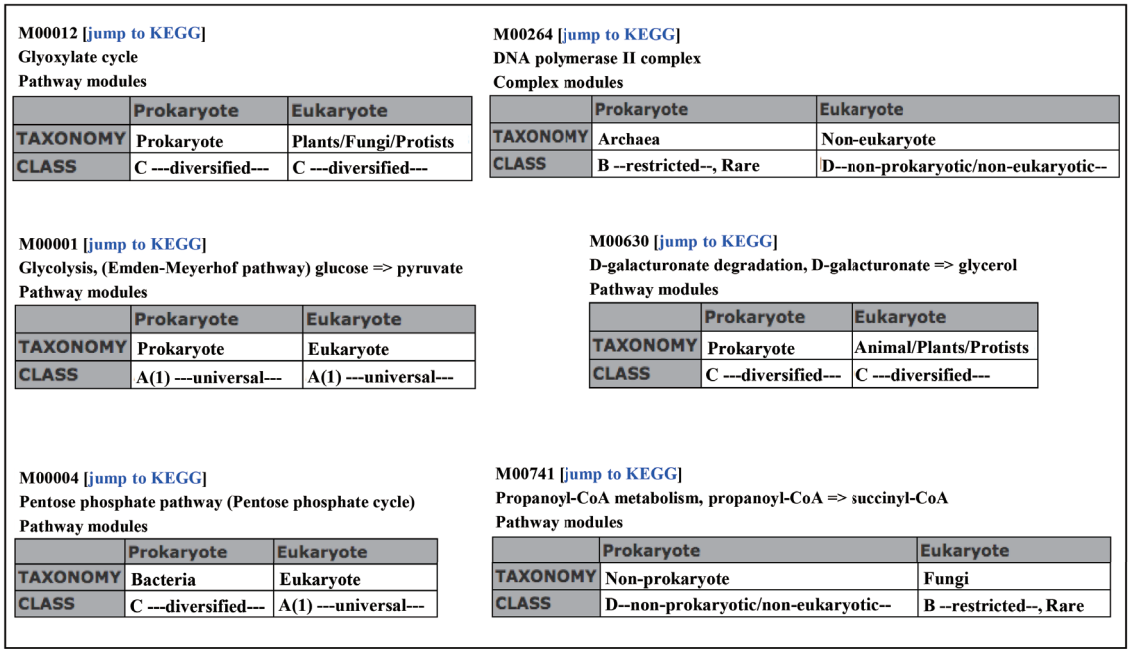


Fig. 10

Pattern C (defined as "diversified") accounted for 33.7% of all the pathway modules and comprised modules ranging widely in completion ratios. M00012 (the glyoxylate cycle, comprising five reactions) is one of the representatives of pattern C (Fig. 10C). One or several KO IDs were assigned to each reaction in this module; however, KO IDs, except for K01637 and K01638 assigned to the third and fourth reactions, were also assigned to other pathway modules such as the tricarboxylic acid (TCA) or Krebs cycle (M00009), reductive TCA cycle (M00173), and C4-dicarboxylate cycle (nicotinamide adenine dinucleotide (NAD)+-malic enzyme type; M00171). Some KO IDs assigned to many of the modules categorized into pattern C were also assigned to several other independent modules [4]. Thus, when the MCR is low, the relationship between the MCR of the targeted module and others for which the same KO IDs were assigned should be considered. Pattern D, which accounted for 32.6% of all pathway modules, comprised nonprokaryotic modules that are not completed for prokaryotic species (Fig. 10D).

Among structural complex modules redefined from modules with various complex patterns, 150 modules were categorized into pattern B (51.7%), and 119 were rare modules. Pattern C accounted for only 18.6% of the structural modules compared with 33.7% of the pathway modules. Thus, it was hypothesized that most of the structural complex modules, except for those conforming to pattern D, are shared only among limited prokaryotic species. Non-prokaryotic modules account for 32.6% of the pathway (e.g., M00741) and 26.6% of structural complex modules, respectively, and other modules were classified into various taxonomic patterns such as prokaryotic, Bacteria-specific, and Archaea-specific based on the module completion profiles (Fig. 11). These four patterns indicate the universal and unique nature of each module and also the versatility of the KO IDs mapped to each module. Thus, the four criteria and taxonomic classifications for each module should be helpful for interpretation of results based on module completion profiles.



M00012 [jump to KEGG]
Glyoxylate cycle
Pathway modules

	Prokaryote	Eukaryote
TAXONOMY	Prokaryote	Plants/Fungi/Protists
CLASS	C ---diversified---	C ---diversified---

M00264 [jump to KEGG]
DNA polymerase II complex
Complex modules

	Prokaryote	Eukaryote
TAXONOMY	Archaea	Non-eukaryote
CLASS	B --restricted--, Rare	D--non-prokaryotic/non-eukaryotic--

M00001 [jump to KEGG]
Glycolysis, (Emden-Meyerhof pathway) glucose => pyruvate
Pathway modules

	Prokaryote	Eukaryote
TAXONOMY	Prokaryote	Eukaryote
CLASS	A(1) ---universal---	A(1) ---universal---

M00630 [jump to KEGG]
D-galacturonate degradation, D-galacturonate => glycerol
Pathway modules

	Prokaryote	Eukaryote
TAXONOMY	Prokaryote	Animal/Plants/Protists
CLASS	C ---diversified---	C ---diversified---

M00004 [jump to KEGG]
Pentose phosphate pathway (Pentose phosphate cycle)
Pathway modules

	Prokaryote	Eukaryote
TAXONOMY	Bacteria	Eukaryote
CLASS	C ---diversified---	A(1) ---universal---

M00741 [jump to KEGG]
Propanoyl-CoA metabolism, propanoyl-CoA => succinyl-CoA
Pathway modules

	Prokaryote	Eukaryote
TAXONOMY	Non-prokaryote	Fungi
CLASS	D--non-prokaryotic/non-eukaryotic--	B --restricted--, Rare

Fig. 11

References

[1] Venter JC, Remington K, Heidelberg JF. Halpern AL, Rusch D, Eisen JA. et al. (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science*, 304: 66–74.

[2] Arai, W., Taniguchi, T., Goto, S., Moriya, Y., Uehara, H. et al. (2018) MAPLE 2.3.0: an improved system for evaluating the functionomes of genomes and metagenomes. *Biosci. Biotechnol. Biochem.,* 82: 1515-1517.

[3] Kanehisa M. Araki M. Goto S, Hattori M, Hirakawa M, Itoh M. et al. (2008) KEGG for linking genomes to life and environment. *Nucleic Acids Res.,* 36: D480–484.

[4] Takami H, Taniguchi T, Moriya Y, Kuwahara T, Kanehisa M, Goto S. (2012) Evaluation method for the potential functionome harbored in the genome and metagenome. *BMC Genomics,* 13: 699.

[5] Takami H, Taniguchi T, Arai W, Takemoto K, Moriya Y, Goto S. (2016) An automated system for evaluation of the potential functionome: MAPLE version 2.1.0. *DNA Res.,* 23: 467–475.

[6] Suzuki S, Kakuta M, Ishida T, Akiyama Y. (2014) GHOSTX: An Improved Sequence Homology Search Algorithm Using a Query Suffix Array and a Database Suffix Array. *PLoS ONE,* 9: e103833.

[7] Liu Z, Lozupone C, Hamady M, Bushman FD, Knight R. (2007) Short pyrosequencing reads suffice for accurate

microbial community analysis. *Nucleic Acids Res.,* 35: e120.

[8] Stoddard SF, Smith BJ, Hein R, Roller BRK, Schmidt TM. (2015) *rrn*DB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. *Nucleic Acids Res.,* 43: D593-598.

[9] Mende DR, Sunagawa S, Aeller G, Bork P. (2013) Accurate and universal delineation of prokaryotic species. *Nat Methods,* 10: 881–884.

Examples of research paper using MAPLE and the MAPLE related paper (2018)

[1] Dangi AK, Sharma B, Hill RT. et al. (2018) Bioremediation through microbes: Systems biology and metabolic engineering approach. *Critical reviews in Biotechnology,* doi.org/10.1080/07388551.2018.1500997.

[2] Maeda AH, Nishi S, Ishii S, Shimane Y. et al. (2018) Complete Genome Sequence of Altererythrobacter sp. Strain B11, an Aromatic Monomer-Degrading Bacterium, Isolated from Deep-Sea Sediment under the Seabed off Kashima, Japan. *Genome Announcements,* 6, e00200-18.

[3] Lu HP, Liu PY, Wang Y, Hsieh JF, Ho HC. et al. (2018) Functional Characteristics of the Flying Squirrel's Cecal Microbiota under a Leaf-Based Diet, Based on Multiple Meta-Omic Profiling. *Frontiers in Microbiology*. 8, 2622.

[4] Collins-Fairclough AM, Ellis MC, Hug LA. (2018) Widespread Antibiotic, Biocide, and Metal Resistance in Microbial Communities Inhabiting a Municipal Waste Environment and Anthropogenically Impacted River.3, *mSphere,* e00346-18.

[5] Espinoza J, Harkins D, Torralba M, Gomezet A. et al. (2018) Supragingival plaque microbiome ecology and functional potential in the context of health and disease. *bioRxiv,* doi: https://doi.org/10.1101/325407.

[6] Ohta Y, Shimane Y, Nishi S, Ichikawa J. et al. (2018) Complete Genome Sequence of Sphingobium sp. Strain YG1, a Lignin Model Dimer-Metabolizing Bacterium Isolated from Sediment in Kagoshima Bay, Japan. *Genome Announcements,* 6, e00267-18.

[7] Richter DJ, Fozouni P, Eisen MB, King N. et al. (2018) Gene family innovation, conservation and loss on the animal stem lineage. *eLife,* 7, e34226.

[8] Campanaro S, Treu L, Kougias PG, Luo G, Angelidaki I. (2018) Metagenomic binning reveals the functional roles of core abundant microorganisms in twelve full-scale biogas plants. *Water Research,* 140, 123-134.

[9] Kogawa M, Hosokawa M, Nishikawa Y, Mori K. et al. (2018) Obtaining high-quality draft genomes from uncultured microbes by cleaning and co-assembly of single-cell amplified genomes. *Scientific reports,* 8: 2059.

[10] Cairns J, Jokela R, Hultman J, Tamminen M. et al. (2018) Construction and characterization of synthetic bacterial community for experimental ecology and evolution. *Frontiers in Genetics,* 9, 312.

(2017)

[11] Suzuki S, Ishii S, Hoshino T, Rietze A, Tenney A. et al. (2017) Unusual metabolic diversity of hyperalkaliphilic microbial communities associated with subterranean serpentinization at The Cedars. *The ISME Journal,* 11, 2584-2598.

[12] Pilgrim J, Ander M, Garros C, Baylis M. et al. (2017) Torix group Rickettsia are widespread in Culicoides biting midges (Diptera: Ceratopogonidae), reach high frequency and carry unique genomic features. *Environmental Microbiology,* 19, 4238-4255.

[13] Zheng HQ, Wu NY, Chow CN, Tseng KC. et al. (2017) EXPath tool—a system for comprehensively analyzing regulatory pathways and coexpression networks from high-throughput transcriptome data. *DNA Research,* 24, 371-375.

[14] Zeng Q, Tian X, Wang L. (2017) Genetic adaptation of microbial populations present in high-intensity catfish production systems with therapeutic oxytetracycline treatment. *Scientific reports,* 7, 17491.

[15] Kuo V, Shoemaker WR, Muscarella ME. et al. (2017) Whole-Genome Sequence of the Soil Bacterium Micrococcus sp. KBS0714. *Genome Announcements,* 32, e00697-17.

[16] Takami H, Toyoda A, Uchiyama I, Itoh T, Takaki Y. et al. (2017) Complete genome sequence and expression profile of the commercial lytic enzyme producer Lysobacter enzymogenes M497-1. *DNA Research,* 24, 169-177.

[17] Nishimura I, Shiwa Y, Sato A, Oguma T. et al. (2017) Comparative genomics of *Tetragenococcus halophilus. The Journal of General and Applied Microbiology,* 63, 369-372.

[18] Hayatsu M, Tago K, Uchiyama I, Toyoda A, Wang Y. et al. (2017) An acid-tolerant ammonia-oxidizing γ-proteobacterium from soil. *The ISME Journal,* 11, 1130-1141.

(2016)

[19] Hiraoka S, Yang C, Iwasaki W. Metagenomics and bioinformatics in microbial ecology: current status and beyond. *Microbes and environments,* 31, 204-212. (2016)

[20] Ngugi DK, Blom J, Stepanauskas R, Stingl U. (2016) Diversification and niche adaptations of Nitrospina-like bacteria in the polyextreme interfaces of Red Sea brines. *The ISME journal,* 10, 1383-1399.

[21] Vikram S, Guerrero LD. et al. (2016) Metagenomic analysis provides insights into functional capacity in a hyperarid desert soil niche community. *Environmental Microbiology,* 18, 1875-1888.

[22] Vaudel M, Barsnes H, Bjerkvig R. et al. (2016) Practical considerations for omics experiments in biomedical sciences. *Current Pharmaceutical Biotechnology,* 17, 105-114.

[23] Kotera M, Goto S. (2016) Metabolic pathway reconstruction strategies for central metabolism and natural product biosynthesis. *Biophysics and Physicobiology,* 13, 195-205.

[24] Tabei Y, Yamanishi Y, Kotera M. (2016) Simultaneous prediction of enzyme orthologs from chemical transformation patterns for de novo metabolic pathway reconstruction. *Bioinformatics,* 32, i278-i287.

[25] Fujinawa K, Asai Y, Miyahara M, Kouzuma A, Abe T. et al. (2016) Genomic features of uncultured methylotrophs in activated-sludge microbiomes grown under different enrichment procedures. *Scientific Reports,* 6, 26650.

[26] Sánchez LFH, Aasebø E, Selheim F, Bervenet FS. et al. (2016) Systemic analysis of regulated functional networks. Proteomics in Systems Biology. pp. 287-310.[27] Higashi K, Kawai Y, Baba T, Kurokawa K, Oshima T. (2016) Essential cellular modules for the proliferation of the primitive cell. Geoscience Frontiers, 9, 1155-1161.

[27] Higashi K, Kawai Y, Baba T, Kurokawa K, Oshima T. (2016) Essential cellular modules for the proliferation of the primitive cell. *Geoscience Frontiers,* 9, 1155-1161.

(2015)

[28] Takami H, Arai W, Takemoto K, Uchiyama I. et al. (2015) Functional classification of uncultured "Candidatus Caldiarchaeum subterraneum" using the MAPLE system. *PLoS one,* 10(7): e0132994.

[29] Obregon-Tito AJ, Tito RY, Metcalf J, Sankaranarayanan K, Clemente JC, Ursell LK. et al. (2015) Subsistence strategies in traditional societies distinguish gut microbiomes. *Nat Com.,* 6: 6505.

Figure legends

Fig.1. Outline of functional metagenomic analysis using the MAPLE system. MAPLE Submission Data Maker (MSDM), for preparing multi-FASTA file, and MAPLE Graph Maker (MGM) and MAPLE Metabolic Map Viewer (MMMV), for visualizing MAPLE results, are available from the website (https://maple.jamstec.go.jp/maple/maple-2.3.1/softdownload/).

Fig. 2. Workflow of the MAPLE system, including the four steps and intermediate results at each step. This diagram was slightly modified from the original version with permission from *DNA Research* [10]. (A) Query sequences submitted to the MAPLE system must be amino acid sequences containing partial genes from metagenomic sequences generated by high-throughput DNA sequencers, such as Illumina MiSeq and HiSeq. (B) The Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthology (KO) assignments are performed by the KEGG Automatic Annotation Server (KAAS) on the basis of results from the BLASTP program. (C) The query sequences should be in multi-FASTA format with unique IDs, and the gene IDs must not include tabs. After the KO assignment to each query sequence is finished, mapping of the KO-assigned sequences to the KEGG functional modules starts, and (D-1) subsequently, the module completion ratio (MCR) is calculated. (D-2) After MCR calculation, KO abundance and module abundance are calculated. (D-3) Finally, the Q-value of the module is calculated based on the MCR results, the abundance, and the similarity score of each KO-assigned sequence mapped to the module. MSDM, MAPLE Submission Data Maker.

Fig. 3. Glycolysis reactions registered in the KEGG pathway database. (A) KEGG reaction map for glycolysis. (B) KEGG functional module for glycolysis corresponding to the reaction map. The module M00001 comprising 10 reactions is defined for the glycolysis module and represented as a Boolean algebra-like equation of KEGG Orthology identifiers or K numbers for computational applications. The relationship between this module and the corresponding KEGG pathway map is also indicated by the corresponding K number sets in the module and Enzyme Commission (EC) numbers in the pathway map using the same index. In each K number set, horizontally arranged K numbers indicate alternatives, related to each other with "Or" or "," in the equation.

Fig. 4. Illustration of the Q-value concept. This image was slightly modified from the original version with permission from *DNA Research* [10] because the M00530 module was redefined by the Kyoto Encyclopedia of Genes and Genomes. (A) Schematic diagram of a reaction module. (B-E) As an example, the reaction module M00529 is shown. The weight of the K number (e.g., K00370) indicates the sequence similarity scores. The Q-value is lower in the following cases. (B) The module is more complete (i.e., it has a higher module completion ratio). (C) The module consists of genes with higher similarity scores. Red numbers indicate high similarity scores. (D) Genes are more abundant. (E) Enzymes or reactions include alternative elements (e.g., isozymes exist). (F) The module has less overlap with the other modules because there are fewer multiple comparisons. Note that the left-side module in (F) is ideal (i.e., M00529 is assumed to be independent from M00530) for the purpose of illustration. MCR, module completion ratio.

Fig. 5. Illustration of the module abundance concept. This diagram was slightly modified from the original version with permission from *DNA Research* [10]. Module M00529 comprises four reaction components and is defined as a denitrification reaction. In each K number set, vertically arranged K numbers indicate a complex whereas horizontally arranged K numbers indicate alternatives. Small numbers in the lower right of the boxes indicate the abundance of the Kyoto Encyclopedia of Genes and Genomes Orthology (KO)-assigned genes normalized by the mean number of genes categorized in each orthologous group (i.e., KO). In the case of a complex (1 and 3), the minimum number is defined as an abundance of a complex. When there are alternatives in the reaction component (i.e., 1, 2, and 3), the sum of both abundances is defined as the abundance of each reaction component. Finally, the minimum abundance of the four reaction components is defined as the module abundance.

Fig. 6. Virtual modules for prokaryotic and all ribosome (M90000 and M91000) and taxonomical characterization of the ribosomal proteins by the MAPLE system. This figure was modified from the original version with permission from *DNA Research* [10]. (A) Organization of the module. The M90000 is composed of 79 (21 bacterial, 27 archaeal, and 31 common) and the M91000 is composed of 130 (M90000+51 eukaryotic) ribosomal proteins, respectively. Accessory KEGG Orthologies (KOs) are not used to define the module. Because most of the genes for broadly conserved ribosomal proteins are single copy within individual genomes and specific for each organism, the taxonomic information for each KO assigned to this module can be used for more precise analysis of the taxonomical composition of microbiomes from various environments. In contrast, 16S rRNA genes are of limited value because their copy number obviously varies among organisms.

Fig. 7. Overview of MAPLE results. This figure was modified from the original version with permission from *DNA Research* [10]. Table containing all MAPLE results (module completion ratio [MCR], Kyoto Encyclopedia of Genes and Genomes [KEGG] Orthology [KO] and module abundances, and Q-value) are displayed by clicking a job ID. The MCR results are displayed as both a table and histogram. Detailed results for each module, such as the mapping pattern and taxonomic information, can be displayed by clicking the module ID in the table or module name in the histogram. The list of KO-assigned genes by the KEGG Automatic Annotation Server (KAAS), MAPLE results, and taxonomic information for KO-assigned sequence are downloadable from the links highlighted by blue boxes on the first results page. An example of a downloaded file is shown in Fig. 8. The data package needed to redisplay all MAPLE results after the expiration of data storage on the MAPLE server is also downloadable from the same page.

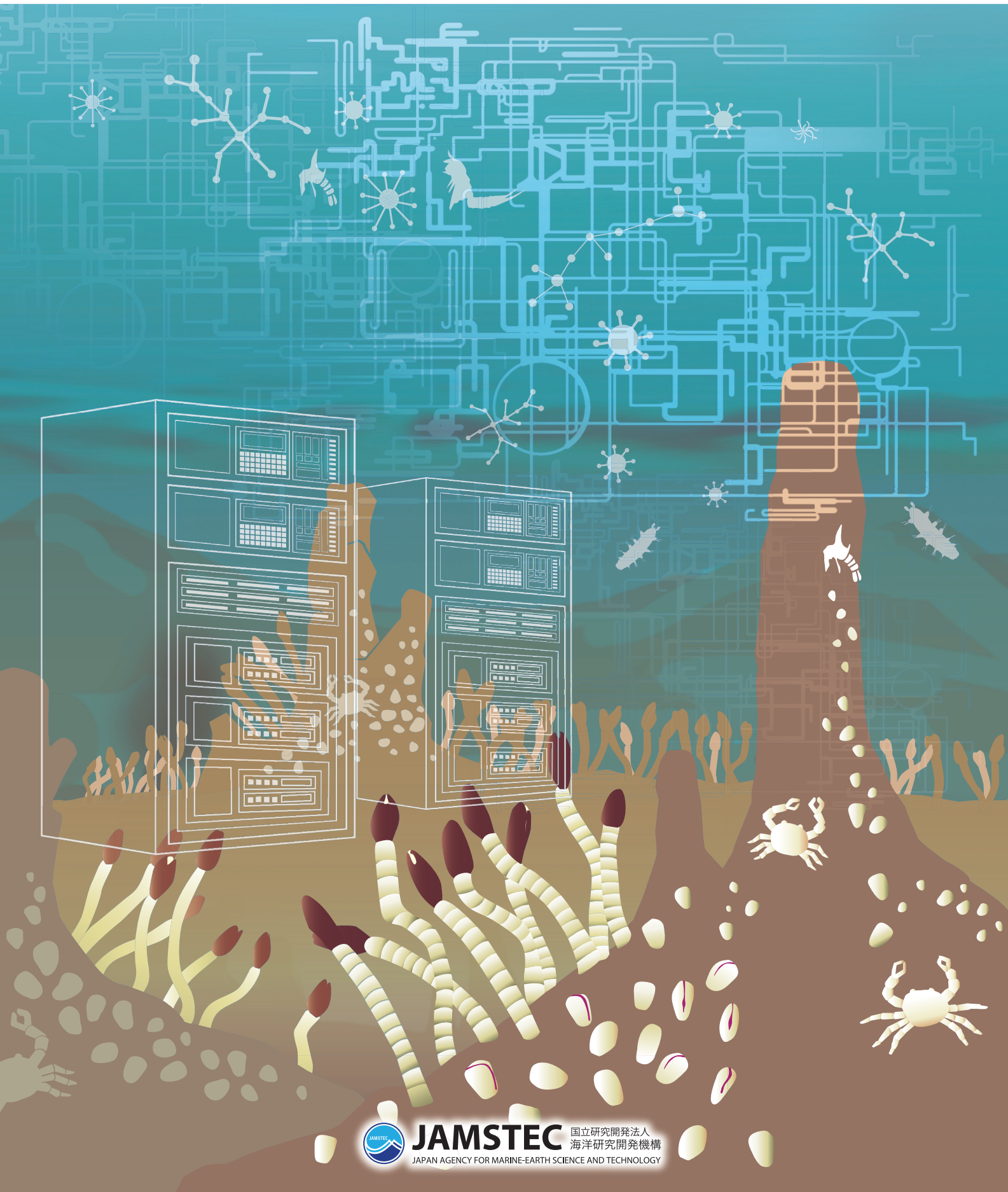
Fig. 8. Examples of downloadable results. All results generated by MAPLE are downloadable in an Excel-compatible format (tab delimited). This figure was reproduced with permission from *DNA Research* [10]. (1) An example of downloaded files containing the Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthology (KO) assignment results for the genes performed by the KEGG Automatic Annotation Server (KAAS). (2) Module completion ratio (MCR), individual taxonomic rank (ITR), Q-value, and module abundance. (3) Taxonomic breakdown of the genes mapped to the KEGG modules. WC, whole community.

Fig. 9. Results from comparative analyses of different metagenomic samples. This figure was modified from the original version with permission from *DNA Research* [10]. Module completion ratio (MCR), patterns of mapping to the modules, abundance of complete modules, and taxonomic information are displayed and available in Excel format. GOS, Global Ocean Sampling; ITR, individual taxonomic rank; KEGG, Kyoto Encyclopedia of Genes and Genomes; KAAS, KEGG Automatic Annotation Server.

Fig. 10. Typical completion patterns of the Kyoto Encyclopedia of Genes and Genomes (KEGG) modules for 2367 prokaryotic species. (A) Universal modules. (A-1) Modules completed by more than 70% of 768 prokaryotic species, such as M00018, which represents threonine biosynthesis (aspartate homoserine threonine). (A-2) Modules for which more than 70% of 2367 prokaryotic species show a MCR of >80%, such as M00096, which represents C5 isoprenoid biosynthesis and is a non-mevalonate pathway. (B) Restricted modules completed by less than 30% of 768 prokaryotic species, such as M00038, which represents tryptophan metabolism (tryptophan kynurenine 2-aminomuconate). (C) Diversified modules, that is modules that vary in their MCRs among 768 prokaryotic species, such as M00012, which represents the glyoxylate cycle. (D) Nonprokaryotic modules completed by no prokaryotic species, such as M00014, which represents the glucuronate pathway (uronate pathway). Some examples of the taxonomic variation that completes each KEGG module are shown in Fig. 4.

Fig. 11. Taxonomic variation that completes Kyoto Encyclopedia of Genes and Genomes (KEGG) modules. Taxonomic variation and completion patterns of each module can be displayed in the MAPLE results page. Some typical examples are shown in this figure.

Version.1
○○○.0.2019



JAMSTEC 国立研究開発法人 海洋研究開発機構
JAPAN AGENCY FOR MARINE-EARTH SCIENCE AND TECHNOLOGY