Supplementary materials and methods

Sample collection and processing

To apply the PathoChip developed in this study to understand dynamics of virulence genes in the environment under various environmental stimuli, three sets of environmental samples were collected and processed from soil, seawater, and human saliva.

a) Soil samples were collected from unclipping subplots in an experimental warming site at the Kessler Farm Field Laboratory (KFFL) located at the Great Plain Apiaries in McClain County, Oklahoma, USA (34°58'54"N, 97°31'14"W). The warming condition (+2°C) has been established since 1999 for six pair of 1 m × 1 m subplots by using infrared radiator (Kalglo Electronics, Bethlehem, PA, USA), suspended 1.5 m above the ground in warming plots. The dummy infrared radiator is suspended in control plots to exclude a shading effect of the device itself on treatments. Twelve soil samples were taken from the 0–15 cm layer of both 6 warming and 6 control plots in July, 2010. Each sample was composited from four soil cores (2.5 cm diameter × 15 cm deep) after being sieved by a 2mm sieve. All samples were transported to the laboratory immediately and stored at -80°C.

b) Oil-contaminated seawater samples were collected from the Gulf of Mexico during two monitoring cruises between 27 May and 2 June in 2010 (http://www.epa.gov/bpspill/dispersants.html#directives). Five oil-contaminated samples (BM053, BM054, BM057, BM058 and BM064) from the MC252 dispersed oil plume at 1099–1219 m depth were collected in Niskin bottles attached to a CTD sampling rosette (Sea-Bird Electronics Inc., Bellevue, WA, USA) after the presence of oil was detected using the WETLabs WETStar fluorometer (WET Labs, Philomath, OR, USA). Five control samples (OV003, OV004, OV009, OV013 and OV014) were collected from a nearby non-contaminated zone. Approximately, 0.8–2 L of water were filtered through 47 mm diameter polyethylsulfone membranes with 0.22 μm pore size (MO BIO Laboratories, Inc., Carlsbad,

CA, USA), and the filters were then immediately frozen and stored at -20°C for the remainder of the cruise. Filters were shipped on dry ice and stored at -80°C.

c) Saliva samples were collected from undergraduate students during an oral health survey at Sun Yat-sen University, Guangzhou, China in September, 2009. All were informed of the nature of the study and provided written informed consent in accordance with the ethical committee of the Stomatology Hospital, Sun Yat-sen University. They were comprised of unrelated individuals from both genders, aged between 18–22 years and shared a relatively homogeneous campus living environment. All reported no antibiotics-intake for at least the previous six months. All were asked to avoid eating or drinking for one hour prior to oral sampling. Those with other oral (e.g. periodontitis or halitosis) or systematic diseases were excluded. After the oral health survey, 2 mL of saliva were collected from 10 caries-free (CF; sample # 1, 3, 4, 5, 6, 8, 9, 10, 11 and 12) and 10 caries-active (CA; sample # 14, 16, 17, 18, 19, 20, 21, 22, 23 and 24) individuals, mixed with an equal volume of lysis buffer in a tube and stored at -80°C.

DNA extraction

a) Soil DNAs were extracted by freeze-grinding mechanical lysis as described previously (Zhou *et al*., 1996) and purified using a low melting agarose gel after phenol extraction. DNA purity was determined based on the ratios of 260/280 nm and 260/230 nm absorbance by a NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies Inc., Wilmington, DE, USA), and its quantity was measured with PicoGreen29 using a FLUOstar Optima (BMG Labtech, Jena, Germany).

b) Seawater DNAs, kindly provided by Dr. T. C. Hazen, were extracted from filters using a modified Miller method (Miller *et al*., 1999). Briefly, one quarter of each filter was cut into small pieces and placed in a Lysing Marix E tube (MP Biomedicals, Solon, OH,

USA). Miller phosphate buffer (300 μL) and Miller SDS lysis buffer (300 μL) were added and further mixed with 600 μL phenol: chloroform: isoamyl alcohol (25:24:1). The tubes were bead-beat at 5.5 m/s for 45 sec in a FastPrep instrument and then centrifuged at 16,000 × g for 5 min at 4°C. A total of 540 μL supernatant was transferred to a 2 mL tube and an equal volume of chloroform was added, mixed and then centrifuged at 10,000 × g for 5 min. Approximately, 400 μL aqueous phase was transferred to another tube, and 2 volumes of Solution S3 (MoBio, Carlsbad, CA, USA) was added and mixed by inversion. The rest of the clean-up procedures followed the manufacturer's instructions. DNA samples were recovered with 60 μL of Solution S5 and checked DNA yield and purity using a NanoDrop spectrophotometer and PicoGreen29, respectively.

c) Saliva DNAs were extracted using a high salt protocol as previously described (Quinque *et al*., 2006). Both 30 μL proteinase K (20 mg/mL, Sigma, St. Louis, MO, USA) and 150 μL SDS (10%) were added to 2 mL of the saliva/buffer mixture and incubated overnight at 53°C. After addition of 400 μL NaCl (5 M), the mixture was incubated on ice for 10 min, distributed equally into two 2 mL tubes and centrifuged at 13,000 rpm for 10 min. The supernatant was transferred into a new tube, and 800 μL isopropanol was added. The tubes were incubated for 10 min at RT and then centrifuged at 13,000 rpm for 15 min. After discarding the supernatant, the DNA pellet was washed once with 500 μL ethanol (70%), dried and then resuspended in 30 μL double-distilled water. DNA concentration was measured using a NanoVue spectrophotometer (GE Healthcare Life Science, Piscataway, NJ, USA). DNA purity was determined by the ratio of 260/280 nm absorbance. DNA integrity was verified by agarose gel electrophoresis.

Microarray analysis

DNA (~2 μg) was labeled with Cy3 using random primers and the Klenow fragment of DNA polymerase I according to Wu *et al*. (2006). A whole community genome amplification (WCGA) was employed for seawater DNA to obtain enough template DNA (~2.5–4.0 μg) for microarray analysis using the TempliPhi Kit (GE Healthcare Life Science) before labeling (Wu *et al*., 2006). Labeled DNA was purified using the QIAquick purification kit (Qiagen, Valencia, CA, USA), measured on a NanoDrop ND-1000 spectrophotometer and then dried down in a SpeedVac (ThermoSavant, Milford, MA, USA) at 45°C for 45 min.

Dried DNA was rehydrated with 2.68 μL sample tracking control (NimbleGen, Madison, WI, USA) to confirm sample identity. Rehydrated DNA samples were incubated at 50°C for 5 min, vortexed for 30 sec and then briefly spun down to collect all liquid at the bottom of the tube. Hybridization buffer (7.32 μL) was then added to the samples, vortexed to mix, spun down, incubated at 95°C for 5 min and then maintained at 42°C on a Hybridization Station (MAUI, BioMicro Systems, Salt Lake City, UT, USA) for at least 5 min (Lu *et al*., 2012). Samples (~6.8 μL) were then loaded onto the array surface and hybridized approximately 16 hours with mixing.

After hybridization, the arrays were washed three times according to the manufacturer's instruction and then scanned by MS 200 Microarray Scanner (NimbleGen) with a laser power of 100% and a photomultiplier tube (PMT) gain of 100%. Scanned images were extracted and quantified using NimbleScan software (NimbleGen) and then proceeded to data pre-processing (Wu *et al*., 2006; He *et al*., 2010). Raw data from NimbleScan were submitted to Microarray Data Manager and analyzed using the data analysis pipeline (http://ieg.ou.edu/microarray/). The data normalization and quality filtering were performed by multiple steps (Liang *et al*., 2010; Deng and He, 2013). First, the average signal intensity of CORS was calculated in each array, and the maximal average value was applied to normalize the signal intensity of samples in each array. Second, the sum of signal intensity of

samples was calculated in each array, and the maximal sum value was then applied to normalize the signal intensity of all spots in an array, which produced a normalized value for each spot in each array. Spots were scored as positive and retained if the signal-to-noise ratio (SNR) was ≥ 2.0 (He and Zhou, 2008), and the coefficient of variation (CV) of the background was < 0.8. In addition, spots with signal intensity less than 1,000 were discarded. Spots that were detected in less than two samples were also removed. Before statistical analysis, logarithmic transformation was carried out for the remaining spots, and the signals of all spots were transferred into relative abundances. All hybridization data of functional gene arrays are available at the Institute for Environmental Genomics, University of Oklahoma (http://ieg.ou.edu/4download/).

Supplementary references

Deng Y, He Z. (2013). Microarray Data Analysis. In: He Z (ed) *Microarrays: Current Technology, Innovations and Applications*. Horizon Scientific Press: Norwich, UK (in press).

He Z, Deng Y, Van Nostrand JD, Tu Q, Xu M, Hemme CL *et al.* (2010). GeoChip 3.0 as a high-throughput tool for analyzing microbial community composition, structure and functional activity. *ISME J* **4:** 1167–1179.

He Z, Zhou J. (2008). Empirical evaluation of a new method for calculating signal-to-noise ratio for microarray data analysis. *Appl Environ Microbiol* **74:** 2597–2566.

Liang Y, He Z, Wu L, Deng Y, Li G, Zhou J. (2010). Development of a common oligo reference standard (CORS) for microarray data normalization and comparison across different microbial communities. *Appl Environ Microbiol* **76:** 1088–1094.

Lu Z, Deng Y, Van Nostrand JD, He Z, Voordeckers J, Zhou A *et al*. (2012). Microbial gene functions enriched in the Deepwater Horizon deep-sea oil plume. *ISME J* **6:** 451–460.

Miller DN, Bryant JE, Madsen EL, Ghiorse WC. (1999). Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples. *Appl Environ Microbiol* **65:** 4715–4724.

Quinque D, Kittler R, Kayser M, Stoneking M, Nasidze I. (2006). Evaluation of saliva as a source of human DNA for population and association studies. *Anal Biochem* **353:** 272–277.

Wu L, Liu X, Schadt CW, Zhou J. (2006). Microarray-based analysis of subnanogram quantities of microbial community DNAs by using whole-community genome amplification. *Appl Environ Microbiol* **72:** 4931–4941.

Zhou J, Bruns MA, Tiedje JM. (1996). DNA recovery from soils of diverse composition. *Appl Environ Microbiol* **62:** 316–322.

Supplementary figures

Titles and legends to figures

Figure S1. Distribution of sequence-specific probes designed at their a) maximal sequence identities, b) maximal stretch lengths, and c) minimal free energy with their non-targets.

Figure S2. Distribution of group-specific probes designed at their a) minimal sequence identities, b) minimal stretch lengths, and c) maximal free energy with their group targets.

Supplementary tables

Table S1. Diversity indices of warming (W) and unwarming control (C) samples

	Richness (*S*)	Shannon Index (*H'*)	Simpson Index (*1/D*)	Simpson evenness (*E*)
W1	1423	6.737	435.966	0.306
W2	1579	6.798	456.400	0.289
W3	1551	6.765	436.953	0.282
W4	1617	6.832	451.361	0.279
W5	1568	6.819	457.541	0.292
W6	1786	7.117	719.494	0.403
C1	1346	6.643	394.301	0.293
C2	1365	6.651	400.941	0.294
C3	1361	6.705	423.286	0.311
C4	1378	6.693	410.477	0.298
C5	1390	6.683	402.695	0.290
C6	1390	6.709	396.636	0.285
P	0.004	0.023	0.121	0.565

Table S2. Diversity indices of oil-contaminated (BM) and non-contaminated control (C) samples

	Richness (*S*)	Shannon Index (*H'*)	Simpson Index (*1/D*)	Simpson evenness (*E*)
BM53	403	5.996	400.837	0.995
BM54	384	5.948	381.959	0.995
BM57	420	6.037	417.541	0.994
BM58	441	6.087	438.844	0.995
BM64	435	6.073	432.794	0.995
C1	354	5.867	352.507	0.996
C2	346	5.844	344.428	0.995
C3	349	5.853	347.597	0.996
C4	350	5.856	348.497	0.996
C5	364	5.895	362.427	0.996
P	0.002	0.002	0.002	0.01

150

Table S3. Diversity indices of caries-active (CA) and caries-free (CF) samples

	Richness (*S*)	Shannon Index (*H'*)	Simpson Index (*1/D*)	Simpson evenness (*E*)
CA14	1046	6.532	453.867	0.434
CA16	1039	6.516	424.189	0.408
CA17	1095	6.632	504.213	0.460
CA18	622	6.078	301.823	0.485
CA19	1099	6.647	542.899	0.494
CA20	942	6.477	434.097	0.461
CA21	952	6.487	442.921	0.465
CA22	867	6.395	415.637	0.479
CA23	830	6.312	357.750	0.431
CA24	740	6.221	328.138	0.443
CF1	1049	6.622	554.675	0.529
CF3	901	6.368	367.634	0.408
CF4	1012	6.549	507.204	0.501
CF5	943	6.430	428.975	0.455
CF6	1038	6.615	534.329	0.515
CF8	990	6.598	552.138	0.558
CF9	832	6.398	443.000	0.532
CF10	768	6.299	369.128	0.481
CF11	841	6.378	406.836	0.484
CF12	909	6.456	440.622	0.485
P	0.914	0.438	0.108	0.017