

**Table S1.** The final estimates of  $S_{ACE}$  and  $S_{Chao1}$ , and the fraction of estimated phylotype richness values actually recovered in the libraries of reference soil, oilfield surface soil and gasfield surface soil.

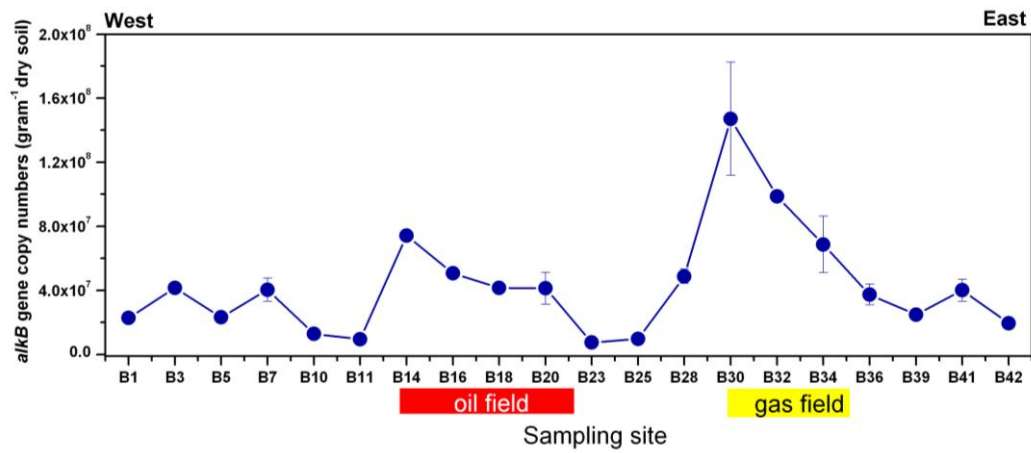
Data Summary	Reference soil	Oilfield soil	Gasfield soil
Number of clones in library	41	48	47
Number of phylotypes observed <sup>a</sup>	12	14	12
Predicted value of $S_{ACE}$	14.45	18.27	13.39
Predicted value of $S_{Chao1}$	13.05	16.77	12.26
Observed phylotypes / predicted $S_{ACE}$	0.84	0.83	0.87
Observed phylotypes / predicted $S_{Chao1}$	0.95	0.90	0.95

<sup>a</sup>The definition of phylotype is same as the definition of OTU (a similarity cutoff of 97%).

**Table S2.** *alkB* gene copy numbers pre gram of the 70-bp T-RF, 74-bp T-RF and 133-bp T-RF<sup>a</sup>

	70 bp ( <i>Alcanivorax</i> )		74 bp ( <i>Acinetobacter</i> & <i>Marinobacter</i> )		133 bp ( <i>Mycobacterium</i> and <i>Rhodococcus</i> )	
	<i>alkB</i> gene copy numbers (gram <sup>-1</sup> dry soil)	Standard deviation	<i>alkB</i> gene copy numbers (gram <sup>-1</sup> dry soil)	Standard deviation	<i>alkB</i> gene copy numbers (gram <sup>-1</sup> dry soil)	Standard deviation
BS1	5.60E+06	6.61E+05	8.19E+06	3.84E+05	1.53E+07	2.71E+06
BS2	3.57E+06	2.52E+05	6.00E+06	9.64E+05	1.86E+07	2.68E+06
BS3	4.67E+06	5.92E+05	6.05E+06	4.86E+05	1.33E+07	1.54E+06
OS1	1.89E+07	2.30E+06	2.69E+07	5.08E+06	8.86E+06	1.12E+06
OS2	1.37E+07	2.49E+06	2.99E+07	5.56E+06	1.05E+07	1.24E+06
OS3	1.59E+07	8.23E+05	2.57E+07	2.14E+05	8.40E+06	1.62E+06
GS1	1.10E+07	7.10E+05	2.16E+07	2.33E+06	1.39E+07	9.49E+05
GS2	1.11E+07	1.46E+06	2.29E+07	1.65E+06	1.24E+07	8.81E+03
GS3	1.25E+07	2.20E+06	1.76E+07	1.74E+06	1.33E+07	1.55E+06

<sup>a</sup>The absolute *alkB* gene copy numbers of each T-RF were calculated by multiplication of its relative abundance by the total *alkB* gene copy number from different samples.



**Figure S1** Biogeographical distribution of *alkB* gene copy numbers of soil samples collected from east-west direction survey line of Shaozhuang oil and gas field ( $n = 3$ ). B3, B14 and B34 sampling site are representing reference soil, oil field soil and gas field soil for T-RFLP and clone library analyses, respectively.