Supplemental Material

Title: A new functional marker gene of polycyclic aromatic hydrocarbons (PAHs) degrading bacteria: *pahE*

Chengyue Liang, Yong Huang and Hui Wang[#]

State Key Joint Laboratory on Environment Simulation and Pollution Control, School of Environment, Tsinghua University, Beijing, 100084, China

Running Head: a new functional marker gene of PAH degraders

[#] Address correspondence to Hui Wang, wanghui@tsinghua.edu.cn, State Key Joint Laboratory on Environment Simulation and Pollution Control, School of Environment, Tsinghua University, Beijing, 100084, China. Tel.: +86 010-62772137, fax: +86 01062771472.

Material and Methods

PAH degradation studies. To determine the degradation potential of the selected samples, degradation studies were initiated in 150-ml flasks using 10 g of ODS soil or BAS sludge or 10 ml ASE bacterium solution and 100 ml of minimal medium (1) spiked individually with 50 mg l^{-1} of naphthalene (Nah) or phenanthrene (Phn) or pyrene (Pyr). Flasks were shaken at 150 rev min^{-1} for up to 15 days at 37 ℃. At each time point, samples were removed and used for PAH extraction and HPLC analysis (2).

DNA extraction of the strains and environmental samples. All cultures and environmental samples used for examination of PCR assay specificity and screened for *pahE* and *pahAc* of PAH-degrading bacteria are listed together with sampling details and DNA extraction methods in Table S3. The pure cultures included reference PAH-degrading bacteria (those which could completely mineralize PAHs or grow on PAHs as the only carbon source), other aromatic hydrocarbon degraders (non-PAH degraders) and control bacteria. Environment samples were collected from the following sites: chronically crude oil contaminated soil (depth, 0 to 5 cm) collected from Shengli oilfield, Shandong, China, activated sludge from Bei Xiaohe urban sewage treatment plant, Beijing, China, and an enrichment culture for PAHs degrading bacteria inoculated (with a mixture of phenanthrene (100 mg l^{-1}) and pyrene (100 mg l^{-1}) was supplied as the only carbon and energy source) from the activated sludge of Qing He urban sewage treatment plant, Beijing, China. DNA extraction from bacteria pure cultures were performed using TIANamp Bacteria DNA kit (TIANGEN BIOTECH (Beijing) co., LTD) according to the manufacturer's protocol. The DNA extracts were finally suspended in 100 μl of TE buffer.

Genomic DNA of environmental samples was extracted by slightly modifying the FastDNA® SPIN Kit for Soil (MP Biomedicals, LLC) manufacturer's instructions.

Briefly, 0.3 g contaminated soil or activated sludge was added into the Lysing Matrix E Tube. For enrichment cultures, cells were pelleted at 4 ℃, 8000 *rpm* centrifugation for 10 min, directly resuspended in 978 μl of Sodium Phosphate Buffer in the kit and then added into the Lysing Matrix E Tube . In order to break cells fully and ensure the integrity of Genomic DNA simultaneously, the tubes were vortexed at a slower speed 5.0 for less time 30s but twice instead once. During the intervals between the two vortexes, tubes were cooled in the ice box to avoid superheat. In addition, the 12^{th} and 13^{th} steps in the official instructions for washing up impurities of salt were replicated three times instead of one time.

Primers design. Primer for PCR amplification of *pahE* genes were designed based on the *pahE* reference database by using Codehop (COnsensus-DEgenerate-Hybrid Oligonucleotide Primer) (3). Length limitations imposed by Illumina amplicon sequencing make it difficult to identify suitable degenerate primers for all *pahE*, so different primer sets were designed to target each of the 4 clades identified in the *pahE* phylogenetic tree (Table 1). The criteria for primer selection were as follows: 1) appropriate amplification length (300 ~ 400 bp) and similar regiospecificity; 2) low degeneracy and higher Tm; 3) GC content ＜60%. The selected primer set for each clade was manually modified to minimize self-complementarity, primer dimers formation and hairpin loop structures. The annealing temperature for each clade-specific primer pair was optimized using genomic DNA extracted from the selected reference bacterial strain for the clade by temperature gradient PCR. The optimum annealing temperature was considered to be the highest temperature at which a strong band was observed.

Specificity of the *pahE* and *pahAc* PCR assay. Primer specificity for each of the *pahE* gene clade and *pahAc* gene was validated via PCR. PCR products of each

reference organism at the determined optimum annealing temperature were checked for size and purity on 1.8% agarose gel and purified using a Gv-High-Efficiency Agarose Gel DNA Purification Kit (GEN-VIEW Scientific INC) according to the manufacturer's instructions. The purified products were cloned using the TA Cloning kit (TaKaRa), and inserts in transformants were checked using the primer M13-47 and RV-M targeting the vector pMD18-T sequence. Further, clones with correct inserts were selected for sequencing in a PRISM ready reaction dye terminator cycle sequencing kit according to the instructions of the manufacturer (TsingKe, Beijing, China). The obtained sequences were compared against the GenBank database using BLAST to confirm that amplicon sequences match target gene sequences.

78 **Table S1 Sequence information of reference PahAc and PahE used in this study.**

Strains	**abbreviation**	**PahAc ID**	**PahE ID**
Pseudomonas stutzeri AN10	Pseudo AN10	AFM32591	EPL61966
Pseudomonas bauzanensis W13Z2	Pseudo W13Z2	EZQ14080	EZQ14085
Pseudomonas stutzeri B1SMN1	Pseudo B1SMN1	EPL61971	AAB62713
Pseudomonas stutzeri 19SMN4	Pseudo 19SMN4	AHY45194	AHY45199
Pseudomonas aeruginosa PaK1	Pseudo PaK1	BAA12240	BAA12246
Pseudomonas stutzeri KOS6	Pseudo KOS6	EWC41262	AAO64280
Pseudomonas putida ND6	Pseudo ND6	AAP44288	AAP44192
Pseudomonas putida NCIB 9816-4-1	Pseudo 9816-4	1O7W:A	AFM32586
Pseudomonas putida BS202	Pseudo BS202	AAB62707	BAE92162
Pseudomonas putida OUS82	Pseudo OUS82	BAA20391	BAA20397
Pseudomonas putida G7	Pseudo G7	BAE92156	EWC41257
Ralstonia sp. U2	Ralsto U2	AAD12610	AAD12616
Acidovorax sp. JHL-9	Acido JHL-9	WP_026437498	WP_026437504
Burkholderia multivorans DDS 15A-1	Burk DDS15A-1	AIO74308	AIO75636
Comamonas testosteroni JC12	Coma JC12	KGH10180	KGH10186
Polaromonas naphthalenivorans CJ2	Polaro CJ2	AAZ93388	AAZ93394
Burkholderia sp. DNT	Burk DNT	AAB09766	*NO*
Burkholderia cepacia R34	Burk R34	AAL50021	*NO*
Comamonas sp. JS765	Coma JS765	2BMO:A	*NO*
Acidovorax sp. JS42	Acido JS42	AAB40383	*NO*
Pseudomonas stutzeri ZWLR2-1	Pseudo ZWLR2-1	ADQ90222	*NO*
Diaphorobacter sp. DS1	Diapho DS1	AGH09221	*NO*
Diaphorobacter sp. DS2	Diapho DS2	AGH09226	*NO*
Alteromonas sp. SN2	Altero SN2	AEF05078	AEF05081
Marinomonas sp. D104	Marino D104	ETI60159	ETI60157
Burkholderia sartisoli RP007	Burk RP007	AAD09872	AAD09869

Burkholderia sp. K24	Burk K24	KFX64020	KFX64016
Nevskia ramosa DSM11499	Nevskia 11499	WP_022978279	WP_028475726
Hydrocarboniphaga effusa AP103T	Hydroca AP103	EIT71332	EIT71336
Algiphilus aromaticivorans DG1253	Algiph 1253	WP_043770866	WP_043765396
Polycyclovorans algicola TG408	Polycy TG408	WP_029889175	WP_029889213
Burkholderia sp. DBT1	Burk DBT1	AAK62353	AER08042
Alcaligenes faecalis AFK2	Alcali AFK2	BAA76323	BAA76332
Delftia sp. Cs1-4	Delftia Cs1-4	AEF88772	AEF88788
Acidovorax sp. NA3	Acido NA3	ACG70971	ART51183
Burkholderia sp. Ch1-1	Burk Ch1-1	EIF28482	EIF28466
Cycloclasticus sp. P1	Cyclo P1	AFT68259	AFT67194
Cycloclasticus sp. PY97M	Cyclo PY97M	EPD12175	EPD12632
Cycloclasticus pugetii PS-1	Cyclo PS-1	WP_016391028	WP_015005964
Novosphingobium aromaticivorans F199	Novosphi F199	ABP64144	ABP64082
Sphingomonas sp. LH128	Sphingo LH128	EJU15002	EJU12841
Hyphomonas oceanitis SCH89	Hypho SCH89	KDA00874	KDA01194
Novosphingobium pentaromativorans US6-1	Novosphi US6-1	AIT82665	AIT82654
Sphingomonas paucimobilis EPA505	Sphingo EPA505	EZP70079	EZP70093
Sphingobium yanoikuyae B1	Sphingo B1	2GBW:A	ABM79813
Novosphingobium sp. PP1Y	Novosphi PP1Y	CCA92480	CCA93880
Sphingobium sp. KK22	Sphingo KK22	WP_025551213	WP_025548219
Pseudomonas sp. P51	Pseudo P51	AAC43632	*NO*
Pseudomonas putida F1	Pseudo F1	AAA26005	*NO*
Pseudomonas pseudoalcaligenes KF707	Pseudo KF707	AAA25743	*NO*
Burkholderia xenovorans LB400	Burk LB400	AAB63425	*NO*
Citreicella aestuarii Strain 357	Citrei 357	EIE49957	EIE49938
Oceanicola sp. MCTG156(1a)	Oceani MCTG156	WP_036563000	WP_036561753

Polymorphum gilvum SL003B-26A1T	Polymo 26A1	ADZ72512	ADZ72499
Thioclava dalianensis DLFJ1-1	Thiocl DLFJ11	KEP68756	KEP68746
Rhodovulum sp. NI22	Rhodovu NI22	KGB81046	KGB81035
Martelella sp. StrainAD-3	Martel AD-3	WP_024706839	WP_024706824
Paenibacillus sp. YK5	Paenibaci YK5	BAE53401	*NO*
Terrabacter sp. DBF63	Terra DBF63	BAB55886	*NO*
Paenibacillus sp. TSY30	Paenibaci TSY30	BAK48593	*NO*
Terrabacter sp. DBF63	Terra DBF63	BAC54156	*NO*
Arthrobacter keyseri 12B	Arthro 12B	AAK16534	*NO*
Mycobacterium vanbaalenii PYR-1	Myco PYR-1	Q44256	*NO*
Rhodococcus sp. TFB	Rhodococcus TFB	AAY57926	*NO*
Rhodococcus sp. DK17	Rhodococcus DK17	ABP48118	*NO*
Arthrobacter phenanthrenivorans Sphe3	Arthro Sphe3	ADX75094	ADX75098
Sciscionella marina DSM 45152	Scisci 45152	WP_020501050	WP_020501058
Amycolatopsis methanolica 239	Amyco 239	AIJ21952	AIJ21944
Rhodococcus opacus R7	Rhodo R7	AII11432	AII11501
Rhodococcus wratislaviensis IFP2016	Rhodo IFP2016	ELB89368	ELB89137
Rhodococcus imtechensis RKJ300	Rhodo RKJ300	EID78828	EID78824
Rhodococcus opacus TKN14	Rhodo TKN14	BAE53376	BAE53379
Rhodococcus opacus B4	Rhodo B4	BAH47212	BAH47216
Rhodococcus opacus SAO101	Rhodo SAO101	BAD02377	BAD02380
Rhodococcus sp. P200	Rhodo P200	AAR05114	AAR05117
Gordonia sp. CC-NAPH129-6	Gordo NAPH129	ACV96857	ACV96860
Rhodococcus sp. P400	Rhodo P400	AAR05106	AAR05109
Rhodococcus opacus M213	Rhodo M213	EKT84394	EKT84398
Rhodococcus sp. BCP1	Rhodo BCP1	KDE09919	KDE09923
Rhodococcus sp. I24	Rhodo I24	AAD25395	AAD25398
Nocardioides sp. KP7	Nocar KP7	BAA94708	BAA94711

Mycobacterium sp. JLS	Myco JLS	ABN97453 ABN98023	ABN97402
Mycobacterium aromaticivorans JS19b1T	Myco JS19b1	KDE96988 KDE97321	KDE97298
Mycobacterium sp. KMS	Myco KMS	ABL90893 ABL94808 ABL94821	ABL90865
Mycobacterium sp. MCS	Myco MCS	ABG07752 ABG11558	ABG07743
Mycobacterium intracellulare MIN 052511 1280	Myco MIN1280	ETZ36022 ETZ38624 ETZ38652	ETZ38640
Mycobacterium rhodesiae NBB3	Myco NBB3	AEV73692 AEV73714 AEV73736	AEV73685
Mycobacterium gilvum PYR-GCK	Myco PYR-GCK	ABP43029 ABP43050 ABP43151	ABP43160
Mycobacterium gilvum Spyr1	Myco Spyr1	ADT96888 ADT96910 ADT96931	ADT96879
Mycobacterium vanbaalenii PYR-1	Myco PYR-1	ABM11333 ABM11390 ABM11369	ABM11319
Comamonas testosteroni BR60	Coma BR60	Q44256	*NO*
Pseudomonas putida NMH102-2	Pseudo NMH102-2	Q05183	*NO*

79 *NO: indicate this strain has no PahE sequences.

Table S3 Environmental samples and bacteria cultures used in this study.

Sample type	Sample location	Sample name	Sample date	DNA extraction protocol	Sequencing technique	Reference (for strain or sample)
Strains						
Pseudomonas stutzeri 1-5				TIANamp Bacteria DNA kit	Sanger sequencing	This study
Delftia sp. Cs4-1				TIANamp Bacteria DNA kit	Sanger sequencing	(4)
Marinomonas profundimaris D104				TIANamp Bacteria DNA kit	Sanger sequencing	(5)
Sphingomonas sp. 1-1				TIANamp Bacteria DNA kit	Sanger sequencing	This study
Novosphingobium pentaromativorans US6-1				TIANamp Bacteria DNA kit	Sanger sequencing	(6)
Rhodococcus sp. B4				TIANamp Bacteria DNA kit	Sanger sequencing	(7)
Mycobacterium vanbaalenii PYR-				TIANamp Bacteria DNA kit	Sanger sequencing	(8)
Thioclava dalianensis DLFJ1-1				TIANamp Bacteria DNA kit	Sanger sequencing	(9)
Pseudomonas stutzeri ZWLR2-1				TIANamp Bacteria DNA kit	Sanger sequencing	(10)
Comamonas sp.CNB-1				TIANamp Bacteria DNA kit	Sanger sequencing	(11)
E.coli DH5α				TIANamp Bacteria DNA kit	Sanger sequencing	
Bacillus subtilis 168				TIANamp Bacteria DNA kit	Sanger sequencing	
Environmental Samples						
Chronically crude oil contaminated soil	Surface soil (depth, 0 to 5 cm), Shengli oilfield, Shandong, China	ODS	Sep, 2015	FastDNA® SPIN Kit	Illumina HiSeq2500	This study

Activated sludge	Bei Xiaohe urban sewage treatment plant, Beijing, China	BAS	Nov, 2015	FastDNA® SPIN Kit	Illumina HiSeq2500	This study
PAH degradation enrichment cultures	Inoculated the activated sludge (taken from Qing He urban sewage treatment plant, Beijing, China) with mixture of phenanthrene (100 mg l^{-1}) and pyrene (100 mg l^{-1}) as the only carbon and energy source.	ASE	Nov, 2015	FastDNA® SPIN Kit	Illumina HiSeq2500	This study

Table S4. Biodegradation of polycyclic aromatic hydrocarbons (PAHs) in studied samples. Residual PAHs analysis was carried out by liquid chromatography and extraction efficiency of all the three compounds was >90%. The values given here are the average of three experiments.

Samples	Substrate	Day 0 (mg l^{-1})	Day 7 (mg l^{-1})	Day 15 (mg l^{-1})
ODS	Nah	50	18.7	N.D
	Phn	50	25.6	4.6
	Pyr	50	37.3	10.4
ASE	Nah	50	9.7	N.D
	Phn	50	10.2	N.D
	Pyr	50	24.3	5.7
BAS	Nah	50	48.2	45.5
	Phn	50	48.9	47.6
	Pyr	50	49.3	48.9

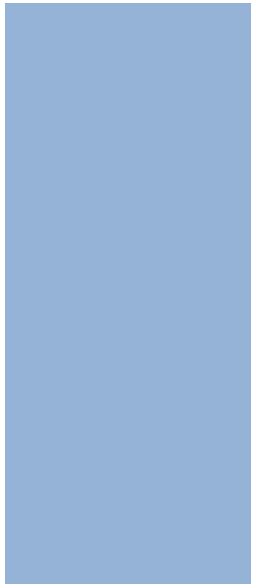
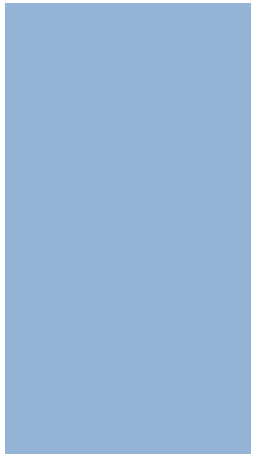
*Nah: naphthalene, Phn: phenanthrene, Pyr: pyrene.

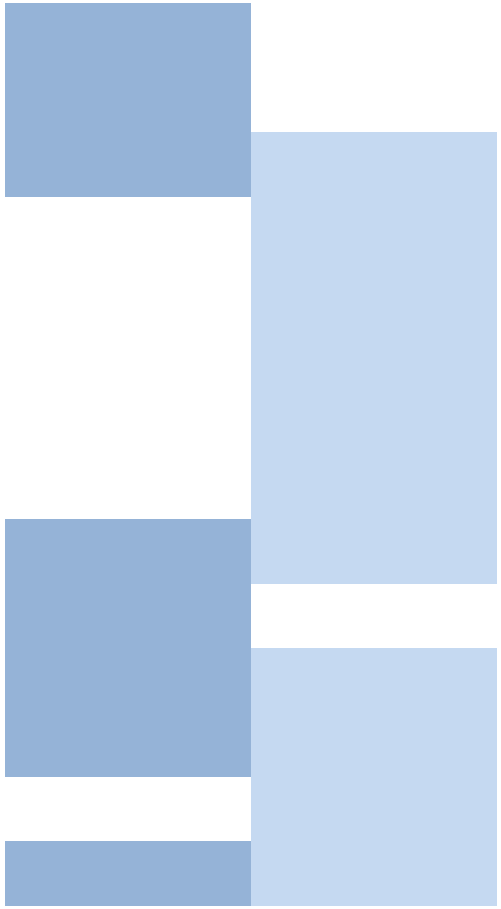
*N. D: not detected.

*ODS: a chronically crude oil contaminated soil, ASE: an enrichment culture of PAHs-degrading bacteria, BAS: an activated sludge of urban wastewater treatment plant.

Table S5 The coverage and specificity of the designed *pahE* primer tested *in silico*.

***pahE* reference sequences**	**Primer sets**			
Nucleotide ID|Strain|Protein ID	**pahE1F&pahE1R**	**pahE2F&pahE2R**	**pahE3F&pahE3R**	**pahE4F&pahE4R**
AB004059|Pseudo_OUS82|BAA20397				
AB237655|Pseudo_G7|BAE92162				
AF491307|Pseudo_9816-4|AAO64280				
AMCZ02000012.1|Pseudo_KOS6|EWC41257				
AMVM01000007|Pseudo_B1SMN1|EPL61966				
AF010471|Pseudo_BS202|AAB62713	N.D	N.D	N.D	N.D
CP003677|Pseudo_AN10|AFM32586				
CP007510|Pseudo_19SMN4|AHY45199				
D84146|Pseudo_PaK1|BAA12246				
CP008730|Burk_DDS15A-1|AIO75636				
AY208917|Pseudo_ND6|AAP44192				
JFHS01000025|Pseudo_W13Z2|EZQ14085				
AF036940|Ralsto_U2|AAD12616				
DQ167474|Polaro_CJ2|AAZ93394				

NZ_AUEX01000033|Acido_JHL-9|WP_026437504				
AWOU01000028|Coma_JC12|KGH10186				
AWTM01000090.1|Coma_DS1|KGH21324				
JH603161|Burk_Ch1-1|EIF28466				
JMIK01000009.1|Burk_K24|KFX64016				
AF061751|Burk_RP007|AAD09869				
AF404408|Burk_DBT1|AER08042				
AYOZ01000018.1|Marino_D104|ETI60157				
CP002339.1|Altero_SN2|AEF05081				
AJKJ01000121.1|Citrei_357|EIE49938	N.D	N.D	N.D	N.D
JQFU01000048.1|Rhodovu_NI22|KGB81035				
CP002568|Polymo_26A1|ADZ72499				
AKGD01000001.1|Hydroca_AP103|EIT71336				
JHEH01000024|Thiocl_DLFJ1-1|KEP68746				
NZ_ATVI01000011|Nevskia_11499|WP_084711626				
NZ_AYGY02000037|Martel_AD-3|WP_024706824				
NZ_JOMH01000001|Polycy_TG408|WP_029889213				

NZ_JPOG01000001|Algiph_1253|WP_043765396				
NZ_JQMY01000001|Oceani_MCTG156|WP_036561753				
ARYL01000034|Hypho_SCH89|KDA01194				
AB024945|Alcali_AFK2|BAA76332				
CP002735|Delftia_Cs1-4|AEF88788				
CP021361|Acidov_NA3|ART51183				
ALVC01000135|Sphingo_LH128|EJU12841				
BATN01000019.2|Sphingo_KK22|WP_0255482				
CP009294|Novosphi_US6-1|AIT82654				
EF151283.1|Sphingo_B1|ABM79813				
FR856862|Novosphi_PP1Y|CCA93880				
JFYY01000019|Sphingo_EPA505|EZP70093				
JNFC01000067|Sphingo_LC363|KGB52059				
CP000676|Novosphi_F199|ABP64082				
ASHL01000007|Cyclo_PY97M|EPD12632	N.D	N.D	N.D	N.D
CP003230|Cyclo_P1|AFT67194	N.D	N.D	N.D	N.D
NZ_ARVU01000001|Cyclo_PS-1|WP_015005964	N.D	N.D	N.D	N.D

GQ848233.3|Gordo_NAPH129|ACV96860

CP008952|Rhodo_R7|AII11501

CP009110|Amyco_239|AIJ21944

AVAE01000030|Rhodo_BCP1|KDE09923

ANIU01000578|Rhodo_IFP2016|ELB89137

AP011117|Rhodo_B4|BAH47216

AY392423.2|Rhodo_P400|AAR05109

AY392424.3|Rhodo_P200|AAR05117

CP002380|Arthro_Sphe3|ADX75098

AB110633|Rhodo_SAO101|BAD02380

AB206671|Rhodo_TKN14|BAE53379

AF121905|Rhodo_I24|AAD25398

AJJH01000096|Rhodo_RKJ300|EID78824

AJYC02000009|Rhodo_M213|EKT84398

NZ_KB905425|Scisci_45152|WP_020501058

AB031319|Nocar_KP7|BAA94711

CP003169|Myco_NBB3|AEV73685

CP000384|Myco_MCS|ABG07743				
CP000511|Myco_PYR-1|ABM11319				
CP000518|Myco_KMS|ABL90865				
CP000580|Myco_JLS|ABN97402				
CP000656|Myco_PYR-GCK|ABP43160				
CP002385|Myco_Spyr1|ADT96879				
JALN02000002|Myco_JS19b1|KDE97298				
JAON01000018|Myco_MIN1280|ETZ38640				
NZ_AUNF01000055|Thermotoga_sp._A7A|WP_029683321	N.D	N.D	N.D	N.D
NZ_JSBN01000007|Sphingomonas_sp._Ant_H11|WP_033863407	N.D	N.D	N.D	N.D
AEVT01000093|Vibrio_sinaloensis_DSM_21326|EGA69073	N.D	N.D	N.D	N.D
AMRV01000001|Alpha_proteobacterium_JLT2015|EMD84320	N.D	N.D	N.D	N.D
L17071|Rhizobium_meliloti_rhizopine|AAA26301	N.D	N.D	N.D	N.D
Targets in Total			68	
Total			73	

Notes: The colored cells indicated that the primers target the sequences; N.D: not targeted; Gray texted cells are non-PAH degrading homologs in the same protein family with *pahE*.

Table S6 The amplicons of *pahAc* or *pahE* from the environment samples for Illumina sequencing.

Environmental samples	*pahAc*	*pahE*			
		pahE1F&E1R	pahE2F&E2R	pahE3F&E3R	pahE4F&E4R
Contaminated-oil field soil (ODS)	√	√	√	√	-
Phenanthrene and Pyrene enrichment culture（ASE）	√	√	-	√	-
Activated sludge of Bei Xiaohe urban sewage treatment plant (BAS)	√	√	-	√	-

√: indicating that right and specific stretch were amplified;

-: indicating that no anything or wrong size fragments were amplified.

Table S7 Identities of the amino acid sequences deduced from the Illumina sequencing reads obtained from environmental samples to the reference sequences (complementary to Fig 5.).

	OTUs	Closest relative amino acid sequence accession number and bacteria	Amino acid Identity (%)
pahAc	OTU_75	BAB55886|Terra_DBF63|dibenzofuran	51.26
	OTU_146	ACG70971|Acido_NA3	50.91
	OTU_151	WP_024706839|Martel_AD-3	53.70
	OTU_152	ETZ38624|Myco_MIN1280	52.78
	OTU_153	ADZ72512|Polymo_26A1	53.70
	OTU_154	WP_024706839|Martel_AD-3	51.85
	OTU_158	BAK48593|Paenibaci_TSY30|Dox/Naph	51.92
	OTU_160	WP_036563000|Oceani_MCTG156	53.70
	OTU_162	BAB55886|Terra_DBF63|dibenzofuran	50.46
	OTU_163	ADT96910|Myco_Spyr1	95.37
	OTU_165	BAK48593|Paenibaci_TSY30|Dox/Naph	64.42
	OTU_167	BAK48593|Paenibaci_TSY30|Dox/Naph	67.31
	OTU_169	KGB81046|Rhodovu_NI22	50
	OTU_170	WP_024706839|Martel_AD-3	51.85
	OTU_171	WP_036563000|Oceani_MCTG156	58.33
	OTU_172	WP_029889175|Polycy_TG408	50.93
	OTU_182	KFX64020|Burk_K24	95.33
	OTU_183	EIT71332|Hydroca_AP103	50.46
	OTU_184	AFM32591|Pseudo_AN10	98.13
	OTU_185	BAK48593|Paenibaci_TSY30|Dox/Naph	51.92
	OTU_187	EIT71332|Hydroca_AP103	75.70
	OTU_188	CCA92480|Novosphi_PP1Y	85.05
	OTU_190	WP_026437498|Acido_JHL-9	98.13
	OTU_192	2GBW:A|Sphingo_B1	61.11

	OTU_199	AAK62353|Burk_DBT1	97.17
	OTU_202	ACG70971|Acido_NA3	70.76
	OTU_204	ACG70971|Acido_NA3	70.76
	OTU_206	EIT71332|Hydroca_AP103	64.49
pahE	E1_OTU_31	EJU12841|Sphingo_LH128	86.29
	E1_OTU_32	KGH10186|Coma_JC12	88.80
	E1_OTU_33	EZQ14085|Pseudo_W13Z2	98.40
	E1_OTU_34	KGH10186|Coma_JC12	76.00
	E1_OTU_35	EZQ14085|Pseudo_W13Z2	95.20
	E1_OTU_36	AER08042|Burk_DBT1	96.77
	E1_OTU_37	AAZ93394|Polaro_CJ2	88.00
	E1_OTU_38	ADZ72499|Polymor_26A1	87.20
	E1_OTU_40	AAD12616|Ralsto_U2	98.40
	E1_OTU_41	KFX64016|Burk_K24	95.16
	E1_OTU_42	KGB52059|Sphingo_LC363	83.87
	E1_OTU_179	EPL61966|Pseudo_B1SMN1	87.20
	E2_OTU_2	ADX75098|Arthro_Sphe3	60.98
	E2_OTU_5	KGB52059|Sphingo_LC363	88.53
	E2_OTU_6	EJU12841|Sphingo_LH128	88.53
	E2_OTU_7	EZQ14085|Pseudo_W13Z2	100
	E2_OTU_8	WP_024706824|Martel_AD3	72.13
	E2_OTU_9	WP_025548219|Sphingo_KK22	88.53
	E2_OTU_10	WP_025548219|Sphingo_KK22	100
	E2_OTU_11	EZP70093|Sphingo_EPA505	87.71
	E2_OTU_12	KDA01194|Hypho_SCH89	96.72
	E2_OTU_13	KGB52059|Sphingo_LC363	86.89
	E2_OTU_14	EZP70093|Sphingo_EPA505	85.25
	E2_OTU_15	ETI60157|Marino_D104	77.87

E2_OTU_19	KDA01194|Hypho_SCH89	95.54
E2_OTU_21	KGB52059|Sphingo_LC363	84.55
E3_OTU_544	KDE09923|Rhodo_BCP1	95.37
E3_OTU_606	AIJ21944|Amyco_239	100
E3_OTU_607	KDE09923|Rhodo_BCP1	64.49
E3_OTU_610	ADX75098|Arthro_Sphe3	97.17
E3_OTU_617	KDE09923|Rhodo_BCP1	64.49
E3_OTU_622	WP_020501058|Scisci_45152	99.07
E3_OTU_623	AIJ21944|Amyco_239	80.37
E3_OTU_803	AII11501|Rhodo_R7	55.34
E3_OTU_805	AII11501|Rhodo_R7	67.96
E3_OTU_825	AII11501|Rhodo_R7	57.28
E3_OTU_826	EIE49938|Citrei_357	51.46
E3_OTU_829	AII11501|Rhodo_R7	59.22
E3_OTU_834	AII11501|Rhodo_R7	66.02
E3_OTU_836	AII11501|Rhodo_R7	58.25
E3_OTU_842	ABP64082|Novosphi_F199	56.31

Table S8. The corresponding PahAc and PahE detected in the environmental samples and the detected higher resolution of *PahE* in *Pseudomonas* and *Sphingomonadaceae* (Complementary with Fig 5.).

The branches in PahAc tree (Fig 5)	***pahAc* Genotypes**		***pahE* Genotypes**	**The branches in PahE tree（Fig 5）**
AAZ93388|Polaromonas naphthalenivorans CJ2 OTU 190 AAD12610|Ralstonia sp. U2 WP_026437498|Acidovorax sp. JHL-9 AIO74308|Burkholderia multivorans DDS 15A-1 KGH10180|Comamonas testosteroni JC12	*Ralstonia* / *Comamonas* | NagAc	Corresponding to each other (indicated by green double sided arrow or green box in Fig 5)	*Ralstonia* / *Comamonas* | NagE	AAD12616|Ralstonia sp. U2 WP_026437504|Acido JHL9 E1 OTU 40 AIO75636|Burkholderia multivorans DDS 15A-1 KGH10186|Comamonas testosteroni JC12 AAZ93394|Polaromonas naphthalenivorans CJ2
OTU 182 AAD09872|Burkholderia sartisoli RP007 KFX64020|Burkholderia sp. K24	*Burkholderia* | PhnAc		*Burkholderia* | PhnE	E1 OTU 41 KFX64016|Burkholderia sp. K24 AAD09869|Burkholderia sartisoli RP007 E1 OTU 37 E1 OTU 32
				KGB81035|Rhodovulum sp. NI22

OTU 183 OTU 187 WP_022978279|Nevskia ramosa DSM11499 EIT71332|Hydrocarboniphaga effusa AP103T WP_043770866|Algiphilus aromaticivorans DG1253 OTU 206 OTU 202 OTU 204 WP_029889175|Polycyclovorans algicola TG408	New potential PahAc		New potential PahE	KEP68746|Thioclava dalianensis DLFJ1-1 WP_028475726|Nevskia ramosa DSM11499 WP_024706824|Martelella sp. StrainAD-3 ADZ72499|Polymorphum gilvum SL003B-26A1T E1 OTU 38 WP_036561753|Oceanicola sp. MCTG156(1a) WP_043765396|Algiphilus aromaticivorans DG1253 EIT71336|Hydrocarboniphaga effusa AP103T WP_029889213|Polycyclovorans algicola TG408 E3 OTU 826 EIE49938|Citreicella aestuarii Strain 357
AAK62353|Burkholderia sp. DBT1 OTU 199 BAA76323|Alcaligenes faecalis AFK2 AEF88772|Delftia sp. Cs1-4 ACG70971|Acidovorax sp. NA3	*Burkholderia* | PhnAc		*Burkholderia* | PhnE	E1 OTU 36 EIF28466|Burkholderia sp. Ch1-1 AER08042|Burkholderia sp. DBT1

EIF28482|Burkholderia sp. Ch1-1				
AFM32591|Pseudomonas stutzeri AN10 OTU 184 EZQ14080|Pseudomonas bauzanensis W13Z2 AHY45194|Pseudomonas stutzeri 19SMN4 EPL61971|Pseudomonas stutzeri B1SMN1 BAA12240|Pseudomonas aeruginosa PaK1 EWC41262|Pseudomonas stutzeri KOS6	*Pseudomonas* | NahAc/PahA3	Higher resolution of *pahE* were detected in *Pseudomonas* and *Sphingomonadaceae* degraders(indicated by the green arc ①,②, Fig 5)	*Pseudomonas* | NahE/PahE	AFM32586|Pseudomonas stutzeri AN10 E1 OTU 179 EZQ14085|Pseudomonas bauzanensis W13Z2 EWC41257|Pseudomonas stutzeri KOS6 EPL61966|Pseudomonas stutzeri B1SMN1 AHY45199|Pseudomonas stutzeri 19SMN4 E2 OTU 7 E1 OTU 33
AAB62707|Pseudomonas putida BS202 AAP44288|Pseudomonas putida ND6 1O7W:A|Pseudo NCIB 9816-4 BAE92156|Pseudomonas putida G7 BAA20391|Pseudomonas putida OUS82	*Pseudomonas* | NahAc/NahA3 /NdoB		*Pseudomonas* | NahE/PahE	AAP44192|Pseudomonas putida ND6 BAA12246|Pseudomonas aeruginosa PaK1 AAO64280|Pseudo NCIB 9816-4 AAB62713|Pseudomonas putida BS202 BAA20397|Pseudomonas putida OUS82 BAE92162|Pseudomonas putida G7 E1 OTU 35

ABP64144|Novosphingobium aromaticivorans F199 EJU15002|Sphingomonas sp. LH128 OTU 188 KDA00874|Hyphomonas oceanitis SCH89 AIT82665|Novosphingobium pentaromativorans US6-1 EZP70079|Sphingomonas paucimobilis EPA505 2GBW:A|Sphingobium yanoikuyae B1 CCA92480|Novosphingobium sp. PP1Y WP_025551213|Sphingo KK22	*Novosphingo* / *Sphingo* | BphA1		*Novosphingo* / *Sphingo* | PahE	E1 OTU 31 E2 OTU 6 ABP64082|Novosphingobium aromaticivorans F199 EJU12841|Sphingomonas sp. LH128 E2 OTU 9 E2 OTU 12 E2 OTU 19 KDA01194|Hyphomonas oceanitis SCH89 AIT82654|Novosphingobium pentaromativorans US6-1 EZP70093|Sphingomonas paucimobilis EPA505 E2 OTU 10 ABM79813|Sphingobium yanoikuyae B1 CCA93880|Novosphingobium sp. PP1Y WP_025548219|Sphingobium sp. KK22

Figures

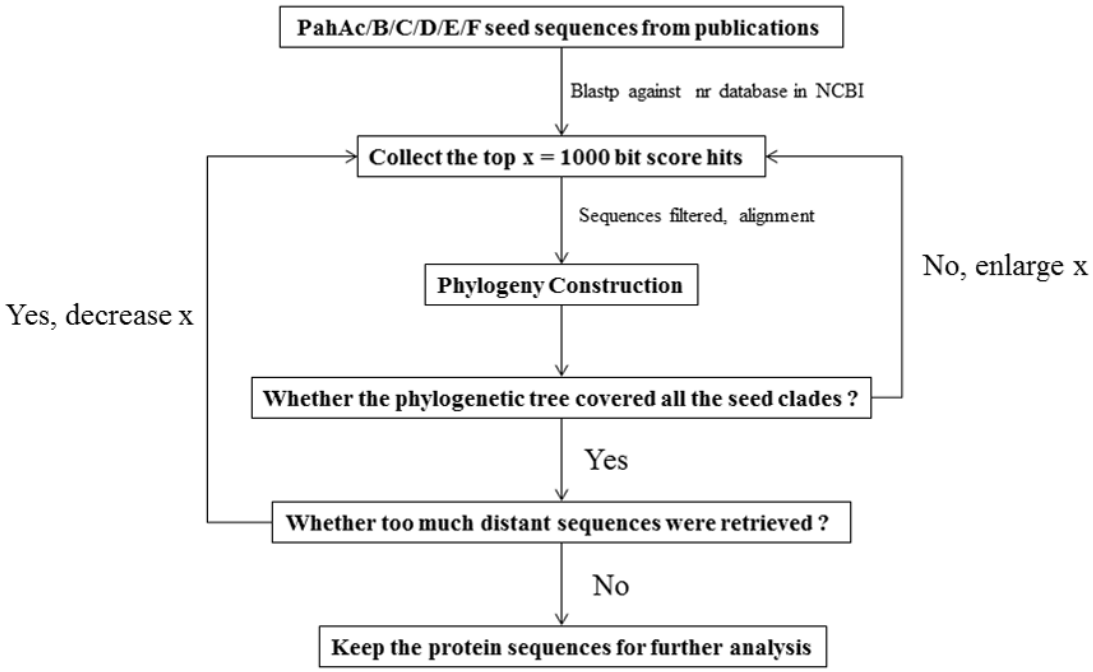


Fig S1. The schedule of collecting PahAc/B/C/D/E/F sequences from NCBI.

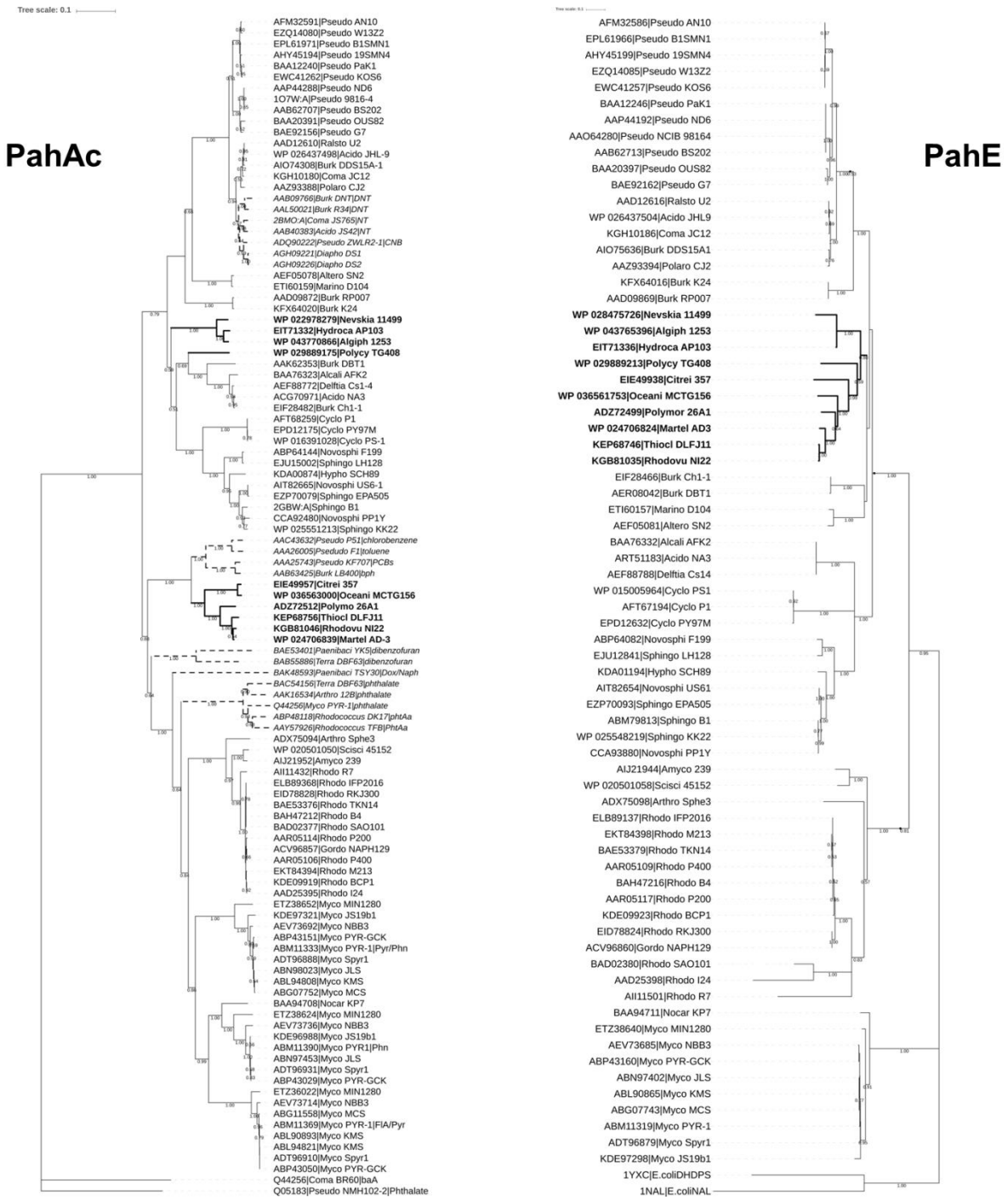


Fig S2. The details of the comparison of PahE- and PahAc-based phylogenies of PAHs degrading bacteria. Both panels show majority rule consensus trees based on neiborhood-joining method. Bootstrap support is indicated at individual branches, only value greater than 0.5 are shown. The bold text in both trees indicate members of the group of newly-defined potential PAH degraders, which also correspond to each other. The other aromatic-ring-hydroxylating dioxygenase clades (other-*ArhAc*, could not degrade PAH) in PahAc trees are indicated by dashed branches. The scale bar indicates 0.1 branch distance.

References

1. **Kumar, M., Lakshmi, C.V. and Khanna, S.** 2008. Biodegradation and biotransformation of endosulfan contaminated soil. Bioresour Technol **99:** 3116–3122.
2. U.S. EPA Method 3540C, Soxhlet extraction. 1996. <http://www.epa.gov/osw/hazard/testmethods/sw846/pdfs/3540c.pdf (14.03.14).
3. **Rose, T. M.; Henikoff, J. G.; Henikoff, S.** 2003. CODEHOP (COnsensus-DEgenerate hybrid oligonucleotide primer) PCR primer design. Nucleic Acids Res **31:** 3763-3766.
4. **Vacca, D.; Bleam, W.; Hickey, W.** 2005. Isolation of soil bacteria adapted to degrade humic acid-sorbed phenanthrene. Appl Environ Microbiol **71:** 3797-3805.
5. **Bai, X.; Lai, Q.; Dong, C.; Li, F.; Shao, Z.** 2014. *Marinomonas profundimaris* sp. nov., isolated from deep-sea sediment sample of the Arctic Ocean. Antonie Van Leeuwenhoek **106:** 449-455.
6. **Sohn, J. H.; Kwon, K. K.; Kang, J. H.; Jung, H. B.; Kim, S. J.** 2004. *Novosphingobium pentaromativorans* sp. nov., a high-molecular-mass polycyclic aromatic hydrocarbon-degrading bacterium isolated from estuarine sediment. Int J Syst Evol Microbiol **54:** 1483-1487.
7. **Grund, E.; Denecke, B.; Eichenlaub, R.** 1992. Naphthalene degradation via salicylate and gentisate by *Rhodococcus* sp. strain B4. Appl Environ Microbiol **58:** 1874-1877.
8. **Khan, A. A.; Kim, S. J.; Paine, D. D.; Cerniglia, C. E.** 2002. Classification of a polycyclic aromatic hydrocarbon-metabolizing bacterium, *Mycobacterium* sp. strain PYR-1, as *Mycobacterium vanbaalenii* sp. nov. Int J Syst Evol Microbiol **52:** 1997-2002.

9. **Zhang, R.; Lai, Q.; Wang, W.; Li, S.; Shao, Z.** 2013. *Thioclava dalianensis* sp. nov., isolated from surface seawater. Int J Syst Evol Microbiol **63:** 2981-2985.
10. **Liu, H.; Wang, S.-J.; Zhou, N.-Y.** 2005. A new isolate of *Pseudomonas stutzeri* that degrades 2-chloronitrobenzene. Biotechnol Lett **27:** 275-278.
11. **Wu, J.; Shen, X.; Zhou, Y.; Liu, S.** 2004. Characterization of p-chloronitrobenzene-degrading *Comamonas* sp. CNB1 and its Degradation of p-chloronitrobenzene. Acta Microbiol Sin **44:** 8-12.