**Supplementary Materials:**

*Hypotheses That Were Tested*

Prior to the differential abundance analysis, cluster analysis combining PAM and PCA was used to identify sites and conditions along the LA River that may be associated with differential abundance of bacteria and fungi. This knowledge, combined with factor analysis, was used to form the alternative hypotheses given below.

1. There was differential abundance of fungi between submerged and frequently submerged samples.
2. There was differential abundance of bacteria between submerged and frequently submerged samples.
3. There was differential abundance of bacteria between soft-bottom and concrete river conditions.
4. There was differential abundance of bacteria between Glendale Narrows and Verdugo Wash.
5. There was differential abundance of fungi between Maywood and Arroyo Seco.

*Chi Square Test of Proportions using Published Data [25]*

The questions of interest were:

1. Within the LA River and global Freshwater datasets, how do the proportions of sequences in each Division compare?
2. Does the LA River have a typical fungal profile, when compared with other Rivers and Freshwater bodies, at the Division level?
3. Could there be another factor at play in the LA River environment that is influencing the proportion sequences in each Division?

*Test 1:* The null hypothesis was that the proportion of LA River Fungi in Ascomycota and Basidiomycota was equal to the proportion of Ascomycota and Basidiomycota in published Freshwater environment datasets [25].

The alternative hypothesis was that the proportion of Ascomycota for the LA River was greater than or less than the proportion for Freshwater environments.

*Test 2:* The null hypothesis was that the proportion of LA River Fungi in Ascomycota and Basidiomycota was equal to the proportion of Ascomycota and Basidiomycota in published River environments [25].

The alternative hypothesis was that the proportion of Ascomycota for the LA River was greater than or less than the proportion for River environments. These follow up questions were generated after the Chi Square test:

1. Which sites have the most Ascomycota species?
2. Which LA river sites contributed to the abundance of Ascomycota?

Figure S1. A Google Map of sampling locations has been provided; sites have been marked with a red dot. Map data © 2021 Google.

Map

Description automatically generated

*Heatmap*

Figure S2 shows the Chi Square Standardized Euclidean Distance Matrix Heatmap Visualizations for four of the markers, 16S, 12S, COI, and FITS. Each one of the assays captured a different picture of Beta diversity, which is diversity between samples. The figure demonstrates for the fungi that one pair of samples is highly distinct, and more than 50% are moderately similar. It appeared that the Chi Square standardized heat map was the least sensitive of the methods which employed a Euclidean distance matrix.

Figure S2 includes the Heatmap Visualization for the 16S marker. One pair of samples is highly dissimilar and there were two pairs of samples with moderate dissimilarity. Most pairs of bacterial samples were similar, according to this visualization method. This emphasized the lack of sensitivity to biodiversity in this method, considering that the bacteria samples reflected the most species richness. The Heatmap Visualization in Figure S2 shows a measure of diversity for the samples run on the COI marker. The COI Visualization shows that most samples were nearly identical. One pair is highly distinct.

The Heatmap Visualization for the 12S marker is shown in Figure S2.  In this assay, there were several pairs of samples that were moderately distinct, and two pairs of samples that were highly distinct from one another.

Figure S2. Euclidean Distance Matrix Heatmap visualization of the LA River samples, standardized by the Chi Square Distribution, for 16S (A), 12S (B), COI (C), and FITS (D) markers.

Shape

Description automatically generated with medium confidence

Figure S3. PCA for Fungal identified sequences from the FITS marker was visualized by sample, and color coded by sampling site. Several samples appear as outliers.

Chart, scatter chart

Description automatically generated

Figure S4. PCA for Fungal identified sequences from the FITS marker was visualized by sample, and color coded by the best PAM clustering.

Chart, scatter chart

Description automatically generated

Figure S5. PCA for Bacterial identified sequences from the 16S marker was visualized by sample, and color coded by sampling site. The evidence of overdispersion was the strongest in the 16S marker count data.

Chart, scatter chart

Description automatically generated

Figure S6. PCA for identified sequences from the 18S marker was visualized by sample, and color coded by sampling site, which also exhibits overdispersion.

Chart, scatter chart

Description automatically generated

Figure S7. PCA for identified sequences from the 18S marker by sample, color coded by the best PAM clustering. There is unconcealed evidence of overdispersion, especially for those samples that were high on both PC 1 and PC 2.

Chart, scatter chart

Description automatically generated

Figure S8. The boxplot of Observed alpha diversity has been provided. The boxplot comparison shows that the species richness for all fungi is the highest for Arroyo Seco, Bull Creek, Compton Creek, Verdugo Wash, and Maywood.

Chart

Description automatically generated

Figure S9. The mosaic plot is given for the Chi square test of proportions. The mosaic plot shows that there is a difference in the proportion of ascomycetes to basidiomycetes in the LA River compared to Freshwater Habitats [25].

A picture containing shape

Description automatically generated

Figure S10. PCA for identified sequences from the COI marker was visualized by sample, and color coded by sampling site.

Chart, scatter chart

Description automatically generated

Figure S11. PCA for identified sequences from the COI marker was visualized by sample, and color coded by the best PAM clustering.

Chart, scatter chart

Description automatically generated

Figure S12. PCA for identified sequences from the 12S marker was visualized by sample, and color coded by sampling site.

Chart, scatter chart

Description automatically generated

Figure S13. PCA for identified sequences from the 12S marker by was visualized by sample, and color coded by the best PAM clustering. Note that most of the samples appear to be nearly identical.

Chart

Description automatically generated

Figure S14. PCA for identified plant sequences from the PITS marker was visualized by sample, and color coded by sampling site.

Chart, scatter chart

Description automatically generated

Figure S15. Comparison of the relative abundance of Fabaceae and Solanaceae sp. is depicted by sampling site.

Which sites contribute to the overall abundance of plant sequences from the *Fabaceae* and *Solanaceae* families? Note that most of the *Fabaceae* sequences were derived from Compton Creek. There is also a sizeable proportion of *Fabaceae* sequences from Bowtie Parcel samples. What is the relative abundance of sequences coming from each site? Note the appearance that a small number of sequences of *Datura stramonium* in Arroyo Seco lead to a large abundance of *Datura* sequences in Maywood.

Chart, bar chart, box and whisker chart

Description automatically generated

Figure S16. Relative abundance of Alnus, Atriplex, Chenopodium, Paspalum, and Salix sp. color coded by LA River site. Which sites contribute to the overall abundance of these 5 genera? What is the relative abundance between sites? Notice that *Salix* abundance has contribution from a multitude of samples, whereas *Chenopodium* was mostly found in Maywood and Sepulveda Dam and *Atriplex* was most abundant in Maywood. Elysian Valley, Arroyo Seco, and Tujunga Wash contributed the highest proportions of *Alnus* sequences. *Paspalum* sequences had a high influence on PC3 but apparently were only predominant in Sepulveda Dam, revealing potential bias in the PCA PAM model.

**Chart, bar chart

Description automatically generated.**

Figure S18. The dispersion estimates for the fungal inter-transcribed spacer (FITS) marker dataset have been shown for the local mean method which implements the locfit package.  Using the local method, several points were considered as outliers. Notice that those outlying points were not hugged toward the mean in the final estimate, which is represented by the fitted red curve.

Chart, scatter chart

Description automatically generated

Figure S19. The dispersion estimates have been shown for the DESeq() method chosen by the algorithm, which appears to be a better fit. Notice that the blue points representing the final estimates were all pulled slightly toward the mean, which is represented by the fitted red curve.

Chart, scatter chart

Description automatically generated

Figure S20. The volcano plot for the DESeq analysis on the 16S marker data is given. This analysis showed 24 taxa that were differentially abundant between Glendale Narrows and Verdugo Wash, based on significant p-values and Log2 Fold Changes.

Chart, scatter chart

Description automatically generated

Figure S21. The volcano plot for differential expression of fungi based on the FITS marker is given, with respect to the contrast between frequently submerged versus fully submerged sites.

The differences in fungal taxa as represented by the FITS marker reflected 27 taxa were differentially expressed, so that some taxa were more likely to be associated with one condition than the other.

Chart, scatter chart

Description automatically generated

Figure S22. The volcano plot shows a visualization of how many taxa were differentially expressed in the soft bottom and concrete environments, with regards to the bacteria and archaea.

The results of the DESeq analysis on the 16S marker seem to reflect two distinct populations associated with soft bottom and concrete environments.

Chart, box and whisker chart

Description automatically generated

Table S5. Summary of the hypothesis tests that were conducted in DESeq2, and the resulting conclusions.

|  |  |  |
| --- | --- | --- |
| **HA** | **Description** | **Conclusion** |
| **Hypothesis 1** | Differential abundance of fungi in submerged vs. freq. submerged samples | Supported |
| **Hypothesis 2** | Differential abundance of bacteria in submerged vs. frequently submerged samples | Supported |
| **Hypothesis 3** | Differential abundance of bacteria in soft-bottom vs. concrete samples | Supported |
| **Hypothesis 4** | Differential abundance of bacteria in Glendale vs. Verdugo Wash | Supported |
| **Hypothesis 5** | Differential abundance of fungi in Maywood vs. Arroyo Seco | Supported |