Supplemental information

Structure of the microbiome types

16S rRNA gene libraries were rarefied to the lowest number of sequences across libraries (excluding one replicate of the TSA inoculum, which had <100 sequences after quality processing), which was 11,667 reads. After sequence processing, 2323 operational taxonomic units (OTUs) (with singletons removed) were recovered in the three microbiome types prior to the incubation. Of these, 2225, 785, and 520 OTUs were observed in the INIT, CO-M, and REG-M microbiomes, respectively (Supplementary Fig. 3a). The cultured microbiome subsets shared 405 OTUs (45%), while 802 OTUs (36%) were shared between the INIT microbiome and the two cultured microbiome subsets. Our high estimates of cultivable bacterial diversity are undoubtedly inflated to some extent by remnant DNA in the irradiated soil, which still contained DNA from a wide range of taxa (Supplementary Fig. 4). However, the amount of DNA extracted from uninoculated sterile soil was more than 20 times lower than what we obtained from our initial microcosms. In addition, the number of OTUs observed in the crude oil + glucose media was more than three times higher than what we observed in the TSA media, which yielded the lowest number of OTUs (Supplementary Fig. 3), and certain phyla observed in the initial and gamma-irradiated soil were either absent or extremely rare in the extracts of the culture media (Supplementary Figs. 3-4). We also found that the relative abundance of phyla and the most abundant OTUs (represented by the first two PCoA axes) in both the TSA media and the TSA + crude oil media was nearly identical (Supplementary Fig. 3), indicating that this easily accessed carbon source, and not petroleum toxicity, determined the dominant cultured bacteria. This is in spite of the fact that petroleum hydrocarbons can have toxic effects on microbial cell walls (1, 2) and can force physiological tradeoffs in bacteria (3). Although abundant bacteria in petroleum-

contaminated soils are likely to be petroleum-tolerant, they may not metabolize petroleum contaminants when other easily-accessed carbon sources are available.

Supplementary Table 1. Mean concentrations of polycyclic aromatic hydrocarbon (PAH) compounds measured at the end of a 6-week incubation with crude oil from five replicates of each treatment and three replicate soil samples that were spiked with crude oil but not incubated (BASELINE).

PAH compound	CO-M	REG-M	INIT	BASELINE
Acenaphthene	0.1 ± 0	0.1 ± 0	0.1 ± 0	0.1 ± 0
Acenaphthylene	0.1 ± 0	0.1 ± 0	0.1 ± 0	0.1 ± 0.1
Anthracene	0.2 ± 0	0.2 ± 0	0.1 ± 0	0 ± 0
Benzo[a]anthracene	0 ± 0	0 ± 0	0 ± 0	0 ± 0
Benzo[a]pyrene	0 ± 0	0 ± 0	0 ± 0	0 ± 0
Benzo[b]fluoranthene	0 ± 0	0 ± 0	0 ± 0	0 ± 0
Benzo[j]fluoranthene	0 ± 0	0 ± 0	0 ± 0	0 ± 0
Benzo[k]fluoranthene	0 ± 0	0 ± 0	0 ± 0	0 ± 0
Benzo[c]phenanthrene	0 ± 0	0 ± 0	0 ± 0	0 ± 0
Benzo[g,h,i]perylene	0 ± 0	0 ± 0	0 ± 0	0 ± 0
Chrysene	0.3 ± 0	0.3 ± 0	0.3 ± 0	0.1 ± 0.1
Dibenzo[a,h]anthracene	0 ± 0	0 ± 0	0 ± 0	0 ± 0
Dibenzo[a,i]pyrene	0 ± 0	0 ± 0	0 ± 0	0 ± 0
Dibenzo[a,h]pyrene	0 ± 0	0 ± 0	0 ± 0	0 ± 0
Dibenzo[a,l]pyrene	0 ± 0	0 ± 0	0 ± 0	0 ± 0
Dimethyl-7,12-benzo[a]anthracene	0 ± 0	0 ± 0	0 ± 0	0 ± 0
Fluoranthene	0 ± 0	0 ± 0	0 ± 0	0 ± 0
Fluorene	0.6 ± 0.1	0.6 ± 0	0.3 ± 0.1	0.3 ± 0
Indeno[1,2,3-cd]pyrene	0 ± 0	0 ± 0	0 ± 0	0 ± 0
Methyl-3-cholanthrene	0 ± 0	0 ± 0	0 ± 0	0 ± 0
Naphthalene	0 ± 0	0.4 ± 0	0 ± 0	0.7 ± 0.1
Phenanthrene	1.2 ± 0.2	1.3 ± 0.1	0.5 ± 0.1	0.5 ± 0
Pyrene	0.1 ± 0	0.1 ± 0	0.1 ± 0	0 ± 0
Methyl-1-naphthalene	1.1 ± 0.1	1.8 ± 0.2	0.4 ± 0.3	1.5 ± 0.1
Methyl-2-naphthalene	0.3 ± 0.2	1 ± 0.1	0.4 ± 0.2	1.9 ± 0.2
Dimethyl-1,3-naphthalene	3.2 ± 0.3	2.7 ± 0.2	2.2 ± 0.7	3.3 ± 0.3
Trimethyl-2,3,5-naphthalene	1.2 ± 0.1	0.9 ± 0.1	1 ± 0.1	0.7 ± 0.1

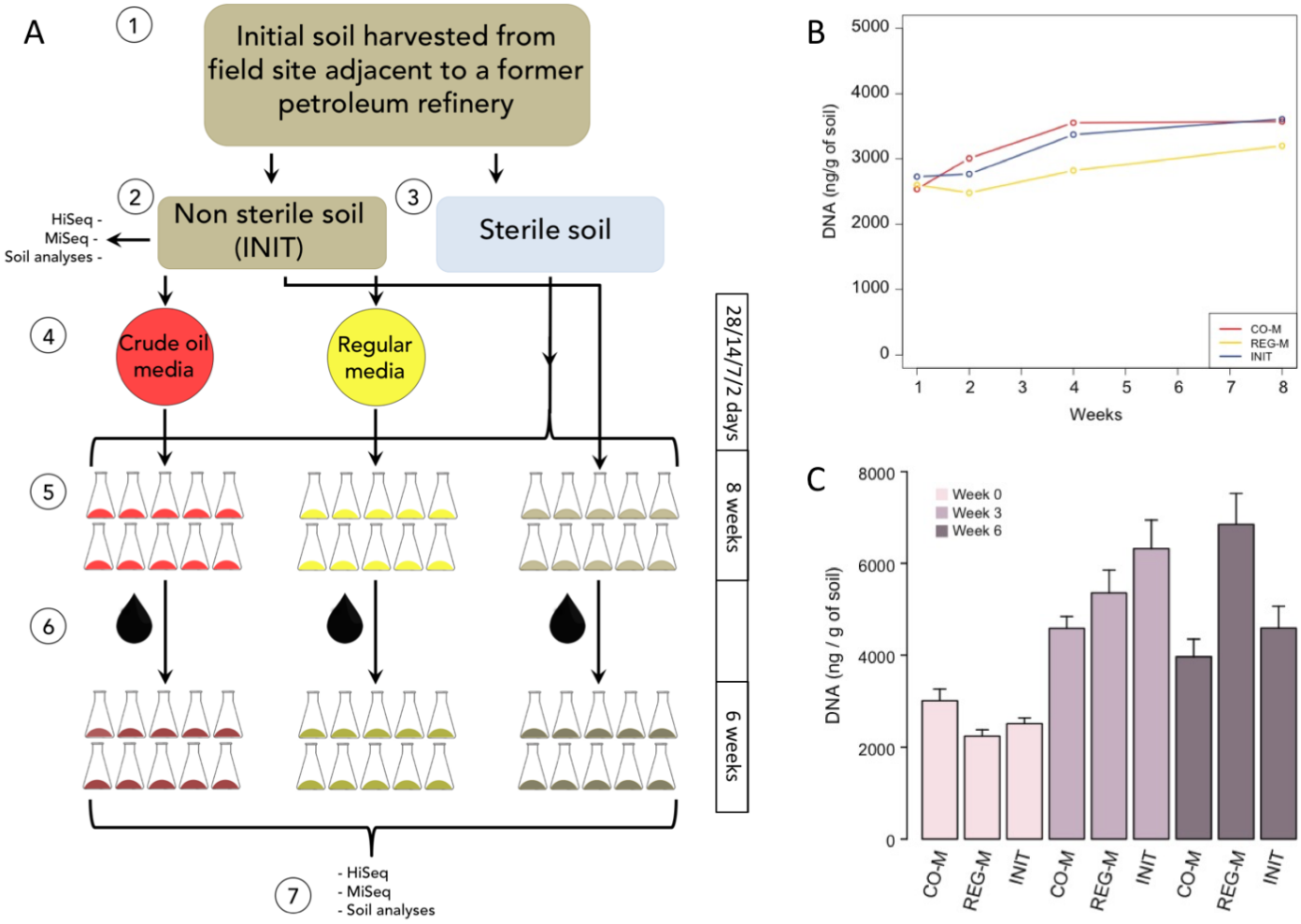
Supplementary Table 2. Summary of sequence statistics from metagenomics sequencing on the Illumina HiSeq. Values reported are means of 5 replicates from week 6 after crude oil addition (REG-M, CO-M, INIT) or of 3 replicates from the original unsterilized soil.

	REG-M	CO-M	INIT	Initial soil
Predicted rRNA features	70781	55795	80550	62315
Predicted protein features	3485623	3540681	5945513	7441989
rRNA/protein	0.021±0.001 **a**	0.016±0.001 **b**	0.014±0.001 **b**	0.008±0.000 **c**
GC[a] %	66.4	64.4	67.2	62.7
Functional categories	1632898	1857715	2449697	2384184
Sequences post QC[b]	10085426	7944149	9676980	8561399
FC[c]/seq	0.16±0.01 **b**	0.23±0.02 **a**	0.25±0.01 **a**	0.28±0.02 **a**

[a] Percent of guanine and cytosine bases in sequences.
[b] QC = quality control
[c] FC = identified functional categories

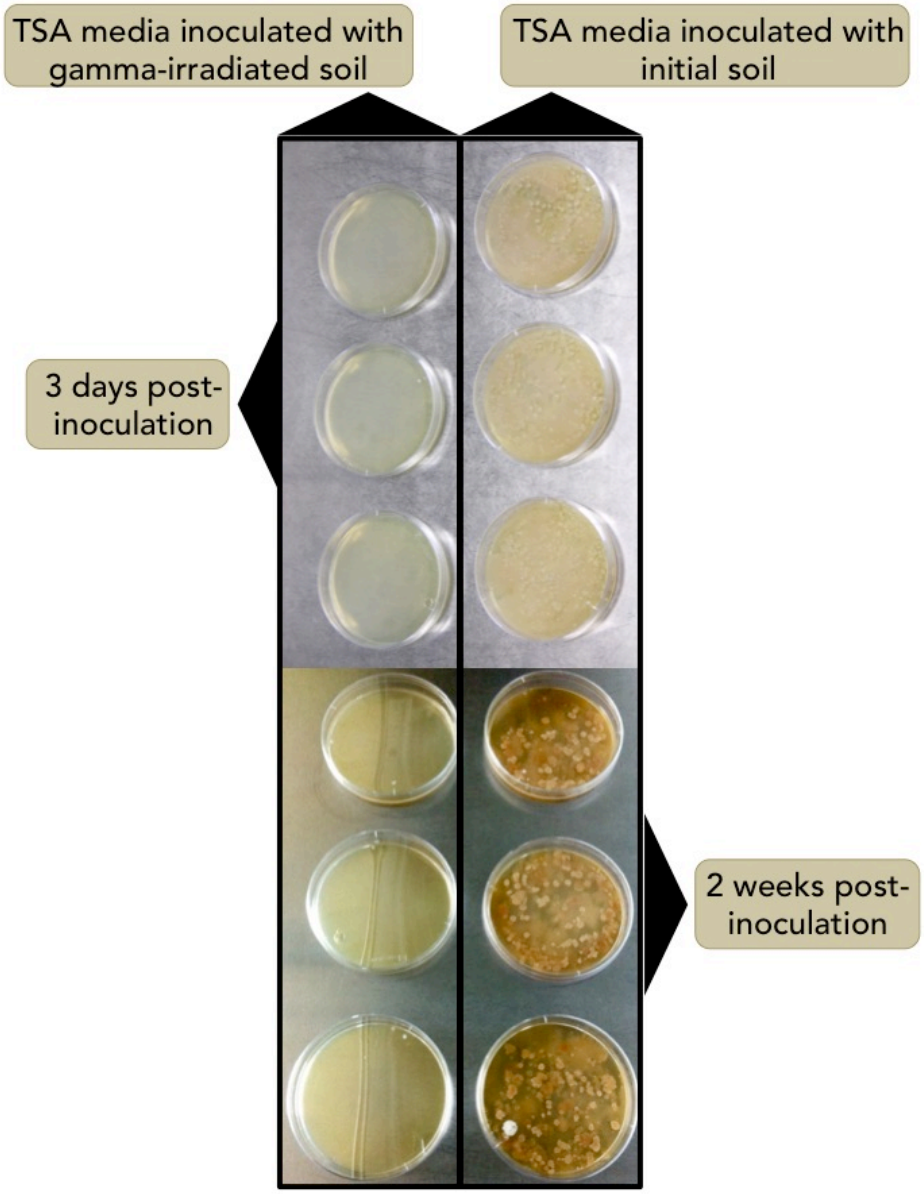
Supplementary Table 3. Genera found in our dataset and in the Minnesota Biodegradation and Biocatalysis Database (MBBD) that are indicated as degrading hydrocarbons of C10 or greater in the MBBD.

Phylum	**Class**	**Order**	**Family**	**Genus**
Actinobacteria	*Actinobacteria*	*Actinomycetales*	*Corynebacteriaceae*	*Corynebacterium*
Actinobacteria	*Actinobacteria*	*Actinomycetales*	*Micrococcaceae*	*Arthrobacter*
Actinobacteria	*Actinobacteria*	*Actinomycetales*	*Mycobacteriaceae*	*Mycobacterium*
Actinobacteria	*Actinobacteria*	*Actinomycetales*	*Nocardia*	*Nocardia*
Actinobacteria	*Actinobacteria*	*Actinomycetales*	*Nocardiaceae*	*Rhodococcus*
Actinobacteria	*Actinobacteria*	*Actinomycetales*	*Nocardioidaceae*	*Nocardioides*
Actinobacteria	*Actinobacteria*	*Actinomycetales*	*Streptomycetaceae*	*Streptomyces*
Bacteroidetes	*Flavobacteriia*	*Flavobacteriales*	*Flavobacteriaceae*	*Flavobacterium*
Firmicutes	*Bacilli*	*Bacillales*	*Bacillus*	*Bacillus*
Firmicutes	*Bacilli*	*Bacillales*	*Staphylococcaceae*	*Staphylococcus*
Proteobacteria	*Alphaproteobacteria*	*Sphingomonadales*	*Sphingomonadaceae*	*Novosphingobium*
Proteobacteria	*Alphaproteobacteria*	*Sphingomonadales*	*Sphingomonadaceae*	*Sphingobium*
Proteobacteria	*Alphaproteobacteria*	*Sphingomonadales*	*Sphingomonadaceae*	*Sphingomonas*
Proteobacteria	*Alphaproteobacteria*	*Sphingomonadales*	*Sphingomonadaceae*	*Sphingopyxis*
Proteobacteria	*Betaproteobacteria*	*Burkholderiales*	*Variovorax*	*Variovorax*
Proteobacteria	*Gammaproteobacteria*	*Pseudomonadales*	*Pseudomonadaceae*	*Pseudomonas*
Proteobacteria	*Gammaproteobacteria*	*Xanthomonadales*	*Xanthomonadaceae*	*Stenotrophomonas*

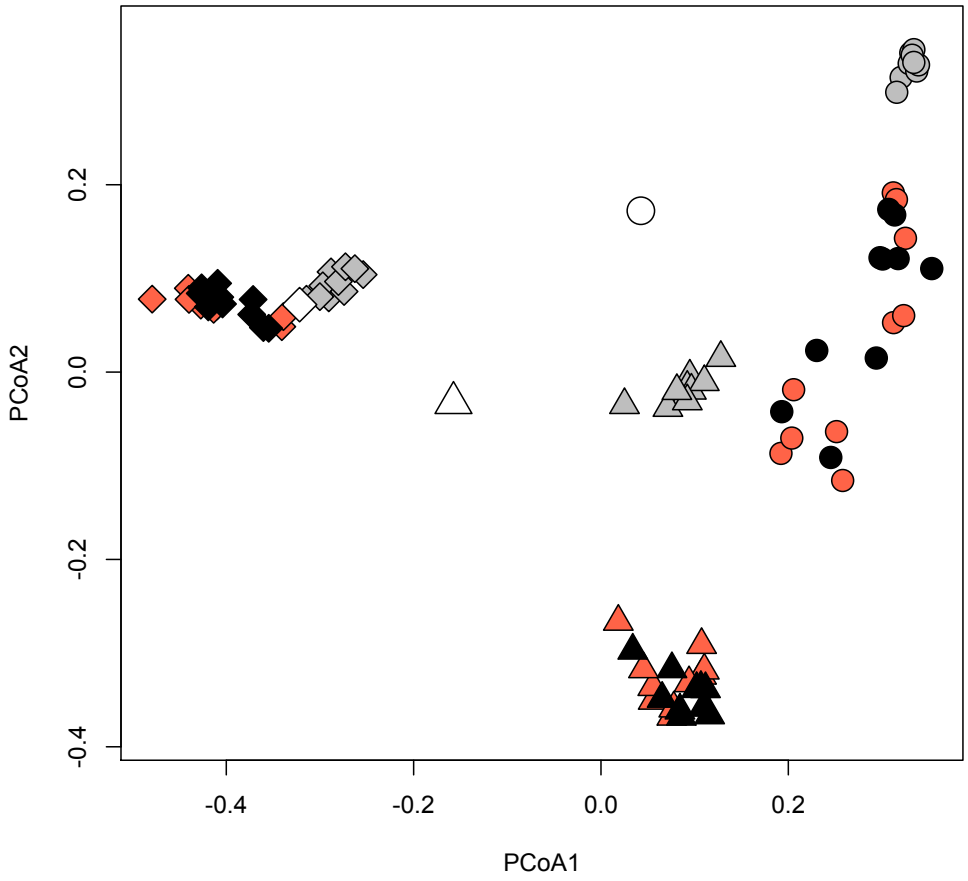


Supplementary Figure 1. (a) Schematic illustration of the experimental design. 1) Collected several kg of the top 20 cm of soil at a site in Varennes, QC, Canada; 2) 16S rRNA gene amplicon sequencing and shotgun metagenomic sequencing were performed on subsamples of the initial unsterilized soil; 3) a portion of the soil collected from Varennes was sterilised with a minimum gamma-irradiation exposure of 50 kGy; 4) dilutions of the initial soil were plated onto regular media and media containing crude oil, and incubated for 2, 7, 14, or 28 days (start dates synchronized so that all incubations ended on the same day); 5) the sterilized soil was used as the matrix in soil microcosms and was inoculated with all bacteria cultivated on regular media (REG-M), all bacteria cultivated media containing crude oil (CO-M), or with the initial unsterilized soil

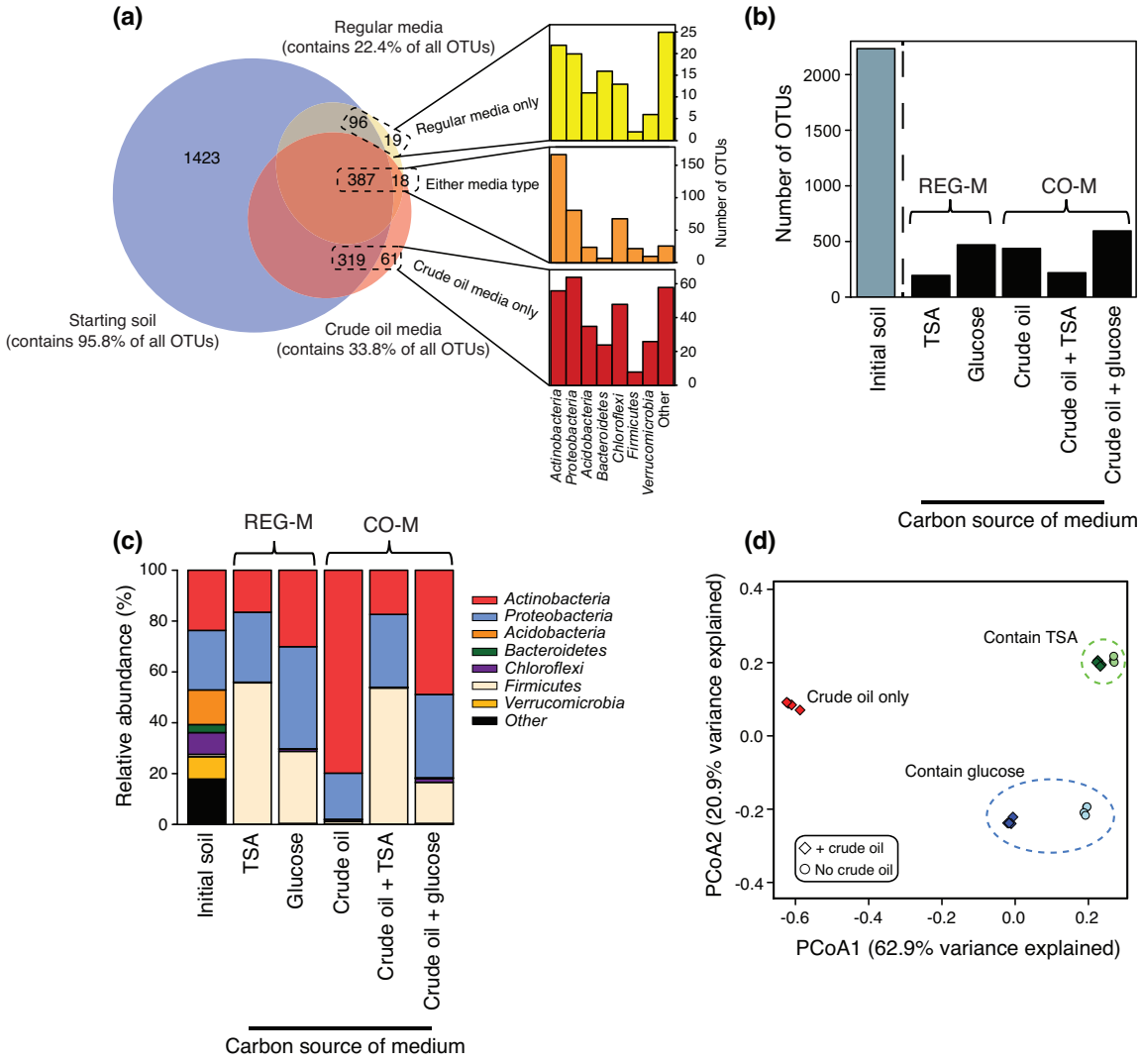
(INIT); 6) the inoculated microcosms were incubated for eight weeks to let the microbiomes stabilize before we added 6000 mg/kg of crude oil to each microcosm ; 7) after six weeks of incubation, microbiomes were examined with 16S rRNA gene amplicon sequencing and shotgun metagenomic sequencing, and total remaining hydrocarbons were quantified in each microcosm. (b) Plot showing microbiome stabilization, estimated by DNA quantification (ng/g soil) in two replicate microcosms per treatment at weeks 1, 2, 4 and 8. By 8 weeks post-inoculation, microbiome growth (estimated by DNA quantification) had mostly levelled off, and was similar between microbiome types. (c) Plot showing microbiome stabilization, estimated by DNA quantification (ng/g soil) in 10 replicate microcosms per treatment at weeks 0, 3, and 6 following the addition of crude oil and monoammonium phosphate. Bars represent standard error.



Supplementary Figure 2. Comparison between TSA plates inoculated with a 10^{-1} dilution of the gamma-irradiated soil and the initial unsterilized soil, both 3 days and 2 weeks post-inoculation. Plates inoculated with the initial soil were covered in colonies, while no colonies were observed in plates inoculated with the irradiated soil.

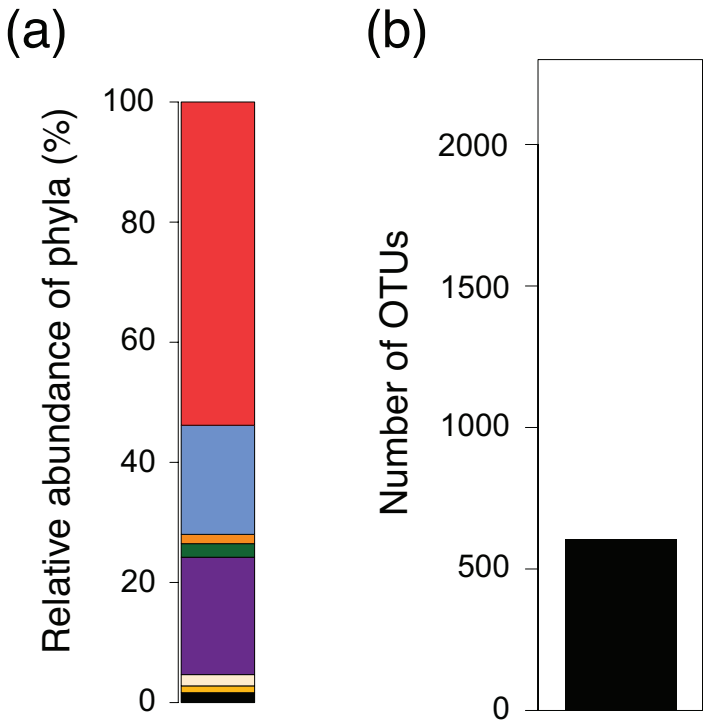


Supplementary Figure 3. Principal coordinate analysis (PCoA) ordinations of Bray-Curtis dissimilarity at the OTU level. Shapes coloured in white represent the mean of replicate DNA extractions from the initial inocula. For the REG-M and CO-M inocula, they represent the mean of the two and three media types used to produce them, respectively, as bacteria from each media type were added in equal amounts. Circles = INIT; triangles = REG-M; diamonds = CO-M; white = initial inocula; grey = week 0; red = week 3; black = week 6.

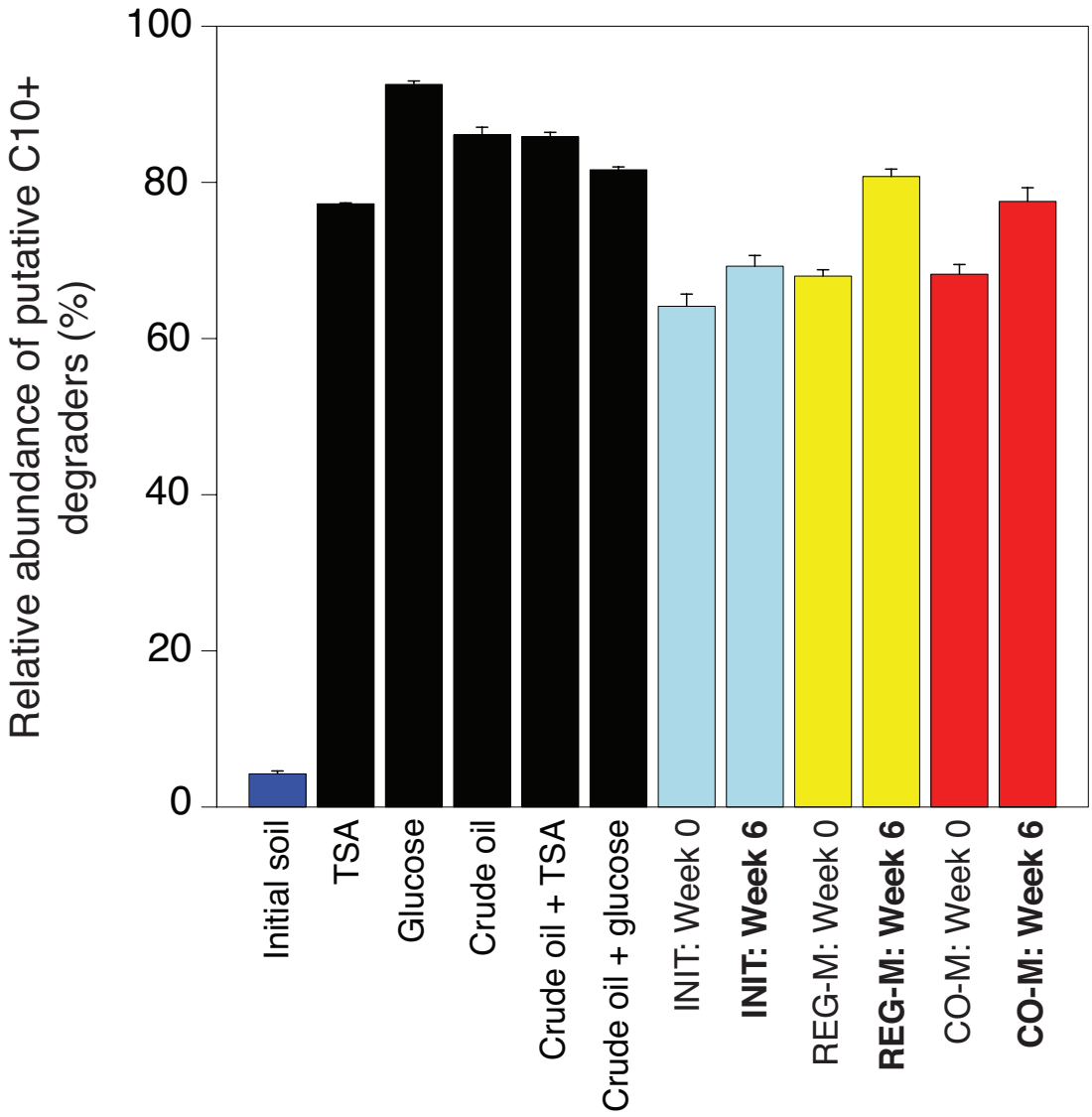


Supplementary Figure 4. Summary of microbiome subsets captured through isolation. A solution containing scraped bacterial colonies was added to gamma-irradiated, at which point DNA was extracted from soil immediately. (a) Proportional Venn diagram showing the number of OTUs found within the initial soil, regular media (TSA, M9 + Glucose), or crude oil-amended media (TSA + Crude oil, M9 + Glucose + Crude oil, M9 + Crude oil). Histograms show the number of OTUs from each major bacterial phylum identified in the regular media, crude oil media, or both. (b) Total number of OTUs identified across the four replicate extractions of the initial soil (replicates pooled) and each medium after rarefying to the lowest number of reads

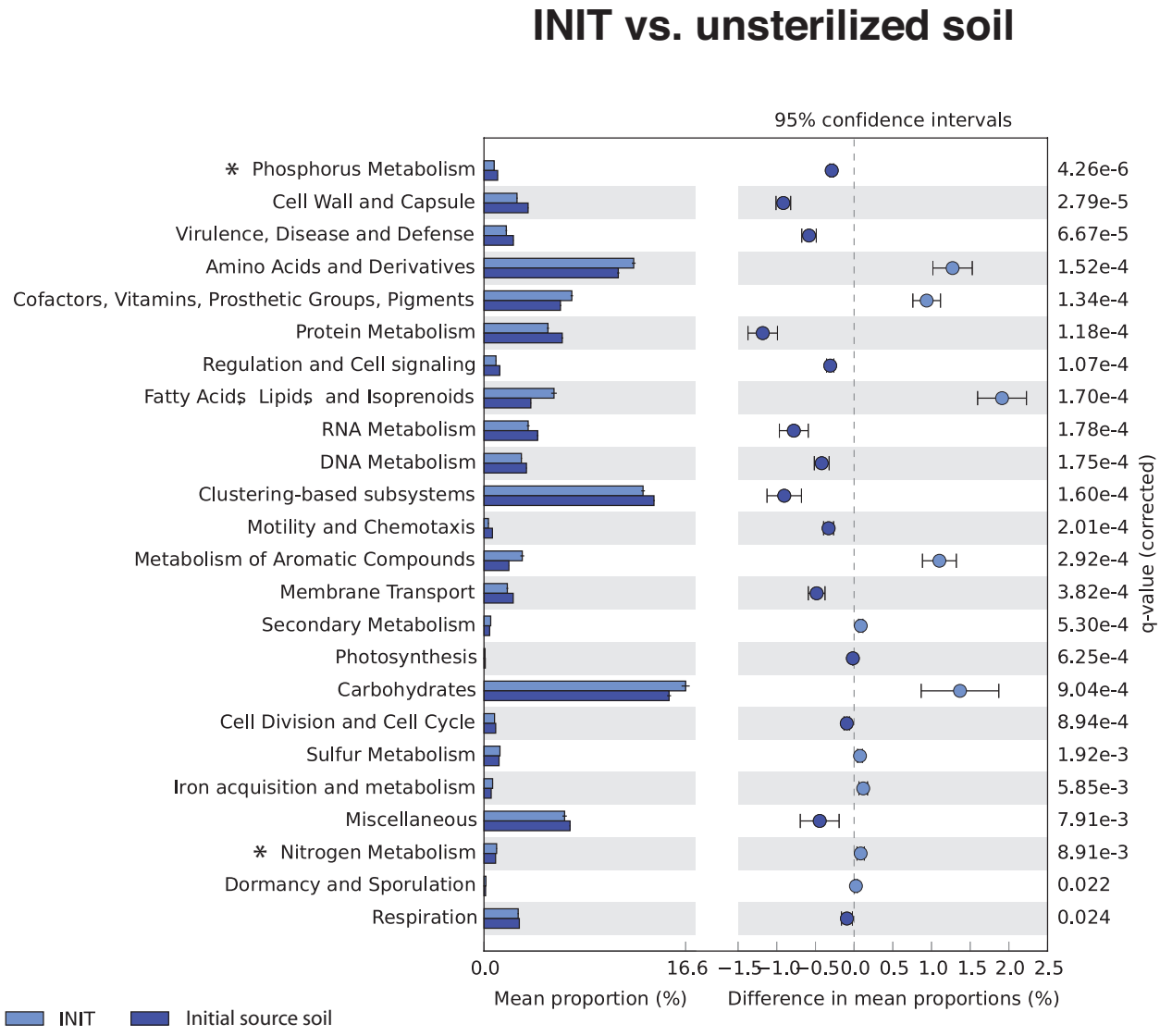
across treatments. (c) Relative abundance of sequences for the initial soil and each medium at the phylum level. Note that most phyla are either absent or extremely rare in cultured media, showing that the influence of residual soil DNA on relative sequence abundance was minimal. The media belonging to the REG-M and CO-M sub-microbiomes are indicated by braces. (d) PCoA ordination of the five media types. Ellipses enclose media containing either glucose or TSA, with or without crude oil added.



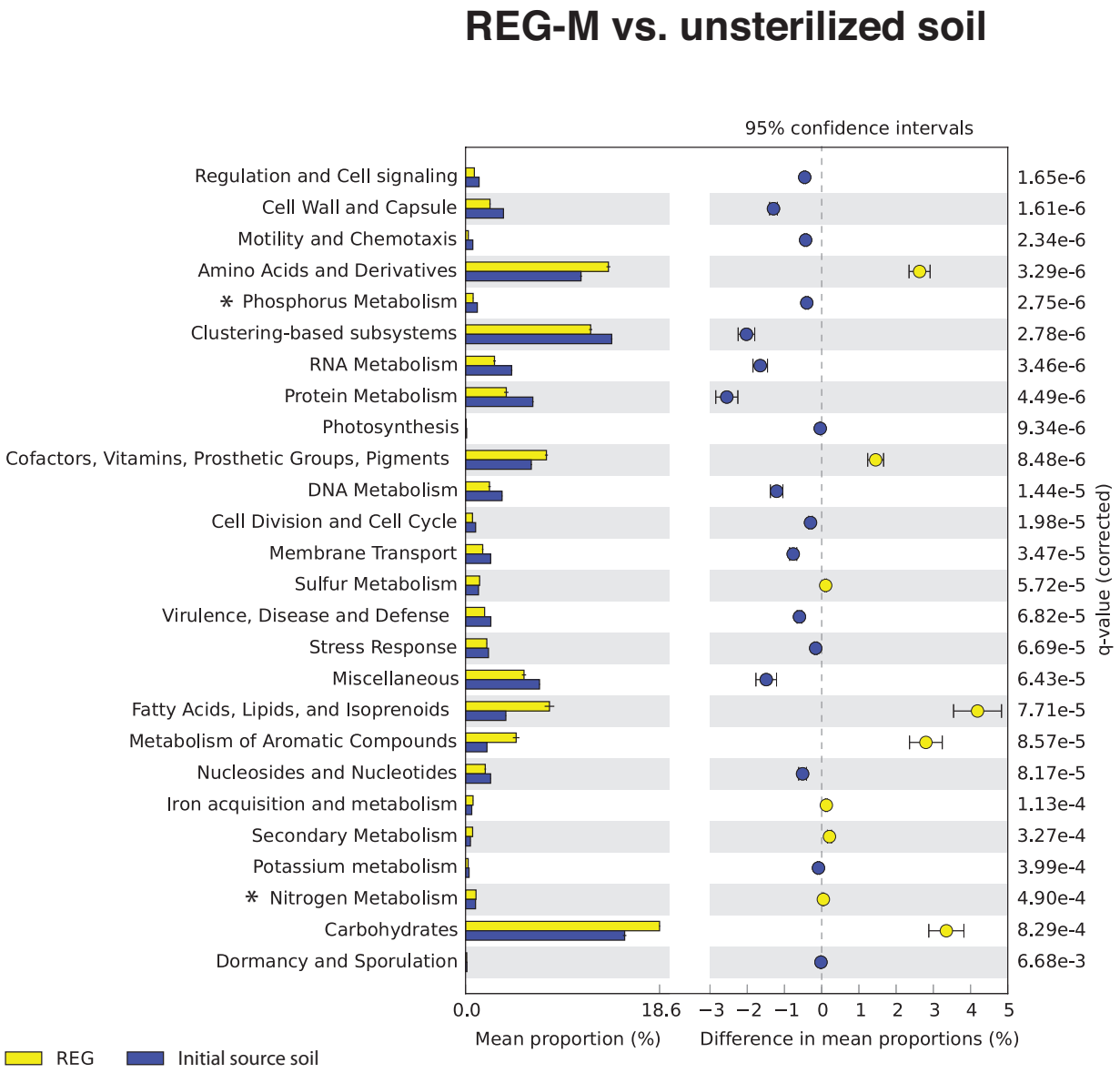
Supplementary Figure 5. (a) 16S rRNA gene taxonomy at the phylum level and (b) OTU count at 97% similarity for the gamma-irradiated soil. Despite a large reduction in DNA from the irradiation process, a small amount of DNA remained.



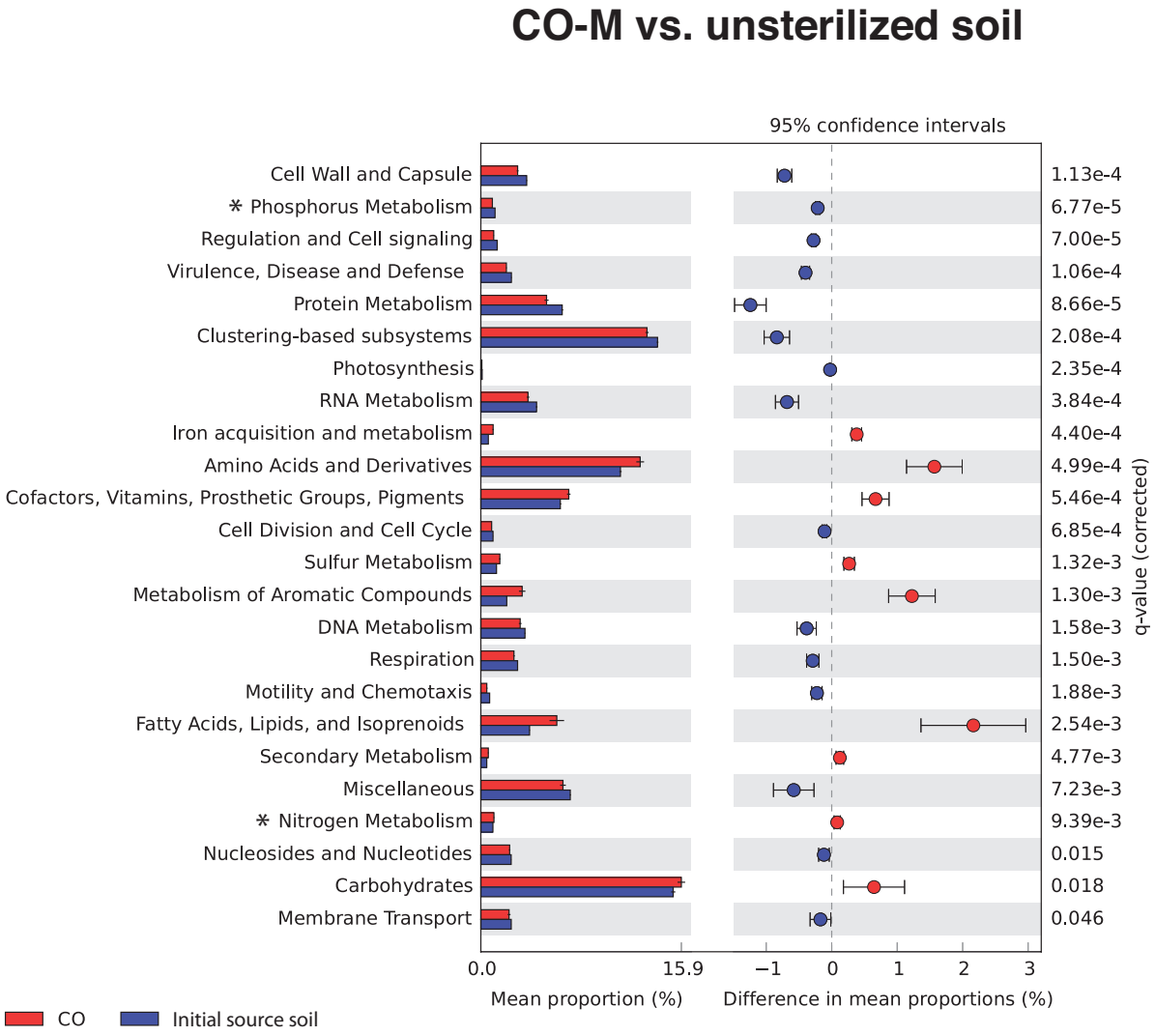
Supplementary Figure 6. Relative abundance of genera identified as putative degraders of C10 or larger organic compounds (see Supplementary Table 5), based on their presence in the Minnesota Biodegradation and Biocatalysis Database (http://eawag-bbd.ethz.ch/) and inclusion in the degradation pathway of at least one organic compound of C10 or greater.



Supplementary Figure 7. Comparison of Level 1 functional annotations from the SEED subsystems database for the INIT treatment (initial soil reinoculated into gamma-irradiated soil) at week 6 and the initial source soil. FDR-corrected *P*-values are reported, and only features that are significantly different in relative abundance ($q<0.05$) are shown. Nitrogen and phosphorus metabolism, categories potentially influenced by monoammonium phosphate addition, are highlighted by asterisks.

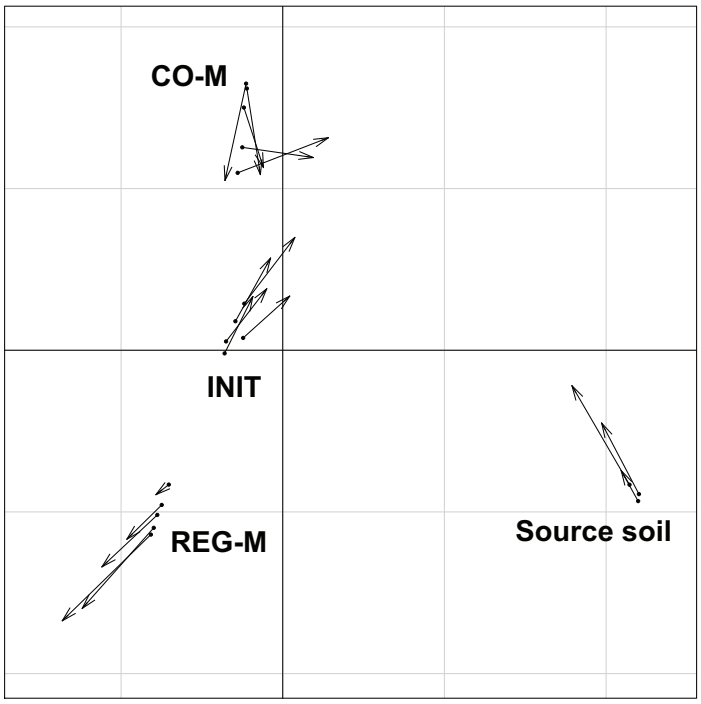


Supplementary Figure 8. Comparison of Level 1 functional annotations from the SEED subsystems database for the REG-M treatment (regular media, i.e. TSA and glucose, reinoculated into gamma-irradiated soil) at week 6 and the initial source soil. FDR-corrected *P*-values are reported, and only features that are significantly different in relative abundance ($q<0.05$) are shown. Nitrogen and phosphorus metabolism, categories potentially influenced by monoammonium phosphate addition, are highlighted by asterisks.

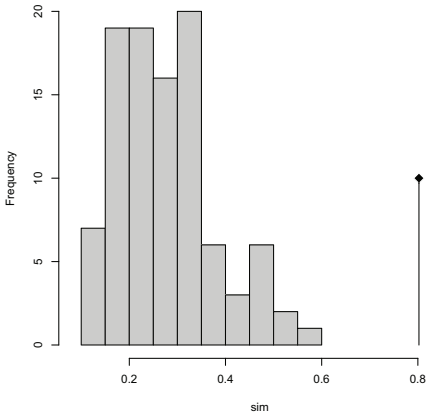


Supplementary Figure 9. Comparison of Level 1 functional annotations from the SEED subsystems database for the CO-M treatment (media containing crude oil, i.e. crude oil + TSA, crude oil + glucose, crude oil alone, reinoculated into gamma-irradiated soil) at week 6 and the initial source soil. FDR-corrected *P*-values are reported, and only features that are significantly different in relative abundance ($q<0.05$) are shown. Nitrogen and phosphorus metabolism, categories potentially influenced by monoammonium phosphate addition, are highlighted by asterisks.

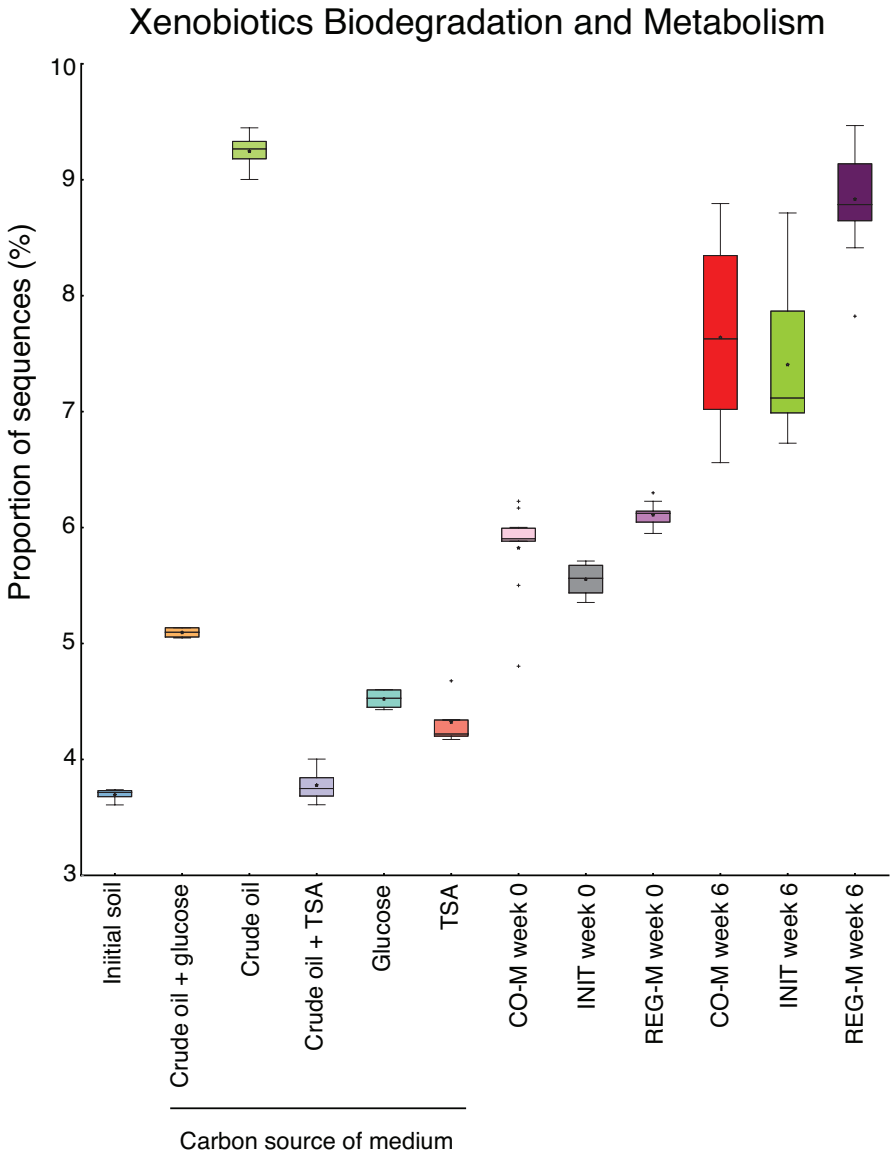
(a)



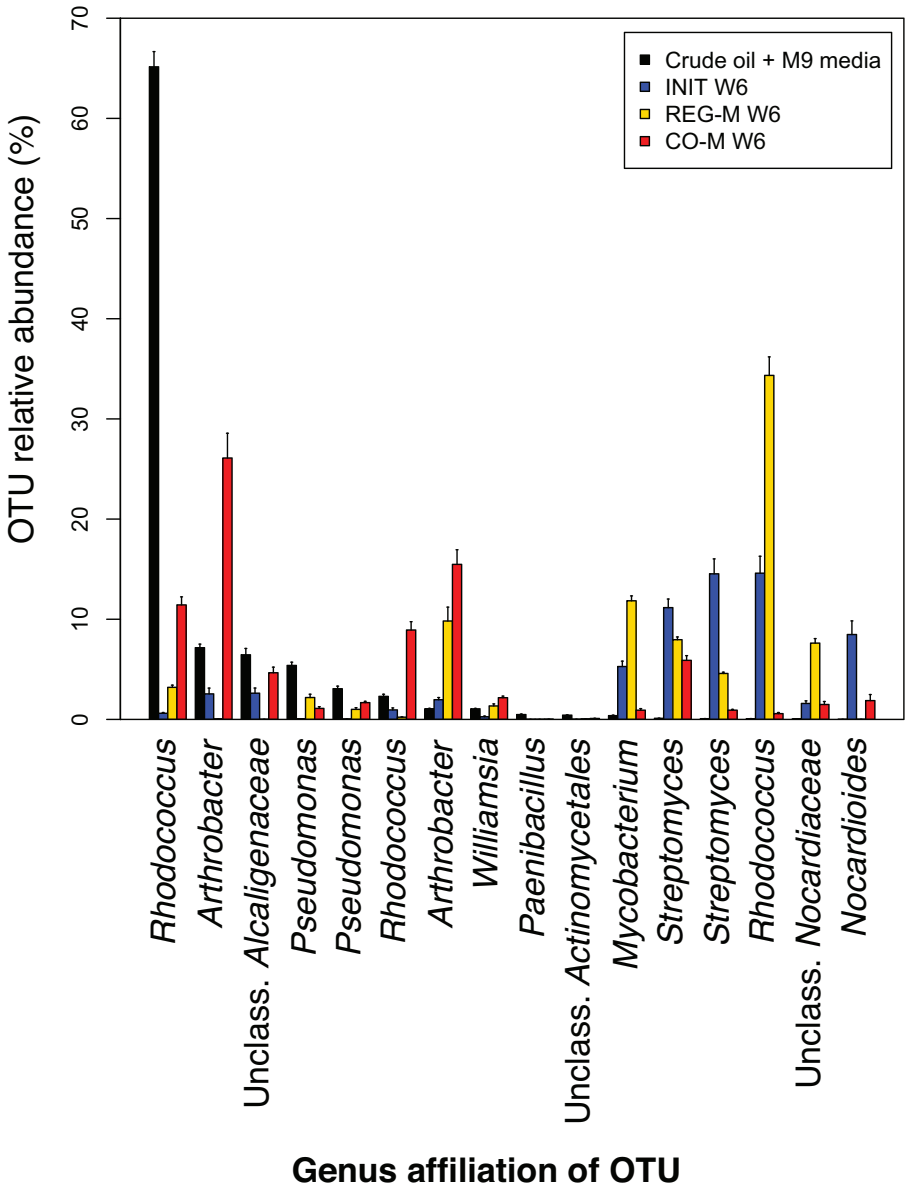
(b)



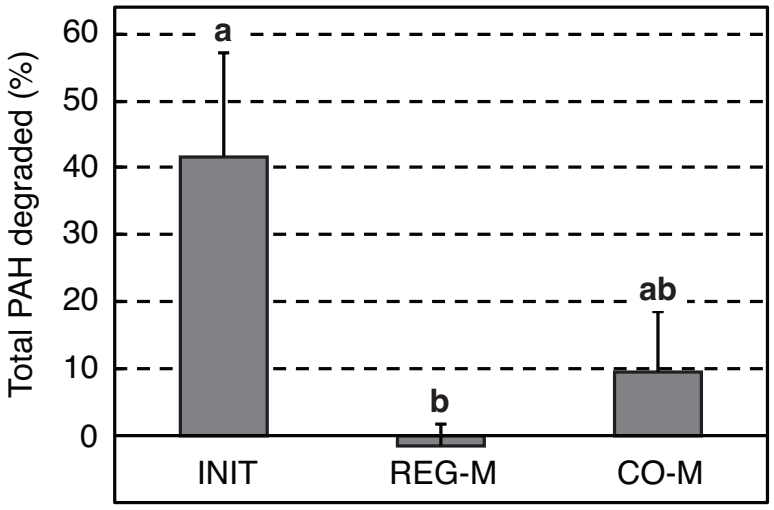
Supplementary Figure 10. Coinertia analysis of the 16S rRNA OTU tables and the functional gene annotations obtained in MG-RAST (SEED Subsystems annotations – Level 3). (a) Coinertia ordination of paired 16S rRNA gene data (points) and functional gene annotations (arrowheads). (b) Permutation test demonstrating that the coefficient of correlation between the 16S rRNA gene data and functional annotation in this analysis (vertical line with dot) is greater than would be expected by chance.



Supplementary Figure 11. Proportion of functional sequences projected to be related to Xenobiotics Biodegradation and Metabolism based on STAMP analysis of PiCrust projections from 16S rRNA gene data. Although the proportion of sequences attributed to this category was lower in MG-RAST annotated metagenomes (for samples Initial soil, CO-M week 6, INIT week 6 and REG-M week 6), the general pattern (e.g. highest mean in REG-M, lowest in Initial soil) was very similar.



Supplementary Figure 12. The 10 most abundant OTUs in the Crude Oil + M9 media and the 10 most abundant OTUs in each soil microbiome type after 6 weeks of incubation with crude oil and monoammonium phosphate (16 OTUs total because of overlap between the two). OTUs are ranked by descending abundance in the Crude Oil + M9 media. Genus affiliations are provided for each OTU. Bars represent standard error.



Supplementary Figure 13. Percent degradation of polycyclic aromatic hydrocarbons (PAHs) after 6 weeks of microcosm incubation following the addition of crude oil and monoammonium phosphate (n=5). Bars represent standard error. Different letters over columns indicate significant differences between treatments based on a Tukey HSD post-hoc test following one-way ANOVA. INIT = sterile soil reinoculated with the initial soil; REG-M = sterile soil reinoculated with bacteria cultured on regular media; CO-M = sterile soil reinoculated with bacteria cultured on media containing crude oil.

1. **Ramos JL, Duque E, Gallegos MT, Godoy P, Ramos-González MI, Rojas A, Terán W, Segura A.** 2002. Mechanisms of solvent tolerance in Gram-negative bacteria. Annu Rev Microbiol **56:**743-768.
2. **Kang YS, Park W.** 2010. Protection against diesel oil toxicity by sodium chloride-induced exopolysaccharides in *Acinetobacter* sp strain DR1. J Biosci Bioeng **109:**118-123.
3. **Whyte LG, Slagman SJ, Pietrantonio F, Bourbonnière L, Koval SF, Lawrence JR, Inniss WE, Greer CW.** 1999. Physiological adaptations involved in alkane assimilation at a low temperature by *Rhodococcus* sp strain Q15. Appl Environ Microbiol **65:**2961-2968.