Supplementary Information

Title: Microbial fermentation of organic carbon substrates drives rapid pH neutralisation and element removal in alkaline wastewater

Authors: Talitha C. Santini, Yong G. Peng

Contents: DNA extraction, sequencing, and data analysis methods, 6 tables, 1 figure, 10 pages

Materials and Methods

DNA extraction, sequencing, and data analysis

Microbial DNA was extracted using the MoBio PowerSoil DNA Isolation kit (MoBio, Carlsbad, CA), and PCR was performed with the Q5 Hot Start High-Fidelity 2X Master Mix Kit (New England Biolabs, Ipswich, MA), targeting the V6-V8 region of the 16S rRNA gene using modified versions of universal primers 926F (5'-AAACTYAAAKGAATTGRCGG-3') and 1392R (5'-ACGGGCGGTGWGTRC-3') (Matsuki *et al.*, 2002), and DNA sequencing was performed using the Illumina MiSeq platform and reagents according to manufacturer's protocols. DNA sequencing data was processed with QIIME (v 1.8.0) (Caporaso *et al.*, 2010). Quality filtering and screening steps involved removal of multiplex identifiers and primers, removal of chimeric sequences, sequences containing ambiguous base calls, sequences ≤ 150 bp in length, and sequences containing homo-polymer runs > 6 bp. Operational taxonomic units (OTUs) were defined by clustering at 97% similarity using an open reference OTU picking strategy. After sequence alignment, phylogenetic trees were created with FastTree (Price *et al.*, 2009) and taxonomy was assigned to OTUs using BLAST against a curated GreenGenes database (DeSantis *et al.*, 2006). An average of 26718 reads per sample were returned after trimming and quality filtering/screening. Samples were rarefied to a uniform depth of 4800 reads per sample, and relative abundances of OTUs were corrected for differences in 16S rRNA gene copy number using CopyRighter (v 0.46) (Angly *et al.*, 2014). Alpha diversity was compared between samples using Shannon (H'), Simpson, and Chao1 metrics. Community composition in each sample was visualized by nonmetric multidimensional scaling based on Bray-Curtis dissimilarities. PERMANOVA (Anderson, 2001) and PERMDISP (Anderson, 2006, Anderson *et al.*, 2006), implemented in PRIMER (v 7.0.10, with PERMANOVA+ v 1 add-in) (Clarke, 2015), were used to test for statistically significant differences in community composition and dispersion among treatments based on Bray-Curtis distance matrices, with permutations of residuals under a reduced model using 9999 permutations. Permutation P-values were used unless low unique permutations necessitated the use of Monte Carlo asymptotic P-values. Significant relationships between microbial community composition and fermentation products were identified using distance-based multivariate multiple regression (DistLM), implemented in PRIMER, with a 'best' selection procedure using 9999 permutations and Bayesian Information Criterion selection criterion for model parsimony. Key OTUs accounting for the majority of variation between communities were identified using SIMPER, implemented in PRIMER.

Additional references

Angly, F. E.; Dennis, P. G.; Skarshewski, A.; Vanwonterghem, I.; Hugenholtz, P.; Tyson, G. W. CopyRighter: a rapid tool for improving the accuracy of microbial community profiles through lineage-specific gene copy number correction. *Microbiome* **2014**, *2*, 11.

Caporaso, J. G.; Kuczynski, J.; Stombaugh, J.; Bittinger, K.; Bushman, F. D.; Costello, E. K.; Fierer, N.; Gonzalez Peña, A.; Goodrich, J. K.; Gordon, J. I.; Huttley, G. A.; Kelley, S. T.; Knights, D.; Koenig, J. E.; Ley, R. E.; Lozupone, C. A.; McDonald, D.; Muegge, B. D.; Pirrung, M.; Reeder, J.; Sevinsky, J. R.; Turnbaugh, P. J.; Walters, W. A.; Widmann, J.; Yatsunenko, T.; Zaneveld, J.; Knight, R. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **2010**, *7*, 335-336.

Matsuki, T.; Watanabe, K.; Fujimoto, J.; Miyamoto, Y.; Takada, T.; Matsumoto, K.; Oyaizu, H.; Tanaka, R. Development of 16S rRNA gene-targeted group specific primers for the detection and identification of predominant bacteria in human feces. *Appl. Environ. Microbiol.* **2002**, *68*, 5445-5451.

Price, M. N.; Dehal, P. S.; Arkin, A. P. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **2009**, *26*, 1641-1650.

Supplementary Information Table 1. Final leachate pH values (day 15) after bioremediation with different carbon sources; and different concentrations of added soil inoculant. Values displayed are the means of three replicates; error bars denote ± 1 standard error of the mean. Lower case letters indicate significant differences between treatments according to repeated measures ANOVA.

Treatment	Final pH
Carbon substrate type	
C	10.18 ± 0.01 e
CS	10.12 ± 0.01 e
G	7.38 ± 0.08 ab
B	7.09 ± 0.06 a
E	8.07 ± 0.09 bcd
W	8.93 ± 0.02 d
Soil inoculant addition level	
0%S	8.39 ± 0.38 cd
0.01%S	8.15 ± 0.41 bcd
0.1%S	7.80 ± 0.11 abc

Supplementary Information Table 2. Major and minor element concentrations in bauxite residue leachates after bioremediation. Values displayed are the means of three replicates ± 1 standard error of the mean. Lower case letters indicate significant differences between treatments according to one-way ANOVA with Tukey's HSD post-hoc test. "-" indicates concentration below detection limits.

Treatment	Ba	Ca	Cu	Fe	K	Mg	Mn	P	Se	Si	Sr	Zn
	mg/L	mg/L	mg/L	mg/L	mg/L	mg/L	mg/L	mg/L	mg/L	mg/L	mg/L	mg/L
Carbon substrate type												
C	0.032 ± 0.001 a	0.72 ± 0.03 a	0.0019 ± 0.002 a	0.005 ± 0.001 a	39.14 ± 0.78 a	-	-	0.84 ± 0.04 a	0.026 ± 0.005 abc	0.52 ± 0.01 a	0.050 ± 0.000 a	0.004 ± 0.001 a
CS	0.198 ± 0.002 b	10.1 ± 0.28 a	0.497 ± 0.017 b	2.33 ± 0.16 b	86.11 ± 2.76 ab	2.22 ± 0.13 a	0.310 ± 0.013 a	5.13 ± 0.04 ab	0.042 ± 0.012 bc	5.64 ± 0.53 ab	0.72 ± 0.027 b	0.13 ± 0.007 a
G	0.243 ± 0.015 bc	22.51 ± 3.83 a	0.249 ± 0.065 ab	7.26 ± 0.69 c	117.68 ± 6.66 ab	6.86 ± 0.63 a	1.25 ± 0.226 b	2.54 ± 0.56 ab	0.082 ± 0.008 d	8.32 ± 0.50 b	1.41 ± 0.047 c	0.32 ± 0.044 a
B	0.252 ± 0.023 bc	64.92 ± 4.91 b	0.310 ± 0.025 ab	7.38 ± 0.97 c	1181.08 ± 33.1 d	23.0 ± 0.82 b	2.390 ± 0.206 c	12.89 ± 0.74 b	0.082 ± 0.002 d	15.1 ± 2.27 c	2.26 ± 0.224 d	0.45 ± 0.088 a
E	0.292 ± 0.016 c	49.81 ± 9.33 b	0.0081 ± 0.0003 a	0.27 ± 0.05 ab	677.65 ± 27.38 c	74.2 ± 5.29 d	0.930 ± 0.100 b	46.05 ± 2.26 d	0.047 ± 0.003 c	39.0 ± 1.44 e	0.73 ± 0.05 b	0.23 ± 0.03 a
W	0.467 ± 0.029 d	141.17 ± 11.8 c	1.006 ± 0.276 c	5.25 ± 0.26 c	144.81 ± 6.86 b	32.8 ± 1.83 c	0.930 ± 0.040 b	24.46 ± 2.54 c	0.051 ± 0.004 c	24.0 ± 1.54 d	3.84 ± 0.26 e	1.32 ± 0.34 b
Soil inoculant addition level												
0%S	0.007 ± 0.002 a	0.43 ± 0.26 a	0.0031 ± 0.0004 a	0.02 ± 0.01 a	104 ± 12.3 ab	0.28 ± 0.14 a	0.001 ± 0.001 a	12.1 ± 3.44 b	0.014 ± 0.002 ab	0.96 ± 0.08 a	0.036 ± 0.001 a	0.087 ± 0.021 a
0.01%S	0.013 ± 0.001 a	0.23 ± 0.18 a	0.0017 ± 0.0003 a	0.18 ± 0.01 a	97.5 ± 6.18 ab	0.08 ± 0.02 a	-	5.93 ± 0.90 ab	0.010 ± 0.004 a	0.83 ± 0.05 a	0.037 ± 0.001 a	0.055 ± 0.009 a
0.1%S	0.015 ± 0.003 a	2.13 ± 0.81 a	0.0035 ± 0.0004 a	1.11 ± 0.42 ab	116.9 ± 20.2 ab	0.50 ± 0.15 a	0.033 ± 0.012 a	11.3 ± 4.38 ab	0.015 ± 0.001 ab	1.20 ± 0.14 a	0.050 ± 0.002 a	0.085 ± 0.011 a

Supplementary Information Table 3. Major and minor element concentrations and total carbon and nitrogen concentrations in carbon substrates and soil inoculum before use in bioreactors. Samples are: S: soil inoculum; B: banana; E: eucalyptus mulch; W: woodchips. Values displayed are the means of three replicates ± 1 standard error of the mean. Samples marked with the same lower case letter in individual element concentration columns are not significantly different according to one-way ANOVA. Tukey's HSD was used to separate means. "-" indicates concentration below detection limits.

Treatment	Al wt %	Ca wt %	Fe wt %	K wt %	Mg wt %	Na wt %	P wt %	S wt %	Si wt %	C wt %	N wt %
S	2.323 ± 0.387 b	0.707 ± 0.055 b	1.782 ± 0.187 b	1.24 ± 0.20 bc	0.05 ± 0.02 a	0.843 ± 0.172 c	0.060 ± 0.001 a	0.051 ± 0.002 ab	32.5 ± 0.66 d	5.17 ± 0.10 a	0.24 ± 0.01 a
B	0.003 ± 0.001 a	0.090 ± 0.008 a	0.004 ± 0.002 a	3.64 ± 0.11 d	0.19 ± 0.01 bc	0.032 ± 0.014 a	0.113 ± 0.006 b	0.031 ± 0.005 a	0.22 ± 0.17 ab	36.8 ± 0.39 b	0.60 ± 0.01 b
E	0.005 ± 0.002 a	0.988 ± 0.030 c	0.008 ± 0.001 a	1.72 ± 0.07 c	0.28 ± 0.04 c	0.700 ± 0.028 bc	0.100 ± 0.001 b	0.216 ± 0.006 ab	0.04 ± 0.01 a	50.1 ± 0.29 c	1.52 ± 0.02 d
W	0.176 ± 0.009 a	1.106 ± 0.031 c	0.429 ± 0.036 a	0.22 ± 0.01 a	0.10 ± 0.01 ab	0.071 ± 0.002 a	0.073 ± 0.002 a	0.080 ± 0.004 ab	5.27 ± 0.38 c	38.4 ± 0.57 b	0.69 ± 0.01 c

Treatment	As ppm	Ba ppm	Cu ppm	Mn ppm	Mo ppm	Se ppm	Sr ppm	V ppm	Zn ppm
S	2.29 ± 0.06 a	207 ± 14.1 c	31.2 ± 3.53 c	368 ± 37.9 c	1.25 ± 0.37 a	0.38 ± 0.03 a	25.6 ± 1.61 a	185 ± 3.31 c	128 ± 4.20 b
B	3.86 ± 0.58 ab	2.78 ± 0.91 a	7.42 ± 0.23 a	11.1 ± 1.37 a	0.96 ± 0.32 a	0.25 ± 0.04 a	16.5 ± 3.47 a	0.99 ± 0.07 a	4.83 ± 0.69 a
E	1.91 ± 0.05 a	6.78 ± 0.74 a	5.61 ± 0.31 a	22.8 ± 0.16 a	1.49 ± 0.24 a	0.32 ± 0.02 a	19.1 ± 1.29 a	0.79 ± 0.05 a	14.2 ± 0.76 a
W	5.13 ± 0.70 b	68.1 ± 4.43 b	29.8 ± 1.22 bc	155 ± 5.43 b	0.67 ± 0.17 a	0.35 ± 0.01 a	66.1 ± 1.64 c	16.3 ± 0.87 b	165 ± 6.25 c

Supplementary Information Table 4. Dissolved inorganic (DIC) and organic carbon (DOC) concentrations in bauxite residue leachates after bioremediation. Values displayed are the means of three replicates ± 1 standard error of the mean. Lower case letters indicate significant differences between treatments according to one-way ANOVA with Tukey's HSD post-hoc test.

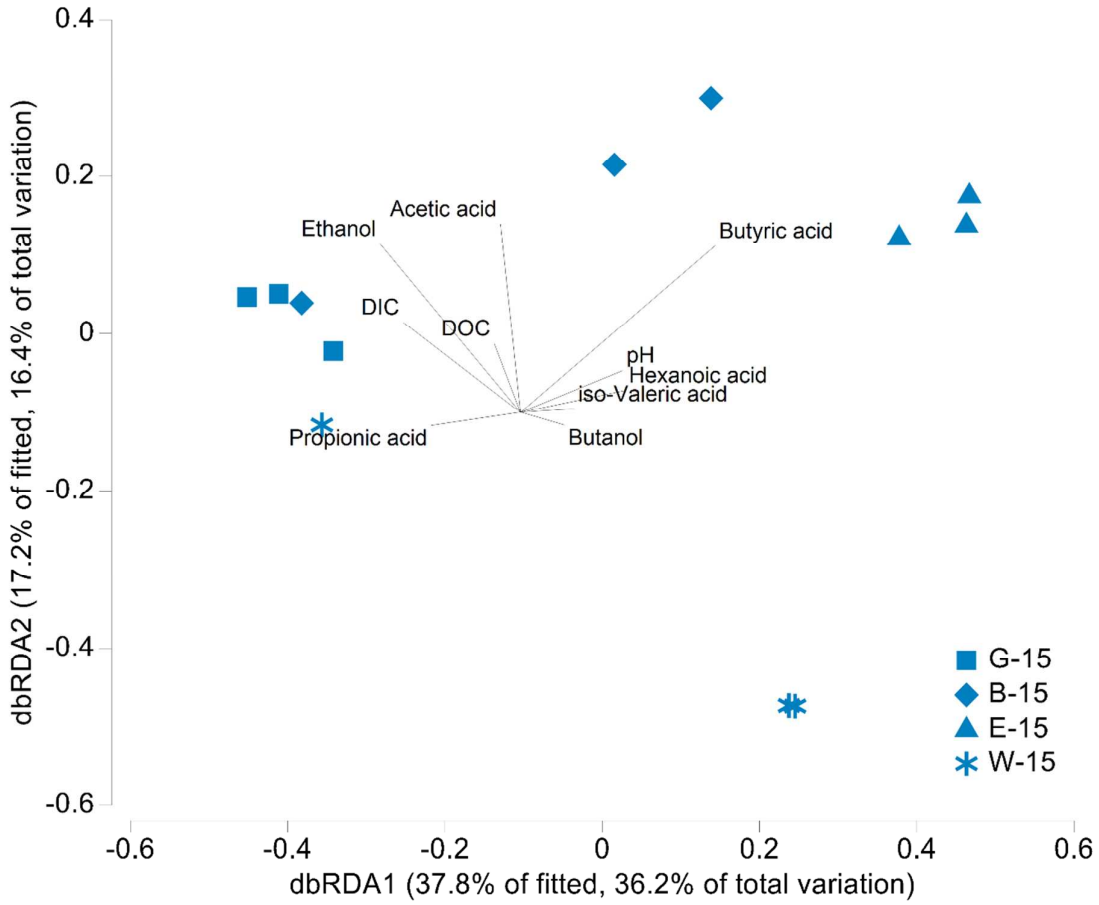
Treatment	Dissolved inorganic carbon (DIC)	Dissolved organic carbon (DOC)
	g/L	g/L
Carbon substrate type		
C	3.55 ± 0.04 c	0.26 ± 0.009 a
CS	3.30 ± 0.04 c	0.65 ± 0.02 ab
G	1.85 ± 0.09 a	5.31 ± 0.11 de
B	1.86 ± 0.02 a	6.55 ± 0.18 e
E	1.75 ± 0.08 a	4.98 ± 0.13 cde
W	2.41 ± 0.12 ab	2.84 ± 0.64 bc
Soil inoculant addition level		
0%S	2.94 ± 0.05 bc	6.00 ± 0.25 de
0.01%S	3.27 ± 0.14 c	4.27 ± 0.26 cd
0.1%S	3.53 ± 0.46 c	4.01 ± 1.10 cd

Supplementary Information Table 5. Alpha diversity metrics for microbial communities in bioreactors before and after incubation, and in initial water and soil samples. Values displayed are the mean of three replicates ± 1 standard error of the mean. Treatments are: C-0: control (leachate only), time 0 days; CS-0: control plus soil inoculant, time 0 days; S-0: soil inoculant, time 0 days; G-0: glucose plus soil inoculant, time 0 days; B-0: banana plus soil inoculant, time 0 days; E-0: eucalyptus mulch plus soil inoculant, time 0 days; W-0: woodchips plus soil inoculant, time 0 days; G-15: glucose plus soil inoculant, time 15 days; B-15: banana plus soil inoculant, time 15 days; E-15: eucalyptus mulch plus soil inoculant, time 15 days; W-15: woodchips plus soil inoculant, time 15 days. Treatments marked with the same lower case letter in individual columns are not significantly different according to one-way ANOVA. Tukey's HSD was used to separate means.

Treatment	Species richness	Shannon diversity	Reciprocal Simpson diversity	Faith's PD
C-0	617 ± 29 b	5.77 ± 0.24 b	10.5 ± 3.22 ab	10.7 ± 1.20 ab
CS-0	1150 ± 58 d	8.18 ± 0.15 d	74.7 ± 13.3 d	10.7 ± 0.53 ab
S-0	2132 ± 28 e	9.62 ± 0.01 e	189 ± 1.90 e	16.9 ± 0.54 bc
G-0	747 ± 15 b	7.77 ± 0.01 cd	59.8 ± 3.07 cd	18.6 ± 0.67 c
B-0	147 ± 45 a	0.82 ± 0.18 a	1.21 ± 0.03 a	4.37 ± 1.15 a
E-0	791 ± 10 bc	7.65 ± 0.06 cd	52.9 ± 7.19 bcd	17.9 ± 0.79 c
W-0	2833 ± 18 f	10.8 ± 0.02 e	974 ± 20.6 f	20.5 ± 1.97 c
G-15	749 ± 24 b	7.64 ± 0.10 cd	61.0 ± 3.79 cd	17.9 ± 0.25 c
B-15	691 ± 62 b	6.36 ± 0.64 b	21.1 ± 6.96 abc	10.1 ± 2.74 ab
E-15	770 ± 8 bc	6.70 ± 0.13 bc	27.6 ± 3.82 abc	5.65 ± 0.48 a
W-15	1065 ± 170 cd	6.96 ± 0.30 bcd	25.8 ± 9.28 abc	14.4 ± 2.12 bc

Supplementary Information Table 6. Relative abundances of *Firmicutes* OTUs involved in fermentation of various organic carbon substrates supplied to bioreactors for pH neutralisation, as identified by SIMPER analysis. Relative abundance is expressed as a percentage of total sequence reads. Values displayed are the mean of three replicates ± 1 standard error of the mean.

OTU	Glucose	Banana	Eucalyptus	Woodchips
Ruminococcaceae sp.	3.52			2.87
Thermoanaerobacter sp.	3.31			
Enterococcus sp. 1	1.94	1.15		
Enterococcus sp. 2		14.4		
Coprococcus sp.		11.0	26.0	8.29
Lachnospiraceae sp. 1		5.85	21.5	
Clostridiaceae sp.		1.65		
Natronincola sp.		1.23	13.0	3.37
Bacillus sp.		1.13		
Lachnospiraceae sp. 2			4.26	
Clostridiales sp.				3.53



Supplementary Information Figure 1. Ordination plot (DistLM results, as visualised by dbRDA) showing relationships between microbial community composition in bioreactors after incubations, and markers (pH, inorganic and organic carbon) and products of fermentation. Treatments are: G-15: glucose plus soil inoculant, time 15 days; B-15: banana plus soil inoculant, time 15 days; E-15: eucalyptus mulch plus soil inoculant, time 15 days; W-15: woodchips plus soil inoculant, time 15 days. Samples were collected and analysed from triplicate bioreactors. Propanol yields were strongly positively correlated with ethanol; and iso-butyric, iso-valeric, and valeric acids were all strongly positively correlated with each other, and therefore were removed from the DistLM analysis to avoid redundancy.