1          *Supporting Materials for*

2

3    **Stochastic extremes but convergent recovery of bacterial and archaeal soil communities**
4                   **to an ongoing subterranean coal mine fire**

5

6      Sang-Hoon Lee, Jackson W Sorensen, Keara L Grady, Tammy C Tobin, and Ashley Shade

7    **Supporting Figure 1**. PCoA showing the variability among technical replicates. Three replicate

8    DNA extractions, amplifications and sequencing reactions were performed per soil, and these

9    sequences were subsequently pooled into one aggregate set of sequences to achieve deep

10   coverage of the community within each soil. Error bars are standard deviation around the mean

11   weighted UniFrac distance among technical replicates, each subsampled to an even 53,000

12   sequences per replicate.

## Weighted UniFrac PCoA

PCoA2: 12.7% var. explained

PCoA1: 76.3% var. explained
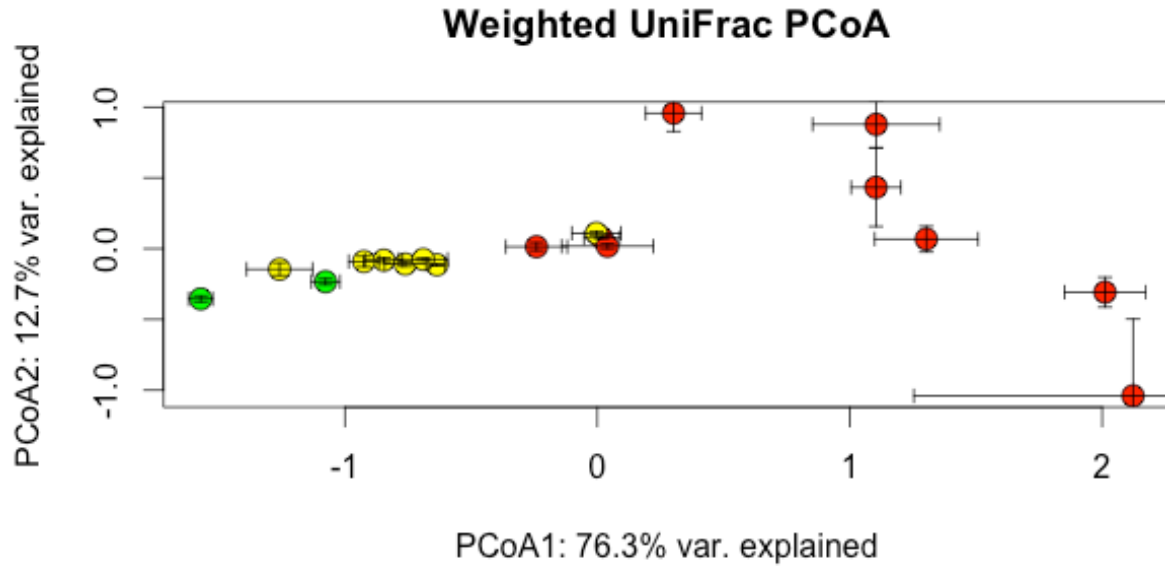
13

14    **Supporting Figure 2**.   Quantification of (**A**) 16S rRNA copies and (**B**) cell counts in fire-
15    affected, recovered, and reference soils. 16S rRNA copies were assessed using quantitative
16    PCR, and cell counts were assessed using cell separation from soil, staining and microscope
17    imaging.
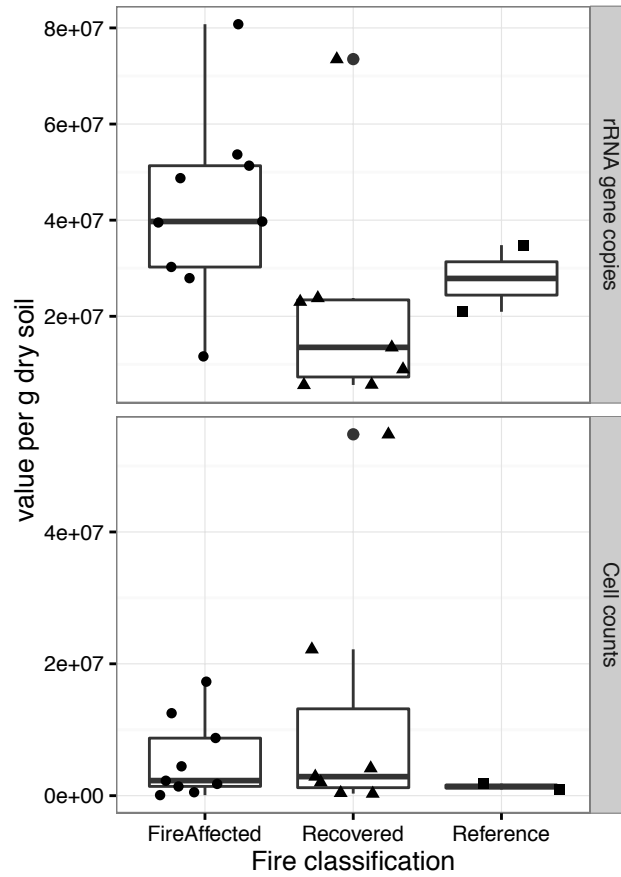


18

19

20  **Supporting Figure 3**. Centralia 16S rRNA amplicon sequencing effort assessed by

21  subsampling/rarefaction of (**A**) richness and (**B**) Faith's phylogenetic diversity with increasing

22  total number of sequences.
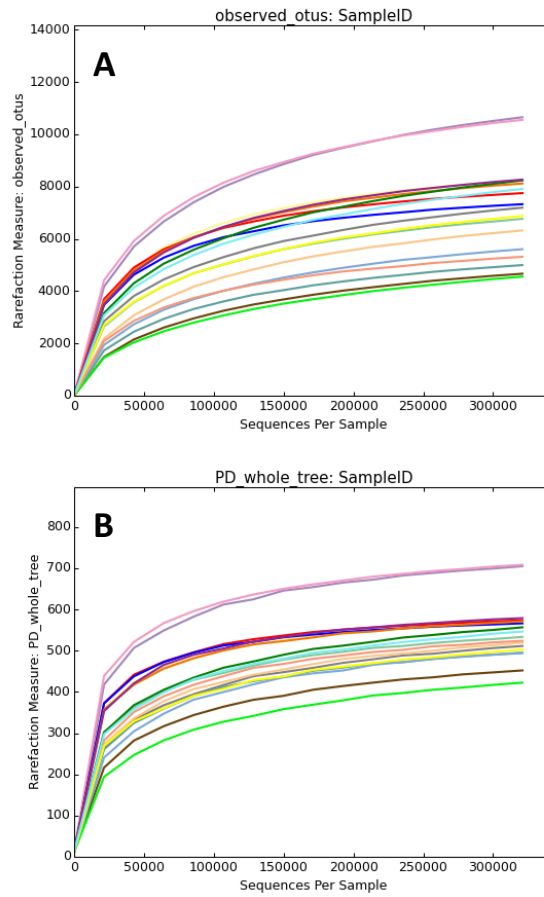


23

24

25   **Supporting Figure 4**. Divergences in fire-affected soils are not well explained by temperature.

26   **(A)** Principal coordinate analysis (PCoA) based on weighted UniFrac distances of phylogenetic

27   bacterial and archaeal community structure in fire-affected soils. The strength of statistically

28   significant (p < 0.10) explanatory variables are shown with blue arrows. **(B)** Constrained

29   analysis based on weighted UniFrac distances, where the explanatory value of temperature is

30   removed from the analysis to understand the influence of the remaining explanatory variables.
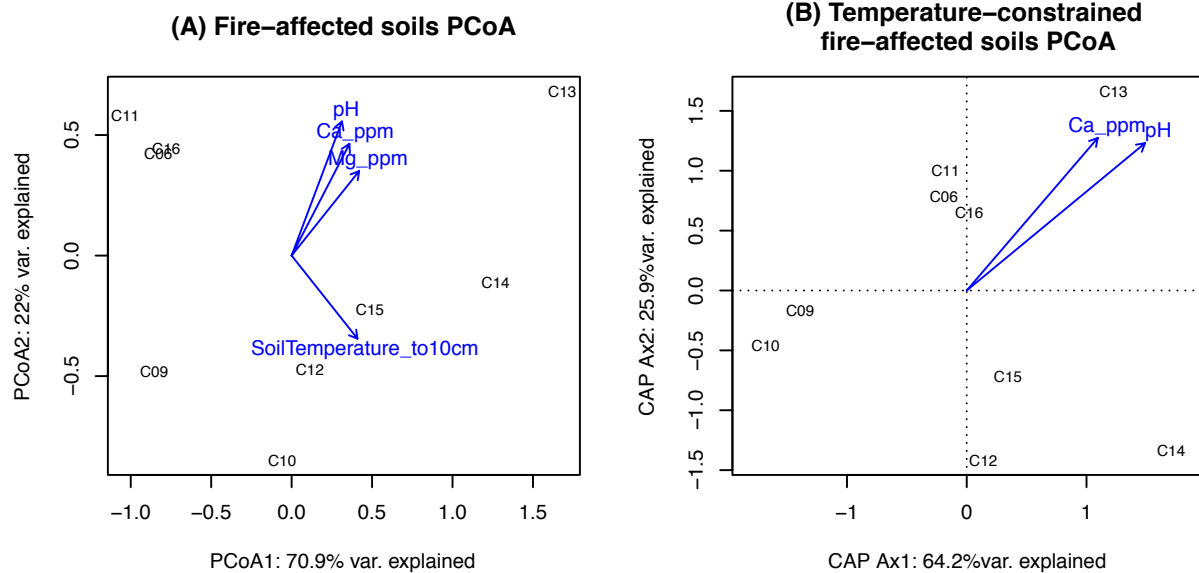


31

32

33    **Supporting Figure 5.** Neutral models of community assembly (abundance v. occurrence) for

34    (**A**) the total community ("All", n= 18), **(B)** fire-affected soils ("Fire_Affected", n=9) and **(C)**

35    recovered soils ("Recovered" n=7).  Red symbols show OTUs that had higher abundance than

36    their prediction, and blue symbols show OTUs that had lower abundance than their prediction.

37    The thick yellow line is the neutral model prediction, and the thin yellow lines show a 95%

38    confidence interval around the prediction.

39



40

41
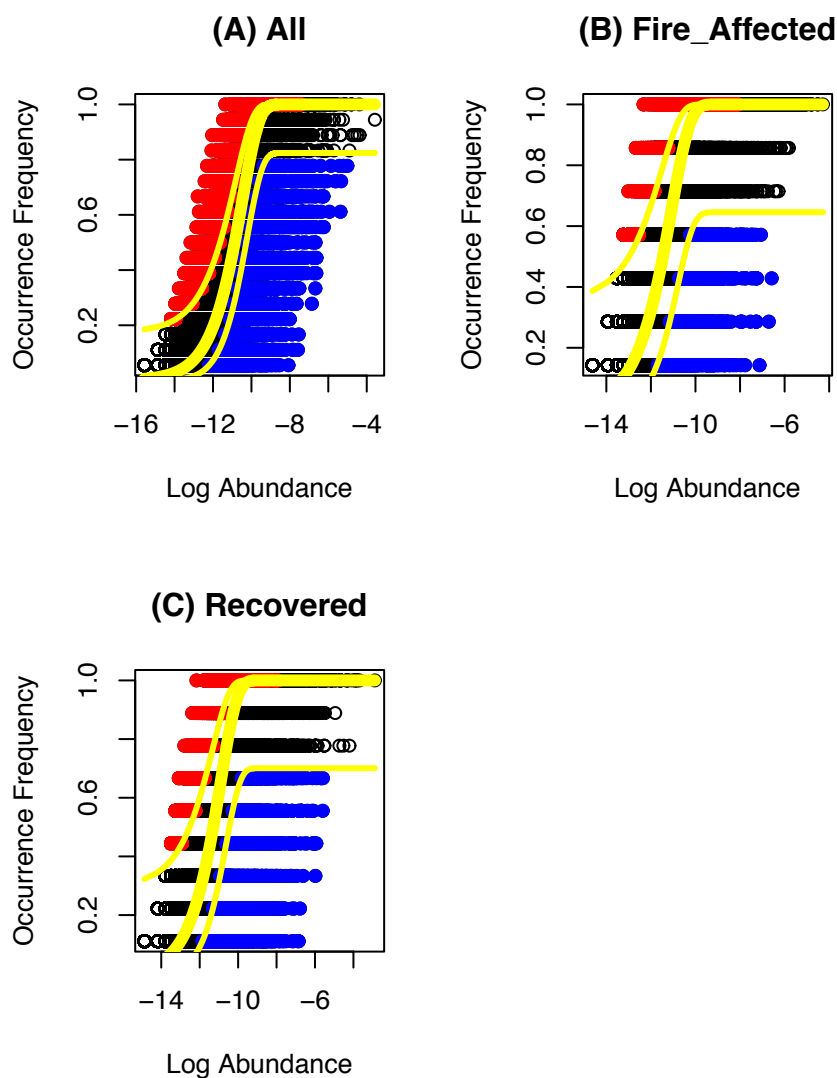
42    **Supporting Table 1.** Primers used in this study.


**Table 1.** Primer set used for this study.

| Primer name | sequence (5' - 3') | Target | target site | Product size (bp) | Tm | Reference |
|---|---|---|---|---|---|---|
| 515F | GTGCCAGCMGCCGCGGTAA | 16S V4 | 515-534 | 291 | 69.5 | Caporaso et al., ISME J. 2012 |
| 806R | GGACTACHVGGGTWTCTAAT | | 787-806 | | 45.1 | |

43

44

45  **Supporting Table 2**. Mean and standard deviation ("sd") phylogenetic diversity and number of

46  OTUs ("richness) across technical sequencing replicates for the un-collapsed dataset (rarefied

47  to 53,000 sequences per sample). Three replicate DNA extractions, amplifications and

48  sequencing reactions were performed per soil, and, after calculating the technical variability,

49  these sequences were pooled into one aggregate set of sequences to achieve deep coverage

50  of the community within each soil.

| SampleID | PD_mean | PD_sd | Richness_mean | Richness_sd |
|---|---|---|---|---|
| C01 | 393.96 | 16.22 | 4073.67 | 55.77 |
| C02 | 392.48 | 9.42 | 3805.00 | 48.50 |
| C03 | 403.12 | 15.25 | 4498.67 | 39.72 |
| C04 | 374.95 | 6.51 | 4420.33 | 89.51 |
| C05 | 405.05 | 14.17 | 4389.33 | 109.25 |
| C06 | 332.89 | 13.26 | 3718.67 | 117.33 |
| C07 | 371.50 | 7.80 | 4253.00 | 67.01 |
| C08 | 525.93 | 5.37 | 6011.67 | 191.04 |
| C09 | 312.71 | 32.40 | 2328.33 | 352.23 |
| C10 | 267.32 | 27.06 | 2128.00 | 225.08 |
| C11 | 343.84 | 12.26 | 3886.67 | 81.56 |
| C12 | 249.92 | 29.65 | 2106.67 | 280.73 |
| C13 | 316.18 | 58.27 | 2471.00 | 816.28 |
| C14 | 307.29 | 16.47 | 2688.67 | 232.20 |
| C15 | 330.40 | 38.06 | 3011.67 | 435.15 |
| C16 | 356.85 | 12.24 | 3546.33 | 83.93 |
| C17 | 506.13 | 19.77 | 5724.00 | 179.43 |
| C18 | 392.64 | 13.98 | 4210.67 | 105.61 |

51

52

53    **Supporting Table 3.** Percent variation explained for PCoA axes 1 and 2 for weighted and
54    unweighted UniFrac, Sorensen-dice, and Bray-Curtis distances/dissimilarities.    Weighted
55    UniFrac was chosen because it was most informative in explaining the variance along the first
56    two axes.

|  | PCoA1 | PCoA2 |
|---|---|---|
| *Weighted UniFrac* | 77.1 | 12.7 |
| *Normalized Weighted Unifrac* | 74.6 | 10.9 |
| *Unweighted UniFrac* | 18.3 | 13.6 |
| *Sorensen-dice* | 20.1 | 15.2 |
| *Bray-Curtis* | 23.9 | 13.7 |

57

58

59    **Supporting Table 4**. Explanatory value of soil contextual data to changes in Centralia soil
60    community structure along PCoA axes for the all soils.  Factors significant at p < 0.10 are in
61    bold.

| | PCoA1 | PCoA2 | R2 | P value | |
|---|---|---|---|---|---|
| % explanation | 77.1 | 12.3 | | | |
| **Soil Temperature** | **0.968** | **-0.252** | **0.787** | **0.002** | ** |
| **NO$_3$N (ppm)** | **0.226** | **-0.974** | **0.290** | **0.067** | . |
| **pH** | **0.185** | **0.983** | **0.649** | **0.008** | ** |
| K (ppm) | -0.813 | 0.582 | 0.006 | 0.946 | |
| Mg (ppm) | -0.148 | 0.989 | 0.123 | 0.374 | |
| Organic matter | 0.812 | -0.583 | 0.002 | 0.984 | |
| **NH$_4$N (ppm)** | **0.194** | **-0.981** | **0.287** | **0.088** | . |
| SulfateSulfur (ppm) | 0.121 | -0.993 | 0.116 | 0.372 | |
| **Ca (ppm)** | **0.182** | **0.983** | **0.529** | **0.022** | * |
| **Fe (ppm)** | **0.253** | **-0.967** | **0.271** | **0.094** | . |
| Fire history | -0.605 | 0.797 | 0.253 | 0.169 | |
| As (ppm) | -0.014 | -1.000 | 0.124 | 0.404 | |
| P (ppm) | 0.435 | -0.900 | 0.093 | 0.462 | |
| **Soil Moisture (%)** | **0.263** | **-0.965** | **0.405** | **0.035** | * |
| Significant codes: '***' 0.001; '**' 0.01; '*' 0.05; '.' 0.1; ' ' 1 | | | | | |
| Number of permutations: 999 | | | | | |

62

63

64  **Supporting Table 5**.  Explanatory value of soil contextual data to changes in Centralia soil

65  community structure along PCoA axes for the fire-affected soils.  Factors significant at p < 0.10

66  are in bold.

| | PCoA1 | PCoA2 | R2 | P value | |
|---|---|---|---|---|---|
| % explanation | 70.9 | 22.0 | | | |
| **SoilTemperature_to10cm** | **0.765** | **-0.644** | **0.578** | **0.088** | . |
| NO3N_ppm | -0.002 | -1.000 | 0.328 | 0.236 | |
| **pH** | **0.490** | **0.872** | **0.823** | **0.002** | ** |
| K_ppm | 0.282 | -0.959 | 0.232 | 0.429 | |
| **Mg_ppm** | **0.767** | **0.641** | **0.604** | **0.058** | . |
| OrganicMatter_500 | 0.407 | -0.913 | 0.218 | 0.498 | |
| NH4N_ppm | -0.021 | -1.000 | 0.342 | 0.155 | |
| SulfateSulfur_ppm | -0.216 | -0.976 | 0.118 | 0.759 | |
| **Ca_ppm** | **0.613** | **0.790** | **0.694** | **0.015** | * |
| Fe_ppm | 0.044 | -0.999 | 0.355 | 0.204 | |
| As_ppm | -0.492 | -0.871 | 0.388 | 0.228 | |
| P_ppm | 0.142 | -0.990 | 0.238 | 0.453 | |
| SoilMoisture_Per | -0.023 | -1.000 | 0.460 | 0.143 | |
| Fire_history | 0.742 | -0.670 | 0.136 | 0.637 | |
| Significant codes: '***' 0.001; '**' 0.01; '*' 0.05; '.' 0.1; ' ' 1 | | | | | |
| Number of permutations: 999 | | | | | |

67

68

69

70  **Supporting Table 6**.  Explanatory value of soil contextual data to changes in Centralia soil
71  community structure along the constrained PCoA axes for the fire-affected soils, after removing
72  the influence of temperature.  Factors significant at p < 0.10 are in bold.
73

| | CAP_A1 | CAP_A2 | R2 | P value | |
|---|---|---|---|---|---|
| % explanation | 64.2 | 25.9 | | | |
| SoilTemperature_to10cm | 1.000 | 0.000 | 0.000 | 1.000 | |
| NO3N_ppm | -0.973 | -0.233 | 0.354 | 0.285 | |
| **pH** | **0.771** | **0.637** | **0.729** | **0.014** | * |
| K_ppm | -0.416 | -0.909 | 0.093 | 0.730 | |
| Mg_ppm | 0.641 | 0.767 | 0.370 | 0.247 | |
| OrganicMatter_500 | 0.070 | -0.997 | 0.128 | 0.613 | |
| NH4N_ppm | -0.962 | -0.273 | 0.367 | 0.240 | |
| SulfateSulfur_ppm | -0.988 | 0.154 | 0.234 | 0.446 | |
| **Ca_ppm** | **0.652** | **0.759** | **0.551** | **0.092** | . |
| Fe_ppm | -0.862 | -0.508 | 0.396 | 0.355 | |
| As_ppm | -0.948 | -0.317 | 0.378 | 0.216 | |
| P_ppm | -0.132 | -0.991 | 0.287 | 0.350 | |
| SoilMoisture_Per | -0.813 | -0.583 | 0.419 | 0.203 | |
| Fire_history | 0.636 | -0.771 | 0.276 | 0.375 | |
| Significant codes: '***' 0.001; '**' 0.01; '*' 0.05; '.' 0.1; ' ' 1 | | | | | |
| Number of permutations: 999 | | | | | |

74

75
76
77 **Supporting Table 7**.  Parameters and fits of neutral models as per Burns et al. 2015.

| Model parameter | all | Fire-affected | Recovered |
|---|---|---|---|
| m | 0.04 | 0.08 | 0.10 |
| m.ci | 0.00 | 0.00 | 0.00 |
| m.mle | 0.04 | 0.08 | 0.10 |
| maxLL | -5838.12 | 1187.68 | -2735.42 |
| binoLL | 475.69 | 1162.47 | -143.93 |
| poisLL | 475.67 | 1162.46 | -143.94 |
| Rsqr | 0.45 | 0.12 | 0.53 |
| Rsqr.bino | -1.19 | -0.86 | -0.47 |
| Rsqr.pois | -1.19 | -0.86 | -0.47 |
| RMSE | 0.20 | 0.26 | 0.21 |
| RMSE.bino | 0.39 | 0.38 | 0.37 |
| RMSE.pois | 0.39 | 0.38 | 0.37 |
| AIC | -11672.24 | 2379.36 | -5466.85 |
| BIC | -11655.75 | 2394.86 | -5451.16 |
| AIC.bino | 955.38 | 2328.94 | -283.86 |
| BIC.bino | 971.88 | 2344.43 | -268.17 |
| AIC.pois | 955.35 | 2328.92 | -283.88 |
| BIC.pois | 971.84 | 2344.42 | -268.19 |
| N | 321000.00 | 321000.00 | 321000.00 |
| Samples | 18.00 | 9.00 | 7.00 |
| Richness | 28220.00 | 17097.00 | 18866.00 |
| Detect | 0.00 | 0.00 | 0.00 |
| %AbovePred | 0.14 | 0.12 | 0.13 |
| %BelowPred | 0.10 | 0.07 | 0.12 |

78

79
80

81

82

83    **Supporting Table 8**. Welch's t-tests comparing the mean relative abundances of phyla across

84    fire-affected and recovered soils.  Bold values are significant at p < 0.05.

| Phylum | T-statistic | DF | p-value |
|---|---|---|---|
| **Crenarchaeota** | **2.80** | **8.36** | **0.02** |
| Euryarchaeota | -0.47 | 11.86 | 0.65 |
| **[Parvarchaeota]** | **-3.31** | **11.34** | **0.01** |
| **k__Bacteria;Other** | **2.33** | **8.22** | **0.05** |
| AD3 | -1.58 | 7.28 | 0.16 |
| Acidobacteria | -1.74 | 13.64 | 0.10 |
| Actinobacteria | -0.22 | 13.12 | 0.83 |
| Armatimonadetes | -0.58 | 13.21 | 0.57 |
| **Bacteroidetes** | **-4.00** | **9.73** | **0.00** |
| Chlamydiae | -1.68 | 10.73 | 0.12 |
| Chlorobi | -0.43 | 10.96 | 0.67 |
| **Chloroflexi** | **2.82** | **9.67** | **0.02** |
| Cyanobacteria | 1.85 | 8.07 | 0.10 |
| **Elusimicrobia** | **-3.45** | **8.01** | **0.01** |
| FCPU426 | -0.79 | 11.28 | 0.45 |
| Firmicutes | 0.60 | 10.97 | 0.56 |
| **Gemmatimonadetes** | **-2.24** | **12.33** | **0.04** |
| Nitrospirae | 0.04 | 12.47 | 0.97 |
| OD1 | -1.28 | 10.05 | 0.23 |
| OP11 | -1.82 | 7.56 | 0.11 |
| **Planctomycetes** | **-3.33** | **11.61** | **0.01** |
| **Proteobacteria** | **-2.42** | **12.89** | **0.03** |
| SBR1093 | 2.02 | 8.00 | 0.08 |
| **Spirochaetes** | **-2.43** | **6.68** | **0.05** |
| **TM6** | **-2.48** | **7.47** | **0.04** |
| Tenericutes | 0.14 | 10.06 | 0.89 |
| **Verrucomicrobia** | **-3.78** | **10.92** | **0.00** |
| WPS-2 | 0.41 | 10.37 | 0.69 |
| WS3 | -2.26 | 6.59 | 0.06 |
| Below_0.01 | -0.27 | 8.39 | 0.79 |

85

86