www.sciencemag.org/content/349/6246/420/suppl/DC1

Supplementary Materials for

Exploring deep microbial life in coal-bearing sediment down to ~2.5 km below the ocean floor

F. Inagaki,* K.-U. Hinrichs,* Y. Kubo, M. W. Bowles, V. B. Heuer, W.-L. Hong, T. Hoshino, A. Ijiri, H. Imachi, M. Ito, M. Kaneko, M. A. Lever, Y.-S. Lin, B. A. Methé, S. Morita, Y. Morono, W. Tanikawa, M. Bihan, S. A. Bowden, M. Elvert, C. Glombitza, D. Gross, G. J. Harrington, T. Hori, K. Li, D. Limmer, C.-H. Liu, M. Murayama, N. Ohkouchi, S. Ono, Y.-S. Park, S. C. Phillips, X. Prieto-Mollar, M. Purkey, N. Riedinger, Y. Sanada, J. Sauvage, G. Snyder, R. Susilawati, Y.Takano, E. Tasumi, T. Terada, H. Tomaru, E. Trembath-Reichert, D. T. Wang, Y. Yamada

*Corresponding author. Email: inagaki@jamstec.go.jp (F.I.); khinrichs@uni-bremen.de (K.-U.H.)

Published 24 July 2015, *Science* **349**, 420 (2015)
DOI: 10.1126/science.aaa6882

This PDF file includes:

Other Supporting Online Material for this manuscript includes the following:
(available at www.sciencemag.org/content/349/6246/420/suppl/DC1)

Materials and Methods

Study site and coring operations

The study site (41°10.5983′N, 142°12.0328′E, fig. S1) is located in the forearc basin offshore the Shimokita Peninsula, Japan, and was explored by two expeditions with the drilling vessel *Chikyu.* In 2006, the Japan Agency for Marine-Earth Science and Technology (JAMSTEC) conducted the cruise CK06-06 as one of the shakedown expeditions of the new drilling vessel *Chikyu* to test technical facilities and scientific procedures for scientific ocean drilling. To this end, several holes were drilled at the study site, which was named JAMSTEC Site C9001 at that time.

During the cruise CK06-06, sediment cores were sampled throughout the upper 365 m below the seafloor (mbsf) in Hole C of JAMSTEC Site C9001 (i.e., "shallow" subseafloor sediment samples), using a hydraulic piston corer system (HPCS) which passes a piston-coring shoe through the throat of the drill bit into the undisturbed sediment in advance to the drill bit at the bottom of the borehole. This coring technique yields well-preserved soft to semi-consolidated sediments with no or little risk of contamination, even though seawater is pumped through the drill string in order to apply torque to the HPCS and to move cuttings up to the seafloor. After HPCS coring, Hole D was spudded at the same location, riser drilling was first tested without coring from 527 mbsf down to a total borehole depth of 647 mbsf, and then steel casing was installed to prepare future riser-drilling operations (*6*, *32*). Up to that point, scientific ocean drilling had relied on non-riser drilling techniques.

In 2012, Expedition 337 of the Integrated Ocean Drilling Program (IODP) re-entered JAMSTEC Site C9001 Hole D, renamed it according to IODP nomenclature into IODP Site C0020 Hole A, resumed drilling and advanced the total borehole depth down to 2466 mbsf with spot-coring of sediments by rotary core barrel (RCB) (i.e., "deep" subseafloor sediment samples). Compared to HPCS used during conventional non-riser drilling, this coring technique during riser drilling goes along with higher contamination risks, as cores are collected by a passive core receptacle above the drill bit while the latter is penetrating the geological formation with the help of high density and viscosity drill mud rather than seawater. The riser-drilling system allows all drilling fluids and mud to rise up from the borehole bottom to the rig floor on the ship and thus enables recycling of artificially prepared drill mud in a closed circuit. The latter is essential for ultra-deep drilling and contains aside from seawater additional salts (KCl, NaCl), colloidal lubricants and viscosifiers (e.g., bentonite, cellulose and starch derivatives, sulfonated asphalt sodium salt), NaOH for pH-adjustment to ~10, and antiseptic agents (*33*).

The geological formation drilled during Expedition 337 contains a series of seventeen 0.3 to 7.3 m-thick coal layers that were deposited in terrigenous to shallow marine environments, such as coastal wetlands or deltaic lagoons, during the late Paleogene and early Neogene. Later, forearc subsidence transformed the depositional setting into a deep-sea environment (*6*). Coals of low maturity (vitrinite reflectance: 0.2 to 0.4%) were recovered from a series of coaly shales, siltstones, and porous sand layers between 1826 and 2046 mbsf (i.e., Lithological Unit III) (*6*). Porosity decreases with increasing depth, and exceptionally low values (<10%) were observed in cemented

carbonate layers (*6*). More detailed geological backgrounds and riser-drilling operations have been described in detail elsewhere (*6, 8, 14, 16, 32-35*).

Quality assessment and control for microbiological samples

During both cruises CK06-06 and Expedition 337, all core sections were scanned with X-ray computed tomography (X-CT) and sampled onboard the *Chikyu* immediately after core recovery (*6, 32*). Based on the X-CT image, locations of non-disturbed core sections for microbiological and biogeochemical samples were identified and dissected as whole-round cores (WRC). In general, the latter were sub-sampled immediately in the shipboard microbiology and geochemistry laboratories, using the innermost part of the WRCs with sterilized sampling devices for aseptic procedures under a clean bench or an anaerobic chamber, where necessary (*6*). Some intact WRCs were stored for shore-based processing: e.g., for molecular analyses, all sediment samples were immediately stored in -80ºC deep freezers. The sub-sampling of frozen cores was performed using a diamond band saw system equipped in a clean booth at the Kochi Institute for Core Sample Research, JAMSTEC (*36*).

We obtained a large amount of DNA from high-biomass ($>10^7$ cells cm^{-3}) "shallow" subseafloor sediment core samples using the hot-alkaline DNA extraction method (*37*). In that experiment, no sequenceable PCR amplicons were obtained from negative controls. In addition, we tested fluorescent microsphere beads to check the potential contamination during non-riser HPCS coring, resulting in no detection from the "shallow subseafloor" core samples. Those observations indicate that experiments were successfully completed under the clean condition with strict controls; i.e., contamination threats for "shallow" samples were not severe for microbiological analyses and have been countered with standard procedures that have successfully demonstrated negligible contamination in past studies (*38, 39*).

During the rise-drilling Expedition 337, the contamination risk was high and hence contamination of core samples from drill fluids was routinely monitored by a perfluorocarbon (PFC) tracer method (*6, 38, 39*) that involved daily addition of PFC to the active drill mud tanks (cf. Table 37 of ref. *6*) and monitoring of PFC concentrations in the outer, middle and center parts of cores. In principle, this method allows calculation of drill mud intrusion into the core, but exact determination of contamination levels suffered from considerable fluctuations of PFC concentrations in drill mud during operations, probably due to variations in drill mud production (cf. Table 38 of ref. *6*). Contamination levels were typically high in the outer part of cores but notably low in the core center (cf. Table 40 of ref. *6*).

In order to evaluate the potential effect of drill mud intrusion on geochemical and microbiological analyses, drill mud samples were taken and investigated with the same methods as sediment core samples (cf. Table 38 in ref. *6*). Daily shipboard cell counts revealed an average cell concentration of 2.7×10^8 cells mL^{-1} of drill mud with little variability throughout drilling operations (1.1×10^8 to 8.4×10^8 cells mL^{-1} of drill mud) (cf. Table 38 of ref. *6*). At 13 time points spanning the entire riser-drilling operation, drill mud samples were taken for shipboard DNA extraction and preliminary monitoring of microbial communities: Shipboard PCR assay demonstrated the presence of *Xanthomonas* and *Halomonas* in drill mud (*6*), which were posed by contaminated microbes linked to mud vicosifiers in the drill mud tanks (*33*). Shipboard analysis of ten

drill mud samples by DNA fingerprinting techniques suggested that the bacterial community composition of drill mud was generally stable throughout the entire drilling period of Expedition 337 (cf. Table 41 of ref. *6*). Out of the pool of drill mud samples, four representative samples were selected as controls for in depth molecular investigations in shore-based laboratories (LMT159 [Unit II, 1256.5 – 1826.5 mbsf, 2012/08/27], LMT214 [Unit III, 1826.5 – 2046.5 mbsf, 2012/08/27], LMT279 [Unit IV, 2046.5-2466 mbsf, 2012/09/06] and LMT372 [Unit IV, termination of drilling, 2012/09/09]). In addition to accessing the contribution of drill mud as possible contamination sources, other possible contamination sources such as contamination introduced during molecular experiments was evaluated by multiple experimental negative controls as described below (see "*DNA extraction, purification, and amplification of 16S rRNA genes*", "*Analysis of 16S rRNA gene-tagged sequences*" and "*Probability-based set relationships*").

Cell count

Samples for cell count were obtained from the innermost part of WRCs, immediately fixed with 2% (w/v) paraformaldehyde for ~12 h at 4ºC, washed twice with phosphate buffered saline (PBS), suspended in PBS-ethanol solution (1:1) at the final dilution volume of 10% (v/v) (*6, 8*), and preserved for shore-based analysis. Cell counts were performed by an image-based cell enumeration technique using SYBR Green I fluorescent dye (*8*). To evaluate small populations below ~10^4 cells cm^{-3}, cells were detached from sediments by a multi-layer density gradient technique (*40*), and then the number was counted by both manual and computer-based microscopic observations (*8, 41*). All filter preparation steps, including sonication of sediment and SYBR Green I fluorescent dye-staining, were carried out in an ultra-clean bench placed in a HEPA-filtered clean booth at the Kochi Institute for Core Sample Research, JAMSTEC (*40*).

Hydrocarbon gas and carbon dioxide analysis

During Expedition 337, the carbon isotopic composition of methane ($\delta^{13}C$-CH_4) and ratios of methane over ethane (C_1/C_2) were continuously monitored in mud gas; i.e., the gas that is transported with drill mud from the borehole to the rig floor during riser drilling (Fig. 1, B and C). A fraction of the incoming mud gas was transferred online to a methane carbon isotope analyzer (MCIA), which measures the concentration and stable carbon isotopic composition of methane on the basis of cavity ring-down spectroscopy technology. MCIA analyses are completed within seconds, and typically 100-200 measurements were conducted per meter of drilled sediment, depending on the rate of penetration and flow rate of drilling mud. Another fraction of the incoming mud gas was directed to a sampling line where the concentrations of higher hydrocarbon gases (C_1-C_5) were analyzed by gas chromatograph (GC)-natural gas analyzer (NGA). The run time of the chromatographic method allowed one real-time measurement every 20 minutes. Discrete samples of mud gas were taken by IsoTube samplers (Isotech Laboratories, Inc.) for shore-based verification of $\delta^{13}C$-CH_4 determined by MCIA and stable hydrogen isotope analysis of methane (δD-CH_4). Mud gas monitoring yielded data and samples throughout the geological formation below 645 mbsf, including intervals that were drilled without coring, but a few data gaps exist where technical reasons hindered the recovery of mud gas. Note that concentrations of hydrocarbon gases can be analyzed precisely in

the incoming mud gas, but deduction of in situ gas concentrations is complicated by the fact that recovery of gas from the formation is influenced by drilling operation (e.g., rate of penetration, mud weight). Therefore, only ratios of mud gas components are presented here. Due to the alkaline nature of drill mud for riser drilling, mud gas samples do not convey reliable information on the carbon isotopic composition of carbon dioxide ($\delta^{13}C$-CO_2) in the geological formation. Drilling operations, gas extraction system, sampling and analytical methods used for online mud gas monitoring have been described in detail elsewhere (*6*). Raw data for carbon isotopic composition and concentration of methane in mud gas can be found in Table 17 of ref. *6*. In order to complement mud gas analysis, sediment cores were sampled for shore-based analysis of δD-CH_4, and $\delta^{13}C$-CO_2 (Fig. 1, B and C) in the following way: from the freshly cut cores, ~5 cm^3 of sediment were transferred into 24 mL headspace vials that were sealed with butyl stoppers and crimp caps. Samples for methane analysis were preserved with 5 mL of 1 N NaOH solution and stored at -20°C. Samples for carbon dioxide analysis were preserved with saturated NaCl solution and stored at +4°C.

Shore-based, $\delta^{13}C$-CH_4, δD-CH_4, and $\delta^{13}C$-CO_2 were analyzed by isotope ratio monitoring gas chromatography/mass spectrometry. For $\delta^{13}C$-CH_4 and $\delta^{13}C$-CO_2 analysis, a Thermo Finnigan Trace GC Ultra was connected to a Thermo Finnigan DELTA Plus XP mass spectrometer via a Thermo Finnigan GC combustion III interface. For δD-CH_4 analysis, a Thermo Scientific Trace GC Ultra was connected to a Thermo Scientific DELTA V Plus mass spectrometer via a Thermo Scientific GC-Isolink interface. In both cases, the Trace GC was equipped with a Carboxen column (30 m length, 0.32 mm inner diameter). For $\delta^{13}C$-CH_4 and $\delta^{13}C$-CO_2 analysis, the carrier gas was helium (3 mL min^{-1}), the split ratio ranged from 1:3 to 1:100 depending on sample concentration, and the temperatures of the GC oven and injector were 40°C and 200°C, respectively. The primary standardization was based on multiple injections of reference CO_2 from a lab tank ($\delta^{13}C$ = -35.25 ± 0.1‰ vs. VPDB, 3.0 ± 0.5 V at m/z 44) at the beginning and end of the analysis of each sample. For δD-CH_4 analysis, the carrier gas was helium (1.2 mL min^{-1}), the split ratio was between 1:8 and 1:40 depending on methane concentration, and the temperatures of the GC oven and injector were 40°C and 200°C, respectively. Analysis of δD-CH_4 involved on-line transfer of samples from a high temperature conversion reactor (containing an empty ceramic tube covered with graphite layer that was kept at a temperature of 1440°C) in which compounds were pyrolyzed to molecular hydrogen, carbon, and carbon monoxide, prior to their transfer into the mass spectrometer via a ConfloIV interface. The primary standardization of the DELTA V Plus was based on multiple injections of reference H_2 from a lab tank (δD = -96.4 ± 0.3‰ vs. VSMOW, 3.2 ± 0.3 V at m/z 2) at the beginning and end of the analysis of each sample. Lab tank H_2 was calibrated against the certified CH_4 standard T-iso2 (2.5 vol.% CH_4 in a balance of dry, synthetic air; $\delta^{13}C$-CH_4 = -38.3 ± 0.2‰ vs. VPDB; δD-CH_4 = -138‰ ± 4‰ vs. SMOW). The analytical precision was better than 0.4‰ (1σ) and 2‰ (1σ) for stable carbon isotope and stable hydrogen isotope analysis, respectively.

During the CK06-06 cruise, samples for $\delta^{13}C$-CH_4 analysis (Fig. 1B) were collected based on the standard headspace gas method: immediately after core recovery, 10 cm^3 sediment were transferred by 5 mL tip-cut plastic syringes into a 25 mL glass vial with 10 mL NaCl saturated water and <0.1 mL mercury chloride (II) to inhibit microbial activity. The vial was capped with rubber stopper and aluminum seal and a headspace

was prepared by displacing 5 mL of water with high purity nitrogen. The glass vials were shaken for >20 min, and then headspace samples were withdrawn with a gas-tight micro-syringe for $\delta^{13}C$-CH_4 analysis. Carbon isotopic composition was determined by Thermo Electron DELTA plus isotope ratio mass spectrometer online to the gas chromatograph (GC-14B, Shimadzu) on a Porapak Q column (GL Sciences) with He as carrier gas at 130 ˚C to separate C_1, C_2, and C_3 gases.

Measurement of $^{13}CH_3D$

Analysis of the multiply substituted "clumped" isotopologue $^{13}CH_3D$ was performed in two samples of formation fluids that were recovered under in situ pressure from ~2 km-deep coal beds during downhole wireline logging operations of Expedition 337 (Fig. 1B) (*6*). Using the Schlumberger Modular Formation Dynamics Tester, formation fluid and gas were collected by the Schlumberger Quicksilver Probe and single-phase multi-sample chamber (SPMC) system, keeping in situ pressure during sample recovery. The dissolved gas in the collected fluid (~250 mL per SPMC bottle) was extracted onboard under vacuum conditions as described elsewhere (*6*). In order to facilitate the extraction of the dissolved gases from the fluid into the headspace, the extraction bottle was ultrasonicated at 25°C for 5 min. The extracted gas was transferred into pre-evacuated stainless steel bottles.

The abundance of stable isotopologues of methane, including $^{13}CH_3D$, was measured using a tunable infrared laser direct absorption spectroscopy technique (*42, 43*). Samples of methane were purified by a preparatory gas chromatography system equipped with a packed column (Carboxen-1000, 5'x1/8", Supelco) and helium as carrier gas. We define $\Delta^{13}CH_3D$ as the deviation of the abundance of $^{13}CH_3D$ in a sample of methane from that expected for a random (stochastic) distribution of isotopes among the isotopologues:

$$\Delta^{13}CH_3D = [^{13}CH_3D][^{12}CH_4]/[^{13}CH_4][^{12}CH_3D] - 1.$$

where the terms in brackets represent the abundances of each isotopologue. For samples of methane that have attained thermodynamic equilibrium, the value of $\Delta^{13}CH_3D$ corresponds to the temperature at which the methane isotopologues were equilibrated. The temperature-dependence of the equilibrium $\Delta^{13}CH_3D$ value was theoretically-estimated by molecular simulation at the B3LYP/6-311G level of theory and applying conventional isotope fractionation theory (*44*), and calibrated experimentally by thermally-equilibrating methane between 150 and 400°C over platinum catalyst (*42, 43*). Methane samples with a wide range of δD values (−629‰ to +61‰ vs. SMOW) were prepared and thermally-equilibrated at 250°C to correct for the nonlinearity of the spectroscopy analysis following the method described in ref. *42*. The reported $\Delta^{13}CH_3D$-based temperatures and associated uncertainties do not include potential calibration errors caused by inaccuracies in the molecular simulation and by physical effects, including anharmonicity, solvation and pressure effects.

Analysis of coenzyme F_{430}

Coenzyme F_{430} and its derivatives were extracted from sediment samples with 1% formic acid by ultrasonication for 30 min on ice, followed by centrifugation (×10,000 g;

30 min at 4°C) to recover the supernatant (*12*). This step was repeated three times. The combined supernatant was introduced to an anion exchange column that had been equilibrated with 50 mM of Tris-HCl (pH 7.5) and washed with deionized water prior to use. The recovered eluent was introduced to a C18 SPE column that had been equilibrated with methanol and conditioned with 1% formic acid. Absorbed F_{430} on the column was eluted with 100% methanol. F_{430} methyl ester (i.e., F_{430M}) was obtained by derivatization of the recovered F_{430} with BF3-methanol (40°C, 3.5 h) and extraction with dichloromethane. Silica gel chromatography was conducted for F_{430M} fraction to remove organic matrices. F_{430M} analysis was conducted by LC-MS/MS (Agilent HPLC 1260 Infinity coupled to a 6460 Triple Quadrupole (QQQ LC/MS system). F_{430M} was analyzed in positive ion mode by electrospray ionization (ESI) and an Agilent jet stream (see Fig. 2A). Source and sheath gas temperatures were set at 300 and 250°C, respectively. Source and sheath gas flow rates were set to 5 and 11 L min^{-1}, respectively. Capillary and nozzle voltages were set at 3500 and 500 V, respectively. For MRM analysis, the fragmentor voltage was 180 V and the collision energy was 0 V. Both precursor and product ions of F_{430} were set to m/z 975.4. Compound separation by HPLC was conducted using a ZORBAX Eclipse XDB-C18 (4.6 × 250 mm; 5 μm p.s.). Mobile phases were 10 mM ammonium acetate (A) and acetonitrile (B). The flow rate was 0.5 mL min^{-1}. The gradient condition was started at 0% B followed by 30% B after 3 min and then 90% B after 90 min. Concentration of F_{430M} in sediment samples was determined by using external F_{430M} standards (*12*) (table S3). The concentrations of intact F_{430} in ~2 km-deep coal-bed samples are 0.6-1.4 fg cm^{-3}, which are at least 1-2 orders of magnitude lower than those previously observed in ~100 m-deep sediment samples (*12, 13*) (table S3).

Cultivation of deep subseafloor microbial communities in a continuous-flow bioreactor

The reactor system used in this study is a down-flow hanging sponge (DHS) bioreactor (*14*). The DHS reactor was constructed from a glass column (diameter 12 cm; length 50 cm) with the autoclaved polyurethane sponge cubes (2 cm × 2 cm × 2 cm, pore size 0.83 mm) as the carrier material for microbial habitat. The sponge cubes were encased in plastic nets to prevent crushing of the sponges. A total of 100 sponge carriers were randomly packed into the glass column. The total pore volume of the sponge was 800 mL, and this volume was used for calculating the hydraulic retention time (HRT).

Three whole round core samples were used for the reactor cultivation: two coal samples (1922 and 1998 mbsf) and a sandstone sample (1978 mbsf). The surfaces of the samples were peeled using sterilized ceramic knives and crushed using sterilized hammers. Then, the crushed samples were pulverized in sterilized tungsten carbide lined mortars (Nichika Inc., Kyoto, Japan). We then mixed the three samples (50 g of 1922 mbsf-coal, 10 g of 1978 mbsf-sandstone, and 50 g of 1998 mbsf-coal) with 890 mL anaerobic medium (described below) without any added artificial organic substrates. All the procedure of inoculum sample preparation was performed in an anaerobic chamber at JAMSTEC (Coy lab products, Grass Lake, MI, USA).

The sponge cubes were soaked with the mixture manually and encased in the plastic net. The encased sponges were then placed in the glass column. This inoculation procedure was performed in a cold room maintained at 4°C, and the sample mixture and glass column were flushed by nitrogen gas at all times. After inoculation, the glass column was tightly closed and installed in an incubator in the dark at 40°C.

The composition of the medium for the DHS reactor was the following (L^{-1}): 5.9 mg sodium acetate, 7.3 mg sodium propionate, 8.7 mg sodium butyrate, 10 mg yeast extract, 0.53 g NH_4Cl, 0.1 g KH_2PO_4, 4 g $MgCl_2 \cdot 6H_2O$, 1 g $CaCl_2 \cdot 2H_2O$, 20 g NaCl, 2 g $NaHCO_3$, 0.1 g $Na_2S \cdot 9H_2O$, 2 mL Ti(III)-nitrilotriacetate (*45*), 1 mL trace element solution (*46*), 1 mL vitamin solution (*47*), and resazurin solution (1 mg mL^{-1}). The medium was purged by nitrogen gas, and pH was adjusted to 7.5. The medium was supplied into the reactor from the top inlet port by a peristaltic pump (Masterflex L/S tubing pump 7550-50, Cole-Parmer, Vernon Hills, IL, USA) with viton tubings (Cole-Parmer). The HRT in the reactor was set at 70 h. The DHS reactor was operated under atmospheric pressure (Fig. 2, B to D).

Analysis of mcrA

In samples of the continuous-flow bioreactor DNA extraction, PCR amplification, and sequencing of the enriched communities were performed as described previously (*47-49*). In addition, DNA was extracted from sediment core and cuttings samples following the chemical lysis method outlined in ref. *6*, and PCR amplifications of methyl-co-enzyme M reductase genes (*mcrA*) were attempted with the general mcrIRD primer pair (forward primer: 5'-TWYGACCARATMTGGYT-3'; reverse primer: 5'-ACRTTCATBGCRTARTT-3') and ANME-1-specific ANME-1-mcrI primer pairs (forward primer: 5'-GACCAGTTGTGGTTCGGAAC-3'; reverse primer: 5'-ATCTCGAATGGCATTCCCTC-3') (*50*). A deduced McrA amino acid sequence-based phylogenetic tree was constructed by the neighbor-joining method in the ARB program (*51*) with 168 amino acid positions and the percentage of acceptance of mutations distance correction (fig. S8).

NanoSIMS ion imaging of a bioreactor enrichment culture

Ten milliliters of anaerobically sampled effluent from the DHS bioreactor were put in a 50 mL tempered hard-glass gas-chromatography vial (SVG-50, Nichiden-rika glass Co. Ltd., Kobe, Japan) and amended with 1 μmol of ^{15}N-labeled ammonium and 10 μmol of either ^{13}C-labeled (99.9 atom%) bicarbonate or acetate, and 80 μmol of hydrogen. The vials were incubated at 40 °C for 14 days, then fixed with 4% paraformaldehyde and filtered onto indium tin oxide-coated polycarbonate membranes for NanoSIMS analysis (*16*). Microbial cells on membranes were stained with SYBR Green I and observed with fluorescence microscope (Olympus BX-51) prior to NanoSIMS analysis (Fig. 2E).

The samples were analyzed by raster ion imaging in an AMETEK CAMECA NanoSIMS 50L ion microprobe at the Kochi Institute for Core Sample Research, JAMSTEC. A focused primary Cs^+ beam of ~1 pA for carbon and nitrogen isotopic analyses was rastered over 25 × 25 μm^2 areas on samples. Negative secondary ions of ^{12}C (EM#2), ^{13}C (EM#3), $^{12}C^{14}N$ (EM#4), $^{12}C^{15}N$ (EM#5), and ^{32}S (EM#6) were measured using five electron multipliers (EMs) in multi-detection mode at a high mass resolution of ~9,000, which is sufficient to separate all relevant isobaric interferences (i.e., ^{13}C on $^{12}C^1H$, $^{12}C^{14}N$ on $^{13}C^{++}$ and $^{12}C^{14}N^1H$). Each run was initiated after stabilization of the secondary ion beam intensity following pre-sputtering of approximately 5–10 min with strong primary ion beam current. Each imaging run repeatedly scanned (20 to 30 times) the same area, with individual images consisting of 256 × 256 or 512 × 512 pixels,

depending on the measured area, having a dwell time of 2,000–3,000 μs. The total acquisition time ranged from 44 min to 4.5 hrs. Recorded images and data were processed using an IDL-based NASA JSC imaging software for NanoSIMS (*52*). Different scans of each image were aligned to correct image drift during acquisition. The final images (Fig. 2, F and G) were generated by adding the secondary ion counts of each recorded secondary ion from each pixel over all scans. We prepared *E. coli* standard samples with different ^{13}C and ^{15}N contents of each 1.0% and 0.36%, 2.0% and 1.36%, and 10.9% and 10.32%, respectively, to evaluate an instrumental mass fractionation for carbon isotopes, as well as finding target mass peaks (^{12}C, ^{13}C, $^{12}C^{14}N$, $^{12}C^{15}N$, and ^{32}S).

DNA extraction, purification, and amplification of 16S rRNA genes

Two DNA extraction methods were used. DNA used for 16S rRNA gene (16S) analyses was extracted from 2 g of deep-frozen sediment samples by hot-alkaline DNA extraction method (*37*). The extracted DNA solutions were purified by NucleoSpin columns (Takara Bio, Shiga, Japan). To obtain enough DNA from some low-biomass samples, >10 g of sediments were additionally sub-sampled from deep-frozen WRCs (3R-4, 1373.1 mbsf; 8L-4, 1606.2 mbsf; 11R-4, 1741.0 mbsf; 14R-2, 1822.0 mbsf; 15R-3, 1920.8 mbsf; 25R-2, 1999.1 mbsf; 26R-6, 2116.6 mbsf; 28R-7, 2308.0 mbsf and 29R-4, 2403.6 mbsf) and the bulk DNA was extracted. These extracted DNA solutions were purified and concentrated using an Aurora instrument (Boreal genomics, Vancouver, BC). DNA from cuttings, which led to the detection of *Methanococcus maripaludis*-related 16S sequences, was extracted following a chemical lysis protocol that was specifically optimized for sediment cuttings (*53*).

Attempts were made to measure the extracted DNA using a NanoDrop3300 (Thermo Scientific, MA, USA) with picogreen fluorescence stain; however, concentrations were found to be unquantifiable and below the detection limit (i.e., ~10 pg μL^{-1}). To obtain 16S fragments, V1-V4 of 16S rRNA was first amplified by polymerase chain reaction (PCR) using domain specific primers for Archaea [21F, 958R (*54*), 912R (*55*)] and Bacteria [27F, 926R (*56*)]. The PCR condition was described previously (*57*). After 30-35 amplification cycles, bacterial 16S sequences were obtained from all samples, whereas only one archaeal 16S sequence was detected from all "deep" subseafloor sediment core samples (fig. S8), indicating that no or a very few amplifiable archaeal 16S fragments were present in those PCR reactions. In addition, archaeal 16S and *mcrA* were not stably detectable even by using a microfluidic digital PCR (*17*) with primer sets targeting short fragment lengths [806F (*58*), 958R], possibly due to extremely low-biomass of methanogenic archaea in situ and/or bias caused by the generally low DNA extractability of the deeply buried archaeal cells even by applying hot-alkaline DNA extraction conditions (*37*).

In addition to 4 drill mud samples for the potential contamination source (i.e., LMT159, LMT214, LMT279 and LMT372), we examined lab negative controls during DNA extraction and PCR amplification steps: NA-HAA-NQ1-3 (9 amplicons from blank samples during DNA extraction and Aurora-purification steps [3 × 3 amplicons per extraction and purification experiment]) and NTC1-5 (3 amplicons from the first round PCR and 3 amplicons from the second round PCR [see "*Analysis of 16S rRNA gene-tagged sequences*"]). For all gel-purified PCR products of bacterial 16S fragments obtained from sediment samples, including 4 amplicons from drill mud samples and 15

amplicons from lab negative controls, the analysis of 16S-tagged sequences was performed as described below.

Analysis of 16S rRNA gene-tagged sequences

Due to the extremely low DNA mass produced upon first round amplification, a nested amplification strategy was employed in which amplicons were uniquely barcoded during the second round of amplification using V1-V3 barcoded-primer sets [27F, 534R (*59*)]. The barcoded primers were designed using a set of algorithms developed at the J. Craig Venter Institute (JCVI) (*60*). The 'A' and 'B' adapters for 454-library construction were included as a part of the PCR primers. Amplification primers were designed with FLX Titanium adapters (A adapter sequence, 5' CCATCTCATCCCTGCGTGTCTCCGACTCAG 3'; B adapter sequence: 5' CCTATCCCCTGTGTGCCTTGGCAGTCTCAG 3') and a sample barcode sequence where applicable. All forward primers included the B' adapter while all reverse primers included the A' adapter. To the 534R primer, 10 nt barcodes were included as part of the primer design (5'- A-adapter-N(10)+16S primer-3'). This design allowed for the inclusion of a unique barcode to each sample at the time of PCR, so that the barcoded samples could be multiplexed for sequencing. Every effort was made to prevent contamination of PCR reactions with exogenous DNA including setting up reactions in a laminar flow hood. The second round of amplification of 16S sequences was completed as follows (per reaction): 2 µL of first round amplification product, 0.75 units of Q5® High Fidelity DNA Polymerase (New England Biolabs, MA) and 1× final concentration of Q5® High-Fidelity Master Mix. Primers were added to a final concentration of 200 nM, with dNTPs at a final concentration of 200 µM, along with the Q5 High GC Enhancer for amplification to a 1× final concentration and nuclease-free water to bring the final volume to 20 µL. PCR cycling conditions were: initial denaturation of 30 seconds at 98°C followed by 30 cycles of 98°C for 10 seconds, 30 seconds at 56°C, and 72°C for 60 seconds, followed by a final incubation at 72°C for 5 minutes before cooling to 4°C. Negative controls included first round amplification product and a second round water blank reaction which were examined after 35 cycles. PCR reactions were visualized on 1% agarose gels. Each reaction was cleaned individually using the Agencourt XP beads (Beckman Coulter, Inc., Indianapolis IN) and resuspended in 30 µL of water prior to normalization and pooling of samples for sequencing. Amplicons were quantitated using the Quant-iT PicoGreen dsDNA Assay Kit (Life Technologies, Grand Island NY) and then normalized amounts of each sample were combined into one pool prior to purification using a QIAQuick PCR Purification column (Qiagen, Valencia CA), and submitted for sequencing with the Roche-454 FLX Titanium platform. The pooled samples were further cleaned using the Agencourt AMPure system (Beckman Coulter Genomics, Danvers MA, USA) prior to emulsification PCR (emPCR). Steps for emPCR, enrichment and 454 sequencing were performed by following the vendor's standard operating procedures. Real-time PCR was used to accurately estimate the number of molecules needed for emPCR.

After sequencing, a read processing pipeline consisting of a set of modular scripts designed at the JCVI along with the mothur pipeline (*61*) were employed for deconvolution, trimming and quality filtering. First reads were deconvoluted or assigned to samples based on their unique 10 nt barcode allowing no more than a one nt mismatch

to the barcode. After deconvolution, barcode and 16S primer sequences were removed allowing a maximum of 6 mismatches to the 16S primer and a maximum primer to barcode distance of 3 nt. Quality trimming of reads was performed using mothur. Reads were trimmed when read quality within a sliding window size of 35 fell below an average QV of 20. Reads with an average length of <100 nt, and reads with 'Ns' or with ambiguous base calls or a homopolymer longer than 8 nt were removed from subsequent analyses. A Blastn quality check was performed against an internal data set of 16S sequence reads to remove any sample reads not consistent with 16S sequences, in which at least 30% of the query must be covered by the alignment (60 nt minimum). Passing reads were subsequently further processed including chimera checking through the mothur pipeline.

To remove sequences that may be of exogenous origin (*62*), all sequence reads from negative control samples for DNA extraction and PCR, as well as drill mud samples, along with sequences from all sediment core samples were used to generate initial OTUs at 92% sequence identity. All reads that were members of an OTU that contained a representative from any of the controls samples (i.e., drill mud and experimental negative control samples) were removed from the data set as "potential contaminants" (fig. S2). The remaining reads were subsequently used to form OTUs at 97% sequence identity and this read set was also taxonomically classified by using the mothur utility package (*61*) with SILVA 115 ribosomal RNA database (*63*). In addition, some bacterial sequences closely related to sequences found in "core microbiomes" from human body habitats (*64*) were manually checked and removed from the sequence assemblage. The reads which remained after this "potential contaminant" filtering process were then treated as "the most conservative sequence reads for indigenous bacterial communities" (Figs. 1A and 3, fig. S3). A phylogenetic tree that includes top 30 OTUs, which contain more than 0.25% of total indigenous sequence reads, was constructed by a neighbor-joining algorithm with the Jukes-Cantor correction (fig. S4). For beta-diversity analysis, a Bray–Curtis dissimilarity matrix was calculated using the vegdist function in the vegan package in R (*65*) from the microbial community composition at the family/genus-level classification (Fig. 3B for both "shallow" and "deep" most conservative bacterial communities) and OTUs (fig. S10 for "deep" most likely indigenous communities).

In addition, multidimensional scaling (MDS) analysis was carried out based on the taxonomic classification of 16S sequences at the genus level (fig. S11). An analysis of variance was performed using Euclidean distance computed between all pairs of samples using permutational multivariate analysis of variance (Permanova). Permanova is a methodology for performing multivariate non-parametric analysis of variance (ANOVA) on groups of samples when their pair-wise relationships can be described by a distance matrix. As with standard ANOVA, Permanova attempts to quantify the significance of the difference between groups by measuring the ratio of group centroid differences (sum of squares between) to within group variance (sum of squares within). This ratio is an F-statistic, however, it may not follow the standard null distribution, and thus bootstrapping is performed to estimate the null distribution, when estimating p-values. The program adonis from the vegan package in R was used to perform Permanova. To estimate the group centers and group variance, a linear model was specified with only one main effect, an indicator variable identifying whether a sample should be considered "shallow" (9.5 to 364.0 mbsf, *Chikyu* cruise CK06-06) or "deep" (1279.1 to 2458.8 mbsf, Expedition 337).

A total of 32 samples were included in the analysis, 5 were "shallow" and 27 were "deep" in origin. There was a statistically significant difference between the two groups, with a p-value of 0.001. The sum of squared differences between the groups was estimated to be 0.126. This value is essentially the unadjusted R^2 of the model, which if interpreted as an effect size as defined by Cohen's η^2, is a medium-sized effect.

Probability-based set relationships

The identification of reads representing indigenous (derived from sediment samples) versus exogenous origins requires assumptions to be made, based on the distribution of reads and the taxa assigned to them from the laboratory negative controls, drill mud, and experimental sediment samples. Reoccurring similarities in the relative distribution of microorganisms identified across the control and drill mud samples, also identified in the experimental sediment samples, may be indicative of contamination during sample acquisition and processing or in the case of the drill mud, legitimate mixing of contaminant and indigenous microorganisms during drilling and extraction of experimental sediment samples. The "most conservative" approach that can be applied considers all taxa that have been identified in the negative laboratory controls or in the drilling mud samples as contaminant. This approach has the advantage of eliminating any false positive discoveries of novel taxa; however, it has the disadvantage of introducing false negative taxa, which may be of interest.

To address this issue, we introduce a "probabilistic" approach to examining the microbial composition based on the consistency (i.e., variance) of the relative distribution of the taxonomic assignments or OTUs across the entire set of samples (i.e., lab controls, drill mud samples, and experimental sediment samples). The "most conservative" approach may be considered a special case in probabilistic space, in which any taxonomic overlap between the lab controls, drill mud samples, or a combination of the two and the experimental sediment samples is disallowed. A more relaxed approach, which may provide taxonomic candidates for further investigation, is a case where a taxon is inconsistently found with low abundance in the control samples (laboratory and/or drilling mud), but consistently found with significant abundance in the experimental sediment samples. By accounting for the consistency (variance) at which any taxon or OTU is found in the controls, drill mud or experimental sediment set of samples, a more nuanced view of the overall data set can be achieved.

First, we define an indigenous OTU as one that will only be found in the experimental sediment samples ("Sediment samples"), and a contaminant as one that will be found in either "Sediment samples" and/or control samples ("Controls"). Therefore, the goal of the analysis is to estimate a likelihood based on observed abundances for these two conditions. As a byproduct, the probability that the OTU may be seen only in the control or not recovered again (neither "Sediment samples" nor "Controls", respectively), will also be generated. The greatest likelihood ("most likely") among these four set relationships will be defined as the best indicator for whether the OTU should be considered indigenous, contaminant, or other. Briefly described, the steps are first, bootstrap (by sampling with replacement) the samples included in "Sediment samples" and "Controls" sets. Next, for each sample selected, resample with replacement each read's taxonomic categorization or OTU in each sample. Both resampling steps are completed to the same sample count and read depth as the original survey. For each

bootstrap iteration, the number of times an OTU is identified only in the "Sediment samples", only in the "Controls", both groups ("Sediment samples" and "Controls"), or neither groups (neither "Sediment samples" nor "Controls"), is accumulated across all bootstrap iterations and the probability of each set category is then computed per taxon. We then define a maximum likelihood set category for each taxon of interest, as the category with the greatest probability. This estimates the *probability* that a particular OTU will be seen across each of the four set relationships if the survey were to be performed again, or alternatively, interpreted as a *likelihood* that the OTU should be considered indigenous or contaminant. In most cases, this would be all the taxa with a probability score of "Sediment samples" greater than 0.5, but also, in some rare circumstances OTUs with probabilities just above 0.25, if all four outcomes are evenly possible. When the probability score drops below 0.5, the taxa is more likely to be found in both "Sediment samples" and "Controls", "Controls" only, or not recovered again (i.e., neither "Sediment samples" nor "Controls"), none of which would imply a very strong candidate for being an indigenous microorganism (true positive). A range of probability cutoffs for the score may be used for selecting OTUs depending on the degree of certainty one would like to enforce, but the "most likely", will be defined as the set classification with the greatest probability among the four, since this gives a specific threshold-free assignment.

To aid in the interpretation of the probability-based set relationships, a series of examples are provided. The following example is an excerpt from the complete analysis performed to compare the experimental sample set to the drill mud sample set. The colored line segments represent the probabilities for the four set relationships for OTU00261 (a Firmicutes). The red and blue segments represent the probabilities of rediscovering this OTU in the "Sediment samples" and "Controls" only, as seen based on their labeled probabilities of 0.050 and 0.469, respectively. The likelihood of discovering this OTU in both samples ("Sediment samples" and "Controls") is 0.421 and is represented in purple. The likelihood of not rediscovering this OTU is 0.060 (neither "Sediment samples" nor "Controls"), and is drawn in black on the far right. This OTU would be considered a contaminant with a low predicted false negative rate as an indigenous OTU because the probability of identifying this OTU without detecting it in the "Controls" as well is only 0.050.

The next result shown is an example (OTU00158, a Burkholderiales) of a potential false negative that might have been removed through the most conservative filtering approach. In this example, the probability of identifying this OTU in the "Sediment samples" without recovering it in the "Controls" samples calculated as 0.511, dominates the other outcomes. Thus, it is "most likely" indigenous.

Of the 257 OTUs that were detected in both "Sediment samples" and "Controls", (i.e., OTUs exceeding an abundance of zero in both samples), only 21 had probabilities of exclusive detection in the experimental samples exceeding 0.5, i.e. "most likely" indigenous. From these results, OTU00203 (an Actinomycetales) had the largest

probability of 0.534. An analysis of the overlap between "Sediment samples" and "Controls" (118 overlapping OTUs) identified only four OTUs with probabilities of exclusive detection in "Sediment samples" exceeding 0.5. From this, the OTU00158 (a Burkholderiales) had the greatest probability of exclusive detection in "Sediment samples" with a likelihood of 0.511.

From the comparison of "Sediment samples" and "Controls", 788 OTUs would probably be redetected upon re-performing the sampling and sequencing survey (1,011 additional OTUs are not expected to unrecovered). Of these 788 OTUs, 552 (70.1%), 77 (9.8%), and 159 (20.2%) are expected to be detected exclusively in "Sediment samples", exclusively in "Controls", and mutually, respectively. From comparisons between the drill mud and experimental samples, 801 OTUs are expected to be redetected (998 additional OTUs are not expected to be recovered). Of these 801 OTUs, 625 (78.0%), 93 (11.6%), and 83 (10.4%) OTUs are expected to be detected exclusively in "Sediment samples", exclusively in "Controls", and mutually, respectively. Resultant values represent the probability of that set relationship being declared upon re-performing the sampling and sequencing process for each taxonomic category (fig. S5).

Estimation of the most conservative indigenous cell concentrations

Based on the raw cell count and the proportion of reads most conservatively identified as members of the indigenous population (Fig. 1A, fig. S7, and table S1), the *minimal* estimate of in situ microbial cell numbers per sample was estimated as follows: n' = a/b*n, where *n'*= indigenous cell count, *n* = raw cell count, *a* = number of sequences remaining after removal of potential contaminant sequence reads, and *b* = total number of reads sequenced. The correction factor *a/b* is the proportion of sequences estimated to be indigenous by the taxonomic filtration of "potential contaminants", which were detected from drill mud and lab negative control samples, with 92% similarity cutoff.

Estimation of the most likely indigenous cell concentrations

Based on the raw cell count and the proportion of reads identified as "most likely" indigenous, we used the probability relationship set analysis (Fig. 1A, fig. S7, and table S1), to calculate an *estimate* of the "most likely" in situ microbial cell numbers per sample as follows: n' = a/b*n, where *n'*= indigenous cell count, *n* = raw cell count, *a* = number of sequences remaining after removal of potential contaminant sequence reads, and *b* = total number of reads sequenced. The correction factor *a/b* is the proportion of sequences within OTUs estimated to be "Sediment sample" (not as "Both" or "Neither") among the total sequence read number (fig. S5). The "most likely" indigenous cell concentration is generally between "raw cell concentrations" and "most conservative indigenous cell concentration".

Sequence accession numbers in public databases

The raw pyrosequencing data and metadata of 16S sequences from "shallow" (9.5 to 364 mbsf, *Chikyu* cruise CK06-06) and "deep" (1279.1 to 2458.8 mbsf, Expedition 337) sediment samples are accessible in the DDBJ database under accession no. DRA001030 (*37*) and the NCBI Sequence Archive (SRA) database through BioProject ID: PRJNA272620, respectively. The *mcrA* and archaeal 16S sequences reported in this study (fig. S8) have been deposited in the GenBank/EMBL/DDBJ database

under accession numbers AB828146, AB828147, KP133074, AB828148, and LC042539.

Analysis of hydrogen gas concentrations

Dissolved hydrogen concentrations (fig. S12) were analyzed by extraction method as reported previously (*6*, *66, 67*); sediment samples (3 cm^3) were collected from the center part of fresh cores immediately after core recovery and control samples of mud fluid were sampled from the core liner during core sectioning to evaluate the contamination of core samples with H_2 from drill mud. The analytical reagent blank (analysis of vials filled with only NaCl solution) was 3.2 nM H_2. Raw data for hydrogen concentrations in sediment samples and control samples of drilling mud are reported in Tables 24 and 25, respectively, of ref. *6*.

Thermodynamic calculation of hydrogenotrophic methanogenesis

The standard Gibbs free energy ($\Delta G°$) of hydrogenotrophic methanogenesis (fig. S12) was calculated by using SUPCRT92 (*68*) and the thermodynamic data of dissolved species from Shock and Helgeson (1990) (*69*) at in situ temperature and the depth-expected pressure. The activities of H^+ and HCO_3^- were computed by use of PHREEQC (*70*) with the input of the major ion concentrations (Na^+, Mg^{2+}, Ca^{2+}, K^+, NH_4^+, Cl^-, Br^-, SO_4^{2-}, HS^-, HCO_3^-) measured in the formation water sample from 1,808 mbsf (*6*). The H_2 threshold of hydrogenotrophic methanogenesis ($HCO_3^- + 4\ H_2 + H^+ \rightarrow CH_4 + 3\ H_2O$) corresponding to free energy (ΔG) values of −10 and −20 kJ per mole of reaction was obtained by recasting the equation $\Delta G = \Delta G° + RT\ lnQ$ and solving for the H_2 term in Q, which is the activity quotient. R is the gas constant, and T is the temperature in Kelvin.

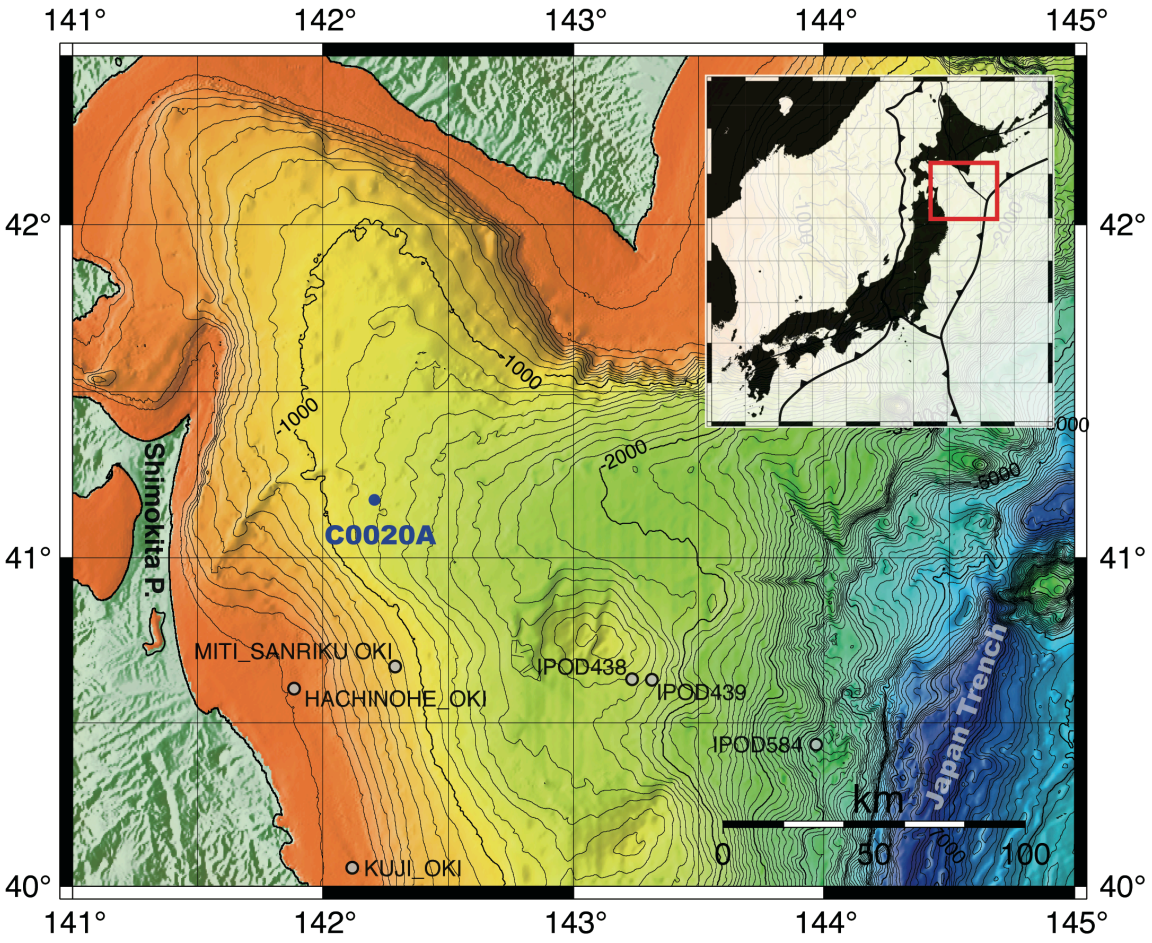

Fig. S1. Bathymetric map showing IODP Site C0020 Hole A (C0020A) drilled by the *Chikyu* in 2006 (JAMSTEC *Chikyu* cruise CK06-06) and 2012 (IODP Expedition 337) together with previously existing drill holes off the Shimokita Peninsula of Japan. Inset map shows plate configuration around Japanese Islands and the location of the index map (red square) (*6*).

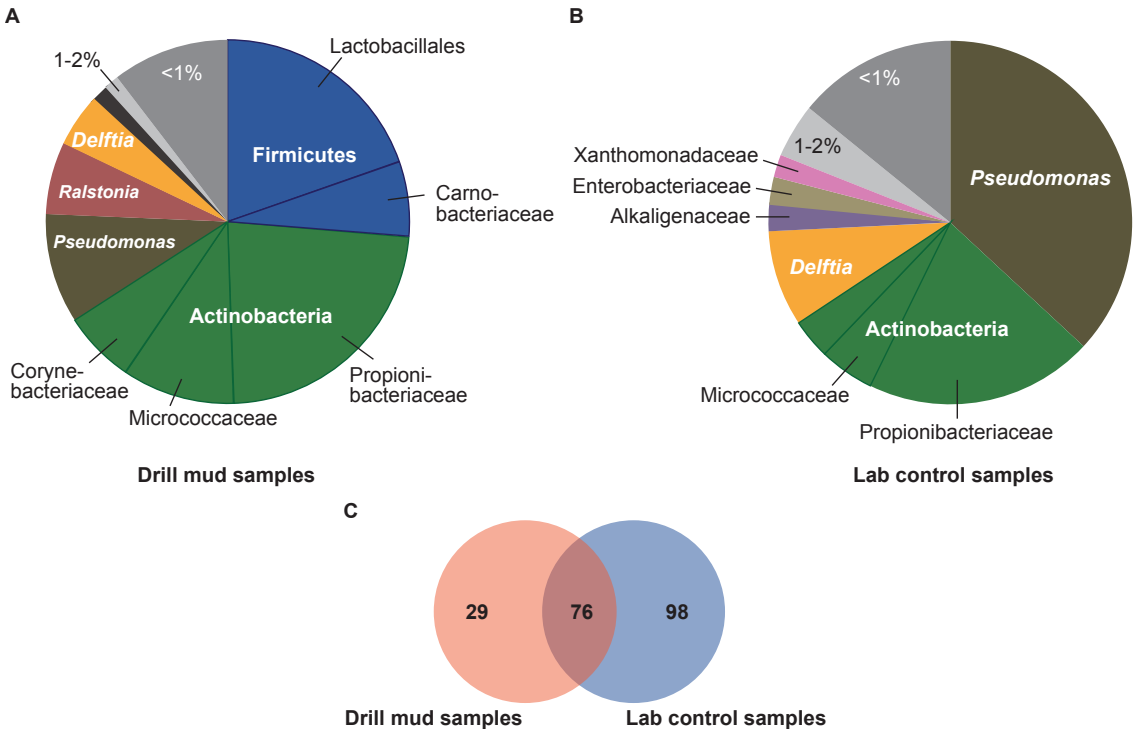

Fig. S2. Taxonomic composition of bacterial communities in drill mud and lab negative control samples during Expedition 337. (**A, B**) Pie charts show family/genus-level taxonomic classification by percent abundance of the compiled 16S sequence reads obtained from drill mud samples used during riser-drilling operations of Expedition 337 (A) and lab negative control samples (B) (*7*). (**C**) Venn diagram indicates the number of genus-level taxonomic overlaps between (A) and (B).

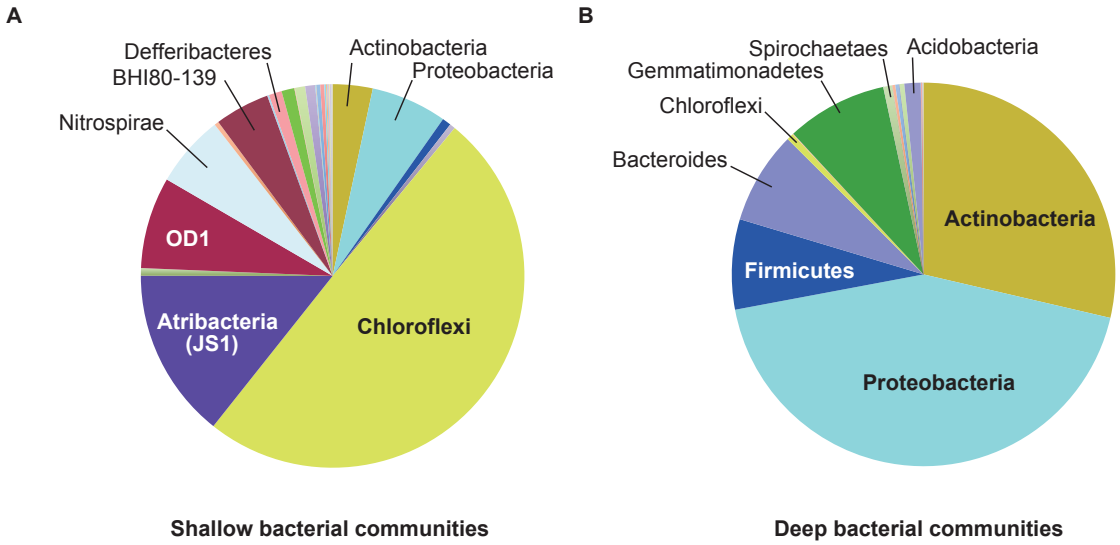

Fig. S3. Taxonomic composition of indigenous bacterial communities in sediments at Site C0020. Pie charts show phylum-level taxonomic classification by present abundance of the complied 16S sequence reads obtained from (**A**) "shallow" [9.5 to 364.0 mbsf, *Chikyu* cruise CK06-06 (*37*)] and (**B**) "deep" (1279.1 to 2458.8 mbsf, Expedition 337) subseafloor sediment samples at Site C0020 (Fig. 3A and table S4) (*7*). The "deep" bacterial communities (B) represent the most conservative indigenous members based on taxonomic classification of the 16S sequences (*7*).

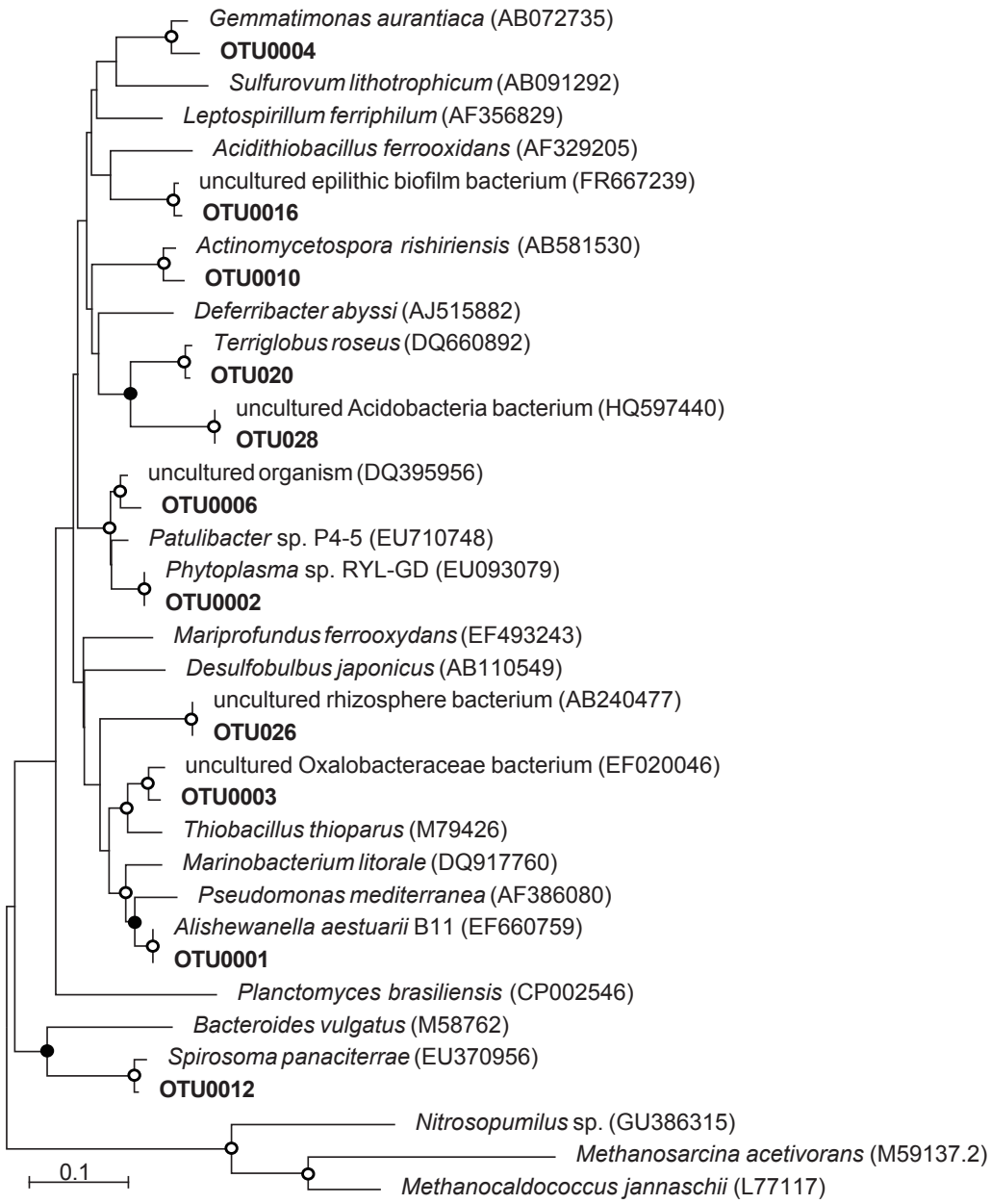

Fig. S4. Phylogenetic tree of major operational taxonomic units (OTUs) obtained from >1.2 km-deep sediment samples at Site C0020 based on the V1-V3 region of 16S rRNA gene sequences. OTUs were defined as the clusters at 97% sequence identity, and only OTUs containing more than 0.25 % of total sequence reads from all "deep" (1279.1 to 2458.8 mbsf, Expedition 337) subseafloor samples examined are shown in the tree. The accession number of each sequence is shown in the parenthesis. The tree was constructed by neighbor-joining analysis with the Jukes-Cantor correction. The bar indicates 10% estimated sequence divergence. The solid and open circles at branch nodes indicate positions where the confidence value of 1,000 bootstrap trials supports more than 50% and 80%, respectively.

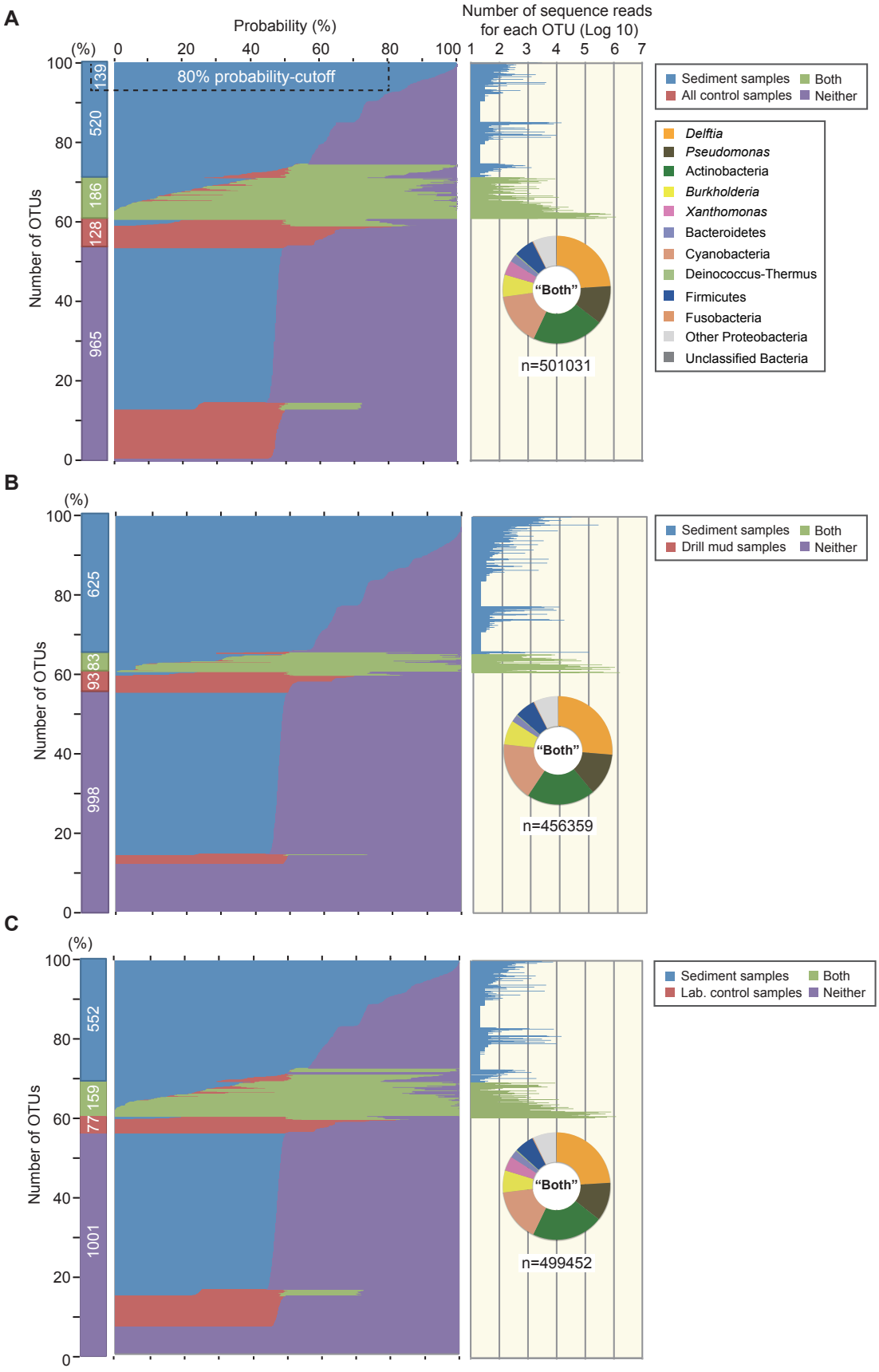
A
Probability (%)
Number of sequence reads for each OTU (Log 10)
(%)
0 20 40 60 80 100
1 2 3 4 5 6 7
80% probability-cutoff
Number of OTUs
100 80 60 40 20 0
139
520
186
128
965
"Both"
n=501031
Sediment samples
Both
All control samples
Neither
Delftia
Pseudomonas
Actinobacteria
Burkholderia
Xanthomonas
Bacteroidetes
Cyanobacteria
Deinococcus-Thermus
Firmicutes
Fusobacteria
Other Proteobacteria
Unclassified Bacteria
B
(%)
Number of OTUs
625
83
93
998
"Both"
n=456359
Sediment samples
Both
Drill mud samples
Neither
C
(%)
Number of OTUs
552
159
77
1001
"Both"
n=499452
Sediment samples
Both
Lab. control samples
Neither

Fig. S5. Probability-based set relationship analysis of most likely indigenous bacterial communities at Site C0020. A probabilistic approach was applied to account for both the consistency of samples (variance) in the control (drill mud and lab negative controls) and the sediment (experimental) sets of samples. Due to the low abundances of reads in many taxonomic categories, it is possible, for example, that if sampling and sequencing of "deep" (1279.1 to 2458.8 mbsf, Expedition 337) sediment samples were repeated, some categories might be recovered only in the experimental samples, some might be shared between experimental and control samples and some might not be recovered at all. Therefore, to account for not only the abundance of taxonomic categories recovered but also the variation across samples, inter-sample and intra-sample probability-based set overlaps between sets was estimated with bootstrapping by using resampling with replacement both the samples in each group and the reads acquired. Control samples ("Controls") performed against the experimental "Sediment samples" were: (**A**) all control samples (laboratory controls and drill mud samples), (**B**) drill mud samples, and (**C**) laboratory negative control samples. These comparisons produce probability-based sets of: 1) "Sediment samples" only, 2) "Both" ("Sediment samples" and "Controls"), 3) "Controls" only (i.e., all controls, drill mud, or laboratory negative controls only) and 4) "Neither" which represents OTUs that would not be expected to be recovered upon repetition of sampling. The colored proportion represents the probability of a set relationship that would be re-established upon re-performing the experiment: i.e., "Sediment samples" in blue, "Controls" (i.e., all, lab controls or drill mud samples) in red, "Both" in green or "Neither" in purple. The numbers of OTUs in each probability category are shown in the left bar chart. In panel A, the number of OTUs with probabilities exceeding the 80%-probability cutoff (boxed within dashed lines) of being exclusively in the experimental set is 139, whereas 520 among a total 1799 OTUs are assigned as "Sediment samples" only based on the probability-set relationship analysis (also see fig. S6). Log_{10} numbers of sequences per OTU are shown in the right panel with the same color indication as used in the box and bar charts, indicating that the total sequence reads of OTUs most likely derived from "Sediment samples" or "Both" are abundant where as those of "Controls" only or "Neither" are notably small. In the left panel, taxonomic compositions of "Both" fraction for each analysis are shown: The results indicate that most sequences within the "Both" OTU-fraction are affiliated to the members of *Delftia*, *Pseudomonas*, Actinobacteria, *Burkholderia*, and *Xanthomonas*, most of which are typical experimental contaminants (cf. ref. *62*) and also abundantly detected in both drill mud samples and laboratory negative controls (fig. S2). The number of "Both" OTUs in panel B ("Sediment samples" vs. Controls [drill mud samples]), as well as the number of "Both" sequence reads represented by "n" below the pie chart, are less than those in panel C ("Sediment samples" vs. Controls [lab. control samples]), suggesting that contamination of samples during experimental processing (e.g., DNA extraction, purification, PCR, etc.) is more critical than drill mud-induced contamination. Hence, these OTUs are subsequently eliminated from the "most likely" indigenous bacterial community for the cell concentration estimate (Fig. 1A).

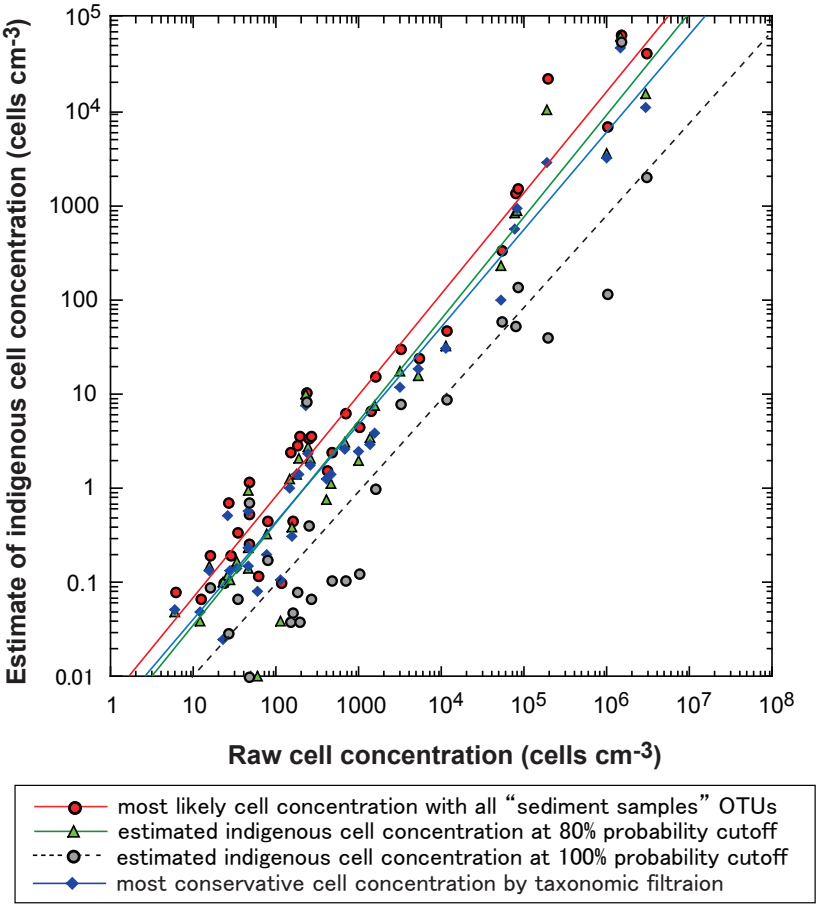

Fig. S6. Correlation between the raw cell concentration and the estimates of "most likely" and "most conservative" indigenous cell concentrations at Site C0020. Using the raw cell count data and the proportion of reads most conservatively identified as members of the indigenous population (Fig. 1A and table S1), the minimal estimate of in situ microbial cell numbers per sample was estimated as follows: n' = a/b*n, where *n'*= indigenous cell count, *n* = raw cell count, *a* = number of sequences remaining after removal of potential contaminant sequence reads, and *b* = total number of reads sequenced. The correction factor *a*/*b* is the proportion of sequences estimated to be indigenous. The numerator, *a*, is based on one of the following approaches: the "most conservative", the "most likely", or a probability threshold (cutoff) (e.g., 80%-probability cutoff; fig. S5A). The resulting (logarithm) regression lines for the raw cell concentrations versus the most likely indigenous cell concentration were: $y = 0.0061504 * x^{(1.0718)}$ ($R^2 = 0.63$) for the estimates based on the 16S sequence reads of all "Sediment samples" OTUs; $y = 0.0030446 * x^{(1.083)}$ ($R^2 = 0.35$) for the estimates based on the 16S sequence reads of 80% probability cutoff-OTUs; $y = 0.0010814 * x^{(0.97976)}$ ($R^2 = 0.18$) for the estimates based on the 16S sequence reads of 100% probability cutoff-OTUs. The regression line for the raw cell concentration versus the most conservative indigenous cell concentration was: $y = 0.004032 * x^{(1.029)}$ ($R^2 = 0.34$) for the estimates based on the manual taxonomic filtration of 16S sequence reads from all control samples. The slope

(m) and intercept (b) for the logarithm regression line is in the form of $y = b * x^{(m)}$. A slope close to 1.0 indicates that there is largely no disproportionate removal of contaminants based on raw cell count magnitudes. The value of the intercept is a measure of the proportion of contaminant that was removed. A smaller intercept indicates more removal. The R^2 indicates how closely the points fit on their respective regression line. A low R^2 only indicates that there were unexplained (i.e., random) factors that were responsible for the proportion of contaminant detected per sample.

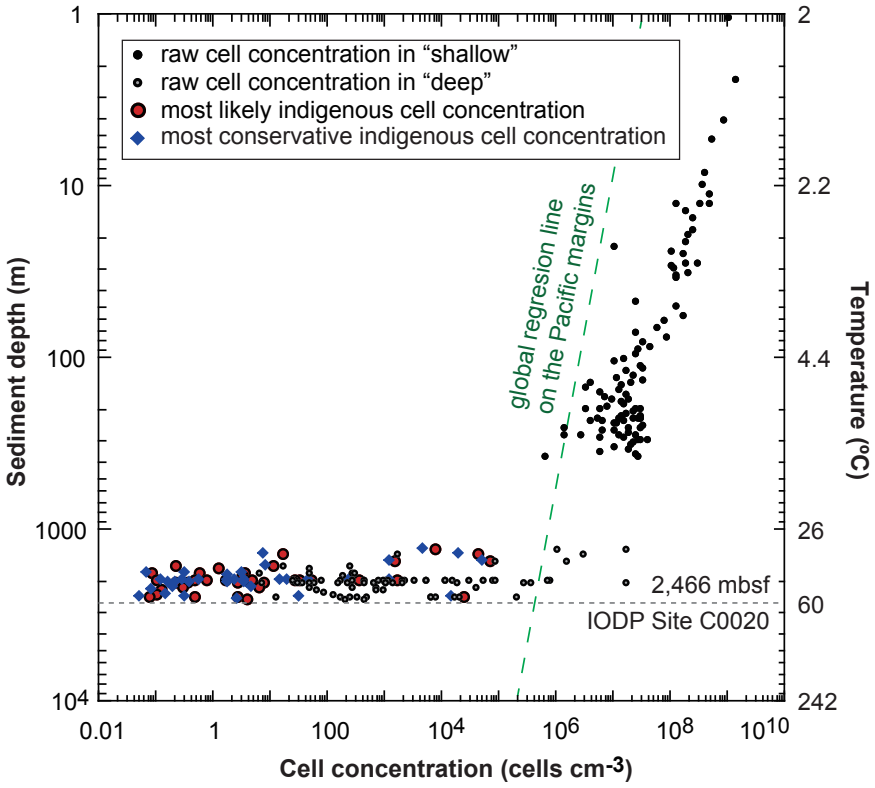

Fig. S7. Cell concentration profile on vertical logarithmic scales of sediment depth and temperature in situ at Site C0020. For cell concentrations in “deep” (1279.1 to 2458.8 mbsf, Expedition 337) subseafloor, raw data of fluorescence image-based cell counts (*8*), the most likely indigenous cell concentrations based on the probability-based set relationship analysis (*7*), and the most conservative indigenous cell concentrations estimated based on the taxonomic classification (*7*) are shown (Fig. 1A and table S1). The cell concentrations in “deep” subseafloor sediments are drastically lower than predicted by the slope of the global regression line (dashed green line) (*2*).

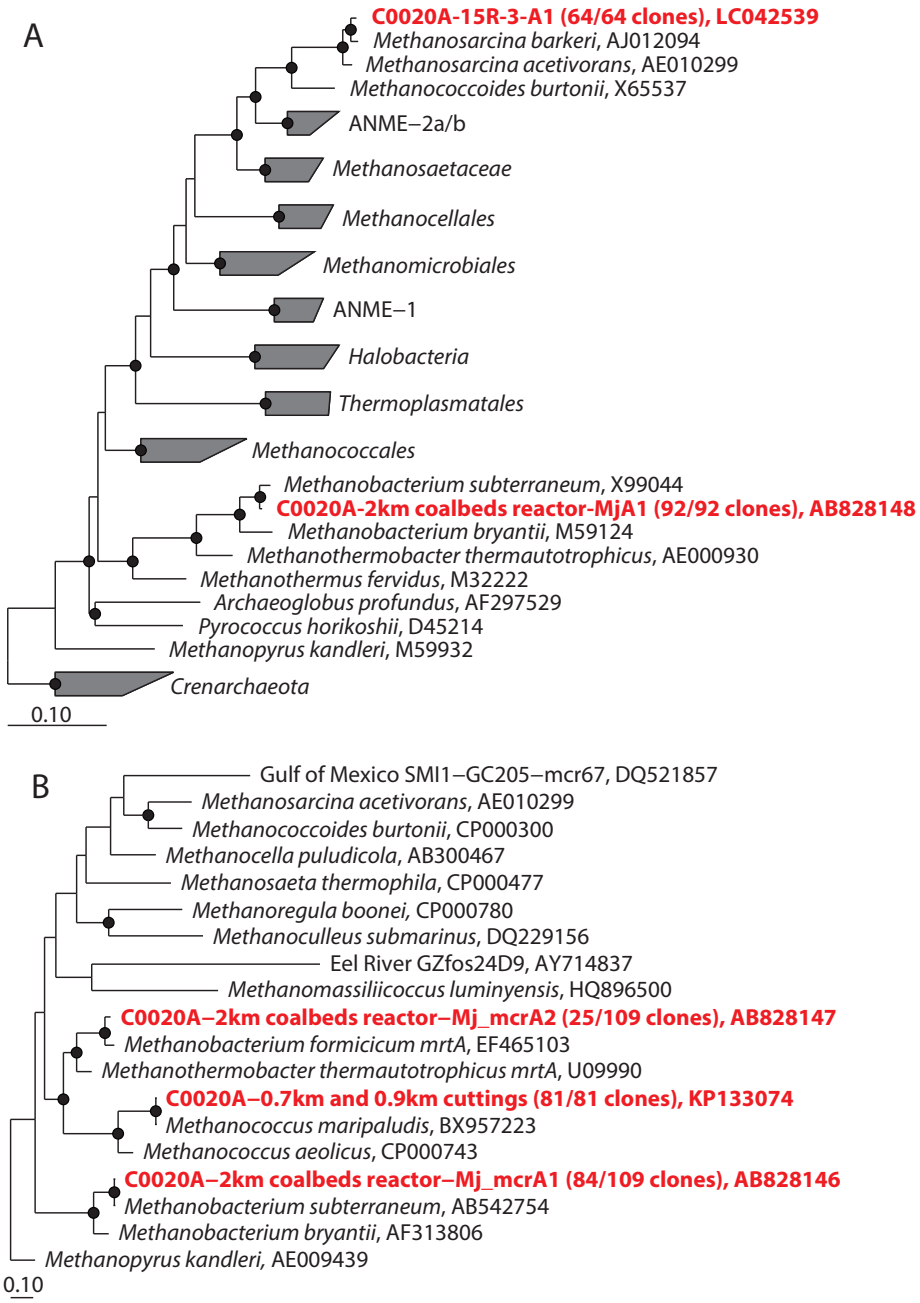

Fig. S8. Phylogenetic trees of archaeal genes obtained from "deep" subseafloor sediments at Site C0020. Phylogenetic trees were constructed by using the neighbor-joining method in the ARB program package (*51*). Phylogenetic positions of clones obtained from a 2 km-deep bioreactor enrichment culture, a coal sample (15R-3, 1920.8 mbsf) and cuttings samples (30SMW, 696.5 mbsf and 61SMW, 946.5 mbsf) are shown in red. The number in parentheses indicates the number of identical clones obtained per number of clone analyzed. (**A**) For the 16S rRNA gene tree, the initial tree was constructed with sequences greater than 1000 nucleotides using the neighbor-joining method. Subsequently, the archaeal 16S sequences obtained in this study were inserted

into the initial tree by using the parsimony insertion tool of the ARB program. (**B**) The McrA tree was constructed based on a distance matrix (i.e., 168 amino acid positions based on *mcrA* sequences; percentage acceptance mutations distance correction) by the neighbor-joining method. The accession number of each sequence is shown after the strain or clone name. To estimate the confidence of the tree topologies, bootstrap resampling analysis with 1000 replicates was performed the neighbor-joining method by using MEGA5 software (*71*). The bar indicates 0.1 changes per sequence position. The solid circles at nodes indicate positions where the confidence value of 1000 bootstrap trials supports more than 80%.

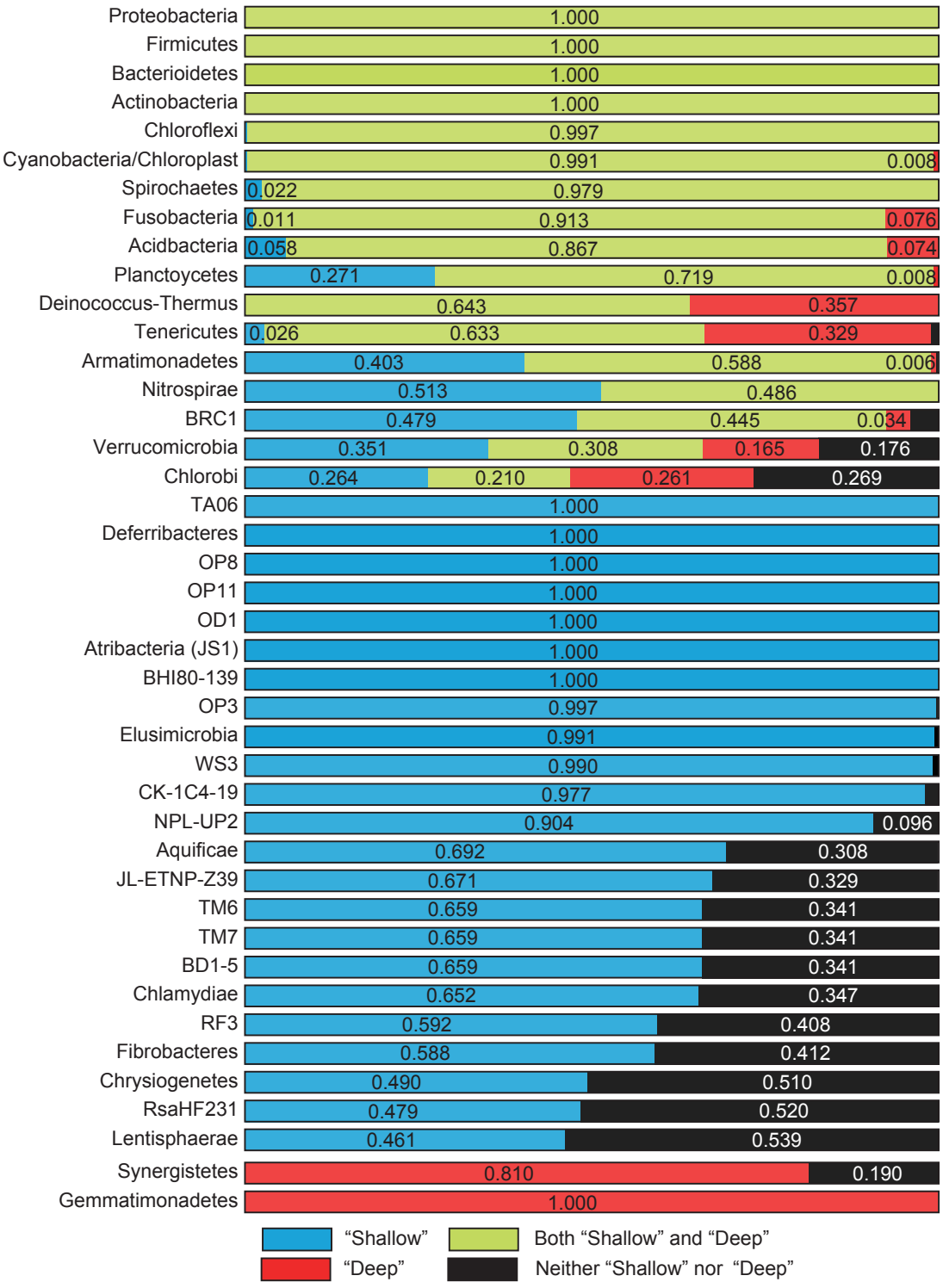

Fig. S9. Probability-based set relationship analysis of indigenous bacterial communities in "shallow" and "deep" subseafloor sediment samples at Site C0020. Due to the low abundances of reads in many taxonomic categories, it is possible that if sampling and sequencing of "shallow" (9.5 to 364.0 mbsf, *Chikyu* cruise CK0-06) and "deep" (1279.1 to 2458.8 mbsf, Expedition 337) sediment samples were repeated, some

categories might not be recovered. Therefore, to account for not only the abundance of taxonomic categories recovered but also the variation across samples, inter-sample and intra-sample probability-based set overlaps between sets (“Shallow”, “Deep”, “Both” or “Neither”) was estimated with bootstrapping by using resampling with replacement both the samples in each group and the reads acquired. The length of the colored proportion of each bar represents the probability that a set relationship would be re-established upon re-performing the experiment. The colors light blue, red, light green and black represent the probability of recovering the taxon in the “Shallow”, “Deep”, “Both”, or “Neither” samples, respectively. For example, in spite of the significantly higher abundance of Chloroflexi in the “shallow” samples compared to in the “deep” samples, there was still a consistent and high enough abundance of that taxon in both groups for it to be recovered mutually with a probability of 99.7%. The abundance of Verrucomicrobia was relatively low in both sample groups, resulting in only a 30.8% chance that the taxon would be recovered in both groups, and 17.6% chance it would not be recovered in either group.

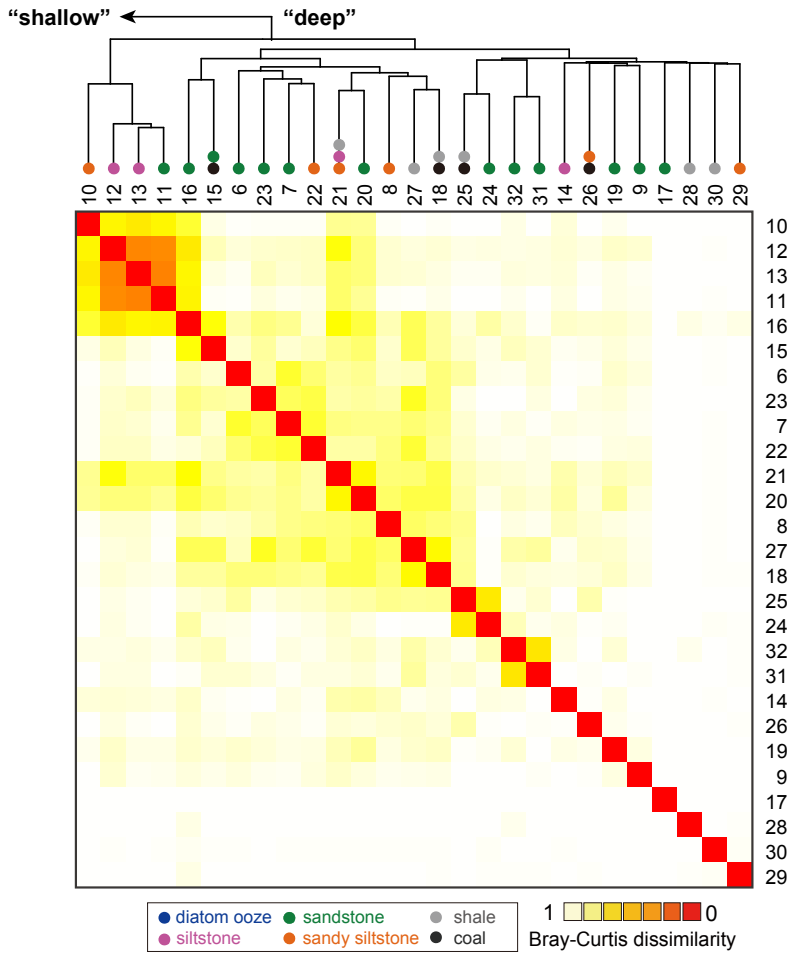

Fig. S10. OTU-based cluster and Bray-Curtis dissimilarity analyses of most likely indigenous bacterial communities in “deep” subseafloor sediment samples (sample numbers 6 to 32, 1279.1 to 2458.8 mbsf, Expedition 337). Colored dots represent sedimentological characteristics of each sample horizon, respectively (*6*). The results based on taxonomic classification (i.e., genus level data in Fig. 3B) and taxonomy-independent data (i.e., OTUs in this figure) reveal very similar patterns upon sample comparison. This is due to the fact that most of the taxonomic classifications (e.g., genera) are largely dominated by one or two OTUs of high abundance within the taxonomic category (see fig. S5): A majority (78%) of the taxonomic categories in our deep subseafloor data set possess a single dominant OTU (where single dominant OTU is defined as an OTU whose reads account for more than 50% of the total reads of the taxa to which they have been classified). For the OTUs whose taxonomy could not be classified confidently to the genus level (e.g., classified only to the phylum level) there are significantly more OTUs compared to those classified to the genus level. However, in spite of the greater OTU diversity contained in these higher level taxonomic categories, the number of reads per OTU was also significantly smaller. Therefore, the contribution on the interprofile distances is negligible.

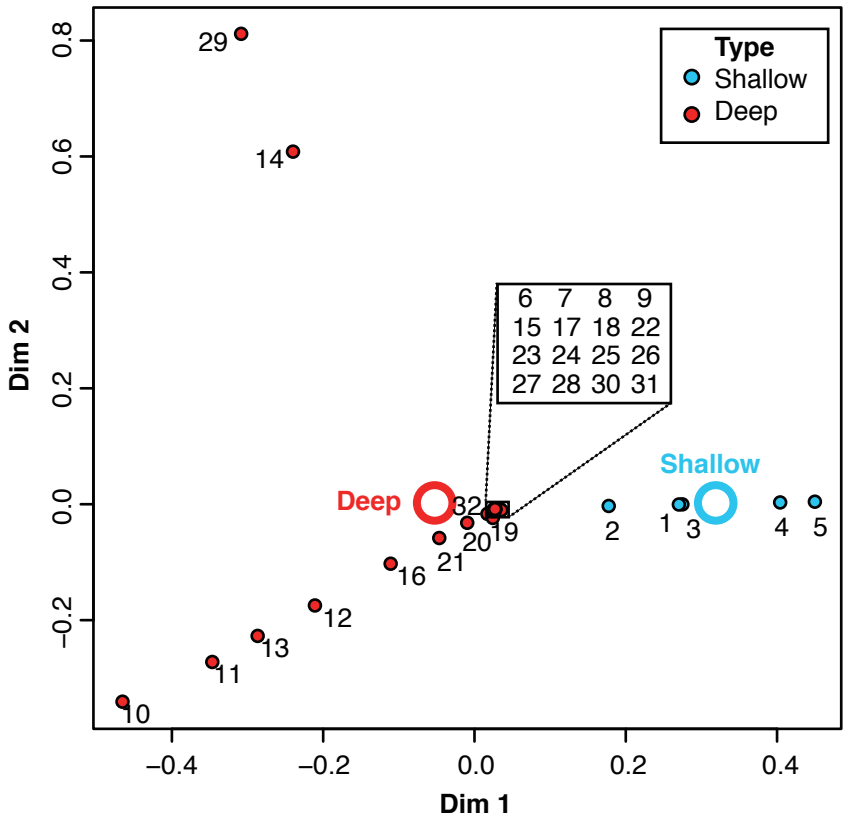

Fig. S11. Multidimensional scaling (MDS) analysis of indigenous bacterial communities at Site C0020. The "shallow" (9.5 to 365.0 mbsf, *Chikyu* cruise CK0-06) and "deep" (1279.1 to 2458.8 mbsf, Expedition 337) subseafloor sediment samples are indicated in blue (sample numbers 1 to 5) and red (sample numbers 6 to 32), respectively. The large colored circles labeled "Shallow" and "Deep" in the plot represent the centroids of each group. A MDS plot was computed based on the Euclidean distance between the abundances characterized by RDP-based taxonomic assignments at the genus level made on sequences assaying the V1-V3 variable region of the 16S rRNA genes (*7*). These distances were also used to perform permutational analysis of variance (PERMANOVA), which indicated a significant difference between the groups with a p-value of 0.001 and a medium effect size of 0.1257, as measured by Cohen's η^2.

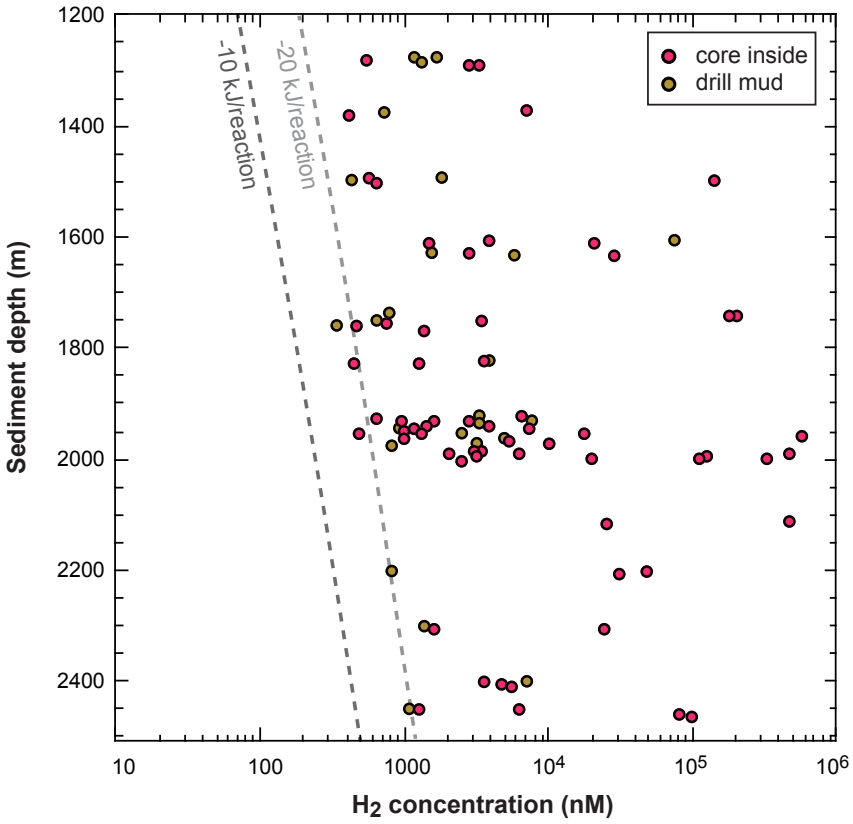

Fig. S12. Downhole profile of hydrogen (H_2) concentrations in deep subseafloor sediments at Site C0020. The dashed lines represent H_2 thresholds of hydrogenotrophic methanogenesis corresponding to free energy (ΔG) values of −10 kJ (in dark gray) and −20 kJ (in light gray) per mole of CH_4 calculated based on in situ temperature and the depth-expected pressure conditions (*6*). Note that H_2 concentrations in core sediments are generally higher than those in drill mud, suggesting that the former is the major source of H_2 rather than the artificial formation during riser-drilling operation.

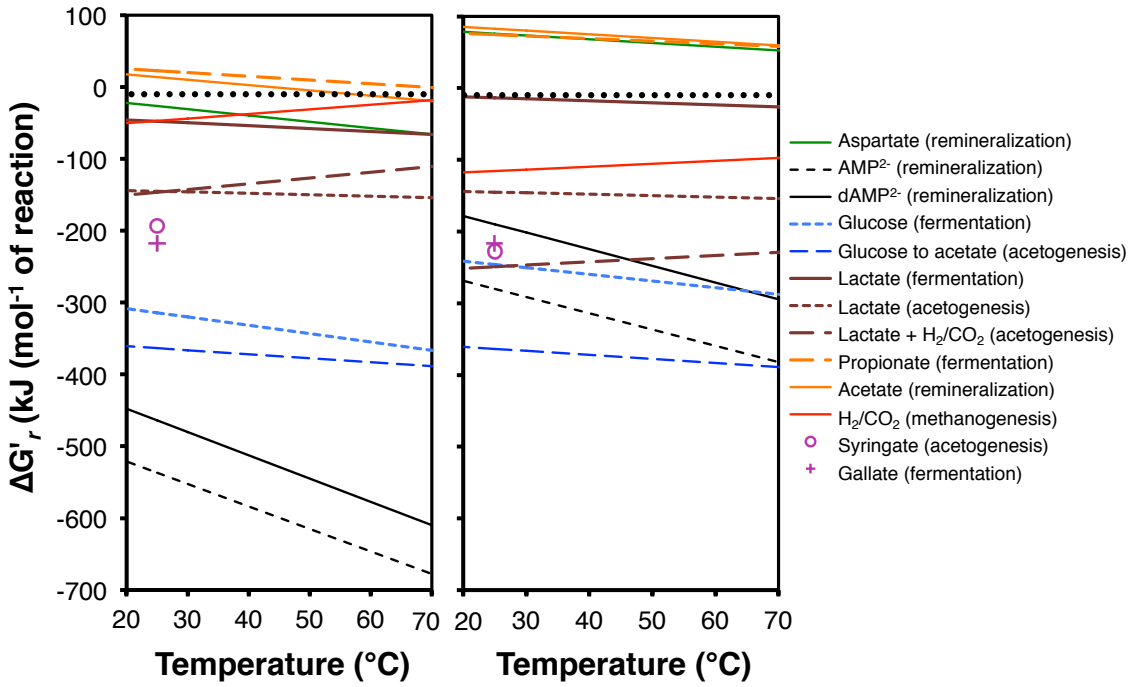

Fig. S13. Modeled Gibbs free energy yields of anaerobic reactions from a range of representative biomass building blocks and metabolic intermediates at dihydrogen activities of 10^{-6} (left panel) and 10^{-3} (right panel) under the in situ temperature regime of the cored interval at Site C0020. Aspartate was chosen as a protein building block, adenosine monophosphate (AMP^{2-}) as an RNA building block, deoxyadenosine monophosphate ($dAMP^{2-}$) as a DNA building block, glucose as a polysaccharide building block, syringate as a lignin building block, gallate as a degradation product of syringate and substrate of primary fermentation, lactate and propionate to represent substrates of secondary fermentation, acetate as a key metabolic intermediate, and H_2/CO_2 as substrates of hydrogenotrophic methanogenesis. For substrates where the catabolic reactions are not known (aspartate, AMP^{2-}, $dAMP^{2-}$), Gibbs free energy values for the full remineralization to inorganic constituents were calculated (table S6) (*72-81*). For substrates where both fermentative catabolism and acetogenic catabolism are possible, Gibbs free energy yields for both reactions are shown (table S7). For reference, we have included a dashed line indicating a Gibbs free energy of -10 kJ mol^{-1}, a value in the range of the minimum Gibbs free energy that can be harvested by microbial cells (*82*). Assumed activities are: aspartate = AMP^{2-} = $dAMP^{2-}$ = glucose = syringate = gallate = 10^{-9}; $H^+ = 10^{-8}$; lactate = propionate = 10^{-6}; acetate = 10^{-5}; $NH_4^+ = HPO_4^{2-} = 10^{-4}$; $HCO_3^- = CH_4 = 10^{-1}$; and $H_2O = 1$. Our calculations indicate the following reactions to be highly energy-yielding under all conditions: remineralization of RNA and DNA building blocks, glucose catabolism by fermentation and acetogenesis, lactate catabolism by acetogenesis reactions with and without H_2/CO_2 as co-substrates. The secondary fermentation of lactate and the hydrogenotrophic production of methane are also exergonic at both H_2 activities. Whether aspartic acid and acetate remineralization are exergonic depends on the assumed H_2 activity. By contrast, secondary fermentation of propionate is never exergonic. The acetogenic demethoxylation of syringate to gallate and the fermentation

of gallate to acetate are thermodynamically highly favorable, however, we could only perform these calculations for standard temperature (25°C) due to absence of published data on free enthalpy of formation. Not shown are calculated data for the remineralization of three additional amino acids (i.e., glutamate, serine, and alanine), which follow the same trend as aspartate. We also performed calculations for the remineralization of the lipid building block palmitate at room temperature; however, this reaction was highly endergonic.

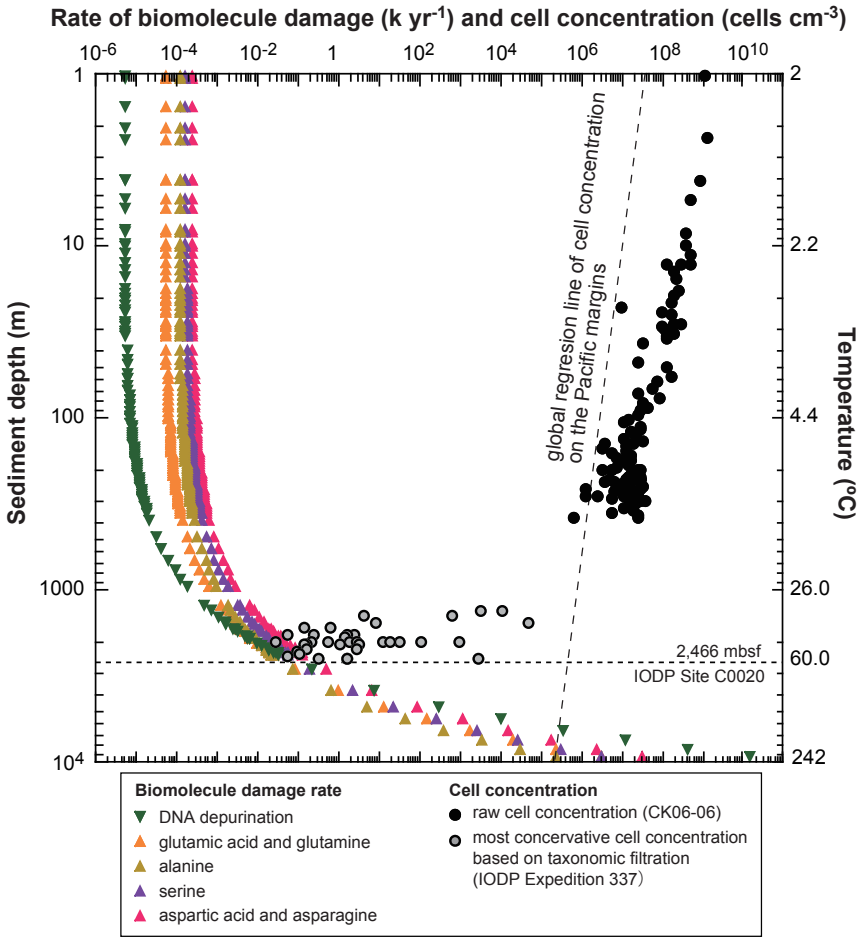

Fig. S14. Depth profiles of cell concentrations and biomolecule damage rates at Site C0020. Dot plots in black and gray show cell concentrations in "shallow" (above 365 mbsf, *Chikyu* cruise CK06-06 cruise) and "deep" (below ~1.2 km, Expedition 337) subseafloor sediment core samples, respectively. The dashed line indicates the global regression line of cell concentrations in sediments on the Pacific margins (*2*). The temperatures in situ are based on the gradient of 24ºC km^{-1} (for linear scale, see Fig. 1) (*6*). Amino acid racemization and DNA depurination rates are based on published relationships with temperature (*31*, *83*).