

A soil-bacterium compatibility model as a decision-making tool for soil bioremediation

Benjamin Horemans^{1*}, Philip Breugelmans¹, Wouter Saeys², Dirk Springael¹

¹ KU Leuven, Division of Soil and Water Management, Kasteelpark Arenberg 20, 3001 Heverlee, Belgium

² KU Leuven, Department of Biosystems, MeBioS, Kasteelpark Arenberg 30, 3001 Leuven, Belgium

Supporting Information

Table S1. Chemical Soil properties	p. S3
Table S2: Chemical constituents of pore water soil extract	p. S4
Table S3: Linear regressions of soil variables and CFU numbers at day 1, day 10 and day 20 and A_{Phe} for phenanthrene degradation	p. S5
Table S4: Probability estimate of LH128 survival in soils at day 20 using PLSDA	p. S7
Table S5: Prediction of phenanthrene degradation activity A_{Phe} of LH128 in different soils	p. S8
Figure S1: Soil texture triangle with USDA soil type classification	p. S9
Figure S2: Carbon and nitrogen content in the used soils	p. S10
Figure S3: Principal component analysis of the 20 soils using all soil variables (top) and all variables except the pore water analysis (bottom)	p. S11
Figure S4: Comparison between measured and predicted survival using PLSR of <i>Novosphingobium</i> sp. LH128 at day 1, day 10 and day 20 and of the phenanthrene degradation activity A_{Phe}	p. S12

Table S1: Chemical soil properties

Label	Exchangeable cations [cmolc/kg soil]						Oxalate extractable [mg/kg soil]			Aqua regia metals [mg/kg soil]					
	Ca	Mg	K	Na*	Mn	Al	Al	Fe	Mn	Zn	Cu	Ni	Pb	Cd	Cr
138	0.372	0.026	0.006	NM	0.001	2.811	564	1481	1	7.06	1.69	0.86	15.38	0.21	2.94
139	1.691	0.277	0.156	NM	0.081	2.008	561	1664	37	50.44	17.26	10.90	55.52	0.19	16.63
140	0.099	0.017	0.017	NM	0.000	0.966	199	492	1	7.59	2.41	1.08	19.16	0.21	3.60
141	7.808	0.678	0.148	NM	0.269	0.539	354	849	26	55.40	13.78	3.97	151.23	0.64	21.08
144	1.315	0.137	0.050	NM	0.012	0.223	1359	957	23	21.22	6.19	2.02	6.05	0.11	4.85
145	6.484	0.856	0.141	NM	0.394	0.061	530	496	116	37.22	31.02	69.07	16.43	0.25	87.60
147	1.558	0.179	0.035	NM	0.092	0.013	162	203	86	16.02	4.59	3.88	13.84	0.05	6.08
149	9.979	2.534	0.580	NM	0.119	0.006	524	984	82	53.03	21.11	19.23	12.99	0.08	55.90
151	14.463	2.385	0.243	NM	0.005	0.011	620	16163	174	98.62	21.78	37.49	54.49	0.97	88.22
152	6.160	0.331	0.404	NM	0.018	0.000	488	2197	272	53.90	21.67	11.09	27.99	0.40	24.82
153	20.720	0.320	0.380	NM	0.022	0.000	1102	3330	1377	402.90	21.09	47.98	94.79	3.25	63.27
155	33.067	0.820	0.436	NM	0.010	0.014	347	460	164	50.80	33.64	82.85	11.88	0.47	52.92
156	18.013	0.302	0.242	NM	0.004	0.008	276	2993	325	80.40	17.60	19.92	24.58	0.45	31.89
157	10.662	0.695	0.496	NM	0.002	0.021	460	1190	179	190.52	88.00	21.08	155.72	0.47	26.70
158	18.996	0.261	0.825	NM	0.060	0.009	369	1646	1043	251.40	30.98	108.75	73.29	3.08	53.52
159	14.460	0.405	0.125	NM	0.001	0.012	307	103	34	27.45	7.32	9.52	12.00	0.11	11.50
277	1.309	0.061	0.081	0.009	0.011	0.239	1012	2188	88	6.00	6.00	1.00	10.00	NM	4.00
278	4.313	0.376	0.311	0.026	0.040	0.013	1061	2758	520	15.00	13.00	3.00	11.00	NM	7.00
283	11.519	1.122	0.434	0.016	0.001	0.001	583	797	259	59.00	45.00	18.00	158.00	NM	22.00
324	1.800	0.700	0.200	NM	NM	NM	1200	1000	60	18.00	14.00	3.00	34.00	0.28	NM

NM: not measured, *not included in multivariate analysis

Table S2: Chemical constituents of pore water soil extracts

Label	Macronutrients [mg/L]										Micronutrients [µg/L]								
	Ortho-	Cl	SO ₄	NO ₃	Na	Mg	K	Ca	S	Al	Cd	Co	Cr	Cu	Fe	Mn	Ni	Pb	Zn
138	0.9	31.8	64.8	55.1	16.7	4.2	9.7	13.0	52.0	NM	5.8	4.9	9.6	12.3	6154.4	101.8	33.8	26.2	289.5
139	2.7	54.3	108.9	34.5	7.0	8.6	11.7	18.8	21.9	10.8	1.4	12.2	8.2	33.0	2835.4	1659.1	30.7	59.7	273.1
140	1.0	17.6	46.2	10.9	4.4	2.6	8.1	4.1	22.3	15.1	4.8	3.3	9.9	13.5	3234.3	38.2	36.1	29.4	247.1
141	12.0	21.6	108.0	15.4	1.7	2.2	2.7	7.7	6.1	0.7	0.3	0.8	1.6	4.7	261.8	348.4	6.1	7.0	52.1
144	1.1	15.3	19.9	127.0	5.6	5.5	8.9	17.4	6.9	1.2	0.1	1.9	0.8	14.7	134.9	262.4	14.4	3.0	38.3
145	NM	11.4	44.3	10.4	1.7	0.7	0.6	4.6	3.3	0.0	0.0	1.7	0.0	0.7	0.0	182.2	5.7	0.0	5.9
147	2.4	33.5	49.8	2.1	11.3	14.1	6.1	55.6	14.7	2.9	1.0	38.4	6.5	77.7	1039.7	4595.9	46.0	14.8	82.2
149	1.1	32.8	34.3	4.3	8.7	1.9	3.2	5.6	3.7	0.4	0.0	0.6	0.4	4.8	305.9	49.6	8.2	0.5	13.2
151	1.9	15.0	144.1	857.3	2.4	11.5	2.5	38.1	10.9	0.0	0.0	0.9	0.5	7.5	35.6	53.1	10.6	0.0	4.7
152	2.7	165.4	136.3	1315.4	7.3	11.6	24.1	125.4	14.9	0.0	0.3	0.9	0.3	26.4	4.6	890.4	6.4	0.0	5.7
153	NM	9.3	38.3	69.1	0.1	0.1	0.1	4.9	0.7	0.0	0.0	0.0	0.0	2.0	0.0	0.0	1.4	0.0	0.6
155	0.9	73.1	47.7	8.3	6.4	2.0	1.1	34.6	4.3	0.0	0.0	0.1	0.0	5.8	7.3	11.4	7.7	0.0	2.6
156	3.1	11.5	NM	47.7	2.8	1.1	1.6	33.3	17.6	0.0	0.0	0.0	0.0	15.0	0.7	17.4	6.2	0.0	1.2
157	3.0	89.8	94.9	351.1	5.2	2.0	2.8	15.3	3.3	0.0	0.0	0.0	0.0	27.0	0.0	10.2	3.8	0.3	0.9
158	0.8	20.6	51.1	38.8	0.5	0.2	1.9	7.6	1.5	0.0	0.0	0.0	0.0	4.2	0.0	0.0	4.5	0.0	1.3
159	0.2	6.6	16.2	80.5	1.8	2.6	2.5	45.1	4.3	0.0	0.0	1.0	0.0	12.1	3.3	5.5	12.0	0.0	3.6
277	0.2	17.8	16.3	59.9	9.6	5.9	20.9	68.9	19.9	2.0	0.7	1.1	4.8	47.5	452.7	571.6	111.2	13.4	88.7
278	3.6	226.1	178.3	1457.3	19.9	56.7	117.4	545.4	78.3	0.2	1.4	0.4	1.9	45.6	73.1	3678.6	25.8	3.2	46.3
283	0.5	5.0	9.2	41.1	0.4	0.6	0.6	5.1	0.5	0.0	0.00	0.0	0.1	4.7	0.0	0.0	1.6	0.0	0.3
324	1.6	10.3	126.3	845.1	3.4	51.5	35.2	114.5	38.8	0.9	2.0	1.6	3.9	52.5	292.6	1855.3	12.6	7.0	113.3

NM: not measured

Table S3: Linear regressions of individual soil variables as independent variables and CFU numbers at day 1, day 10 and day 20 and A_{Phe} for phenanthrene degradation as dependent variable. For each soil variable, R², correlation coefficient corr and ANOVA p-value of F test for the slope coefficient is provided for the linear regressions with dependent variables.

	CFU day 1			CFU day 10			CFU day 20			A _{Phe}		
	R ²	corr [†]	p	R ²	corr [†]	p	R ²	corr [†]	p	R ²	corr [†]	p
pH	0.62	0.78	0.00	0.25	0.50	0.05	0.67	0.82	0.00	0.24	0.48	0.03
total C (%)	0.03	-0.18	0.47	0.02	-0.15	0.63	0.02	-0.14	0.51	0.01	-0.09	0.80
Organic C (%)	0.17	-0.41	0.07	0.10	-0.32	0.21	0.15	-0.38	0.10	0.04	-0.21	0.44
CaCO ₃ (%)	0.14	0.38	0.11	0.07	0.26	0.24	0.15	0.39	0.13	0.03	0.17	0.44
total N (%)	0.00	-0.06	0.82	0.00	-0.03	0.88	0.00	-0.04	0.88	0.01	0.09	0.57
Corg/N (-)	0.49	-0.70	0.00	0.26	-0.51	0.06	0.40	-0.63	0.00	0.31	-0.56	0.02
Sand (%)	0.42	-0.65	0.00	0.17	-0.41	0.09	0.33	-0.58	0.01	0.16	-0.40	0.07
Silt (%)	0.32	0.56	0.01	0.09	0.29	0.24	0.27	0.52	0.01	0.05	0.22	0.33
Clay (%)	0.30	0.54	0.01	0.17	0.41	0.07	0.22	0.46	0.05	0.24	0.49	0.03
WHC (%)	0.07	0.26	0.25	0.03	0.18	0.46	0.05	0.22	0.35	0.07	0.26	0.22
CEC (cmolc/kg)	0.39	0.62	0.00	0.17	0.42	0.08	0.32	0.57	0.01	0.22	0.47	0.02
Ca (cmolc/kg)	0.43	0.65	0.00	0.19	0.43	0.06	0.38	0.62	0.00	0.21	0.46	0.02
Mg (cmolc/kg)	0.17	0.41	0.07	0.13	0.35	0.21	0.10	0.31	0.21	0.19	0.43	0.04
K (cmolc/kg)	0.30	0.55	0.01	0.14	0.38	0.12	0.37	0.61	0.00	0.25	0.50	0.04
Na* (cmolc/kg)	na	na	0.10	na	na	0.09	na	na	0.15	na	na	0.11
Mn (cmolc/kg)	0.04	-0.20	0.43	0.00	-0.04	0.82	0.14	-0.38	0.11	0.00	0.06	0.88
Al (cmolc/kg)	0.42	-0.65	0.00	0.27	-0.52	0.04	0.38	-0.61	0.01	0.40	-0.63	0.00
Al (mg/kg)	0.04	-0.21	0.37	0.12	-0.35	0.11	0.06	-0.25	0.27	0.05	-0.22	0.34
Fe (mg/kg)	0.11	0.34	0.15	0.09	0.31	0.29	0.09	0.30	0.18	0.08	0.29	0.15
Mn (mg/kg)	0.20	0.45	0.04	0.03	0.17	0.55	0.24	0.49	0.03	0.07	0.27	0.30
Zn (mg/kg)	0.12	0.35	0.12	0.00	0.00	0.80	0.19	0.43	0.04	0.02	0.13	0.58
Cu (mg/kg)	0.09	0.31	0.18	0.00	0.04	0.79	0.14	0.37	0.08	0.02	0.16	0.58
Ni (mg/kg)	0.15	0.38	0.08	0.12	0.35	0.14	0.11	0.33	0.16	0.19	0.43	0.05
Pb (mg/kg)	0.00	0.01	0.99	0.09	-0.30	0.07	0.02	0.13	0.52	0.02	-0.14	0.49
Cd (mg/kg)	0.09	0.30	0.22	0.00	0.05	0.93	0.14	0.38	0.12	0.03	0.18	0.52
Cr (mg/kg)	0.23	0.48	0.03	0.13	0.36	0.22	0.09	0.31	0.21	0.19	0.43	0.05
OrthoP (mg/kg)	0.02	0.13	0.61	0.00	0.06	0.92	0.01	0.12	0.54	0.01	0.07	0.61
Inorganic P (mg/L)	0.02	-0.15	0.53	0.01	-0.12	0.64	0.02	-0.16	0.59	0.00	0.02	0.90
Cl (mg/L)	0.12	0.35	0.14	0.14	0.38	0.08	0.12	0.35	0.14	0.15	0.38	0.15
SO ₄ (mg/L)	0.02	0.14	0.57	0.05	0.23	0.32	0.02	0.13	0.53	0.04	0.19	0.40
NO ₃ (mg/L)	0.11	0.33	0.17	0.12	0.35	0.15	0.09	0.30	0.20	0.08	0.28	0.26
Na (mg/L)	0.00	-0.06	0.78	0.02	0.14	0.40	0.00	-0.03	0.86	0.02	0.13	0.69
Mg (mg/L)	0.00	0.01	1.00	0.02	0.13	0.51	0.00	0.01	1.00	0.01	0.08	0.75
K (mg/L)	0.01	0.11	0.67	0.03	0.19	0.36	0.01	0.08	0.77	0.03	0.16	0.63
Ca (mg/L)	0.08	0.28	0.25	0.12	0.34	0.11	0.06	0.24	0.32	0.09	0.31	0.26
S (mg/L)	0.02	-0.15	0.51	0.00	0.00	0.84	0.03	-0.17	0.48	0.01	-0.12	0.56
Al (mg/L)	0.29	-0.54	0.02	0.16	-0.40	0.14	0.24	-0.49	0.04	0.23	-0.48	0.04
Cd (μg/L)	0.28	-0.53	0.02	0.15	-0.39	0.14	0.24	-0.49	0.03	0.27	-0.52	0.02
Co (μg/L)	0.01	-0.10	0.60	0.02	0.15	0.48	0.00	0.03	0.89	0.03	0.17	0.47
Cr (μg/L)	0.43	-0.65	0.00	0.18	-0.43	0.10	0.32	-0.57	0.01	0.27	-0.52	0.02
Cu (μg/L)	0.03	-0.17	0.41	0.00	0.05	0.72	0.00	-0.03	0.93	0.00	0.01	1.00
Fe (μg/L)	0.28	-0.53	0.01	0.16	-0.40	0.12	0.24	-0.49	0.03	0.26	-0.51	0.02
Mn (μg/L)	0.00	0.03	0.98	0.07	0.26	0.21	0.01	0.11	0.66	0.08	0.27	0.29
Ni (μg/L)	0.19	-0.44	0.05	0.06	-0.24	0.43	0.14	-0.38	0.10	0.08	-0.29	0.20
Pb (μg/L)	0.37	-0.60	0.00	0.20	-0.44	0.08	0.30	-0.54	0.01	0.27	-0.52	0.02
Zn (μg/L)	0.49	-0.70	0.00	0.27	-0.52	0.04	0.41	-0.64	0.00	0.39	-0.62	0.00
As (μg/L)	0.06	-0.25	0.26	0.00	-0.04	0.97	0.03	-0.17	0.48	0.04	-0.21	0.41

- † correlation coefficient corr indicates the negative (dark red) or positive (dark green) relation between the two properties and the extent of this relation.
- * P value of F test for slope coefficient is lower than 0.05
- n.a. data for exchangeable N.a. was absent for many soil samples and as such omitted from regression analysis

Table S4: Probability estimate of LH128 survival in soils at day 20 using PLSDA based on different sets of soil variables. For each soil, the PLSDA model gives a value between 0 and 1 with a value of 0 for no survival and a value of 1 for survival.

Soil N°	Location	Survival	All*	Day 1		Day 20	
				iPLS†	iPLS	iPLS†	iPLS
				basic††		basic††	
138	Gudow (D)	No	0.0	0.0	0.0	0.0	0.0
139	Nottingham (UK)	No	0.0	0.0	0.0	0.0	0.0
140	Houthalen (B)	No	0.0	0.0	0.0	0.0	0.0
141	Rhydtalog (UK)	No	0.2	0.0	0.2	0.0	0.0
144	Kövlinge I (SU)	No	0.2	0.0	0.0	0.0	0.1
145	Souli (GR)	Yes	1.0	1.0	0.9	1.0	1.0
147	Montpellier (F)	Yes	0.1	0.0	0.1	0.0	0.2
149	Aluminusa (I)	Yes	1.0	1.0	1.0	1.0	1.0
151	Woburn (UK)	Yes	1.0	1.0	1.0	1.0	1.0
152	Leuven (B)	Yes	1.0	1.0	1.0	1.0	1.0
153	Vault de Lugny (F)	Yes	1.0	1.0	1.0	1.0	1.0
155	Souli (GR)	Yes	1.0	1.0	1.0	1.0	1.0
156	Marknesse (NL)	Yes	1.0	1.0	1.0	1.0	1.0
157	Barcelona (ES)	Yes	1.0	1.0	1.0	1.0	1.0
158	Brecy (F)	Yes	1.0	1.0	1.0	1.0	1.0
159	Guadalajara (ES)	Yes	1.0	1.0	0.9	1.0	1.0
277	Jyndevad (DK)	No	0.0	0.0	0.0	0.0	0.1
278	Borris (DK)	Yes	0.5	0.0	0.5	0.4	0.8
283	Cordoba (S)	Yes	1.0	1.0	1.0	1.0	1.0
324	Zwijnaarde (B)	No	0.1	0.0	0.3	0.0	0.2

* PLSDA model based on all variables

† PLSDA model based on the variables selected by iPLS

†† PLSDA model based on the variables selected by iPLS without including the pore water chemistry properties.

Table S5: Prediction of phenanthrene degradation activity A_{Phe} of LH128 in different soils and its ranking from low to high A_{Phe} (high to low activity). Soils showing no survival were excluded from analysis.

		Measured		PLS predicted		iPLS predicted	
		Soil n°	A_{Phe} [%]	Soil n°	A_{Phe} [%]	Soil n°	A_{Phe} [%]
Survival, high activity		151	68	152	66	151	69
		155	66	155	58	155	65
		149	66	145	57	149	63
		158	63	153	57	158	60
Survival, Medium to limited activity		152	51	151	53	152	51
		156	38	149	50	156	40
		145	33	158	46	145	35
		157	26	157	46	153	27
		153	24	159	39	157	27
		283	23	156	37	159	24
		159	23	283	19	283	22
No or limited survival, no activity		138	-5	138	NA	138	NA
		139	0	139	NA	139	NA
		140	-2	140	NA	140	NA
		141	23	141	NA	141	NA
		144	27	144	NA	144	NA
		277	4	277	NA	277	NA
		324	0	324	NA	324	NA
Outlier		278	70	278	NA	278	NA
		147	73	147	NA	147	NA

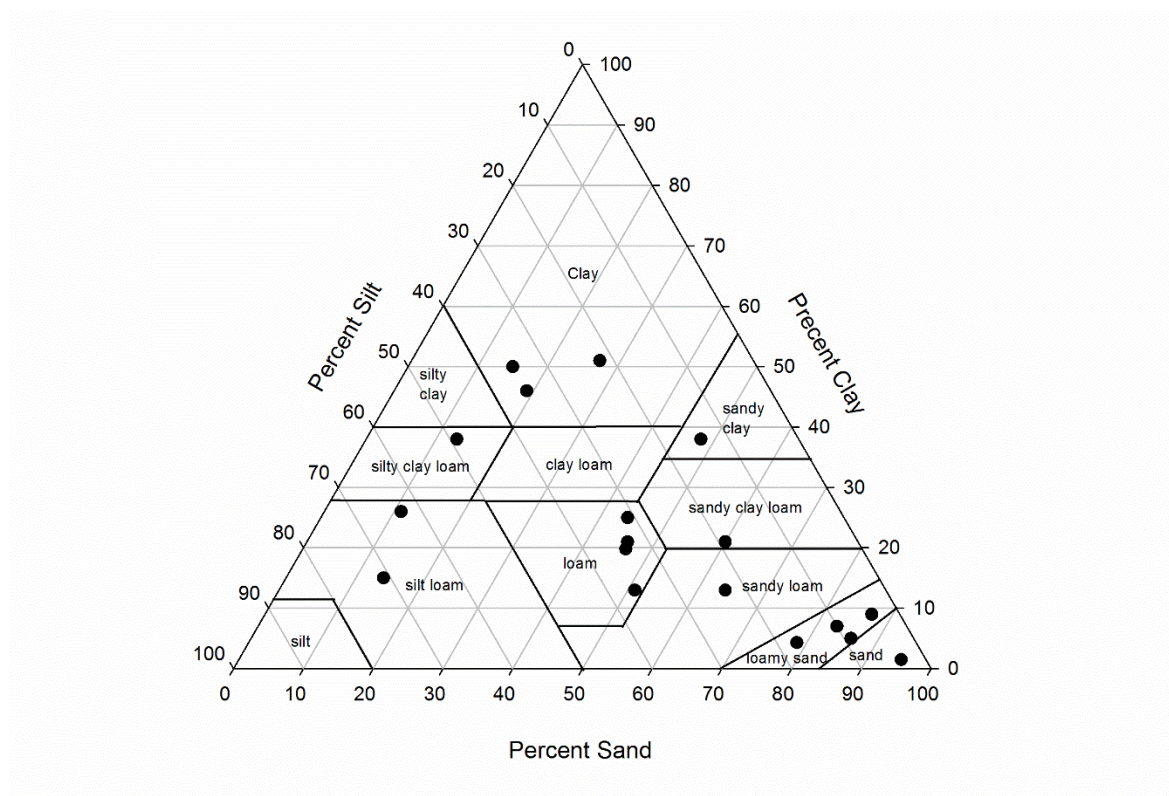


Figure S1: Soil texture triangle with USDA soil type classification of the 20 soil samples used in this study

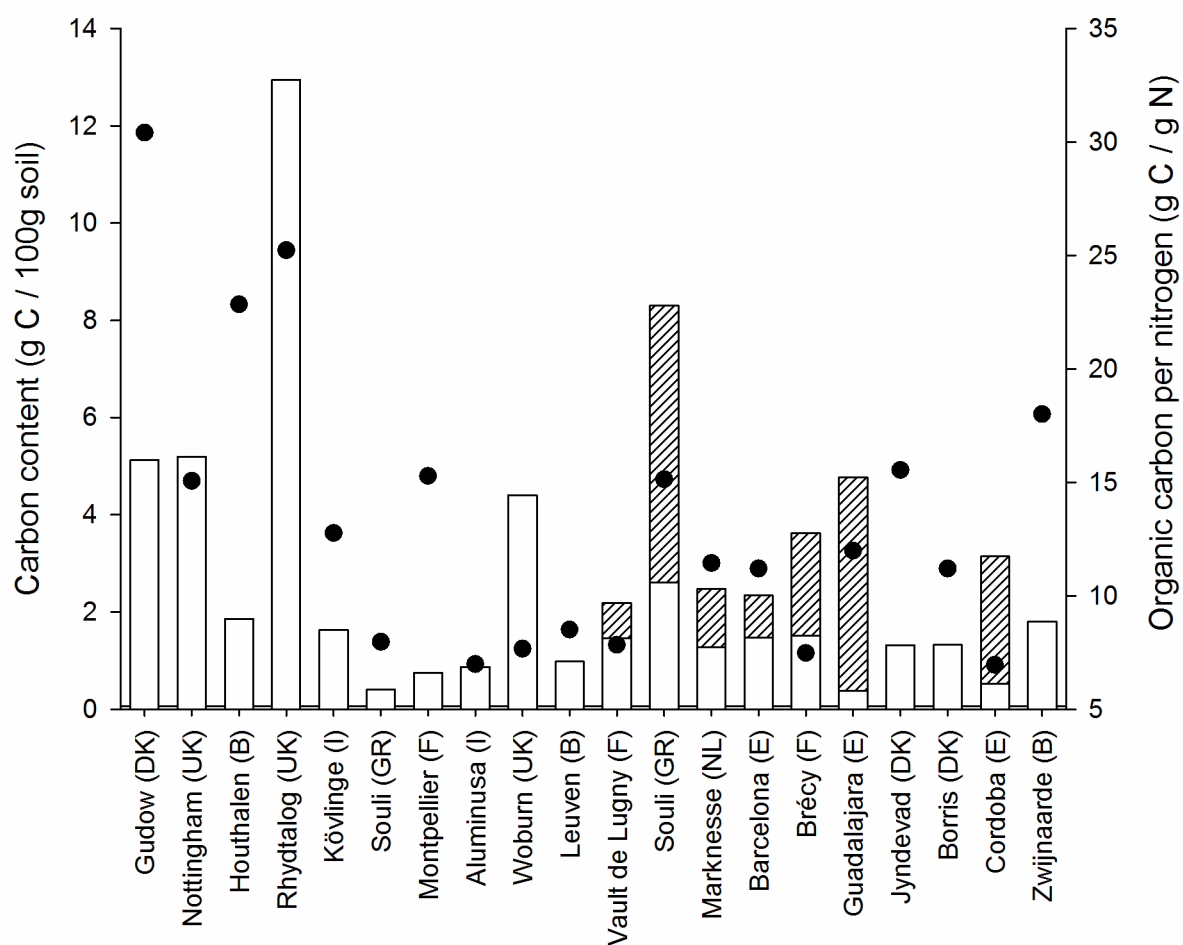


Figure S2: Carbon and nitrogen content in the used soils. Organic (white bar) and inorganic (grey bar) carbon content is indicated on the left axis. The organic carbon/nitrogen ratio (●) is indicated on the right axis.

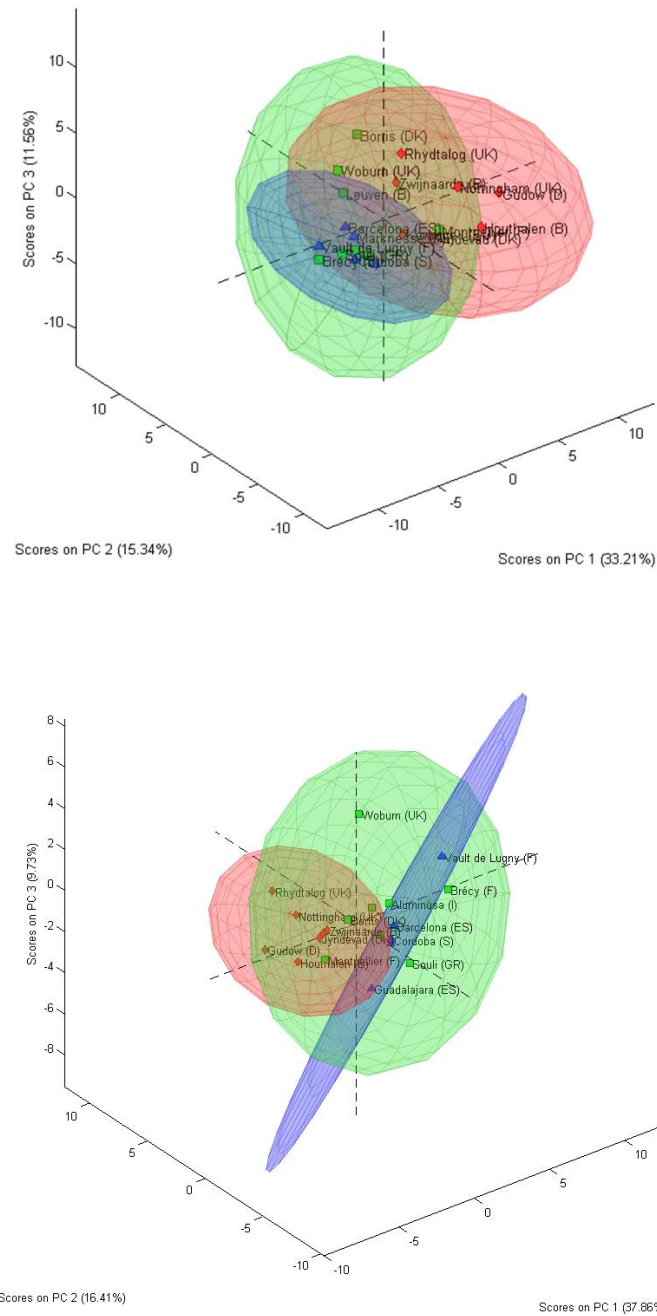
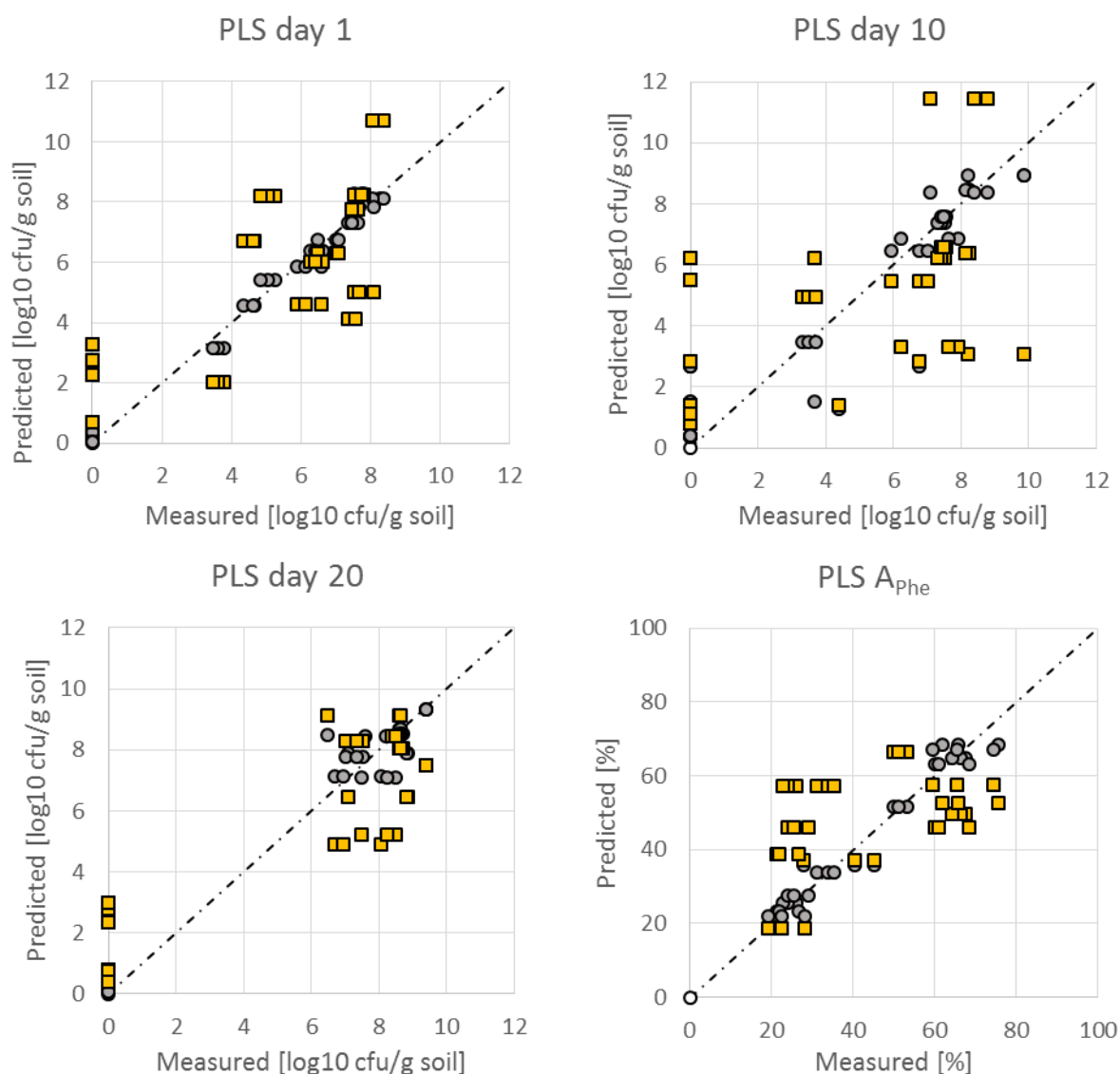


Figure S3: Principal component analysis of the 20 soils using all soil variables (top) and all variables except the pore water analysis (bottom). PC1 versus PC2 versus PC3 is shown. Ten principal components captured 97% of all variance and were selected based on the increase in captured variance by adding an extra PC (>2%). Soils with no survival of LH128 (red), with impaired survival (blue) and survival (green) are indicated and confidence ellipse at 95% is indicated for each cluster.



	Day 1	Day 10	Day 20	A _{Phe}
Variance covered (%)	71	68	70	68
<i>RMSE</i>				
Calibration	0.30	1.07	0.51	4.07
Cross-validation	2.12	2.93	2.06	17.98
<i>R</i> ²				
Calibration	0.99	0.92	0.98	0.95
Cross-validation	0.67	0.43	0.76	0.18

Figure S4: PLSR predicted survival at day 1, day 10 and day 20 and phenanthrene degrading activity A_{Phe} in function of the measured survival/activity of *Novosphingobium* sp. LH128. Circles represent calibration data and square symbols cross-validation data. Full line is the 1:1 relation between predicted and measured values in case of a perfect model. PLSR models were based on 47 soil variables. Soil data of 18 soils and of 11 soils was used for the PLSR models of CFU at day 1, 10 and 20

and the A_{Phe} , respectively. The table shows the covered variance within the soil data and the RMSE and R^2 for the calibration and cross-validation.