ARTICLE

OPEN ACCESS

www.the-innovation.org

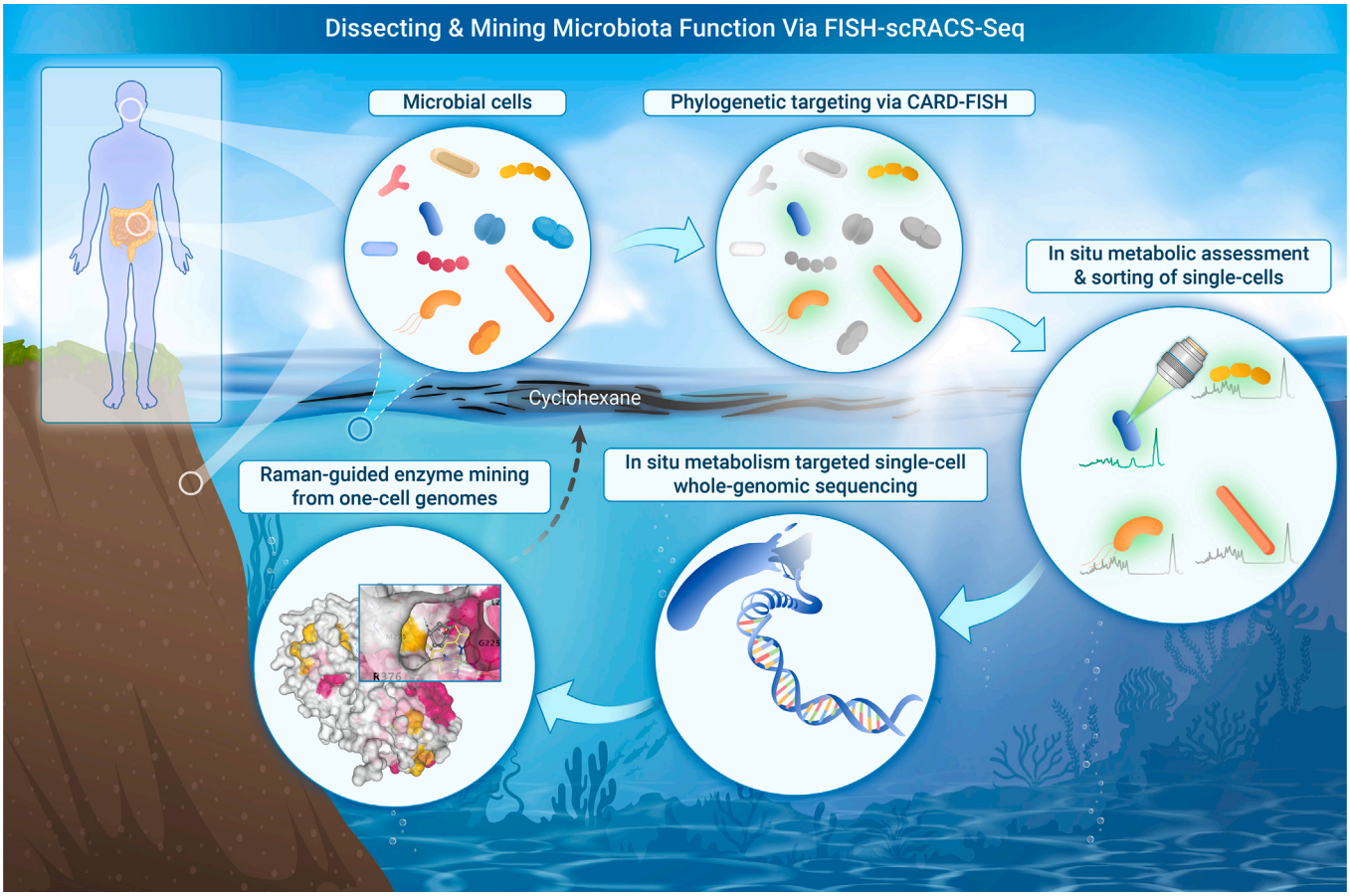
Phylogeny-metabolism dual-directed single-cell genomics for dissecting and mining ecosystem function by FISH-scRACS-seq

Xiaoyan Jing,[1,6,7,8,10] Yanhai Gong,[1,6,7,8,10] Zhidian Diao,[1,6,7,8,10] Yan Ma,[1,2,10] Yu Meng,[1,6,7,8] Jie Chen,[1,6,7,8] Yishang Ren,[1,6,7,8] Yuting Liang,[5] Yinchao Li,[3] Weihan Sun,[4] Jia Zhang,[1,6,7,8] Yuetong Ji,[9] Zhiqi Cong,[1,6,7,8] Shengying Li,[4] Bo Ma,[1,6,7,8] Zhisong Cui,[3,*] Li Ma,[4,*] and Jian Xu[1,6,7,8,*]

*Correspondence: czs@fio.org.cn (Z.C.); maliqd@sdu.edu.cn (L.M.); xujian@qibebt.ac.cn (J.X.)

Received: June 9, 2024; Accepted: December 9, 2024; Published Online: January 2, 2025; https://doi.org/10.1016/j.xinn.2024.100759

GRAPHICAL ABSTRACT



PUBLIC SUMMARY

- A single-cell genomics approach based on phylogeny and metabolism was introduced.
- γ-Proteobacteria that degrade cyclohexane in the sea was identified and sequenced.
- A novel yet globally found P450 system for cyclohexane degradation was discovered.
- Metabolism and genomes were linked to ecosystem function at single-cell precision.

OPEN ACCESS

ARTICLE

The Innovation

Phylogeny-metabolism dual-directed single-cell genomics for dissecting and mining ecosystem function by FISH-scRACS-seq

Xiaoyan Jing,[1,6,7,8,10] Yanhai Gong,[1,6,7,8,10] Zhidian Diao,[1,6,7,8,10] Yan Ma,[1,2,10] Yu Meng,[1,6,7,8] Jie Chen,[1,6,7,8] Yishang Ren,[1,6,7,8] Yuting Liang,[5] Yinchao Li,[3] Weihan Sun,[4] Jia Zhang,[1,6,7,8] Yuetong Ji,[9] Zhiqi Cong,[1,6,7,8] Shengying Li,[4] Bo Ma,[1,6,7,8] Zhisong Cui,[3,*] Li Ma,[4,*] and Jian Xu[1,6,7,8,*]

[1]Single-Cell Center, CAS Key Laboratory of Biofuels, Qingdao Institute of BioEnergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao 266000, China

[2]Department of Biotechnology, College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan 430000, China

[3]Marine Bioresource and Environment Research Center, Key Laboratory of Marine Eco-Environmental Science and Technology, First Institute of Oceanography, Ministry of Natural Resources of China, Qingdao 266000, China

[4]State Key Laboratory of Microbial Technology, Shandong University, Qingdao 266000, China

[5]State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Science, Nanjing 211300, China

[6]University of Chinese Academy of Sciences, Beijing 100000, China

[7]Shandong Energy Institute, Qingdao 266000, China

[8]Qingdao New Energy Shandong Laboratory, Qingdao 266000, China

[9]Qingdao Single-Cell Biotechnology, Co., Ltd., Qingdao 266000, China

[10]These authors contributed equally

*Correspondence: czs@fio.org.cn (Z.C.); maliqd@sdu.edu.cn (L.M.); xujian@qibebt.ac.cn (J.X.)

Received: June 9, 2024; Accepted: December 9, 2024; Published Online: January 2, 2025; https://doi.org/10.1016/j.xinn.2024.100759

Citation: Jing X., Gong Y., Diao Z., et al., (2025). Phylogeny-metabolism dual-directed single-cell genomics for dissecting and mining ecosystem function by FISH-scRACS-seq. The Innovation **6(3)**, 100759.

Microbiome-wide association studies (MWASs) have uncovered microbial markers linked to ecosystem traits, but the mechanisms underlying their functions can remain elusive. This is largely due to challenges in validating their *in situ* metabolic activities and tracing such activities to individual genomes. Here, we introduced a phylogeny-metabolism dual-directed single-cell genomics approach called fluorescence-*in situ*-hybridization-guided single-cell Raman-activated sorting and sequencing (FISH-scRACS-seq). It directly localizes individual cells from target taxon via an FISH probe for marker organism, profiles their *in situ* metabolic functions via single-cell Raman spectra, sorts cells of target taxonomy and target metabolism, and produces indexed, high-coverage, and precisely-one-cell genomes. From cyclohexane-contaminated seawater, cells representing the MWAS-derived marker taxon of γ-Proteobacteria and that are actively degrading cyclohexane *in situ* were directly identified via FISH and Raman, respectively, then sorted and sequenced for one-cell full genomes. In such a *Pseudoalteromonas fuliginea* cell, we discovered a three-component cytochrome P450 system that can convert cyclohexane to cyclohexanol *in vitro*, representing a previously unknown group of cyclohexane-degrading enzymes and organisms. Therefore, by unveiling enzymes, pathways, genomes, and their *in situ* cellular functions specifically for those organisms with ecological relevance at one-cell resolution, FISH-scRACS-seq is a rational and generally applicable approach to dissecting and mining microbiota functions.

INTRODUCTION

Microbial consortia, with their rich and diverse metabolic activities, underpin numerous critical ecological processes on Earth, such as geochemical cycling of elements, environmental remediation, and nutrient utilization in hosts.[1–4] To dissect their functioning mechanisms and also mine the underlying bioresources (e.g., useful chassis cells or enzymes), two primary strategies are usually adopted. One is driven by "genotype". For example, microbiome-wide association studies (MWASs) identify DNA-sequence-based taxonomic or functional-gene markers associated with an ecosystem trait by correlating metagenomes with the trait.[5–7] MWAS, a statistical framework, identifies associations between microbial taxa or functions and environmental or clinical variables. While MWAS is commonly employed in clinical research to elucidate associations between microbial communities and health or disease states, it can also be applied to environmental samples, including but not limited to soil,[8–10] wastewater,[11] and marine.[12] One strength of such undirected approaches is the high-throughput and exhaustive discovery of ecosystem-trait-associated organisms or genes, which ensures ecological significance of these markers. However, due to the lack of information for metabolic activities and the challenge in reconstructing individual genomes from those highly heterogeneous metagenomes, it is usually difficult to validate the *in situ* functions of marker taxa and to trace the functions to the underlying genomes, pathways, or enzymes. Consequently, although numerous taxonomic markers have been identified for many microbiota-mediated processes, their functioning mechanisms remain elusive,[13,14] particularly for those involving not yet cultured marker organisms.

The other strategy starts with "metabolic phenotype". For example, Raman-activated cell sorting and sequencing (RACS-seq), a metabolism-directed approach, can directly identify individual cells of target metabolism that corresponds to the ecosystem trait and then track the metabolic activity to the underpinning single-cell genomes.[15–18] Specifically, individual cells of microbiota are profiled for single-cell Raman spectra (SCRS), which serve as a proxy of *in situ* metabolic phenome,[18] and those cells of target phenotypes are sorted via an RACS instrument and then sequenced for their single-cell full genomes at an indexed, one-cell-one-tube manner.[18–20] When coupled with stable isotope probing (SIP-Raman), SCRS can measure cellular intake rate of substrates (e.g., ^{13}C, ^{15}N, ^{18}O, and D),[15,20–26] and the D_2O-intake based cellular vitality can be used to model degradative activity of the carbon source.[23,27,28] In addition, SCRS can reveal the biosynthetic profile of cells (e.g., carotenoids,[29–32] proteins,[33] triacylglycerols,[33–35] and other Raman-sensitive compounds) and characterize cellular response to environmental changes (e.g., susceptibility to drugs or other types of stresses).[18,36,37] A core strength of this strategy is the ability to actually measure *in situ* metabolic activity and directly trace it to genomes, both at single-cell resolution.[31,32,36,38–40] However, due to the sheer number of cells in microbiota yet the comparably low throughput of RACS-seq at present, such a metabolism-directed single-cell genomics approach is usually shallow in sampling depth and narrow in investigative scope. As a result, *in situ* metabolism of those selected cells of interest such as marker organisms identified by MWASs cannot be probed in a targeted manner; i.e., whether the single-cell metabolic phenomes and genomes produced by RACS-seq are of ecological relevance is usually not clear.

To tackle these challenges, we introduce a phylogeny-metabolism dual-directed single-cell omics approach called fluorescence-*in situ*-hybridization-guided single-cell Raman-activated sorting and sequencing (FISH-scRACS-seq). Based on an FISH probe designed from MWAS-derived taxonomical markers, the method directly localizes individual cells of the target taxon in a microbiota, profiles their *in situ* metabolic functions via SCRS, sorts for those cells of target taxonomy and target metabolism, and finally produces their indexed, high-coverage, and precisely-one-bacterial-cell genomes. The method was evaluated via a series of mock community experiments and then demonstrated for soil and seawater microbiota. Coupling of FISH-scRACS-seq to upstream MWAS allowed efficient, culture-independent tracing of cyclohexane degradation in cycloalkane-contaminated seawater from a condensate gas field to the one-cell genomes of uncultured *Pseudoalteromonas fuliginea* and further to a previously unknown group of cytochrome P450-based cyclohexane monooxygenases that can convert cyclohexane to cyclohexanol *in vitro*. Therefore, FISH-scRACS-seq is a rational and generally applicable strategy for dissecting and mining microbiota function.

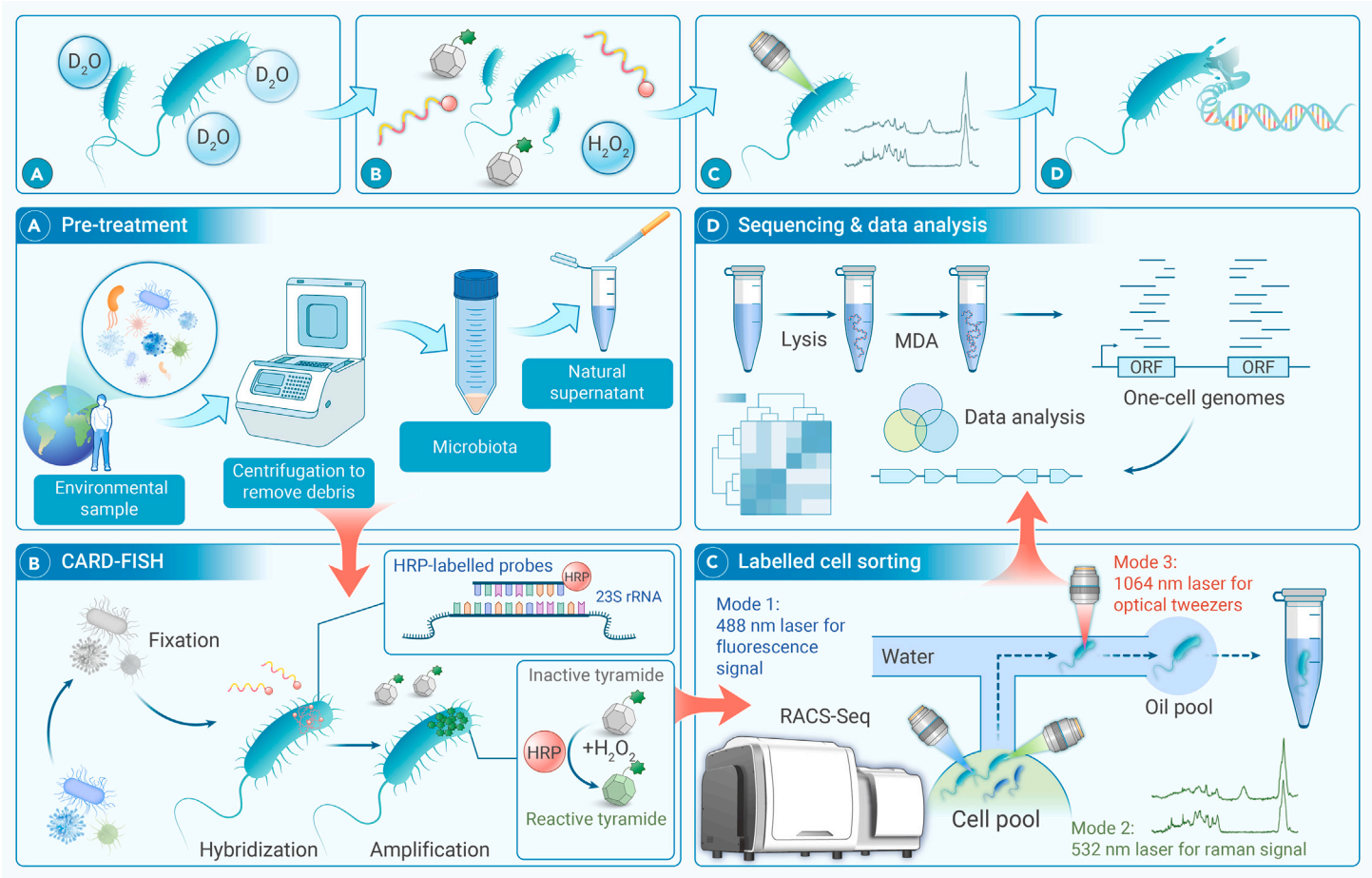


Figure 1. The FISH-scRACS-seq strategy for taxonomy-guided, *in situ*-function-driven profiling of microbiota at single-cell resolution (A) Environmental microbial extraction. Microbial cells are extracted from environmental samples, and deuterium oxide (D_2O) labeling is initiated to assess *in situ* metabolic activity. (B) Phylogenetic targeting via CARD-FISH. Cells are fixed and hybridized with taxon-specific CARD-FISH probes to target γ-Proteobacteria, followed by fluorescent signal amplification. (C) *In situ* metabolic activity assessment and cell sorting. Cells of the targeted phylogeny are identified by their fluorescence signal via a 488-nm laser, followed by acquisition of single-cell Raman spectra that reveal their metabolic phenomes via a 532-nm laser. The cells of targeted phylogeny and targeted metabolic phenomes are sorted using the RAGE chip and then encapsulated as individual droplets. (D) Single amplified genomes (SAGs) sequencing and analysis. Sorted cells are lysed, and their genomic DNA are respectively amplified using multiple displacement amplification (MDA). The MDA products are then respectively subjected to precisely one-cell genome sequencing.

RESULTS

Overview of FISH-scRACS-seq for phylogeny-metabolism dual-directed single-cell omics

The FISH-scRACS-seq workflow for microbiota analysis involves three steps (Figure 1). In step 1 (i.e., FISH), individual cells of a target taxon are directly localized in a microbiota sample via a taxon-specific catalyzed reporter deposition (CARD)-FISH probe, which increases the signal-to-noise ratio.[41–43] Importantly, the FISH process usually sacrifices cellular vitality,[44,45] and pretreatments are performed before hybridization to support downstream SCRS-based profiling of metabolic phenome, such as the feeding of stable-isotope-labeled substrates (e.g., D_2O, $H^{13}CO_3^-$, $^{15}N_2$).[18,46]

In step 2 (i.e., scRACS), post-FISH cells are distinguished and sorted based on both the target phylogeny (via FISH probe) and the target metabolic phenome (via SCRS). CARD-FISH-labeled cells in suspension are trapped and analyzed for SCRS individually in an RAGE chip via a 532-nm laser, which generates high signal-to-noise ratio SCRS. Then, via a technique called single-cell RAGE coupled with sequencing (scRAGE-seq) in an RACS-seq instrument (materials and methods),[38] those CARD-FISH-labeled cells with SCRS profiles matching target metabolic phenotypes are individually captured and moved with a 1,064-nm laser to form one-cell-encapsulated droplets that are then sequentially exported.

In step 3 (i.e., seq), the post-FISH-RACS cells in droplets undergo cell lysis, MDA, and genome sequencing in an indexed, one-cell-one-tube manner (Figure 1). Notably, as the one-cell droplets already carry a mineral-oil phase, an emulsion reaction for MDA (i.e., the process whereby the MDA reaction takes place within tiny droplets of water-in-oil emulsion) are formed simply via vortexing the tube after introducing the lysis buffer. After quality assessment, the one-cell MDA products are shotgun sequenced individually, followed by *de novo* assembly and *in silico* genome analysis. In this way, specifically for those cells of target phylogeny in a microbiota, the target metabolic activity *in situ* is directly traced to genome sequence at single-cell resolution.

Validation of FISH-scRACS-seq using pure-cultured *Escherichia coli* cells

To benchmark the method, we started from a pure culture of *E. coli* K-12 DH5α. Cells were fed 50% D_2O to simulate metabolic profiling, resulting in a Raman band peaking at 2,157 cm^{-1}, indicating the C-D stretching vibrations.[22,23] Following CARD-FISH labeling with the GAM42a probe targeting γ-Proteobacteria (Figure 2A), cells were sorted by scRAGE-seq based on both the taxon-specific fluorescence signal and the C-D band (Figure 2B). Cell identity was validated by 16S rRNA gene PCR from MDA products (Table S1 and Figure S1A). Among 20 sorted cells, 12 yielded successful MDA and PCR results. Then ~1 Gb of raw sequencing data was produced for each cell (E04, E07, E08, E09, E10, E11, E12, E13, E14, and E16; the other two failed in library construction due to severe degradation) (Table S2).

The completeness of these FISH-scRACS-seq-derived single-cell genome (SAG) assemblies ranged from 65.74% to 95.53%, with 60% of SAGs exceeding 80% completeness (Table S2), demonstrating the feasibility of producing high-quality SCRS plus high-coverage one-cell genomes via FISH-scRACS-seq. In addition to RAGE whereby cells are captured with a 1,064-nm laser for droplet formation to preserve cellular vitality in an aquatic environment, we have also tested Raman-activated cell ejection (RACE-seq), which immobilizes and dries cells on a slide and then uses a 532-nm laser for cell ejection via laser-induced

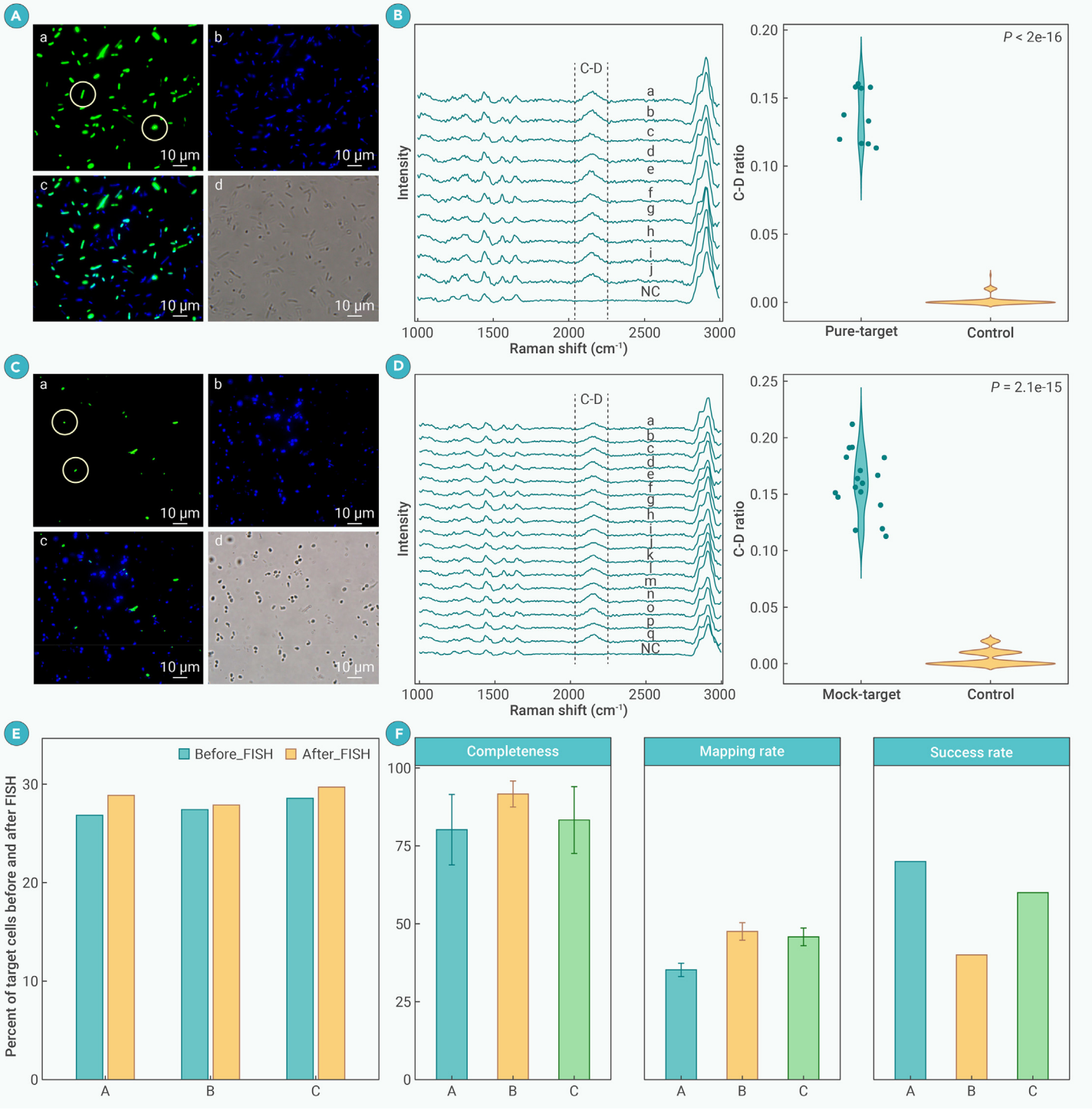


Figure 2. FISH-scRACS-seq accurately and efficiently recovers SAGs from metabolically active, pure-cultured γ-Proteobacteria cells (A) Photomicrographs of CARD-FISH-stained pure-cultured *Escherichia coli* K-12 DH5α. (a) Photomicrographs of cells hybridized with HRP-labeled oligonucleotide probes GAM42a (green); (b) DAPI staining (blue) of cells shown via a color-combined image recorded by epifluorescence microscopy (c); (d) phase-contrast photomicrograph. (b) is the negative control (i.e., without any probes). Scale bar, 10 μm. (B) The C-D peaks (left panel) and their corresponding C-D ratios (right panel) of the target cells, which were sorted via both "taxon-specific" and "metabolic" phenotypes of *E. coli* K-12 DH5α for single-cell genomes. Letters a to j represent cells with C-D peaks in SCRSs of samples E04, E07, E08, E09, E10, E11, E12, E13, E14, and E16, respectively, whereas the "control" SCRSs represent cells without C-D bands. (C) CARD-FISH photomicrographs of four-species mock microbiota hybridized with probe GAM42a. Each series shows identical microscopic fields. (a)–(d) show photomicrographs of the mock bacteria hybridized with γ-Proteobacteria targeting probe GAM42a, DAPI staining of DNA, overlay images of probe signal (green) and DAPI staining (blue), and phase-contrast image, respectively. Scale bars, 10 μm. (D) The C-D peaks (left panel) and their corresponding C-D ratios (right panel) of the target cells from these four species, which were mixed for cell sorting via both "taxon-specific" and "metabolism-specific" features for single-cell genomes. (E) Statistics on the percentage of the target *E. coli* K-12 DH5α cells in this four-species mock microbiota before and after the CARD-FISH experiment. Three batches of experiments (A, B, and C) were performed. (F) Performance validation of FISH-scRACS-seq via statistical analysis of FISH-scRACS-derived SAGs. Three batches of experiments (A, B, and C) were performed. Completeness is shown as the percentage of bases with sequencing reads and total bases in the reference genome. Mapping rate is shown as the percentage of sequencing reads that can be mapped to the reference genome. Success rate is shown as "the number of successful runs/total number of attempted runs". Success was defined based on sequence-based verification of 16S rDNA genes amplified from the gene-specific primer pairs. Statistical analyses (B and D) were performed using Wilcoxon test.

forward transfer.[47,48] To determine whether RAGE or RACE is more suitable for FISH-scRACS-seq, the ten one-cell SAGs from FISH-scRACS-seq were aligned to three five-cells-pooled *E. coli* SAGs derived by RACE (one-cell RACE-seq is not available due to the very low success rate of scRACE-seq; Sequence Read Archive accessions SRA: SRR10549451–SRR10549453).[47] SAGs of FISH-scRACS-seq consistently outperformed those of RACE-seq across all assembly

Table 1. Benchmarking the performance of FISH-scRACS-seq using a four-species mock microbiota, consisting of *Escherichia coli* K-12 DH5α (*Ec*), *Micrococcus luteus* D11 (*Ml*), *Bacillus subtilis* H6 (*Bs*), and *Saccharomyces cerevisiae* BY4742 (*Sc*) in a 1:1:1:1 ratio

		WGS reads mapped to reference genome			One-cell WGS assembly				
Experiment series	Sample ID	Taxonomy of bins based on sequence	Mapping rate of the WGS reads (%)	Average mapping rate (%)	Genome completeness (%)	Average genome completeness (%)	Contamination (%)	WGS consistent with sorting criteria	Success rate (%)
A	FME01	*Ec*	43.0	35.2	99.03	80.18	0.16	yes	70
	FME03	*Ec*	32.7		99.38		0.25	yes	
	FME04	*Ec*	31.7		96.67		0.77	yes	
	FME06	*Ec*	34.2		83.31		5.70	yes	
	FME08	*Ec*	39.5		99.57		0.52	yes	
	FME09	*Ec*	26.3		19.12		0.00	yes	
	FME10	*Ec*	38.9		64.21		4.15	yes	
	NC_A	–	–	–	–	–	–	–	–
B	SME01	*Ec*	43.0	46.6	99.10	91.66	0.75	yes	40
	SME02	*Ec*	39.3		80.86		2.66	yes	
	SME07	*Ec*	49.3		89.46		2.74	yes	
	SME08	*Ec*	54.7		97.21		0.84	yes	
	NC_B	–	–	–	–	–	–	–	–
C	TME02	*Ec*	54.8	45.8	99.76	83.30	0.78	yes	60
	TME03	*Ec*	43.5		99.60		1.50	yes	
	TME05	*Ec*	48.3		99.35		0.21	yes	
	TME06	*Ec*	35.0		37.07		1.72	yes	
	TME08	*Ec*	50.5		98.19		0.65	yes	
	TME09	*Ec*	42.6		65.81		5.09	yes	
	NC_C	–	–	–	–	–	–	–	–

Ec is γ-Proteobacteria.

metrics (genome completeness, sequencing bias, and assembly contiguity; Figures S2A–S2C), recovering more genes (Figure S2, left). This aligns with previous findings that RACE-seq suffers from lower sensitivity and genome coverage due to cell-vitality loss and direct laser exposure.[47,48] In contrast, scRAGE-seq achieves near-complete one-cell genome coverage across various ecosystems.[31,32,36,38,39,49] Therefore, we chose RAGE for FISH-scRACS-seq in subsequent experiments.

We further assessed the impact of the FISH step on downstream scRAGE-seq results by comparing ten SAGs from FISH-scRACS-seq to seven SAGs from scRAGE-seq (SRA: PRJNA574296 in the NCBI SRA database) from the same *E. coli* culture. (1) For read coverage, FISH-scRACS-seq reaches on average ~80.01%, lower than scRAGE-seq (~97.48%; Wilcoxon test, $p < 0.001$) but much higher than RACE-seq (~11.09%; Figure S2A, left). Results from reconstructed draft genomes are similar (Figure S2A, right panel). (2) Regarding uniformity of sequence coverage, FISH-scRACS-seq shows similar mapping uniformity on the genomes to scRAGE-seq and bulk, respectively (Figure S2B, left; SD: Wilcoxon test, 3.65 vs. 2.13, $p < 0.05$; 3.65 vs. 2.06, $p < 0.01$), information entropy (Figure S2B, middle; entropy: Wilcoxon test, 8.54 vs. 8.94, $p > 0.05$; 8.54 vs. 10.11, $p < 0.05$), and dropout rate (Figure S2B, right; dropout: Wilcoxon test, 39.20% vs. 37.60%, $p > 0.05$; 39.20% vs. 49.7%, $p > 0.05$). (3) For assembly continuity, FISH-scRACS-seq produces shorter contigs (NGA50, 9.72 vs. 78.01; Wilcoxon test, $p < 0.001$) than scRAGE-seq (Figure S2C, left), potentially due to FISH-related effects on single-cell DNA extraction or MDA reaction. (4) As for number of recovered genes, FISH-scRACS-seq recovers 86.17% of genes, though 12.98% fewer than scRAGE-seq (from the number of coding sequences, 3,998 vs. 4,725; Wilcoxon test, $p < 0.001$) (Figure S2D, left), reflecting lower genome completeness. Despite the potential impact of FISH on genome quality, FISH-scRACS-seq can still produce functionally informed, high-coverage, one-cell genomes for *E. coli*.

FISH-scRACS-seq demonstrates high specificity and high sensitivity in dissecting a mock microbiota

To evaluate performance of FISH-scRACS-seq for microbiota, we constructed a four-species mock consortium that consists of *Bacillus subtilis* H6 (*Bs*, class Bacillus), *E. coli* K-12 DH5α (*Ec*, class γ-Proteobacteria), *Micrococcus luteus* D11 (*Ml*, class Actinomycetia), and *Saccharomyces cerevisiae* BY4742 (*Sc*, class Saccharomycetes) in a 1:1:1:1 ratio (materials and methods). The sensitivity and specificity of FISH-scRACS-seq were then evaluated by sorting the mock microbiota based on fluorescence signal (i.e., the targeted taxon) and Raman signal (i.e., the targeted metabolic function of vitality via C-D band).

We started from the FISH step by subjecting each respective culture to hybridization with the CARD-FISH probe of GAM42a, which specifically targets *Ec*. Microscopic examination showed that all cells in the *Ec* culture, but no cells in the other three cultures, were labeled with fluorescence (Figures 2C and S3). Moreover, the four cultures of *Bs*, *Ec*, *Ml*, and *Sc* were respectively labeled with 50% D_2O prior to CARD-FISH labeling and then mixed in a 1:1:1:1 ratio (Figure 2D). Cell counting under a fluorescence microscope revealed that mean percentage of *Ec* cells before and after the FISH labeling was ~28.8% and ~27.6%, respectively, showing no significant difference (Wilcoxon test, $p = 0.37$; Figure 2E). This is consistent with the predetermined ratio of *Ec* in microbiota and supports high sensitivity and specificity of CARD-FISH.

Furthermore, triplicate experiments were performed for sorting the mock microbiota based on not just fluorescence (Figure 2E) but the D_2O peak of SCRS (Figures 2D–2F and Table 1). In each run, ten target cells were sorted via the presence of taxon-specific fluorescence and C-D peaks, respectively. One-cell genome sequencing results suggested that (Table 1 and Figure 2F): (1) Sanger sequencing of all the 16S PCR products (from 17 MDA-positive cells in 30 sorted cells) yield only *Ec*-specific 16S rDNA sequences, and the genome completeness

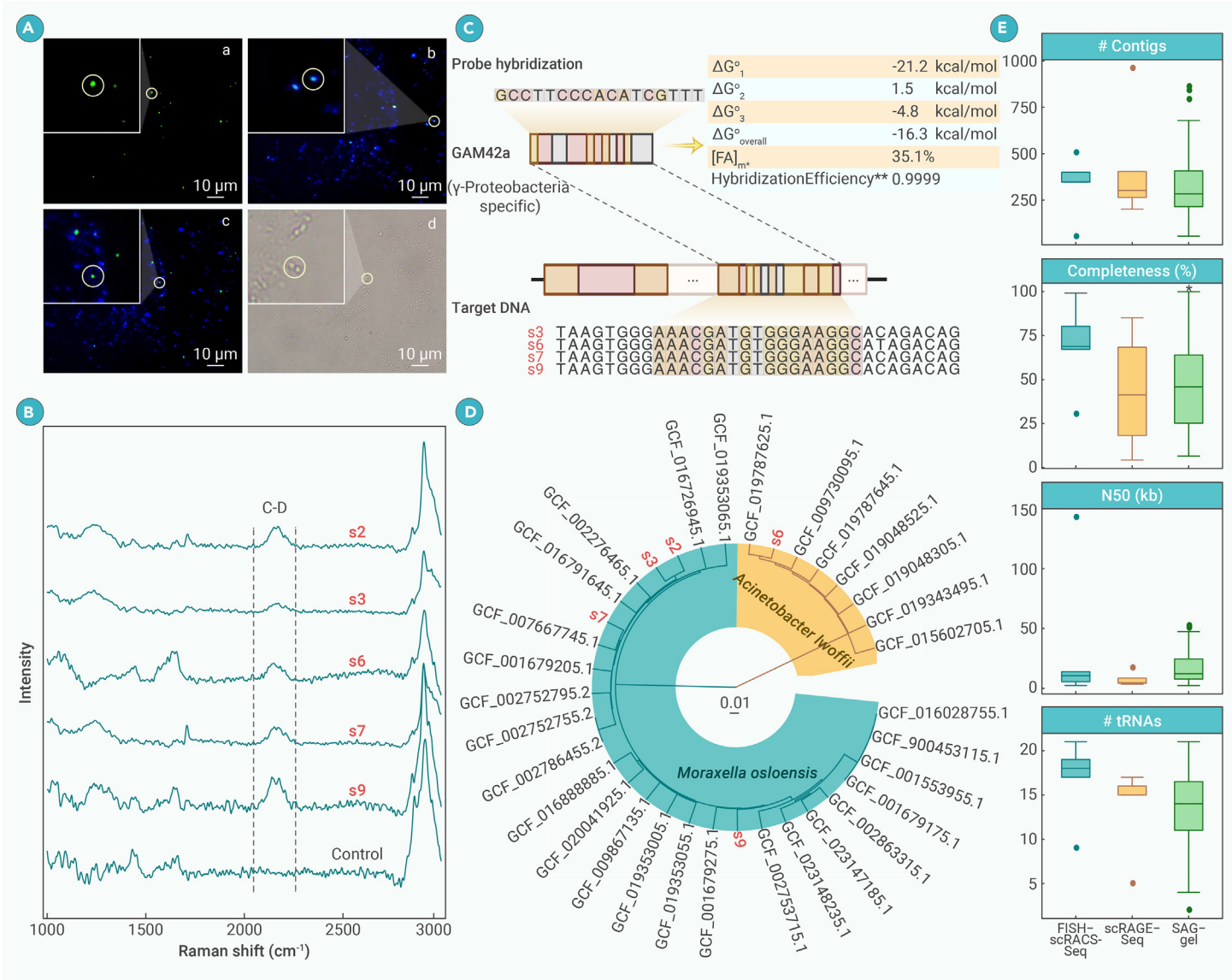


Figure 3. Application of FISH-scRACS-seq in identifying, sorting, and sequencing of metabolic active γ-Proteobacteria cells at one-cell resolution in soil microbiota (A) Photomicrographs of γ-Proteobacteria targeted by CARD-FISH probe from soil bacteria. Panels a to d are photomicrographs of soil bacteria hybridized with γ-Proteobacteria targeting probe GAM42a, DAPI staining of DNA, overlay images of probe signal (green) and DAPI staining (blue), and phase-contrast image, respectively. Each series shows identical microscopic fields. Scale bars, 10 μm. (B) SCRS of the target cells in soil sample, sorted via both "taxon-specific" and "metabolic" phenotypes of microbes for single-cell genomes. (C) The FISH probe perfectly matches those contained genomic regions recovered from SAGs. (D) Phylogenetic tree constructed using unweighted pair group method with arithmetic mean (UPGMA) based on ANI matrix of SAGs and collected genomes from NCBI RefSeq database. (E) Continuity and completeness comparison of soil-derived γ-Proteobacteria SAGs for FISH-scRACS-seq, scRAGE-seq, and SAG-gel. No significant differences were observed for FISH-scRACS-seq and scRAGE-seq. FISH-scRACS-seq and SAG-gel had comparable performance, except that the completeness of FISH-seq-derived SAGs was significantly higher than that of SAG-gel-derived SAGs (*$p < 0.05$, Wilcoxon test).

of the SAGs ranged between 19.12% and 99.76%, with average of 80.18%, 91.66%, and 83.30% for each run, respectively; (2) 26.26%–54.86% of the shotgun reads were mapped to *Ec*, and the average mapping rate was 35.21%, 47.53%, and 45.82%, respectively (the abundant unmapped reads are likely due to contaminated environmental DNA or non-specific amplification in the MDA reaction); (3) for each run, the success rate for a given species ranged from 40% to 70% (average of 56.67%). Reasons behind this particular success rate include: (1) complete lysis of cell walls and thus efficient release of genomic DNA for each and every cell can be difficult, due to the enormous diversity of microbial cell wall composition and structures; and (2) potential bias in the multiple displacement amplification reactions may result in incomplete genome representation. As a result, although the optimization of reaction conditions or key enzymes for amplification (e.g., the phi29 DNA polymerase[49]) can improve the success rate,[31,32,36,38–40] 100% success in single-cell genome amplification for natural microbiota is far from guaranteed. Notably, empty droplets derived from the aqueous phase around the target cells, which served as the negative controls, produce only negative results in 16S rRNA gene validation (Figures S1B–S1D). This suggests stringency against introducing contamination in the aqueous sorting workflow of RAGE-seq.[50,51] Collectively, results from mock microbiota also support high specificity and high sensitivity of FISH-scRACS-seq.

Validation of FISH-scRACS-seq on natural soil microbiota

To assess FISH-scRACS-seq's performance in an actual environmental sample, we employed soil, which harbors arguably the most metabolically and genetically heterogeneous microbiota on Earth.[52] Samples of shallow soil were collected from grassland at a depth <3 cm in the campus of Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, China (36°9'19"N, 120°28'50"E). Next, γ-Proteobacteria, which are metabolically active but of low abundance in soil and can serve as responder/indicator of soil pollution,[53–56] were labeled via CARD-FISH as above (Figure 3A). Finally, ten individual cells that carry the fluorescent CARD-FISH signal (i.e., of the target taxon) and the C-D peak (i.e., with the target metabolic vitality) were isolated from soil samples and sequenced via FISH-scRACS-seq (Figure 3B; materials and methods).

After quality control, clean reads from the five FISH-scRACS-seq reactions were *de novo* assembled into five SAGs (s2, s3, s6, s7, and s9; Table S3). Four of the five SAGs recovered 23S rRNA gene fragments that harbor the exact

Table 2. Performance of FISH-scRACS-seq in profiling metabolic phenome and genome of γ-Proteobacteria in soil and seawater samples

FISH-scRACS-sorted samples		Taxonomic classification	Genome completeness (%)	Contamination (%)
C-D peak-containing γ-Proteobacteria cells in soil	s2	*Moraxella*	41.79	7.85
	s3	*Moraxella*	75.60	6.19
	s6	*Acinetobacter*	85.13	1.57
	s7	*Moraxella*	99.14	0.10
	s9	*Moraxella*	72.97	2.38
C-D peak-containing and cycloalkane-degrading γ-Proteobacteria cells in seawater	m1	*Pseudoalteromonas*	94.35	1.02
	m4	*Pseudoalteromonas*	40.49	2.98
	m7	*Pseudoalteromonas*	70.88	4.95

hybrid site of GAM42A probe (with estimated 99.99% hybridization efficiency; Figure 3C). Taxonomic annotations of the five SAGs also pinpoint them as from γ-Proteobacteria (*Moraxella* spp. for s2, s3, s7, s9, and *Acinetobacter* spp. for s6; Figure 3D and Table 2). GC% of the assembled contigs (>200 bp; after decontamination; materials and methods) exhibit normal distribution (Figure S4A). Completeness of reconstructed one-cell genomes ranges from 41.79% to 99.14% (average of ~74.93%; Table 2), as estimated via lineage-specific marker genes by CheckM.[57] These results support the feasibility of FISH-scRACS-seq on soil microbiota.

The FISH-scRACS-seq derived SAGs were further compared to scRAGE-seq-derived[31] and SAG-gel-derived[58] γ-Proteobacteria SAGs (contigs over 1,000 bp were selected for comparison; materials and methods and Figure 3E). Although the datasets were from different sources, with different sizes, microbial species, and numbers of SAGs obtained, the contiguity (quantity and N50 of contigs) and completeness (genome completeness and number of unique tRNAs recovered) of SAGs among the three groups showed no significant difference (Wilcoxon test, $p > 0.05$; Figure 3E). This observation suggests that incorporating FISH into scRACS-seq may not necessarily downgrade the quality of SAGs when analyzing complex microbiomes. Besides, in the s9 cell, the taxon-specific SAG obtained via FISH-scRACS-seq unraveled a plasmid that harbors a gene encoding class A β-lactamases (Figure S4C), enzymes that can inactivate β-lactam antibiotics including carbapenems. Such genes represent a major challenge in treating bacterial infections, as they are highly diverse, rapidly evolving to acquire new resistance mechanisms, and easily transferred between bacteria through the spreading of plasmids.[59,60] Therefore, FISH-scRACS-seq is able to profile, directly from complex natural microbiota and in a phylogenetically directed manner, both metabolic activities and high-coverage (e.g., >99% for s7; Table 2) genomes at precisely one-cell resolution.

Unraveling *in situ* cycloalkane-degrading γ-Proteobacteria and their genomes at single-cell resolution in contaminated seawater by FISH-scRACS-seq

Cyclic alkanes, abundant in both subsurface hydrocarbon reservoirs and gas condensates,[61] are highly toxic to aquatic organisms and recalcitrant to degradation, thus posing significant ecological risks when large-scale oil spills occur.[62,63] Although little is known about cycloalkane biodegradation mechanisms in marine ecosystems,[64] an MWAS has associated a group of uncultured, psychrophilic, and oligotrophic γ-Proteobacteria with cycloalkane degradation in China's marginal seas.[61]

To further probe the mechanistic link between those γ-Proteobacteria and cycloalkane degradation, seven metagenomes were profiled by whole-genome sequencing (WGS) from each of seven seawater samples of different dilution ratios. The seawater was sampled from a large condensate gas field in the Bohai Sea where cycloalkanes are released from gas mining (Figure 4A). Metagenomic assembly and binning reconstructed seven MAGs of γ-Proteobacteria (phylogenetically annotated as *P. fuliginea*; one MAG in each sample; materials and methods), which are of 4.69–4.82 Mb in size, 99.66%–100.00% in estimated completeness, and contain 4,111–4,266 genes. However, none of the seven MAGs encode cycloalkane degradation pathways. Specifically, in the MAGs belonging to the newly defined proteobacterial genus of C1-B045, most of the key enzymes involved in methylcyclohexane degradation (e.g., alkane monooxygenase, cyclohexanone monooxygenase, and 6-hexanolactone hydrolase) were identified and annotated as such.[65] However, in the seven MAGs of *Pseudoalteromonas*, we failed to identify any of the aforementioned enzymes computationally and found merely one P450-encoding gene. Notably, the identification of this putative P450-encoding gene does not ensure its hosting genomes/cells actually perform the cyclohexane-degrading activity *in situ*. In fact, like many MWASs, due to the difficulty in measuring target taxa's metabolic activities *in situ* and unambiguously assigning them to individual genomes, the specific organisms, pathways, or enzymes responsible for the ecological trait (e.g., cycloalkane degradation) have remained speculative.

To solve this puzzle, we developed an MWAS-coupled FISH-scRACS-seq workflow by employing γ-Proteobacteria-targeted CARD-FISH probes to rapidly identify individual cells shown by MWAS to be associated with cycloalkane degrading, then specifically profiling their cycloalkane-degrading activity *in situ* based on D_2O-intake rate of a cell (which indicates such activity when cycloalkane is the only carbon source available) via the C-D band in SCRS. The phylogeny-metabolism dual-targeted cells were then sorted and sequenced at one-cell resolution to mine the pathways and genes responsible for the function (Figure 4A).

Specifically, the seawater samples were incubated with cyclohexane (as sole source of carbon and energy source) plus 50% D_2O (for tracking metabolic vitality of microbes) at 10°C (temperature of the sampled ocean site; materials and methods). A change in phylogenetic profile of microbiota before and after the cyclohexane treatment is prominent as suggested by 16S rRNA amplicon sequencing, supporting the association of γ-Proteobacteria with cyclohexane utilization (Figure S5 and Data S1). When the dissolved oxygen (DO) level was reduced to 0 μM by microbial hydrocarbon respiration, the marine microbiota was sampled to undergo the FISH procedure that employs the GAM42a probe pair, which is specific to γ-Proteobacteria (Figure 4A; materials and methods). A proportion of cells was successfully labeled, suggesting the presence of a considerable population of γ-Proteobacteria, which serves as the basis for subsequent phylogenetically directed screening of metabolic function via SCRS (Figure 4B). Notably, based on image analysis, the proportion of γ-Proteobacteria cells recognized by FISH is approximately 10.6%, which is far lower than that derived from 16S rDNA-based survey (Figure S6). This discrepancy can be due to (1) the inability of metagenome sequencing to distinguish between live cells and cell-free DNA and (2) the variation of copy number of 16S rDNA loci in bacterial genomes.

To identify those γ-Proteobacteria cells that are degrading cycloalkane *in situ*, cells with fluorescent signals were selected for SCRS acquisition. Among these, those showing C-D bands in SCRS, which indicate cycloalkane-degrading activity, were sorted. A total of ten cells were processed through lysis, genome amplification, and sequencing using the FISH-scRACS-seq method, each in a separate tube (Figure 4C, with one cell-free sample as the negative control in each batch). Out of these, three one-cell MDA products, each exhibiting clear MDA bands and positive 16S rDNA PCR results, were selected for further 16S rDNA and WGS. To maximize the sequence coverage for the one-cell genomes, ~3 Gb of raw sequencing data were generated for each cell (m1, m4, and m7; Table S3). The risk of contamination with non-bacterial sequences was assessed, in which few human or viral contigs were included. For each cell, GC% of the assembled contigs (>200 bp, after decontamination; materials and methods) exhibits a normal distribution (Figure S4B), consistent with a clean assembly. Based on lineage-specific marker genes,[57] 94.35%, 40.49%, and 70.88% genome fractions were recovered for m1, m4, and m7, respectively (Table 2). Notably, the genomic regions targeted by

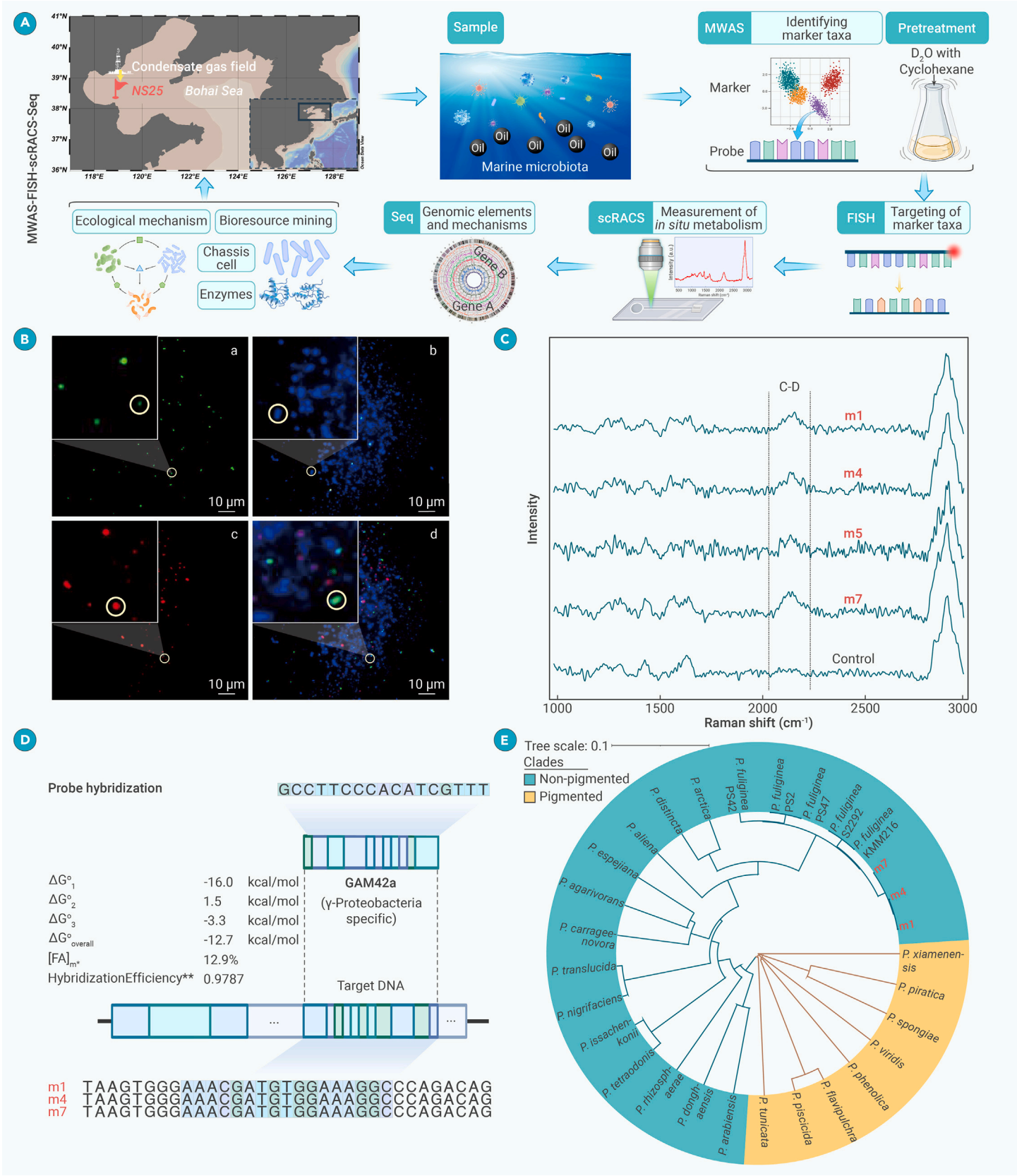


Figure 4. Application of FISH-scRACS-seq on identifying, sorting and sequencing cycloalkane-degrading γ-Proteobacteria in cycloalkane-contaminated seawater samples (A) The overall strategy of MWAS-FISH-scRACS-seq. (B) Photomicrographs of γ-Proteobacteria targeted by CARD-FISH probe from marine microbiota. Identical microscopic fields are displayed for each series. (a) γ-Proteobacteria were detected using GAM42a probe; (b) and (c) depict corresponding DAPI staining and autofluorescence of marine bacteria, respectively; (d) displays overlay images of probe signal (green), DAPI staining (blue), and autofluorescence (red); green corresponds to γ-Proteobacteria labeled with Alexa 488 using CARD-FISH; blue represents DAPI staining; and red indicates autofluorescence of marine bacteria. Cells appear magenta because of an overlay of DAPI staining and autofluorescence. Scale bars, 10 μm. (C) SCRS of the target cells in seawater sample, which were sorted via both taxon-specific and metabolism-specific features for single-cell genomes. (D) FISH-probe-containing genomic regions recovered from SAGs. (E) Phylogenetic tree constructed using UPGMA method based on ANI matrix of SAGs and *Pseudoalteromonas* spp. genomes collected from NCBI RefSeq database.

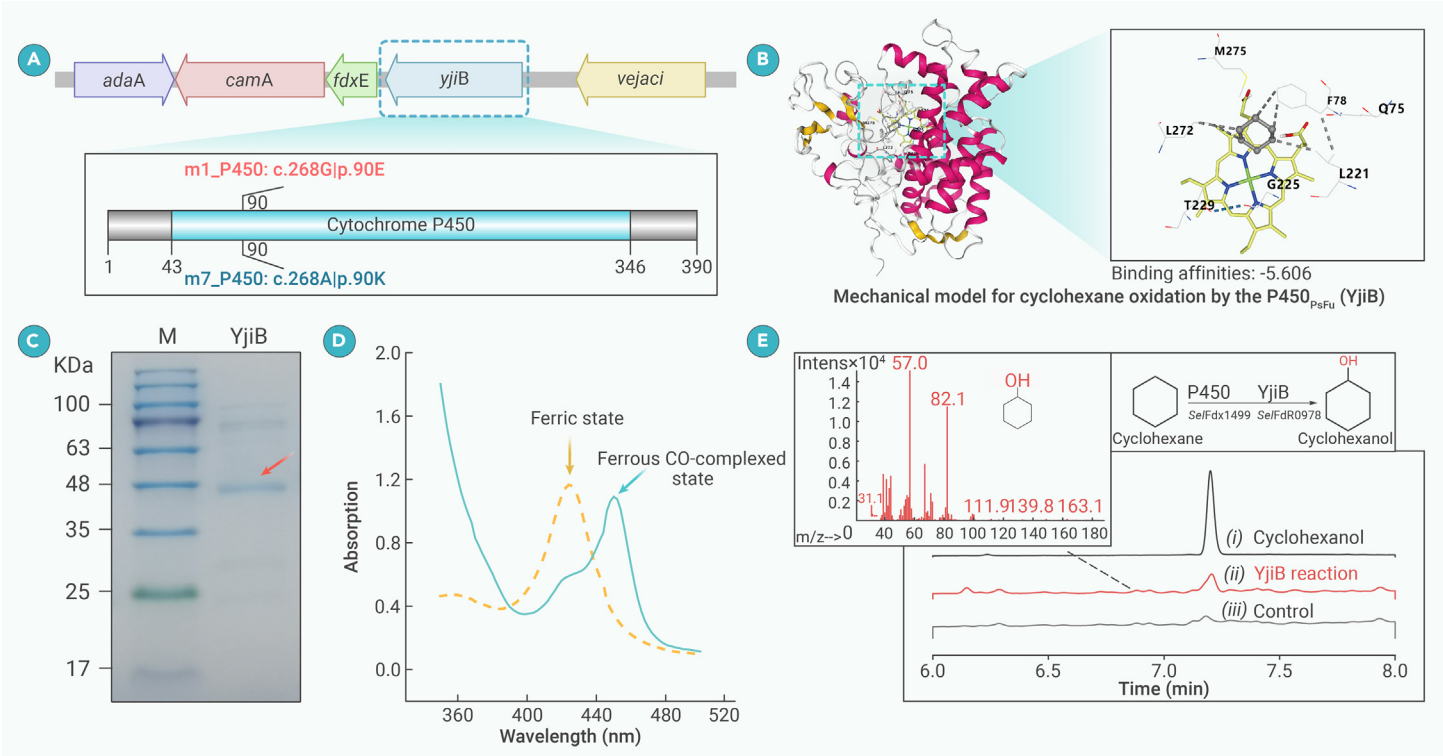


Figure 5. FISH-scRACS-seq-based discovery and experimental validation of cytochrome P450$_{PsFu}$ from *Pseudoalteromonas fuliginea* that catalyses cyclohexane degradation in cyclohexane-contaminated seawater (A) A three-component cytochrome P450 system was recovered in FISH-scRACS-seq-derived SAGs of m1 and m7. The *yji*B gene encodes a cytochrome P450 protein. Genes of *cam*A, *fdx*E, and *yji*B constitute the three-component P450 system. There is one amino acid mutation at position 90 for m7_P450. (B) Molecular docking analysis supports P450$_{PsFu}$ as a cyclohexane-degrading enzyme. The protein-ligand docking simulation reveals that cyclohexane (colored gray) can conjunct with P450$_{PsFu}$. The potential active sites of cyclohexane to P450$_{PsFu}$ are marked. (C) SDS-PAGE analysis of purified cytochrome P450$_{PsFu}$. (D) CO-bound reduced difference spectra of cytochrome P450$_{PsFu}$. Solid line, absorbance spectra of cytochrome P450$_{PsFu}$ in ferrous CO-complexed state; dotted line, absorbance spectra of P450 in ferric state. This assay was also used to determine the concentration of functional cytochrome P450 enzyme using the extinction coefficient of 91,000 M^{-1} cm^{-1}. (E) GC and GC-MS of product from the P450$_{PsFu}$-mediated reaction. (*i*) Authentic standard of cyclohexanol; (*ii*) reaction of cyclohexane with YjiB in the presence of *Sel*FdR0978 and *Sel*Fdx1499; (*iii*) Negative control of (*ii*) with boiled YjiB.

the CARD-FISH probes were also recovered from each SAG, showing hybridization efficiency of up to 98% (estimated by mathFISH)[66] (Figure 4D).

Based on the Genome Taxonomy Database (GTDB), the SAGs of m1, m4, and m7 were all classified as *P. fuliginea*, members of the non-pigmented clade (Figure 4E). Sequence comparison with publicly available *Pseudoalteromonas* genomes via average nucleotide identity (ANI) (Figure 4E; materials and methods) revealed >97% similarities of the one-cell *P. fuliginea* genomes with reference genomes and >99.5% similarities among m1, m4, and m7. To evaluate the capability of FISH-scRACS-seq in accurately reconstructing genomes of the individual cyclohexane-degrading cells, we compared these three SAGs to the seven *P. fuliginea* MAGs aforementioned. The nine SAG-unique genes (four in m1, three in m4, and two in m7) are supported by metagenomic sequencing reads but not recovered in the MAGs (e.g., m1 vs. MAG1 in Figure S7A), highlighting the importance of higher individual-genome coverage in SAG than in MAGs. Moreover, comparison of core-genome SNPs between SAG m1 and MAGs (based on high-quality SNPs; materials and methods) reveals many cell-specific mutations in the SAG that are absent from the MAGs, with some carrying biological consequences such as introduction of premature stop codons that lead to unstable or nonfunctional proteins (Figures S7B and S7C). Furthermore, the SAGs recover many more insertion sequences than MAGs (in percent of genomic length, 0.78% vs. 0.36%; Wilcoxon test, $p < 0.05$; Figure S7D). Therefore, the FISH-scRACS-seq-derived SAGs more completely and accurately reconstruct the genomes of individual cyclohexane-degrading cells.

Based on these SAGs of *P. fuliginea* that degrade cyclohexane *in situ*, annotation of carbohydrate-active enzymes (CAZymes; via dbCAN2a)[67] produced a global view of the glycobiome that potentially underpins the cellular function. In m1, 91 CAZyme genes were identified, which suggests the ability to utilize various carbohydrates including pectin, galactoside, glucan, peptidoglycan, chitin, trehalose, porphyran, agarose, and alginate, among others (Data S2). To probe the selection force that shapes these cyclohexane-degrading cells, dN/dS of 2,748 single-copy ortholog genes between m1 and m7 were calculated.[68] The genes under the strongest positive selection, i.e., with "dN != 0 and dS = 0" and considered "cataclysmic" (225, accounting for 8.2%), are enriched for CAZymes (11 out of 91, Table S4; 12% vs. ~5% in permutations; $p < 0.05$). Strikingly, an alginate lyase harbors four amino acid mutations between m1 (m1_02445) and m7 (m7_06510), suggesting a strong ecological pressure for these organisms to adapt to different types of carbon sources, including but perhaps not limited to alkanes.

A cytochrome P450$_{PsFu}$ from an *in situ* cycloalkane-degrading *P. fuliginea* cell catalyzes cyclohexane degradation

Currently, it is unknown whether *Pseudoalteromonas* spp. can degrade cyclohexane. However, each of the three *P. fuliginea* cells exhibits robust metabolic vitality *in situ* with cycloalkane as the sole carbon source (Figure 4C). To investigate this, we focused on the 4.42-Mb, 3,853-gene m1 genome, which is of the highest completeness (94.35%; Tables 2 and S3). Given the Raman-derived cyclohexane-degrading phenotype of the m1 cell, we speculate the presence of cyclohexanone monooxygenases in m1. In fact, in the one-cell genome of m1, a three-component cytochrome P450 system (class I/class B) was discovered (also in m7; Figure 5A), which consists of *yji*B ("m1_05218"; encoding a cytochrome P450 protein of P450$_{PsFu}$; Figure 5A), *cam*A ("m1_05216"; putidaredoxin reductase), and *fdx*E ("m1_05217"; ferredoxin).

In protein sequence, P450$_{PsFu}$ shows very low similarity (25.8%) to the CYP450cha from *Acidovorax* spp. (AKJ87746.1),[69] which is the only P450 known to transform cyclohexane to cyclohexanol so far. In fact, P450$_{PsFu}$ belongs to the CYP236A subfamily, which reportedly uses 6-*O*-methyl-D-galactose (G6Me; an abundant monosaccharide of algal agarose and porphyrin) as a substrate.[70] However, considering the promiscuity of substrate specificity in P450 enzymes,[71,72] we hypothesize that P450$_{PsFu}$ can oxidize cyclohexane. Molecular docking analysis (materials and methods; Figure 5B) support this, showing that cyclohexane binds to P450$_{PsFu}$ through hydrogen bonds and strong electrostatic interactions with higher affinity than to CYP450cha (i.e., the positive

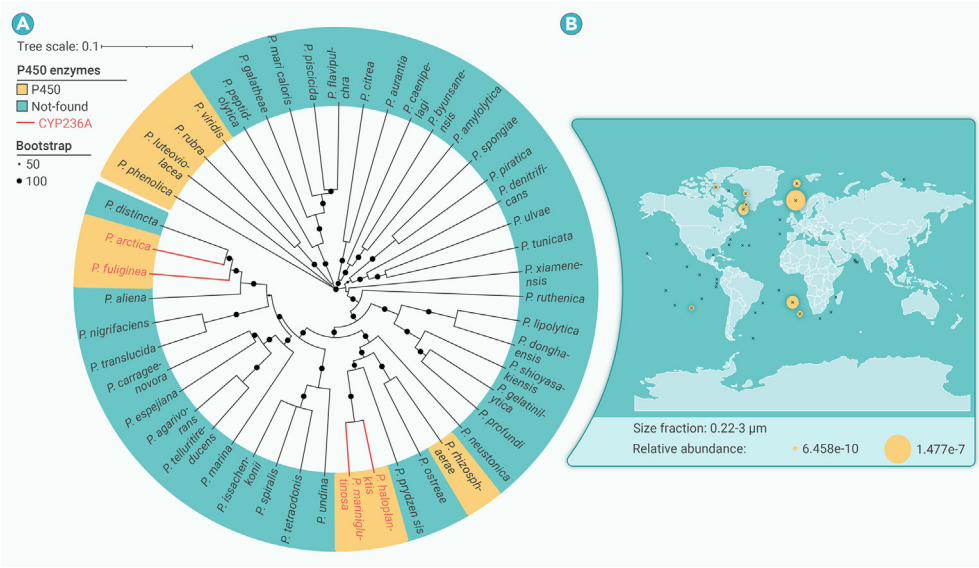


Figure 6. Evolutionary and ecological significance of the CYP236A subfamily that harbors the newly discovered cycloalkane monooxygenase of $P450_{PsFu}$ (A) Genes of CYP236A subfamily found in reference genomes of the *Pseudoalteromonas* genus. Among the reference genomes of 47 *Pseudoalteromonas* species, only four harbor CYP236A P450 subfamily genes. (B) Distribution of the CYP236A subfamily genes in global marine ecosystems. CYP236A enzymes are of low abundance and mostly found in the Arctic Ocean.

control; Figure S8). Intriguingly, the binding sites/residues of cyclohexane in $P450_{PsFu}$ are identical to those in $P450_{ZoGa}$, another member of the CYP236A subfamily (Figure S9). In fact, most of the binding sites/residues of cyclohexane/G6Me are conserved across CYP236A-family proteins (Figure S10). Thus, the CYP236A subfamily including $P450_{PsFu}$ may be a previously unknown group of P450 monooxygenases for cyclohexane.

To validate this hypothesis, we conducted an *in vitro* enzyme activity assay. The sequences of $P450_{PsFu}$, plus its native redox partners of *cam*A and *fdx*E, were optimized based on *E. coli* codon preferences and then expressed and purified to homogeneity (Figure 5C). The CO-bound reduced difference spectra of $P450_{PsFu}$ protein display a characteristic peak at 450 nm, confirming the expression of functional P450 enzymes (Figure 5D). As expression of CamA and FdxE in *E. coli* was unsuccessful, *Sel*FdR0978 and *Sel*Fdx1499 from the cyanobacterium *Synechococcus elongatus* PCC 7942 was employed as surrogate redox partner proteins instead to reconstitute the *in vitro* activity of $P450_{PsFu}$ (Figure 5E). The absorption spectra of *Sel*FdR0978 displayed characteristic peaks at around 456 nm and a shoulder at 396 nm, indicative of functional FAD enzymes. *Sel*Fdx1499 exhibited the typical UV-visible spectra with maximum peaks at approximately 420 nm, which are characteristic of UV-visible spectra for proteins containing the Fe_2S_2 cluster (Figure S11). Thus, these two surrogate redox partners are functionally active.

An *in vitro* enzymatic assay for $P450_{PsFu}$ activity was therefore established (materials and methods). Notably, as the low efficiency of NAD(P)H coupling is frequently a significant constraint on the activity of a reconstituted P450 system (due to the extra drain of NAD(P)H during reaction),[73] we employed an NADPH regeneration system based on glucose dehydrogenase/glucose in order to detect the P450 activity. Gas chromatography-mass spectrometry (GC-MS) analysis of the product profiles of reaction reveal a compound with retention time and ionized fragments identical to those of the pure cyclohexanol (Figure 5E). Thus, $P450_{PsFu}$ is able to convert cyclohexane to cyclohexanol, which is the first and the rate-limiting step of cyclohexane degradation, with the support of *Sel*FdR0978 and *Sel*Fdx1499.

Although members of the CYP236A P450 subfamily such as $P450_{FoAg}$ and $P450_{ZoGa}$ can oxidize the methyl group on G6Me,[70] it is much more difficult to oxidize the inert hydrocarbon bonds, due to their much higher reaction energy requirements. Thus, the metabolic activities of $P450_{PsFu}$ enzyme and of a *Pseudoalteromonas* spp. in oxidizing cyclohexane on the cycloalkyl group are new and surprising. These findings also reveal the diverse substrate specificity, and the ecological versatility (i.e., in macromolecule degradation), of the CYP236A P450 subfamily.

Ecological significance of the newly discovered cycloalkane monooxygenase of $P450_{PsFu}$

Enzymes from the CYP236A P450 subfamily are primarily found in Bacteroidetes and γ-Proteobacteria.[70] Among 47 sequenced *Pseudoalteromonas* spp. genomes, only four contain members of this subfamily, spread across multiple organismal branches (Figure 6A). These genes, located on chromosomes and not plasmids, are "accessory" genes, present in only a very small subset of genomes, such as a few *Pseudoalteromonas* spp., and likely originating via horizontal gene transfer or independent evolution.[70] Two of the four species (*P. fuliginea* PS2 and *P. mariniglutinosa* NCIMB 1770) were originally isolated from the surface of marine algae, suggesting a symbiotic relationship underpinned by P450-mediated bacterial utilization of algal polysaccharides. Interestingly, while $P450_{PsFu}$ was discovered in a *P. fuliginea* from the Bohai Sea in the northeastern Pacific Ocean, another *Pseudoalteromonas* spp., *P. arctica* A 37-1-2, was isolated from 4°C seawater of the Arctic Sea in Spitzbergen (Norway), a region rich in oil resources and known for oil spills.[74,75] Therefore, cyclohexane degradation by the CYP236A P450 subfamily enzymes is likely of ecological significance at a global scale.

A search of the KMAP metagenomic database, which includes MAG-derived genes from diverse environments (such as marine, soil, rhizosphere, lake, wastewater, and saltmarsh),[76,77] identified only eight CYP236A P450 genes, all from seawater, marine sediment, or freshwater (Table S5). Moreover, from the Ocean Microbial Reference Catalog v.2 (OM-RGC.v2)[78] and Ocean Gene Atlas v.2.0[79] (Figure 6B; materials and methods), 32 genes encoding such enzymes were found in 22 Tara Oceans projects (Table S6), with 20 of the 22 projects being from the Arctic Ocean and with temperatures of sampled sites ranging from −1.5°C to 8.5°C. The predominant distribution of these enzymes in low-temperature marine ecosystems is consistent with that of the oceanic cyclohexane sink.[80,81]

Collectively, these results suggest the CYP236A P450 subfamily enzymes, as represented by $P450_{PsFu}$, are numerically rare in their microcosms and spatially constrained to specific environments, yet they likely contribute to cyclohexane degradation in low-temperature oceans globally. These ecological features hold promise for bioaugmented removal of hydrocarbon pollutants during oil spills.

DISCUSSION

In many ecosystems, crucial functions can be mediated by numerically rare members of the microbiota that are not yet cultured,[82–84] yet validation of such roles and mining the underlying pathways or enzymes are usually difficult, due to the inability to profile their functions and the corresponding genomes *in situ*. For instance, in cyclohexane biodegradation in contaminated seawater from a condensate gas field at the Bohai Sea, MWAS identified the group of γ-Proteobacteria as marker organisms,[61] but these functional cells remain uncultured and their functional genes are too low in abundance for detection via WGS of the microbiota. Although metagenomic analyses reveal putative functional genes (e.g., P450 enzymes), it is unclear which are responsible for cyclohexane degradation. Experimental testing of each of these candidate genes for cyclohexane degradation activity *in vitro* or *ex vivo* may not be feasible or advisable, because (1) it would be overly tedious as there are simply too many of such candidate genes and (2) perhaps more importantly, such activities, even validated in particular genes, might not be ecologically relevant, as this observation does not guarantee that the cells encoding such genes are actually exhibiting the cyclohexane degradation activity *in situ*. By establishing FISH-scRACS-seq and coupling it to MWASs, we were able to, in a phylogeny and metabolism dual-directed manner, trace the *in situ* cyclohexane activity to individual *P. fuliginea* cells and then further to $P450_{PsFu}$, which represents a previously unknown group of cyclohexane monooxygenases and of cyclohexane-degrading genus.

The phylogeny directedness of MWAS-derived FISH probes ensures that the metabolic activities of cells (and pathways and enzymes encoded by their genomes) discovered by FISH-scRACS-seq are of ecological relevance. On the other hand, the metabolism-directedness of scRACS-seq ensures that the individual cells recovered are actually performing the targeted *in situ* function. This dual directionality is critical, as genome-based metabolic reconstruction or functional assays in pure culture often fail to reliably predict *in situ* function of a cell in a microbiota at its native state.[25,85] Therefore, FISH-scRACS-seq greatly elevates the efficiency in dissecting "who is doing what" in a microbiota, not only in enriched samples but also in low-abundance yet functionally potent members.

Cyclic alkanes, common in hydrocarbon reservoirs and gas condensates, are highly toxic to aquatic life and pose an ecological risk during oil spills. Despite this, the mechanisms underlying their biodegradation in marine ecosystems remain poorly understood.[61] In particular, cyclohexane, a common organic compound found in various industrial and environmental settings, is biotoxic, yet its biodegradation is difficult due to its chemical structure.[86] In cyclohexane degradation, its oxidation to the non-biotoxic cyclohexanol represents the first and the rate-limiting step. The discoveries of *P. fuliginea in situ* and of $P450_{PsFu}$ for this crucial step are unexpected, as no *Pseudoalteromonas* spp. or members of the CYP236A P450 subfamily were known to have this talent. Notably, as the *P. fuliginea* genome lacks enzymes for subsequent utilization of cyclohexanol, which, however, is of much higher bioavailability to microbial degraders than cyclohexane, it is possible that *P. fuliginea* is a keystone species at the start of a food chain and collaborates with other symbiotic bacteria for the complete mineralization of cyclohexane. Therefore, although numerically rare in their microcosms and spatially constrained to specific environments, they might have contributed to cyclohexane degradation in low-temperature oceans at a global scale.

Further development of FISH-scRACS-seq can take multiple directions. (1) Although CARD-FISH probes improve the detection of target cells,[87] multiplexing of probe hybridization can allow simultaneous interrogation of phenome-genome-gene links for multiple marker organisms.[88,89] Moreover, in addition to taxonomical markers, functional genes can also be targeted by an FISH probe via nucleotide sequence, extending the method to dissect a target gene's *in vivo* function in mutant libraries, microbiota, or even plant or animal tissues. (2) In microbiota RACS-seq, by creating an RAGE chip that preserves cell vitality[38] and designing a Hotja Phi29 enzyme that reduces bias in DNA amplification in one-cell MDA reactions,[49] we have demonstrated production of high-quality SCRS plus corresponding high-coverage genomes at precisely one-cell resolution directly from diverse ecosystems such as urine,[38] gastric biopsy,[36] soil,[31] seawater,[32] wastewater,[39] and probiotics products.[40] However, the sorting throughput (3–8 cells/min[38]) should be elevated, possibly by improving RAGE-chip design[90] and incorporating AI-based image analysis and automation.[91] Flow-mode RACS systems that sort at much higher throughput can also be adopted[92–94] to enable much deeper sampling of complex microbiota for a diverse set of cellular functions via SCRS.[18,25] (3) Cultivation of the target cells after FISH-scRACS operation is highly desirable, particularly since cells can remain viable after RACS, as demonstrated in scRACS-culture for phosphate-solubilizing bacteria in wastewater[39] and pool-based RACS-culture for mucin-degrading microbes from mouse colon microbiota.[85] Live-FISH techniques followed by RACS-seq should be explored to avoid the detrimental effects of chemical crosslinking or fixation on cell viability.[44]

In summary, mechanistic dissections of microbiota function have greatly lagged behind the explosive pace of MWASs generating marker organisms (and marker genes) for ecosystem traits. FISH-scRACS-seq can bridge this long-standing gap by efficiently unveiling enzymes, pathways, genomes, and *in situ* metabolic functions specifically targeting those cells revealed by MWAS as of ecological relevance, regardless of their cultivability. Therefore, we propose the complete MWAS-FISH-scRACS-seq (e.g., Figure 4A) as a rational and generally applicable strategy to systematically and thoroughly dissect and mine microbiota function from the plethora of ecosystems on Earth.

MATERIALS AND METHODS

Experimental design

The FISH-scRACS-seq workflow consists of three steps (Figure 1). In step 1 (i.e., FISH), individual cells of a target taxon are directly localized in a microbiota sample via a taxon-specific FISH probe. In step 2 (i.e., scRACS), post-FISH cells are distinguished and sorted based on not just the target phylogeny (via the FISH probe) but also the target metabolic phenome (via the SCRS). In step 3 (i.e., seq), the post-FISH-RACS cells in droplets undergo cell lysis, multiple displacement amplification (MDA), and genome sequencing in an indexed, one-cell-one-tube manner. In this way, specifically for those cells of target phylogeny in a microbiota, the target metabolic activity *in situ* is directly traced to genome sequence at single-cell resolution, thus accomplishing phylogeny-metabolism dual-directed single-cell genomics for dissecting and mining microbiota function.

For the Raman-activated gravity-driven encapsulation (RAGE) experiments, we employed the RACS-seq instrument system from Qingdao Single-Cell Biotech (SCB), China. Core to the instrument is the RAGE chip, which has been published by this team in collaboration with SCB.[38] The ability of the instrument and the chip to produce both high-quality Raman spectra and high-coverage genome sequences at precisely one-bacterial-cell resolution has been validated with a wide variety of microbiota samples, including gastrointestinal biopsy,[36] urine,[38] soil,[31] seawater,[32] wastewater,[39] and composite probiotics.[40] Therefore, the methodology can be readily implemented in a regular microbiology or molecular biology laboratory.

For other parts of the methodology, please refer to supplemental information.

DATA AND CODE AVAILABILITY

The sequence data reported in this study have been deposited in the NCBI SRA database under BioProject: PRJNA890413, PRJNA891066, PRJNA814381, and PRJNA855277. All data are available in the main text or the supplemental information.

REFERENCES

1. Konopka, A., Lindemann, S. and Fredrickson, J. (2015). Dynamics in microbial communities: unraveling mechanisms to identify principles. *ISME J.* **9**:1488–1495. DOI:https://doi.org/10.1038/ismej.2014.251.
2. Gupta, S., Ross, T.D., Gomez, M.M. et al. (2020). Investigating the dynamics of microbial consortia in spatially structured environments. *Nat. Commun.* **11**:2418. DOI:https://doi.org/10.1038/s41467-020-16200-0.
3. Kang, C.W., Lim, H.G., Won, J. et al. (2022). Circuit-guided population acclimation of a synthetic microbial consortium for improved biochemical production. *Nat. Commun.* **13**:6506. DOI:https://doi.org/10.1038/s41467-022-34190-z.
4. Li, C., Gillings, M.R., Zhang, C. et al. (2024). Ecology and risks of the global plastisphere as a newly expanding microbial habitat. *Innovation* **5**:100543. DOI:https://doi.org/10.1016/j.xinn.2023.100543.
5. Wang, J. and Jia, H. (2016). Metagenome-wide association studies: fine-mining the microbiome. *Nat. Rev. Microbiol.* **14**:508–522. DOI:https://doi.org/10.1038/nrmicro.2016.83.
6. Kishikawa, T., Maeda, Y., Nii, T. et al. (2020). Metagenome-wide association study of gut microbiome revealed novel aetiology of rheumatoid arthritis in the Japanese population. *Ann. Rheum. Dis.* **79**:103–111. DOI:https://doi.org/10.1136/annrheumdis-2019-215743.
7. Zhou, L.-W. (2023). Microbial taxonomy with DNA sequence data as nomenclatural type: How far should we go? *Innov. Life* **1**:100017. DOI:https://doi.org/10.59717/j.xinn-life.2023.100017.
8. Pérez-Enciso, M., Zingaretti, L.M., Ramayo-Caldas, Y. et al. (2021). Opportunities and limits of combining microbiome and genome data for complex trait prediction. *Genet. Sel. Evol.* **53**:65. DOI:https://doi.org/10.1186/s12711-021-00658-7.
9. Chang, H.-X., Haudenshield, J.S., Bowen, C.R. et al. (2017). Metagenome-wide association study and machine learning prediction of bulk soil microbiome and crop productivity. *Front. Microbiol.* **8**:519. DOI:https://doi.org/10.3389/fmicb.2017.00519.
10. Deng, Z., Zhang, J., Li, J. et al. (2021). Application of deep learning in plant–microbiota association analysis. *Front. Genet.* **12**:697090. DOI:https://doi.org/10.3389/fgene.2021.697090.
11. Wang, H., Wang, Y., Zhang, G. et al. (2023). Temporal dynamics and performance association of the Tetrasphaera-enriched microbiome for enhanced biological phosphorus removal. *Engineering* **29**:168–178. DOI:https://doi.org/10.1016/j.eng.2022.10.016.
12. Yang, P., Yang, J., Long, H. et al. (2023). MicroEXPERT: Microbiome profiling platform with cross-study metagenome-wide association analysis functionality. *iMeta* **2**:e131. DOI:https://doi.org/10.1002/imt2.131.
13. Locey, K.J. and Lennon, J.T. (2016). Scaling laws predict global microbial diversity. *Proc. Natl. Acad. Sci. USA* **113**:5970–5975. DOI:https://doi.org/10.1073/pnas.1521291113.
14. Berg, G., Rybakova, D., Fischer, D. et al. (2020). Microbiome definition re-visited: old concepts and new challenges. *Microbiome* **8**:103. DOI:https://doi.org/10.1186/s40168-020-00875-0.
15. Huang, W.E., Griffiths, R.I., Thompson, I.P. et al. (2004). Raman microscopic analysis of single microbial cells. *Anal. Chem.* **76**:4452–4458. DOI:https://doi.org/10.1021/ac049753k.
16. Huang, W.E., Ward, A.D. and Whiteley, A.S. (2009). Raman tweezers sorting of single microbial cells. *Environ. Microbiol. Rep.* **1**:44–49. DOI:https://doi.org/10.1111/j.1758-2229.2008.00002.x.
17. Lee, K.S., Wagner, M. and Stocker, R. (2020). Raman-based sorting of microbial cells to link functions to their genes. *Microb. Cell* **7**:62–65. DOI:https://doi.org/10.15698/mic2020.03.709.
18. He, Y., Wang, X., Ma, B. et al. (2019). Ramanome technology platform for label-free screening and sorting of microbial cell factories at single-cell resolution. *Biotechnol. Adv.* **37**:107388. DOI:https://doi.org/10.1016/j.biotechadv.2019.04.010.
19. Zhang, Q., Zhang, P., Gou, H. et al. (2015). Towards high-throughput microfluidic Raman-activated cell sorting. *Analyst* **140**:6163–6174. DOI:https://doi.org/10.1039/c5an01074h.

20. Li, M., Ashok, P.C., Dholakia, K. et al. (2012). Raman-activated cell counting for profiling carbon dioxide fixing microorganisms. *J. Phys. Chem. A* **116**:6560–6563. DOI:https://doi.org/10.1021/jp212619n.
21. Wang, Y., Huang, W.E., Cui, L. et al. (2016). Single cell stable isotope probing in microbiology using Raman microspectroscopy. *Curr. Opin. Biotechnol.* **41**:34–42. DOI:https://doi.org/10.1016/j.copbio.2016.04.018.
22. Tao, Y., Wang, Y., Huang, S. et al. (2017). Metabolic-activity-based assessment of antimicrobial effects by D2O-labeled single-cell Raman microspectroscopy. *Anal. Chem.* **89**:4108–4115. DOI:https://doi.org/10.1021/acs.analchem.6b05051.
23. Berry, D., Mader, E., Lee, T.K. et al. (2015). Tracking heavy water (D_2O) incorporation for identifying and sorting active microbial cells. *Proc. Natl. Acad. Sci. USA* **112**:194–203. DOI:https://doi.org/10.1073/pnas.1420406112.
24. Cui, L., Yang, K., Li, H.Z. et al. (2018). Functional single-cell approach to probing nitrogen-fixing bacteria in soil communities by resonance Raman spectroscopy with $^{15}N_2$ Labeling. *Anal. Chem.* **90**:5082–5089. DOI:https://doi.org/10.1021/acs.analchem.7b05080.
25. Hatzenpichler, R., Krukenberg, V., Spietz, R.L. et al. (2020). Next-generation physiology approaches to study microbiome function at single cell level. *Nat. Rev. Microbiol.* **18**:241–256. DOI:https://doi.org/10.1038/s41579-020-0323-1.
26. Azemtsop Matanfack, G., Pistiki, A., Rösch, P. et al. (2021). Raman ^{18}O-labeling of bacteria in visible and deep UV-ranges. *J. Biophot.* **14**:e202100013. DOI:https://doi.org/10.1002/jbio.202100013.
27. Ge, X., Pereira, F.C., Mitteregger, M. et al. (2022). SRS-FISH: A high-throughput platform linking microbiome metabolism to identity at the single-cell level. *Proc. Natl. Acad. Sci. USA* **119**:e2203519119. DOI:https://doi.org/10.1073/pnas.2203519119.
28. Yasuda, M., Takeshita, N. and Shigeto, S. (2021). Deuterium-labeled Raman tracking of glucose accumulation and protein metabolic dynamics in Aspergillus nidulans hyphal tips. *Sci. Rep.* **11**:1279. DOI:https://doi.org/10.1038/s41598-020-80270-9.
29. Jing, X., Gou, H., Gong, Y. et al. (2018). Raman-activated cell sorting and metagenomic sequencing revealing carbon-fixing bacteria in the ocean. *Environ. Microbiol.* **20**:2241–2255. DOI:https://doi.org/10.1111/1462-2920.14268.
30. Song, Y., Kaster, A.K., Vollmers, J. et al. (2017). Single-cell genomics based on Raman sorting reveals novel carotenoid-containing bacteria in the Red Sea. *Microb. Biotechnol.* **10**:125–137. DOI:https://doi.org/10.1111/1751-7915.12420.
31. Jing, X., Gong, Y., Xu, T. et al. (2021). One-cell metabolic phenotyping and sequencing of soil microbiome by Raman-Activated Gravity-Driven Encapsulation (RAGE). *mSystems* **6**:e0018121. DOI:https://doi.org/10.1128/msystems.00181-21.
32. Jing, X., Gong, Y., Xu, T. et al. (2022). Revealing CO_2-fixing SAR11 bacteria in the ocean by Raman-based single-cell metabolic profiling and genomics. *Biodes. Res.* **2022**:9782712. DOI:https://doi.org/10.34133/2022/9782712.
33. He, Y., Zhang, P., Huang, S. et al. (2017). Label-free, simultaneous quantification of starch, protein and triacylglycerol in single microalgal cells. *Biotechnol. Biofuels* **10**:275. DOI:https://doi.org/10.1186/s13068-017-0967-x.
34. Wang, T., Ji, Y., Wang, Y. et al. (2014). Quantitative dynamics of triacylglycerol accumulation in microalgae populations at single-cell resolution revealed by Raman microspectroscopy. *Biotechnol. Biofuels* **7**:58. DOI:https://doi.org/10.1186/1754-6834-7-58.
35. Wu, H., Volponi, J.V., Oliver, A.E. et al. (2011). In vivo lipidomics using single-cell Raman spectroscopy. *Proc. Natl. Acad. Sci. USA* **108**:3809–3814. DOI:https://doi.org/10.1073/pnas.1009043108.
36. Liu, M., Zhu, P., Zhang, L. et al. (2022). Single-cell identification, drug susceptibility test, and whole-genome sequencing of Helicobacter pylori directly from gastric biopsy by clinical antimicrobial susceptibility test Ramanometry. *Clin. Chem.* **68**:1064–1074. DOI:https://doi.org/10.1093/clinchem/hvac082.
37. Aljakouch, K., Lechtonen, T., Yosef, H.K. et al. (2018). Raman microspectroscopic evidence for the metabolism of a tyrosine kinase inhibitor, Neratinib, in Cancer Cells. *Angew. Chem. Int. Ed.* **57**:7250–7254. DOI:https://doi.org/10.1002/anie.201803394.
38. Xu, T., Gong, Y., Su, X. et al. (2020). Phenome-genome profiling of single bacterial cell by Raman-activated gravity-driven encapsulation and sequencing. *Small* **16**:e2001172. DOI:https://doi.org/10.1002/smll.202001172.
39. Jing, X., Gong, Y., Pan, H. et al. (2022). Single-cell Raman-activated sorting and cultivation (scRACS-Culture) for assessing and mining in situ phosphate-solubilizing microbes from nature. *ISME Commun.* **2**:106. DOI:https://doi.org/10.1038/s43705-022-00188-3.
40. Zhang, J., Ren, L., Zhang, L. et al. (2023). Single-cell rapid identification, in situ viability and vitality profiling, and genome-based source-tracking for probiotics products. *iMeta* **2**:e117. DOI:https://doi.org/10.1002/imt2.117.
41. Amann, R. and Fuchs, B.M. (2008). Single-cell identification in microbial communities by improved fluorescence in situ hybridization techniques. *Nat. Rev. Microbiol.* **6**:339–348. DOI:https://doi.org/10.1038/nrmicro1888.
42. Wagner, M., Horn, M. and Daims, H. (2003). Fluorescence in situ hybridisation for the identification and characterisation of prokaryotes. *Curr. Opin. Microbiol.* **6**:302–309. DOI:https://doi.org/10.1016/s1369-5274(03)00054-7.
43. Wagner, M. and Haider, S. (2012). New trends in fluorescence in situ hybridization for identification and functional analyses of microbes. *Curr. Opin. Biotechnol.* **23**:96–102. DOI:https://doi.org/10.1016/j.copbio.2011.10.010.
44. Batani, G., Bayer, K., Böge, J. et al. (2019). Fluorescence in situ hybridization (FISH) and cell sorting of living bacteria. *Sci. Rep.* **9**:18618. DOI:https://doi.org/10.1038/s41598-019-55049-2.
45. Aune, T.E.V. and Aachmann, F.L. (2010). Methodologies to increase the transformation efficiencies and the range of bacteria that can be transformed. *Appl. Microbiol. Biotechnol.* **85**:1301–1313. DOI:https://doi.org/10.1007/s00253-009-2349-1.
46. Rinke, C., Lee, J., Nath, N. et al. (2014). Obtaining genomes from uncultivated environmental microorganisms using FACS-based single-cell genomics. *Nat. Protoc.* **9**:1038–1048. DOI:https://doi.org/10.1038/nprot.2014.067.
47. Su, X., Gong, Y., Gou, H. et al. (2020). Rational Optimization of Raman-Activated Cell Ejection and Sequencing for Bacteria. *Anal. Chem.* **92**:8081–8089. DOI:https://doi.org/10.1021/acs.analchem.9b05345.
48. Wang, Y., Ji, Y., Wharfe, E.S. et al. (2013). Raman activated cell ejection for isolation of single cells. *Anal. Chem.* **85**:10697–10701. DOI:https://doi.org/10.1021/ac403107p.
49. Zhang, J., Su, X., Wang, Y. et al. (2023). Improved single-cell genome amplification by a high-efficiency phi29 DNA polymerase. *Front. Bioeng. Biotechnol.* **11**:1233856. DOI:https://doi.org/10.3389/fbioe.2023.1233856.
50. Han, H.-S., Cantalupo, P.G., Rotem, A. et al. (2015). Whole-Genome Sequencing of a Single Viral Species from a Highly Heterogeneous Sample. *Angew. Chem. Int. Ed.* **54**:13985–13988. DOI:https://doi.org/10.1002/anie.201507047.
51. Marín, R.M. and Vanícek, J. (2011). Efficient use of accessibility in microRNA target prediction. *Nucleic Acids Res.* **39**:19–29. DOI:https://doi.org/10.1093/nar/gkq768.
52. Young, I.M. and Crawford, J.W. (2004). Interactions and self-organization in the soil-microbe complex. *Science* **304**:1634–1637. DOI:https://doi.org/10.1126/science.1097394.
53. Zhang, Q., Zhang, Z., Lu, T. et al. (2021). Gammaproteobacteria, a core taxon in the guts of soil fauna, are potential responders to environmental concentrations of soil pollutants. *Microbiome* **9**:196. DOI:https://doi.org/10.1186/s40168-021-01150-6.
54. Kurm, V., van der Putten, W.H., de Boer, W. et al. (2017). Low abundant soil bacteria can be metabolically versatile and fast growing. *Ecology* **98**:555–564. DOI:https://doi.org/10.1002/ecy.1670.
55. Thompson, D.K. and Wickham, G.S. (2018). Gammaproteobacteria and firmicutes are resistant to long-term chromium exposure in soil. *Adv. Microbiol. Res.* **2**:002. DOI:https://doi.org/10.24966/AMR-694X/100002.
56. Köberl, M., Dita, M., Martinuz, A. et al. (2017). Members of Gammaproteobacteria as indicator species of healthy banana plants on Fusarium wilt-infested fields in Central America. *Sci. Rep.* **7**:45318. DOI:https://doi.org/10.1038/srep45318.
57. Parks, D.H., Imelfort, M., Skennerton, C.T. et al. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**:1043–1055. DOI:https://doi.org/10.1101/gr.186072.114.
58. Nishikawa, Y., Kogawa, M., Hosokawa, M. et al. (2022). Validation of the application of gel beads-based single-cell genome sequencing platform to soil and seawater. *ISME Commun.* **2**:92. DOI:https://doi.org/10.1038/s43705-022-00179-4.
59. Pemberton, O.A., Noor, R.E., Kumar, M.V.V. et al. (2020). Mechanism of proton transfer in class A β-lactamase catalysis and inhibition by avibactam. *Proc. Natl. Acad. Sci. USA* **117**:5818–5825. DOI:https://doi.org/10.1073/pnas.1922203117.
60. Palzkill, T. (2018). Structural and Mechanistic basis for extended-spectrum drug-resistance mutations in altering the specificity of TEM, CTX-M, and KPC β-lactamases. *Front. Mol. Biosci.* **5**:16. DOI:https://doi.org/10.3389/fmolb.2018.00016.
61. Cui, Z., Luan, X., Li, S. et al. (2022). Occurrence and distribution of cyclic-alkane-consuming psychrophilic bacteria in the Yellow Sea and East China Sea. *J. Hazard Mater.* **427**:128129. DOI:https://doi.org/10.1016/j.jhazmat.2021.128129.
62. Carrillo, J.C., Adenuga, M.D., Momin, F. et al. (2018). The sub-chronic toxicity of a naphthenic hydrocarbon solvent in rats. *Regul. Toxicol. Pharmacol.* **95**:323–332. DOI:https://doi.org/10.1016/j.yrtph.2018.04.002.
63. Qiao, F., Wang, G., Yin, L. et al. (2019). Modelling oil trajectories and potentially contaminated areas from the Sanchi oil spill. *Sci. Total Environ.* **685**:856–866. DOI:https://doi.org/10.1016/j.scitotenv.2019.06.255.
64. Tremblay, J., Yergeau, E., Fortin, N. et al. (2017). Chemical dispersants enhance the activity of oil- and gas condensate-degrading marine bacteria. *ISME J.* **11**:2793–2808. DOI:https://doi.org/10.1038/ismej.2017.129.
65. Cui, Z., Li, Y., Jing, X. et al. (2024). Cycloalkane degradation by an uncultivated novel genus of Gammaproteobacteria derived from China's marginal seas. *J. Hazard Mater.* **469**:133904. DOI:https://doi.org/10.1016/j.jhazmat.2024.133904.
66. Yilmaz, L.S., Parnerkar, S. and Noguera, D.R. (2011). MathFISH, a web tool that uses Thermodynamics-based mathematical models for in silico evaluation of oligonucleotide probes for fluorescence in situ hybridization. *Appl. Environ. Microbiol.* **77**:1118–1122. DOI:https://doi.org/10.1128/AEM.01733-10.
67. Zhang, H., Yohe, T., Huang, L. et al. (2018). dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **46**:W95–W101. DOI:https://doi.org/10.1093/nar/gky418.
68. Drost, H.-G., Gabel, A., Grosse, I. et al. (2015). Evidence for active maintenance of phylotranscriptomic hourglass patterns in animal and plant embryogenesis. *Mol. Biol. Evol.* **32**:1221–1231. DOI:https://doi.org/10.1093/molbev/msv012.
69. Salamanca, D., Karande, R., Schmid, A. et al. (2015). Novel cyclohexane monooxygenase from Acidovorax sp. CHX100. *Appl. Microbiol. Biotechnol.* **99**:6889–6897. DOI:https://doi.org/10.1007/s00253-015-6599-9.
70. Robb, C.S., Reisky, L., Bornscheuer, U.T. et al. (2018). Specificity and mechanism of carbohydrate demethylation by cytochrome P450 monooxygenases. *Biochem. J.* **475**:3875–3886. DOI:https://doi.org/10.1042/bcj20180762.
71. Wong, L.-L., Westlake, A.C.G. and Nickerson, D.P. (1997). Protein engineering of cytochrome $P450_{cam}$. In Metal Sites in Proteins and Models: Iron Centres, H.A.O. Hill, P.J. Sadler, and A.J. Thomson, eds., pp. 175–207. DOI:https://doi.org/10.1007/3-540-62870-3_6.
72. Bell, S.G., Orton, E., Boyd, H. et al. (2003). Engineering cytochrome P450cam into an alkane hydroxylase. *Dalton Trans.* **11**:2133. DOI:https://doi.org/10.1039/B300869J.

73. Liu, X., Li, F., Sun, T. et al. (2022). Three pairs of surrogate redox partners comparison for Class I cytochrome P450 enzyme activity reconstitution. *Commun. Biol.* **5**:791. DOI:https://doi.org/10.1038/s42003-022-03764-4.
74. Nagy, J. (2009). Oil exploration in Spitsbergen. *Polar Rec.* **12**:703–708. DOI:https://doi.org/10.1017/S0032247400059490.
75. Gulliksen, B. and Taasen, J.P. (1982). Effect of an oil spill in Spitzbergen in 1978. *Mar. Pollut. Bull.* **13**:96–98. DOI:https://doi.org/10.1016/0025-326X(82)90200-4.
76. Alam, I., Kamau, A.A., Ngugi, D.K. et al. (2021). KAUST Metagenomic Analysis Platform (KMAP), enabling access to massive analytics of re-annotated metagenomic data. *Sci. Rep.* **11**:11511. DOI:https://doi.org/10.1038/s41598-021-90799-y.
77. Priyam, A., Woodcroft, B.J., Rai, V. et al. (2019). Sequenceserver: A modern graphical user interface for custom BLAST databases. *Mol. Biol. Evol.* **36**:2922–2924. DOI:https://doi.org/10.1093/molbev/msz185.
78. Salazar, G., Paoli, L., Alberti, A. et al. (2019). Gene expression changes and community turnover differentially shape the global ocean metatranscriptome. *Cell* **179**:1068–1083.e21. DOI:https://doi.org/10.1016/j.cell.2019.10.014.
79. Vernette, C., Lecubin, J., Sánchez, P. et al. (2022). The Ocean Gene Atlas v2.0: online exploration of the biogeography and phylogeny of plankton genes. *Nucleic Acids Res.* **50**. W516-w526. https://doi.org/10.1093/nar/gkac420.
80. Beck, L.J., Sarnela, N., Junninen, H. et al. (2021). Differing Mechanisms of New Particle Formation at Two Arctic Sites. *Geophys. Res. Lett.* **48**:e2020GL091334. DOI:https://doi.org/10.1029/2020GL091334.
81. Tomy, G.T., Pleskach, K., Arsenault, G. et al. (2008). Identification of the novel cycloaliphatic brominated flame retardant 1,2-Dibromo-4-(1,2-dibromoethyl)cyclohexane in Canadian Arctic Beluga (Delphinapterus leucas). *Environ. Sci. Technol.* **42**:543–549. DOI:https://doi.org/10.1021/es072043m.
82. Shu, W.-S. and Huang, L.-N. (2022). Microbial diversity in extreme environments. *Nat. Rev. Microbiol.* **20**:219–235. DOI:https://doi.org/10.1038/s41579-021-00648-y.
83. Hua, Z.S., Han, Y.J., Chen, L.X. et al. (2015). Ecological roles of dominant and rare prokaryotes in acid mine drainage revealed by metagenomics and metatranscriptomics. *ISME J.* **9**:1280–1294. DOI:https://doi.org/10.1038/ismej.2014.212.
84. Zhang, Z., Gao, Q., Ren, X. et al. (2023). Characterization of intratumor microbiome in cancer immunotherapy. *Innovation* **4**:100482. DOI:https://doi.org/10.1016/j.xinn.2023.100482.
85. Lee, K.S., Pereira, F.C., Palatinszky, M. et al. (2021). Optofluidic Raman-activated cell sorting for targeted genome retrieval or cultivation of microbial cells with specific functions. *Nat. Protoc.* **16**:634–676. DOI:https://doi.org/10.1038/s41596-020-00427-8.
86. Lee, E.H. and Cho, K.S. (2008). Characterization of cyclohexane and hexane degradation by Rhodococcus sp. EC1. *Chemosphere* **71**:1738–1744. DOI:https://doi.org/10.1016/j.chemosphere.2007.12.009.
87. Schmidt, H., Eickhorst, T. and Tippkötter, R. (2012). Evaluation of tyramide solutions for an improved detection and enumeration of single microbial cells in soil by CARD-FISH. *J. Microbiol. Methods* **91**:399–405. DOI:https://doi.org/10.1016/j.mimet.2012.09.021.
88. Cao, Z., Zuo, W., Wang, L. et al. (2023). Spatial profiling of microbial communities by sequential FISH with error-robust encoding. *Nat. Commun.* **14**:1477. DOI:https://doi.org/10.1038/s41467-023-37188-3.
89. Shi, H., Shi, Q., Grodner, B. et al. (2020). Highly multiplexed spatial mapping of microbial communities. *Nature* **588**:676–681. DOI:https://doi.org/10.1038/s41586-020-2983-4.
90. Xu, T., Li, Y., Han, X. et al. (2022). Versatile, facile and low-cost single-cell isolation, culture and sequencing by optical tweezer-assisted pool-screening. *Lab Chip* **23**:125–135. DOI:https://doi.org/10.1039/d2lc00888b.
91. Diao, Z., Kan, L., Zhao, Y. et al. (2022). Artificial intelligence-assisted automatic and index-based microbial single-cell sorting system for One-Cell-One-Tube. *mLife* **1**:448–459. DOI:https://doi.org/10.1002/mlf2.12047.
92. Wang, X., Xin, Y., Ren, L. et al. (2020). Positive dielectrophoresis–based Raman-activated droplet sorting for culture-free and label-free screening of enzyme function in vivo. *Sci. Adv.* **6**:eabb3521. DOI:https://doi.org/10.1126/sciadv.abb3521.
93. Lee, K.S., Palatinszky, M., Pereira, F.C. et al. (2019). An automated Raman-based platform for the sorting of live cells by functional properties. *Nat. Microbiol.* **4**:1035–1048. DOI:https://doi.org/10.1038/s41564-019-0394-9.
94. Nitta, N., Iino, T., Isozaki, A. et al. (2020). Raman image-activated cell sorting. *Nat. Commun.* **11**:3452. DOI:https://doi.org/10.1038/s41467-020-17285-3.

ACKNOWLEDGMENTS

This work was supported by grants from the National Key R&D Program Young Scientists Project of China (2021YFD1900400) and the Natural Science Foundation of China (32030003, 32270109, 32370097, 42076165, and 32071266). We are grateful to Zongze Shao (from the Third Institute of Oceanography, Ministry of Natural Resources of China) for his insightful comments. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

J.X. and X.J. designed the research. X.J., Y.Ma., Z.D., J.C., Y.R., Y. Li, and W.S. performed the experiments. Y.G. and X.J. performed the sequencing and data analysis for single-cell genomes. Z.D., B.M., and J.X. introduced and validated the RAGE method and chip. J.X., X.J., Y.G., L.M., and Y.Meng. analyzed data and interpreted results. Y. Liang, J.Z., Y.J., B.M., Z. Cong, and S.L. provided critical suggestions. X.J., J.X., Y.G., L.M., and Z. Cui wrote the manuscript.

DECLARATION OF INTERESTS

J.X. and B.M. are founders of Qingdao Single-Cell Biotech. Co., Ltd.

SUPPLEMENTAL INFORMATION

It can be found online at https://doi.org/10.1016/j.xinn.2024.100759.