Supplementary Materials and Methods

Reactor setup

Experiments were conducted using single chamber reactors made of polycarbonate as previously described (1). Previous studies demonstrated that the single chamber MEC is reliable approach for producing hydrogen and is widely used in the scientific field (2-3). Recently, MEC reactors have been developed as a promising technology to practical wastewater treatment (4-5). Methane production is always accompanied hydrogen in single chamber MECs because exoelectrogens and methanogens can grow well as main functional communities in the given niche (6-7). Efforts were made to regulate and control energy gas production in special operations, but gas generation varied little when reactors were set up under stable conditions (8-9). Thus, the single chamber reactors were used in this study. All the reactors used in this study were manufactured at the same time by the same vendor using the same batches of materials so that the new reactors manufactured should be identical.

The total volume was 43 mL, consisting of a 28.5 ml chamber (3 cm inner diameter and 4 cm long) and a tube attached to the top of the reactor (1.6 cm inner diameter and 7.5 cm length; 14.5 mL capacity). The anode was a graphite brush (25 mm diameter × 25 mm length; 0.22 m^2 surface area; fiber type: PANEX 33 160K, ZOLTEK), with a specific surface area of 18,200 m^2/m^3 and a porosity of 95%. The cathode was made from carbon cloth (type B; E-TEK); the surface area was 7 cm^2 with a Pt catalyst layer (0.5 mg/cm^2) in one side. The cathode was placed opposite to anode brush while it was glued and sealed completely from the beginning for MEC operation (6).

Reactor operation and measurements

Fourteen single chamber reactors were set up and inoculated with a mixture of wastewater (50%) from the Wastewater Treatment Plant in Norman, Oklahoma, and growth medium (50%). The collected wastewater was settled down to remove big particles with no flocs observed, and hence the wastewater was homogeneous. The growth medium contained a phosphate buffer (PBS, 50 mM; pH = 7.0, conductivity = 7.5 mS cm^{-1}), nutrient solution (NH_4Cl, 310 mg L^{-1}; KCl, 130 mg L^{-1}; a trace nutrient minerals: Nitrilotriacetic acid 1.5 g ; $MgSO_4 \cdot 7H_2O$ 3.0 g ; $MnSO_4 \cdot H_2O$ 0.5 g ; NaCl 1.0 g ; $FeSO_4 \cdot 7H_2O$ 0.1 g ; $CoCl_2 \cdot 6H_2O$ 0.1 g ; $CaCl_2$ 0.1 g ; $ZnSO_4 \cdot 7H_2O$ 0.1 g ; $CuSO_4 \cdot 5H_2O$ 0.01 g ; $AlK(SO_4)_2 \cdot 12H_2O$ 0.01 g ; H_3BO_3 0.01 g ; $Na_2MoO_4 \cdot 2H_2O$ 0.01 g ; Distilled water 1.0 L.), and glucose (1000 mg L^{-1}) as the carbon source. All reactors were started up as replicates in direct MEC mode (10) at fixed applied voltages of 0.7 V (model 3645A; Circuit Specialists, Inc.) .During startup, all reactors were fed initially with a mixture of the wastewater inoculum (~13 ml) and growth medium (~ 13 ml). The reactors were incubated at a room temperature (~22 °C) for 48 hours to allow microorganisms in the wastewater to randomly colonize the anode brushes. To enhance biofilm establishment on the anode, the original reactor solution were replaced with the same fresh mixture of the wastewater and growth medium and incubated under the same condition for another 48 hours. After biofilm establishment, all reactors were operated in a fed-batch mode with a 24-hour cycle, that is, replacing the reactor solution with a fresh growth medium every 24 hours for about two months. After medium change each time, the chambers were purged using extreme pure N_2 (99.998%) for 10 minutes to remove oxygen. All of these reactors were run under identical conditions in terms of carbon substrates, medium compositions, medium pH, incubation temperature, voltage controls, and other operational conditions. Thus, technical variations of these bioreactors should be very small.

Gas was collected in gas bags (0.1 L capacity; Cali- 5-Bond, Calibrated Instruments) for analysis. Voltages were measured over a 5 ohm resister in each circuit using a multimeter (model 2700; Keithley Instruments). The gas samples were taken for gas chromatography analysis of hydrogen, carbon dioxide, methane from the gas bag using a syringe (1 mL, Sample-Lock Syringe). The concentration was determined and used to calculate gas yields.

DNA extraction, purification and quantitation

At the end of each batch cycle, the planktonic biomass was collected by centrifugation (8,000 ×g, 10 min) from all discharged solution (~25 mL) of each reactor. The samples were stored at -80^{o}C prior to DNA extraction and analysis. The community DNA was extracted by freeze-grinding mechanical lysis as described previously (11), and was purified using and the Wizard DNA Clean-up System (Promega, Madison, WI). DNA quality was assessed by the ratios of 260 nm/280 nm, and 260/230 nm using a NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies Inc., Wilmington, DE), and final DNA concentrations were quantified with PicoGreen (12) using a FLUOstar Optima (BMG Labtech, Jena, Germany).

Template amplification and labeling

In order to produce consistent hybridizations from all samples, a whole community genome amplification (WCGA) was used to generate approximately 3.0 μg of DNA with 100 ng purified DNA as the template using the TempliPhi Kit (GE Healthcare, Piscataway, NJ) following the manufacturer's instructions. Single-strand binding protein (267 ng μL^{-1}) and spermidine (0.1 mM) were also added to the reaction mix to improve the amplification efficiency. The reactions were incubated at 30 °C for ~3 hours and stopped by heating the mixtures at 65 °C for 10 min.

After amplification, all products were labeled with the fluorescent dye Cy-5 using random priming method as follows. First, the whole amplified products were mixed with 20 μL random primers, denatured at 99.9 °C for 5 min, and then immediately chilled on ice. Following denaturation, the labeling master mix containing 2.5 μL dNTP (5 mM dAGC-TP, 2.5 mM dTTP), 1 μL Cy-5 dUTP (Amersham, Piscataway, NJ), 80 U of the large Klenow fragment (Invitrogen, Carlsbad, CA), and 2.5 μL water were added and then incubated at 37 °C for 3 hours, followed by heating at 95 °C for 3 min. Labeled DNA was purified using the QIA quick purification kit (Qiagen, Valencia, CA) according to the manufacturer's instructions, measured on a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies Inc., Wilmington, DE), and then dried down in a SpeedVac (ThermoSavant, Milford, MA) at 45 °C for 45 min.

Hybridization and imaging processing

The labeled target was resuspended in 120 μl hybridization solution containing 50% formamide, 3 x SSC, 10 μg of unlabeled herring sperm DNA (Promega, Madison, WI), and 0.1% SDS, and the mix was denatured at 95°C for 5 min and kept at 50°C until it was deposited directly onto a microarray. Hybridizations with GeoChip 3.0 (13) were performed for 12 hours at 42ºC and 50% formamide with a TECAN Hybridization Station HS4800 Pro (TECAN, US) according to the manufacturer's recommended method. After washing and drying, the microarray was scanned by ScanArray Express Microarray Scanner (Perkin Elmer, Boston, MA) at 633 nm using a laser power of 90% and a photomultiplier tube (PMT) gain of 75%. The ImaGene version 6.0 (Biodiscovery, El Segundo, CA) was then used to determine the intensity of each spot, and identify poor-quality spots. Many of our previous studies indicated that GeoChip hybridization-based detection is quantitative (14-18).

Data pre-processing

Raw data from ImaGene were submitted to Microarray Data Manager in our website (http://ieg.ou.edu/microarray/) and analyzed using the data analysis pipeline with the following major steps: (**i**) The spots flagged as 1 or 3 by Imagene and with a signal to noise ratio (SNR) less than 1.5 were removed as poor-quality spots; (**ii**) After removing the bad spots, normalized intensity of each spot was calculated by dividing the signal intensity of each spot by the mean intensity of the microarray; (**iii**) If any of replicates had (signal–mean) more than two times the standard deviation, this replicate was moved as an outlier. This process continued until no such replicates were identified; and (**iv**) At least 0.34 time of the final positive spots (probes), or a minimum of two spots was required for each gene.

Dissimilarity metrics

To determine the site-to-site variability in microbial community composition of these reactors, known as β-diversity, six different metrics for measuring β-diversity were evaluated, including Jaccard, Sorensen, Bray-Curtis, Chao's Jaccard, Chao's Sorensen, and Morisita indexes. Since most of the methods yielded similar results in PCA ordination, the complement of two commonly used similarity metrics, Sorensen's incidence-based (D_S) and Bray-Curtis's abundance-based (D_B) methods were presented. If abundance is ignored, Bray-Curtis measurement will reduce to Sorensen measurement. Therefore, Bray-Curtis index is so-called "Sorenson's quantitative index" (19).

Statistical comparison of community functional structure among different groups of reactors

Because dissimilarity values within a group are not independent and they may not follow normal distributions, standard MANOVA test is not appropriate. Thus, three nonparametric multivariate permutation tests, MRPP, ANOSIM and adonis, were performed to assess whether there are significant difference of microbial community functional structure among the four reactor groups defined by DCA ordination analysis. Notice that adonis function in R-program is also named as Permutation Multivariate Analysis of Variance (PERMANOVA) proposed by Anderson (20). All three methods are based on both Sorensen and Bray-Curtis dissimilarities among samples and their rank order in different ways to calculate test statistics, and the Monte Carlo permutation (1,000 permutations) is used to test the significance of statistics. All three tests (anosim, adonis and mrpp) were performed with the Vegan package (v.1.15-1) in R v. 2.8.1 (R Development Core Team, Vienna, Austria). In addittion, SIMPER (similarity percentage) analysis in PAST package (PAlaeontological Statistics) was used to quantify the contribution of each gene/functional group to the Bray-Curtis dissimilarity.

Comparison of the β-diversity among reactors to the null expectation.

To determine what types of the assembly mechanisms cause high β-diversity in these reactor communities, a null model analysis was performed. If stochastic processes or the interactions of stochastic processes and priority effects leading to multiple stable equilibria play predominant roles in community assembly, the β-diversity observed will be statistically indistinguishable from the random null expectation. Otherwise, the β-diversity observed will be significantly lower than the null expectation (21). Similar to the randomization approach with EcoSim (21, 22), the null community is generated by randomly shuffling the original community with the independent swap algorithm by holding the number of genes/populations in each reactor and the number of

reactors in which each gene/population appears constant. Because our microarray dataset is too large to efficiently compute in EcoSim, the randomizeMatrix function of R-program was carried out to generate 1000 null communities. Then, the Sorensen and Bray-Curtis dissimilarities were calculated based on these null communities. The average of β-diversity over these 1000 null communities is considered as the null-expected β-diversity. Finally, Permutation Analysis of Multivariate Dispersions (PERMDISP) was used to test the significance of the difference between the observed β-diversity in these reactors and the average null-expected β-diversity.

Multivariate and direct gradient analysis

Detrended correspondence analysis (DCA) was used to determine the overall functional changes in the microbial communities. DCA is an ordination technique that uses detrending to remove the arch effect, where the data points are similar to a horse-shoe-like shape, in Correspondence analysis (23). The DCA analysis was performed using the package vegan in R project (www.r-project.org). Also standard analysis of variance (ANOVA) was used to test the functional differences (i.e., H_2, CH_4, and CO_2) among the four groups of these reactors.

To determine the potential linkages between microbial community structure and functioning, Mantel test was performed. Both Sorensen and Bray-Curtis dissimilarities were used to the measure the differences in community structure whereas Euclidean distance index was used to measure the differences of the production of various gases. The function 'mantel' in the packages of vegan and ecodist of *R* (www.r-project.org) was used to perform these tests with 10,000 permutations.

To determine the relative contributions of stochastic and deterministic factors to the assembly of these reactor communities, a canonical correspondence analysis (CCA)-based

variation partitioning analysis (VPA) were performed. We began with a global analysis, which included all genes detected and 5 environmental variables (effluent pH, current generation efficiency, H_2, CH_4, and CO_2). CCA test with 1,000 times of permutations was used to select most significant deterministic variables. Four of these variables were significant predictors of community functional composition in this global analysis (forward selection, α=0.1), and only these were retained for VPA analysis. Then, VPA was performed determine the relative importance of abiotic and biotic factors in shaping the functional structure of these microbial communities. We partitioned the variations in community composition into four partitions: (**i**) deterministic biotic portion (H_2, CH_4, CO_2); (**ii**) deterministic abiotic portion (effluent pH); (**iii**) abiotic-biotic interactive portion; and (**iv**) unexplained portion. All these analyses were carried out using the functions in the vegan package (v.1.15-1) in R.

Fitting neutral community models

To differentiate the influences of pure stochastic processes from the interactions of both stochastic and deterministic processes on the assembly of these MEC communities, the experimental data was tested against the neutral community models (24, 25). If these reactor communities are solely controlled by ecological drift (e.g., stochastic colonization, birth, death, and extinction), the gene abundance distributions follow neutral dynamics as predicted by neutral community models (25, 26). The Etienne's model (25) contains three key parameters: the taxon (here a taxon means gene/population or species) diversity in the regional taxon pool (θ), the immigration probability (m), and N samples from N local communities, each of these communities contains J_i individuals, ($\vec{J} = (J_1, J_2, \ldots, J_N)$). These parameters are estimated using the maximum likelihood function of the Etienne's model

$$P(\vec{D}|I,\theta,\vec{J}) = \frac{1}{\prod_{\vec{l}} \Phi_{\vec{l}}!} \prod_{i=1}^{N} \frac{J_i!}{(I_i)_{J_i} \prod_{k=1}^{S} n_{ik}!} \sum_{A} M(\vec{D},A) \frac{I^A \theta^S}{(\theta)_A}$$

Where $\vec{D} = (n_{11}, n_{12}, \ldots, n_{NS})$ is the taxon abundance distribution for S taxon, n_{ik} is the number individuals of taxon k in sample i, $\Phi_{\vec{l}}$ is the number of taxon with abundance vector $\vec{l}$ across the samples, I_i is the number of immigrants ($I_i = \frac{m(J_i-1)}{(1-m)}$) of the i^{th} community, and A is the number of total ancestors. $M(D, A)$ is defined as

$$M(D,A) := \sum_{\{a_{11},a_{12},\ldots a_{NS} | \sum_{i,k} a_{ik} = A\}} \prod_{k=1}^{S} \left[(a_k - 1)! \prod_{i=1}^{N} \overline{s}(n_{ik}, a_{ik}) \right],$$

where a_{ik} represent the number of ancestors of taxon k in sample i. Etienne combined Ewens sampling distributions (with parameter θ and I_i) by summing over all abundance vectors of the ancestors in the local community to get the sampling formula $P(\vec{D}|I,\theta,\vec{J})$. The algorithm for estimating these parameters is to firstly calculate the maximum likelihood estimator for θ by numerical computation, and then use the estimator of θ to calculate the estimator of immigration rate m. All of the above processes were carried out with the software PARI, coded by Etienne (25).

After obtaining the fitted parameters, we used Volkov et al.'s formula (24, 25) to calculate the expected frequency of taxon based on the fitted parameters for the local community. It was assumed that the number of taxon in a metacommunity is extremely large, and derived the relative taxon abundance for the i^{th} local community is

$$\langle \phi(n) \rangle_i = \theta \frac{J_i!}{n!\,(J_i - n)!} \frac{\Gamma(I_i)}{\Gamma(J_i + I_i)} \int_0^{I_i} \frac{\Gamma(n+y)}{\Gamma(1+y)} \frac{\Gamma(J_i - n + I_i - y)}{\Gamma(I_i - y)} \exp(-y\theta/I_i) \mathrm{d}y$$

where $\langle \phi(n) \rangle_i$ is the average number of taxon containing *n* individuals in the local community *i*, and $\Gamma(\cdot)$ is a gamma function $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$.

Finally, we used both χ^2-test and Kolmogorov-Smirnov (KS) test to assess the difference between the observed distribution and expected frequencies distribution generated from neutral model. χ^2-test is popular to assess the neutral model fitting (24, 25), whereas KS test is based on empirical cumulative distribution and it is independent of sample size. Thus, we used both χ^2-test and KS test in this study.

***χ^2-test*:**

The χ^2-test is defined by

$$Q = \sum_{n=1}^{S} \frac{(O_n - E_n)^2}{E_n},$$

where E_n is the expected frequency of gene with *n* abundance obtained by Volkov's approach in this study, and O_n is the observed frequency of gene with *n* abundance obtained from the observed dataset. The null hypothesis is rejected at level α if $Q > C_\alpha$, where C_α is the critical value of chi-square distribution.

Kolmogorov-Smirnov (KS) test:

Suppose there are S_1 genes with ordered abundance $(X_1, X_2, \ldots, X_{S_1})$ in the observed sample and S_2 genes with ordered abundance $(Y_1, Y_2, \ldots, Y_{S_2})$ in the expected sample. To compare the difference between observed and expected distribution by KS test, first we calculated the

empirical cumulative frequency (normalized by the sample size) of the observations as a function of classes $E_X(k)=Z_X(k)/S_1$, where $Z_X(k)$ is the number of points less than X_k. Similarly, we have the empirical cumulative frequency distribution for the expected abundance sample $E_Y(k)=Z_Y(k)/S_2$. The KS test statistics for two samples comparison is defined as

$$d = \max_k |E_X(k) - E_Y(k)|.$$

The null hypothesis is rejected at level α if $\sqrt{\frac{S_1 S_2}{S_1+S_2}} d > K_\alpha$, where K_α is the critical value of Kolmogorov distribution.

The above process is available as volkov function for model fitting and ks.boot function for testing in R-program.

Reference

1. **Call, D. and B. E. Logan**. 2008. Hydrogen production in a single chamber microbial electrolysis cell lacking a membrane. Environ. Sci. Technol. **42**(9):3401-3406.
2. **Logan, B.E. and K. Rabaey**. 2012. Conversion of wastes into bioelectricity and chemicals using microbial electrochemical technologies. Science, 337:686-690.
3. **Liu H, HQ Hu, J Chignell and YZ Fan**. 2010. Microbial electrolysis: novel technology for hydrogen production from biomass. Biofuels. 1(1), 129–142
4. **Rozendal, R. A.; Hamelers, H. V. M.; Rabaey K.; Keller J.; Buisman, C. J. N.** 2008.Towards practical implementation of bioelectrochemical wastewater treatment. Trends Biotechnol. 26: 450-459
5. **Logan BE.** 2010. Scaling up microbial fuel cells and other bioelectrochemical systems. Applied Microbiol Biotechnol, 85: 1665–1671
6. **Cheng, S. H. Liu and B. E. Logan.** 2006. Increased performance of single-chamber microbial fuel cells using an improved cathode structure. Electrochem. Commun. **8**(3):489-494.
7. **Wang AJ, Liu WZ, Cheng SA, Xing DF, Zhou JZ, Logan BE.** 2009. Source of methane and methods to control its formation in single chamber microbial electrolysis cells. International Journal of Hydrogen Energy. 34: 3653-3658
8. **Cheng SA, DF Xing, D Call, BE Logan**. 2009. Direct biological conversion of electrical current into methane by electromethanogenesis. Environmental Science & Technology, 43(10): 3953-3958.

9. **Lu L, NQ Ren, X Zhao, H Wang, D Wu and DF Xing**. 2011. Hydrogen production, methanogen inhibition and microbial community structures in psychrophilic single-chamber microbial electrolysis cells. Energy & Environmental Science, 4, 1329-1336
10. **Liu, W., A. Wang, S. Cheng, B. E. Logan, H. Yu, Y. Deng, J. D. Nostrand, L. Wu, Z. He and J. Zhou.** 2010. Geochip-Based Functional Gene Analysis of Anodophilic Communities in Microbial Electrolysis Cells under Different Operational Modes. Environ. Sci. Technol. **44**(19):7729-7735.
11. **Zhou, J. Z., M. A. Bruns and J. M. Tiedje**. 1996. DNA recovery from soils of diverse composition. Appl. Environ. Microb. **62**(2):316-322.
12. **Ahn, S. J., J. Costa and J. R. Emanuel.** 1996. PicoGreen quantitation of DNA: Effective evaluation of samples pre- or post-PCR. Nucleic Acids Res. **24**(13):2623-2625.
13. **He, Z. L., Y. Deng, J. D. Van Nostrand, Q. Tu, M. Xu, C. L. Hemme, X. Li. L. Wu, T. J. Gentry, Y. Yin, J. Liebich, T. C. Hazen and J. Zhou.** 2010. GeoChip 3.0 as a high-throughput tool for analyzing microbial community composition, structure and functional activity. ISME J. **4**(9):1167-1179.
14. **He, Z., M. Xu, Y. Deng, S. Kang, L. Kellogg, L. Wu, J. D. Van Nostrand, S. E. Hobbie, P. B. Reich and J. Zhou.** 2010. Metagenomic analysis reveals a marked divergence in the structure of belowground microbial communities at elevated CO2. *Ecol. Lett.* **13**(5):564-575.
15. **Wang, F., H. Zhou, J. Meng, X. Peng, L. Jiang, P. Sun, C. Zhang, J. D. Van Nostrand, Y. Deng, Z. He, L. Wu, J. Zhou and X. Xiao.** 2009. GeoChip-based analysis of metabolic diversity of microbial communities at the Juan de Fuca Ridge hydrothermal vent. P. Natl. Acad. Sci. USA **106**(12):4840-4845 .
16. **Rhee, S. K., X. Liu, L. Wu, S. C. Chong, X. Wan and J. Zhou.** 2004. Detection of genes involved in biodegradation and biotransformation in microbial communities by using 50-mer oligonucleotide microarrays. Appl. Environ. Microbiol. **70**(7):4303-4317.
17. **Tiquia, S. M, L. Wu, S. C. Chong, S. Passovets, D. Xu, Y. Xu, J. Zhou.** 2004. Evaluation of 50-mer oligonucleotide arrays for detecting microbial populations in environmental samples. Biotechniques **36**(4):664.
18. **Wu, L., X. Liu, C. W. Schadt and J. Zhou**. 2006. Microarray-based analysis of subnanogram quantities of microbial community DNAs by using whole-community genome amplification. Appl. Environ. Microbiol. **72**(7):4931-4941.
19. **Magurran AE** ed. 2004. *Measuring Biological Diversity* (Blackwell, Oxford , U.K.).
20. **Anderson, M. J**. 2001. A new method for non-parametric multivariate analysis of variance. Austral. Ecol. **26**(1):32-46.
21. **Chase, J. M**. 2010. Stochastic Community Assembly Causes Higher Biodiversity in More Productive Environments. Science 328(5984):1388-1391.
22. **Gotelli, N. J. and G. L. Entsminger**. 2003. Swap algorithms in null model analysis. Ecology 84(2):532-535.
23. Hill, M. O. and H. G. Gauch. 1980. Deterended correspondence analysis, an improved ordination technique. Vegetatio 42:47-58.
24. **Volkov, I., J. R. Banavar, S. P. Hubbell and A. Maritan.** 2003. Neutral theory and relative species abundance in ecology. Nature **424**(6952):1035-1037.
25. **Etienne, R. S.** 2007. A neutral sampling formula for multiple samples and an 'exact' test of neutrality. Ecol. Lett. 10(7):608-618.

26. **Woodcock, S., C. J. van der Gast, T. Bell, M. Lunn, T. P. Curtis, I. M. Head, W. T. Sloan.** 2007. Neutral assembly of bacterial communities. FEMS Microbiol. Ecol. **62**(2):171-180.