SUPPLEMENTARY MATERIAL

Quantifying the importance of the rare biosphere for microbial community adaptation to organic pollutants in a freshwater ecosystem

by

Yuanqi Wang, Janet K. Hatt, Despina Tsementzi, Luis M. Rodriguez-R, Carlos A. Ruiz-Pérez, Michael Weigand, Heidi Kizer, Gina Maresca, Raj Krishnan, Rachel Poretsky, Jim C. Spain, and Konstantinos T. Konstantinidis

Supplementary Figure S1.

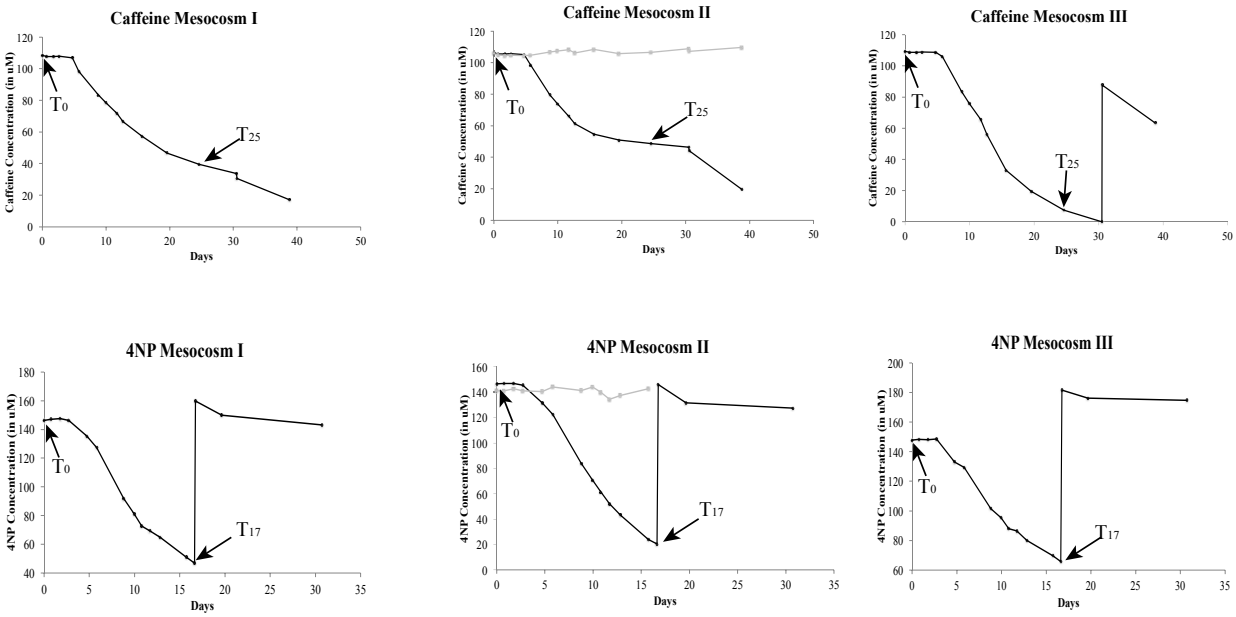

Figure S1. Degradation profiles of caffeine and 4-nitrophenol (4-NP) mesocosms. The black lines indicate caffeine and 4-NP concentrations over time, and the gray line represents the abiotic (negative) control (caffeine and 4-NP in sterilized lake water). Black arrows indicate the sampling time points at which DNA was extracted. Isolation and DNA sequencing were attempted only on T_{17} and T_{25} samples.

Supplementary Figure S2.

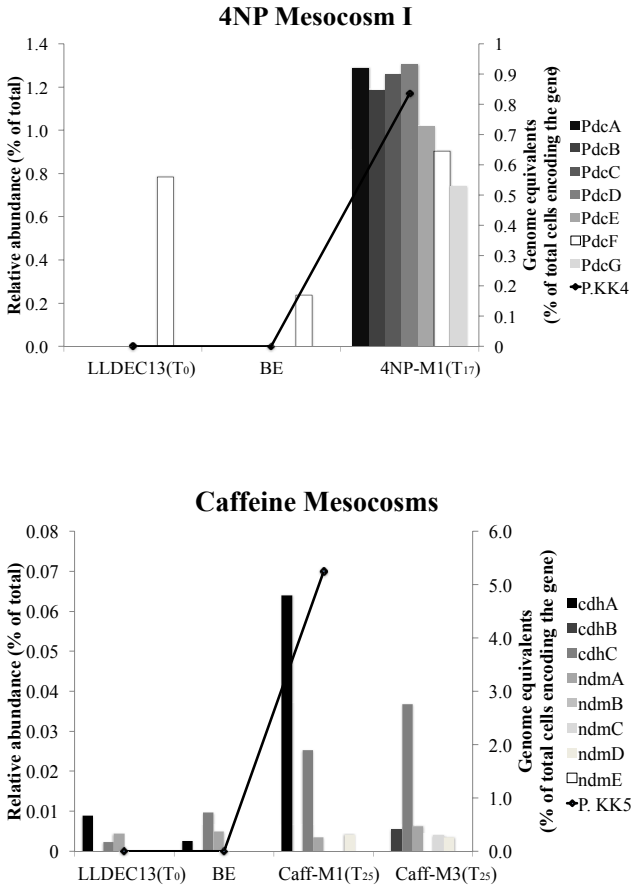

Figure S2. Relative abundance of 4-NP and caffeine degraders and degradation genes in the mesocosm they originated from. Note that both isolates (*Pseudomonas* sp. strain KK4 for 4-NP and *Pelomonas* sp. strain KK5 for caffeine) were rare at LLDEC13 (T_0), but became abundant at the last sampling time point, i.e., 4NP-M1 (T_{17}) and Caff-M1 (T_{25}), respectively (lines; y-axis), similar to the 2,4-D degrader profiles shown in Figure 2. The abundance of 4-NP and caffeine biodegradation genes were measured by genome equivalents (% of total cells encoding the gene) at each time point (bars; secondary Y-axis). BE: Bottle-Effect metagenome (control).

Supplementary Figure S3.

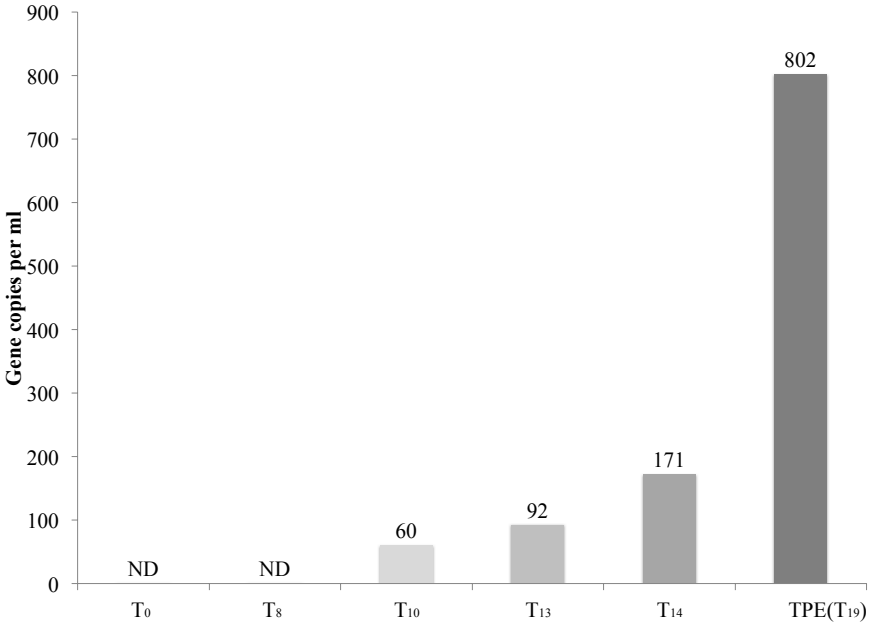

Figure S3. Quantification of *Burkholderia* sp. KK1 *tfdA* genes in 2,4-D mesocosm I based on qPCR. ND: Not detected, i.e., product -if any- was out of range of standard curve (see Material and Methods for details).

Supplementary Figure S4.

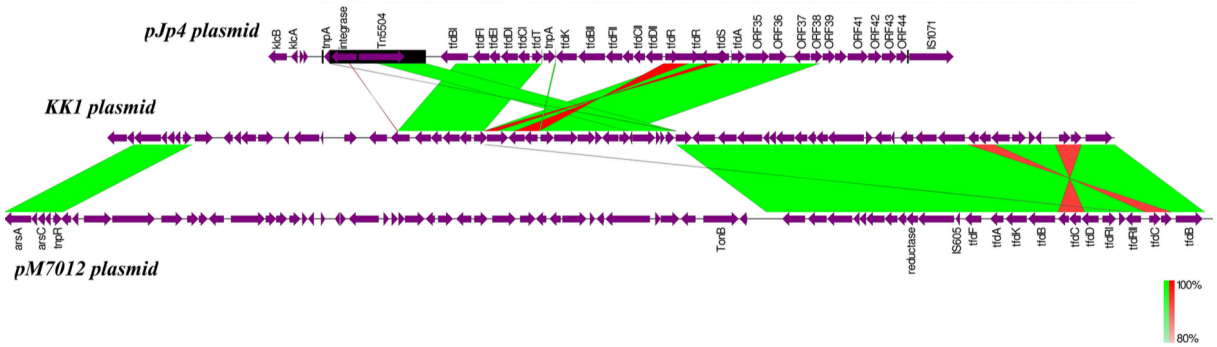

Figure S4. Comparison of the *tfd* operons from the identified KK1 plasmid and the reference pJp4 and pM7012 plasmids. The *tfd* genes and flanking genes are shown in purple color and mobile elements in black background. Lines connecting the plasmids represent ~100% nucleotide identity; green denoting same orientation and red denoting opposite orientation.

Supplementary Figure S5.

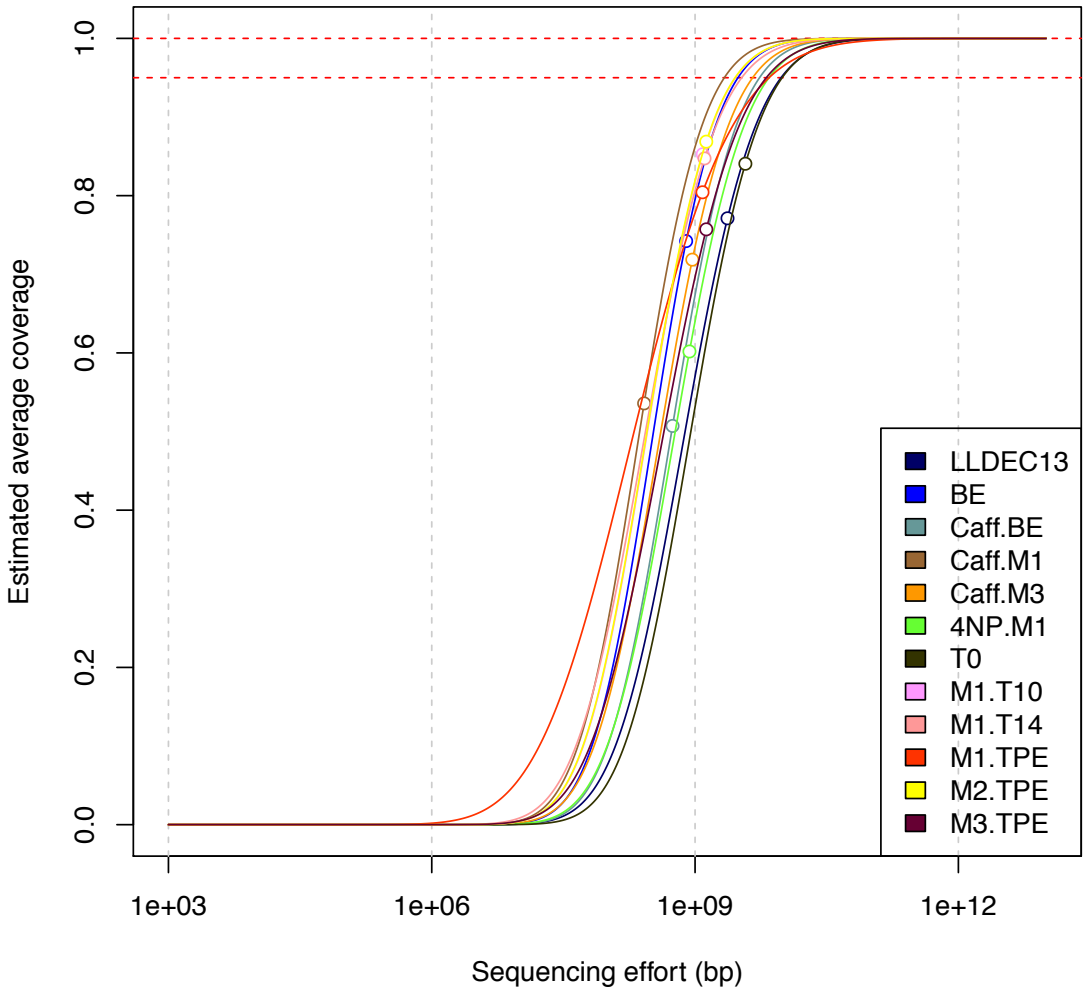

Figure S5. Complexity of 2,4-D, 4-NP and caffeine mesocosm microbial communities as assessed by Nonpareil curves. Empty circles in each Nonpareil curve denote the estimated coverage for each sample (y-axis), given the sequenced depth achieved (x-axis). Note that coverage was above ~55% for all metagenomes.

Supplementary Figure S6.

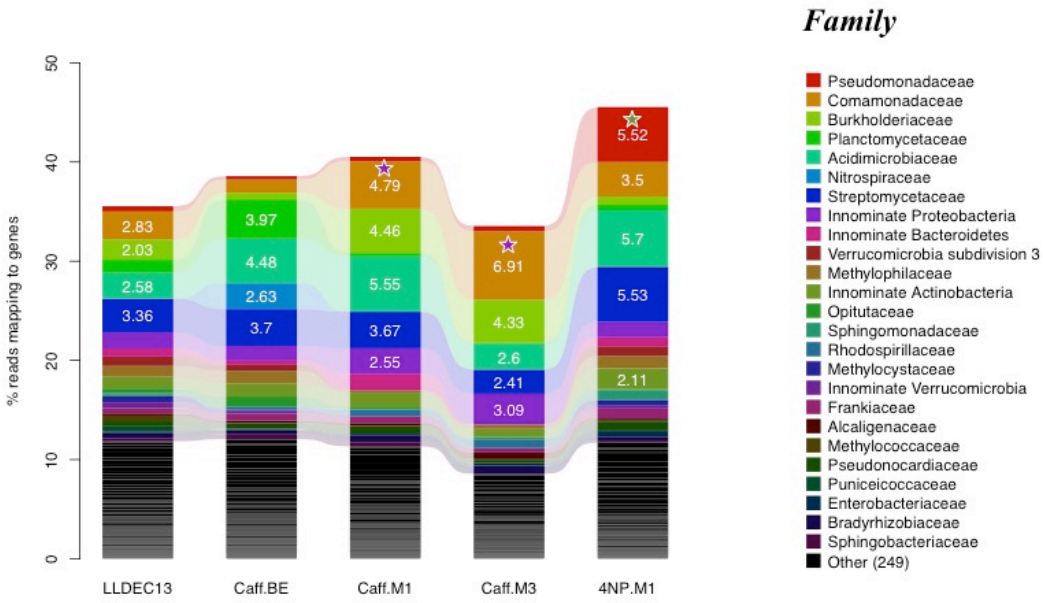

Figure S6. Shifts in microbial community composition in 4-NP mesocosm I (4NP.M1), and caffeine mesocosm I and III (Caff.M1, Caff.M3) relative to the bottle effect (Caff.BE), and the original lake water (LLDEC13). The purple asterisk denotes the family that our caffeine-degrading *Pelomonas* sp. strain KK5 isolate belongs to based on 16S rRNA gene identity (100% nucleotide identity). The green asterisk denotes the family that our 4NP-degrading *Pseudomonas* sp. strain KK4 isolate belongs to (100% nucleotide identity).

Supplementary Figure S7.

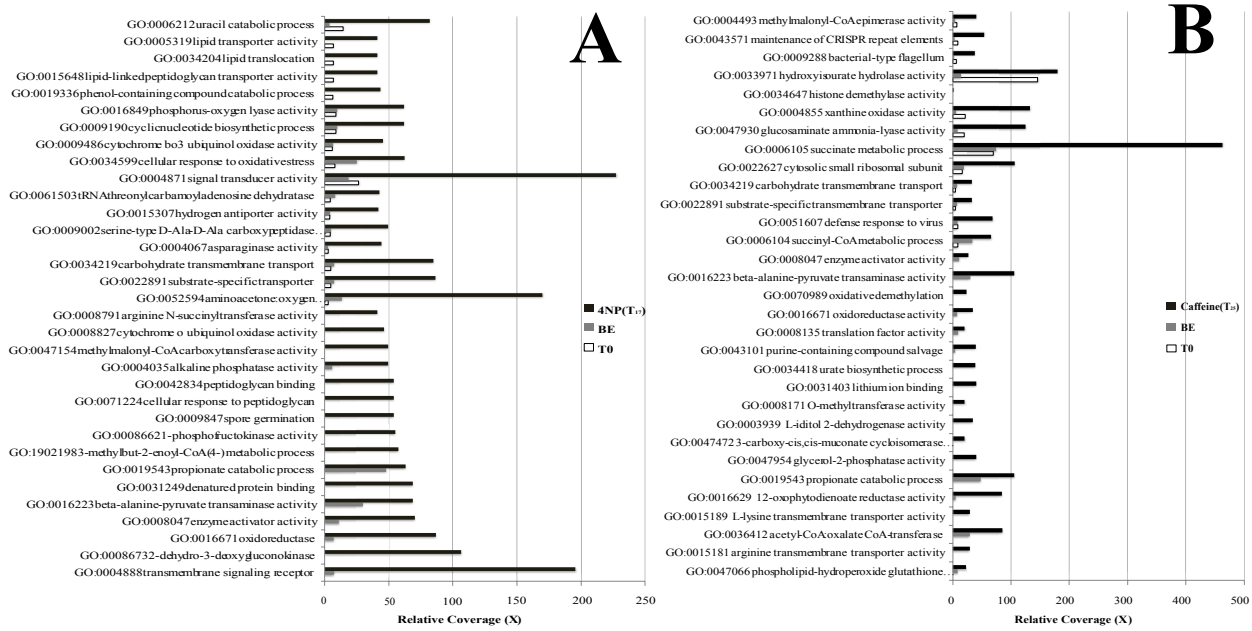

Figure S7. Microbial community gene content shifts in 4NP (Panel A) and caffeine (Panel B) mesocosms. The relative coverage (x-axes) of genes (y-axes) was calculated by summing the length of all reads mapping on the gene (a minimum cut-off for a match of ≥80 bp alignment length and ≥97% nucleotide identity was used) and dividing by the length of the gene sequence. The most differentially abundant genes between T_0 and post-enrichment metagenomes (T_{17} for 4NP and T_{25} for Caffeine) were grouped by their GO terms (y-axes) and are shown on the graph. BE: Bottle-Effect metagenome (control incubation with no caffeine or 4NP added).

Supplementary Figure S8.

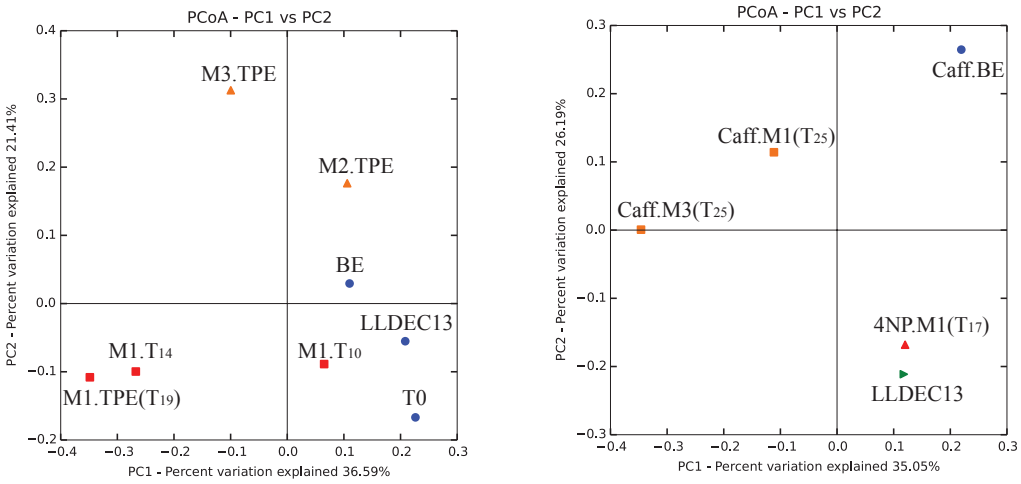

Figure S8. Comparison of microbial community composition among 2,4-D (left panel), 4-NP and caffeine (right panel) mesocosms. Graph shows the Principal Coordinates Analysis (PCoA) of the beta-diversity among the communities. Values were calculated using Bray-Curtis metric.

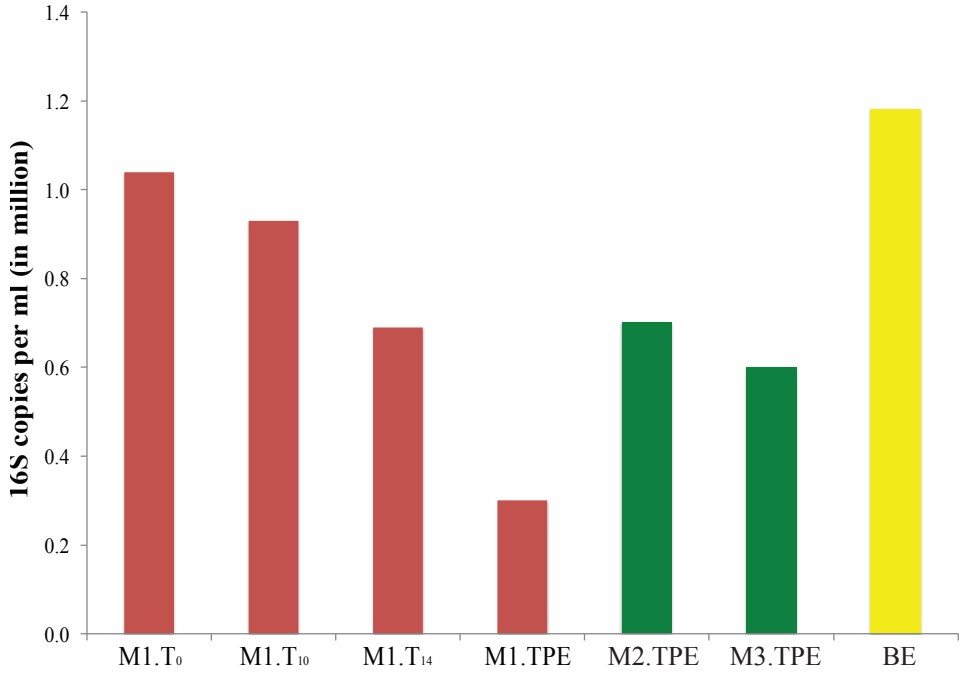

Figure S9. qPCR-based 16S rRNA gene copy number counts for selected metagenomes. qPCR was performed as described in the main text. Assuming no major changes in the average rRNA copy number of the microbial communities over the incubation time, these counts represent the shifts in cell counts as well. Note that 16S counts decreased with incubation time in the 2,4-D mesocosms, consistent with the reduction in OTU numbers in the corresponding metagenomes (see main text). In contrast, no substantial reduction was observed for the negative control sample (Bottle Effect or BE), consistent with the maintenance of OTU diversity in the corresponding BE metagenome. The original T=0 sample had about $5x10^5$/ml, determined by microscopy counts.

Table S1. Statistics for each metagenomic dataset used in this study.

	Trimmed read length	Total number of trimmed reads	Number of contig (>500bp)	Mean contig size	Contig N50	Longest contig	Nonpareil coverage
LLDEC13	131.203	18,220,737	173,541	1,029	1,062	54,874	0.755
2,4-D mesocosm I at T=0 (M1.T0)	131.16	29,145,453	233,906	1,077	1,135	51,509	0.809
2,4-D mesocosm I at T=10 ($M1.T_{10}$)	133.787	9,361,741	80,221	1,158	1,254	58,554	0.825
2,4-D mesocosm I at T=14 ($M1.T_{14}$)	136.583	9,600,373	92,891	1,067	1,123	47,834	0.825
2,4-D mesocosm I at T=19 (M1.TPE)	131.699	9,417,412	73,753	1,292	1,506	68,557	0.817
2,4-D mesocosm II at T=39 (M2.TPE)	134.059	9,495,080	83,309	1,250	1,448	40,480	0.852
2,4-D mesocosm III at T=23 (M3.TPE)	134.86	10,212,412	92,898	1,129	1,224	47,767	0.749
2,4-D BE mesocosm at T=14 (BE)	135.999	5,934,275	55,695	960	979	20,554	0.738
Caffeine Mesocosm I at T_{25} (Caff-M1)	212.33	1,252,602	54,248	941	932	44,283	0.524
Caffeine Mesocosm III at T_{2s5} (Caff-M3)	187.527	4,900,175	116,196	1,025	1,054	42,438	0.698
Caffeine Mesocosm Bottle Effect (Caff-BE)	192.103	2,924,031	99,190	897	868	31,718	0.495
4-NP mesocosm I at T_{17} (4NP-M1)	189.511	4,481,582	136,637	880	872	46,465	0.586

Table S2. Statistics of each isolate genome used in this study.

Genome	Total bp	No. of contigs	Mean contig size	contig N50	Longest contig	Completeness	Contamination	Strain heterogeneity
Burkholderia sp. KK1	9,517,307	8	1,214,494	1,487,370	3,241,870	100	0	0
Sphingopyxis sp. KK2	4,333,434	309	14,024	24,828	85,219	98.28	0	0
Variovorax sp. KK3	7,177,164	197	36,432	68,108	217,155	100	0	0
Pseudomonas sp. KK4	6,374,739	262	24,331	49,190	164,660	98.28	0	0
Pelomonas sp. KK5	6,854,870	558	12,284	24,572	82,945	98.28	0	0

Table S3. OTU counts and diversity indices of 2,4-D, 4-NP and caffeine mesocosms.

Sample	Observed OTUs	Estimated OTUs (Chao1)	Chao1 Lower-Boundary ($CI_{95\%}$)	Chao1 Upper-Boundary ($CI_{95\%}$)	Chao-Shen estimator of Shannon diversity	Bayesian-corrected Shannon diversity
2,4-D Mesocosms						
LLDEC13	1972	2410	2329	2510	6.21	6.10
M1.T_0	2640	2942	2884	3012	6.22	6.13
BE	1098	1480	1395	1588	5.92	5.76
M1.T_{10}	987	1348	1263	1457	5.30	5.20
M1.T_{14}	940	1207	1174	1353	5.12	5.04
M1.TPE	896	1175	1130	1309	5.19	5.08
M2.TPE	1157	1513	1432	1618	5.90	5.80
M3.TPE	925	1267	1186	1372	5.72	5.56
4-NP and Caffeine Mesocosms						
Caff-BE	740	1041	967	1141	5.95	5.73
Caff-M1	377	661	570	793	5.38	5.16
Caff-M3	675	895	836	977	5.30	5.16
4NP-M1	1028	1358	1282	1456	5.87	5.73

Table S4. qPCR primers used in this study.

	Primer sequences	Reference
Forward primer tfdA-F	5'- CTC GAA GGC GGT TTC ATC A -3'	This study
Reverse primer tfdA-R	5' -GTT GAT TCG CGA AGT TCC C -3'	
Forward primer 1055yF	5' -ATG GYT GTC GTC AGC T -3'	(1)
Reverse primer 1392R	5' -ACG GGC GGT GTG TAC -3'	

The theoretical growth yield of the *Burkholderia* sp. KK1 in 2,4-D mesocosm I

Assumptions

1. The 2,4-D degrader, *Burkholderia* sp. KK1, is present as one single cell at time=0 in the lake mesocosm.

2. Once 2,4-D is added, the cell immediately enters exponential growth phase.

3.The 2,4-D is completely biodegraded to CO_2, and no intermediates are present at the end of the experiment.

4. The KK1 strain shows balanced growth (i.e., each cell has the same weight and protein contents). In addition, the composition formula can be approximated as $C_5H_7O_2N$.

5. During the 2,4-D spikes, and throughout the course of the experiment, cells are constantly growing and do not enter stationary phase.

6. The bacteria growth can be represented as a first-order, irreversible chemical reaction. The volume mesocosm volume is also constant.

7. The fraction of electron donor used for energy production equals to that used for cell synthesis ($f_e = f_s = 0.5$).

Method I

Calculation based on growth yield coefficient

Total 2,4-D added: $S_0 = 0.0442$g/L, S = 0

Growth yield, Y, estimated as 0.664 (an estimated value for *E. coli*)

Total cell biomass gained: $X = Y \times S_0 = 0.664 \times 0.0442$g/L = 0.0293 g/L

Exponential phase growth in a batch reactor can be represented as:

$$\frac{dx}{dt} = \mu x$$

where x=cell concentration[g/L], μ = growth rate [d^{-1}].

Integrating the above differential equation results in

$$\ln\frac{X}{X_0} = \mu t \qquad (0.1)$$

X_0 = cell concentration at t=0 [g/L].

Assuming that one cell is present at t=0, we can estimate that $X_0 = 1.5\times10^{-12}$ g/L^2.

By substituting X and X_0 to equation (0.1), t=19 days, we can estimate the growth rate

$\ln\frac{0.0293}{1.5\times10^{-12}} = \mu\times19$, $\mu = 1.247d^{-1}$.

Therefore the doubling time is:

$$t_D = \frac{\ln 2}{\mu} = 0.556\text{d}$$

Assuming exponential growth:

$$N = N_0 \times 2^{t/t_D}$$

Substituting $N_0 = 1$, results in $N = 1.936\times10^{10}$. The total volume of the sample was ~5500ml, so the cell number per ml is 3.52×10^{6}.

Method II

Calculation based on stoichiometry of chemical reaction

$$R_e = R_a - R_d \qquad (2.1)$$

$$R_s = R_c - R_d \qquad (2.2)$$

where

R_e : energy reaction

R_a : acceptor half-reaction

R_d : donor half-reaction

R_s : synthesis reaction

R_c : cell half-reaction

$$R = f_e R_a + f_s R_c - R_d \tag{2.3}$$

applied $f_e = f_s = 0.5$, the f_eR_a, f_sR_c and R_d are as follows:

f_eR_a:

$$\frac{1}{8}O_2 + \frac{1}{2}H^+ + \frac{1}{2}e^- = \frac{1}{4}H_2O \tag{2.4}$$

f_sR_c:

$$\frac{1}{10}CO_2 + \frac{1}{40}HCO_3^- + \frac{1}{40}NH_4^+ + \frac{1}{2}H^+ + \frac{1}{2}e^- = \frac{1}{40}C_5H_7O_2N + \frac{9}{40}H_2O \tag{2.5}$$

$-R_d$:

$$\frac{1}{30}C_8H_6Cl_2O_3 + \frac{13}{30}H_2O = \frac{8}{30}CO_2 + \frac{2}{30}Cl^- + \frac{32}{30}H^+ + e^- \tag{2.6}$$

Combining (2.4), (2.5) and (2.6), results in:

$$\frac{1}{30}C_8H_6Cl_2O_3 + \frac{1}{8}O_2 + \frac{1}{40}HCO_3^- + \frac{1}{40}NH_4^+ = \frac{1}{40}C_5H_7O_2N + \frac{5}{30}CO_2 + \frac{1}{15}Cl^- + \frac{1}{15}H^+ + \frac{1}{24}H_2O \tag{2.7}$$

Given the initial substrate concentration was 0.0442g/L, and MW=221.1g/mol, the cell biomass formed can be estimated according to equation 2.7:

$$\frac{1}{30} \times \frac{0.0442 g/L}{221 g/mol} = \frac{1}{40} \times \frac{X}{113 g/mol}$$

Solve for X = 0.03g/L. By using this value, substitute equation 1.1, 1.2 and 1.3, we can get N=4.5 $\times 10^6$/ml.

REFERENCES CITED

1. **Ritalahti KM, Amos BK, Sung Y, Wu Q, Koenigsberg SS, Löffler FE.** 2006. Quantitative PCR targeting 16S rRNA and reductive dehalogenase genes simultaneously monitors multiple *Dehalococcoides* strains. Applied and Environmental Microbiology **72:**2765-2774.