Partitional Clustering and Differential Abundance Analysis Reveal the Community Structure of eDNA in the Los Angeles River

Savanah Senn *, Sharmodeep Bhattacharyya, Gerald Presley, Anne E. Taylor, Rayne Stanis, Kelly Pangell, Daila Melendez, Jillian Ford

Posted Date: 20 March 2023

doi: 10.20944/preprints202303.0337.v1

Keywords: metabarcoding; statistical modeling; urban river ecology

Article

Partitional Clustering and Differential Abundance Analysis Reveal the Community Structure of eDNA in the Los Angeles River

Savanah Marie Senn [1,2], Sharmodeep Bhattacharyya [1,3], Gerald Presley [1,4], Anne E. Taylor [5], Rayne Stanis [2], Kelly Pangell [2], Daila Melendez [2,5] and Jillian Ford [2]

[1] Environmental Sciences Graduate Program, Oregon State University, Corvallis, OR 97331, USA; sharmodeep.bhattacharyya@oregonstate.edu (S.B.); gerald.presley@oregonstate.edu (G.P.); anne.taylor@oregonstate.edu (A.E.T.)
[2] Department of Agriculture Sciences, Los Angeles Pierce College, 6201 Winnetka Avenue, PMB 553, Woodland Hills, CA 91304, USA; fordjm7164@student.laccd.edu (J.F.); WAGGOND1440@STUDENT.LACCD.EDU (D.M.); pangelkl4875@student.laccd.edu (K.P.); RAYNE-LACE@GMAIL.COM (R.S.)
[3] Department of Statistics, Oregon State University, Corvallis, OR 97331, USA
[4] Department of Wood Science & Engineering, Oregon State University, Corvallis, OR 97331, USA
[5] Department of Crop and Soil Sciences, Oregon State University, Corvallis, OR 97331, USA
* Correspondence: stclais@piercecollege.edu

Abstract: In this study we sought to investigate the impact of urbanization, presence of concrete river bottom, and nutrient pollution on microbial communities along the L.A. River. Six molecular markers were evaluated for identification of bacteria, plants, fungi, fish, and invertebrates in 90 samples. PCA (principal components analysis) was used with PAM (partitioning around medoids) clustering to reveal community structure and an NB (Negative binomial) model in DESeq2 was used for differential abundance analysis. PCA and factor analysis exposed the main axes of variation but were sensitive to outliers. Differential abundance of *Proteobacteria* was associated with soft bottom sites, and there was an apparent balance in the abundance of organisms responsible for nitrogen cycling. Nitrogen cycling was explained by differential abundance of ammonia oxidizing archaea, the complete ammonia oxidizers *Nitrospira* sp., nitrate reducing bacteria *Marmoricola* sp., and nitrogen fixing bacteria *Devosia* sp. which were differentially abundant at soft-bottom sites ($p\ adj < 0.002$). In contrast, differential abundance of several *Cyanobacteria* and other anoxygenic phototrophs was associated with the concrete bottom sites, which suggested the accumulation of excess nitrogen. The soft bottom sites tended to be represented by differential abundance of aerobes, whereas the concrete-associated species tended to be alkaliphilic, saliniphilic, calciphilic, sulfate dependent, and anaerobic. In Glendale Narrows, downstream from multiple water reclamation plants, there were differential abundance of cyanobacteria and algae, however indicator species for low nutrient environments and ammonia-abundance were also present. There was differential abundance of ascomycetes associated with Arroyo Seco and a differential abundance of *Scenedesmaceae* green algae and cyanobacteria in Maywood, in the analysis which compared suburban with urban river communities. The proportion of Ascomycota to Basidiomycota within the LA River differed from the expected proportion based on published worldwide freshwater and river 18S data; the shift in community structure was most likely associated with the extremes of urbanization. This study indicates that extreme urbanization can result in overrepresentation of cyanobacterial species that could cause reductions in water quality and safety.

Keywords: metabarcoding; statistical modeling; urban river ecology

1. Introduction

The Los Angeles River has the potential to influence systems beyond its boundaries such as estuarine environments at its outlet to the Pacific Ocean. In 2020 the County of Los Angeles Gross Domestic Product was $6.5 billion [1] and population was over 10 million [2]. Well-documented issues with contamination such as heavy metals, excess nutrients, coliform bacteria, and cyanide [3] have resulted from industrialization and high population. The LA River is also a habitat for bacteria, fungi, fish, plants, and invertebrates that are all sensitive to pollution. More recently, efforts have focused on protection and recognition of the river as a natural ecosystem, and part of that effort has been assessing the impacts of urbanization on the LA River ecosystems through eDNA sampling [4].

There have been few studies which have aimed to characterize the biome of the LA River, however interest in characterizing microbial communities in this biome has increased in recent years [5,6]. The diversity of life, including

 doi:10.20944/preprints202303.0337.v1

fungi, bacteria, plants, fish, and invertebrates is indicative of ecosystem health. The presence or absence of certain "indicator" species reflect health and the presence of oxygen, or degradation and pollution [7–9,93]. By investigating microbial community composition and identifying relative species abundance, ecosystem health can be compared among different locations subject to different pollutant profiles.

By investigating microbial community composition and identifying relative species abundance, ecosystem health can be compared among different locations subject to different pollutant profiles. The L.A. River presents a unique opportunity to measure the impact of various types of urban pollution and infrastructure on microbial communities. The river runs through rural, suburban and urban areas and the impact of population density can be assessed. In addition, the Los Angeles River was highly modified to facilitate flood control [10]. The city could not have grown to be such a metropolitan center if it was not for the engineering of the LA River [11,12]. The difficulty was that although the climate was dry for most of the year, when the rains did come there was often precipitation greater than 2 inches per hour, which led to flooding that could be catastrophic. Another crucial question documented by Wenger et al refers to the relationship between urbanization and the structure and function of microbial communities, which has not been well-studied [13]. The question of how microbial communities may differ from one another in different land use areas, and how urbanization may affect the proportions of different classes of microbes, remains vital. The importance of this type of investigation was also underscored in Antwis et al's perspective on the most important areas of inquiry in microbial ecology [14]. In terms of urbanization, the modification or toxification of the environment may have influenced which organisms were present.

Ecological habitat was believed to have been diminished due to most of the LA River bottom being impervious concrete [15]. According to Wenger et al, inquiry into the characteristics of piped or concrete paved tributaries as they influence biogeochemical processes represents one of the most important questions in urban stream ecology [13]. Presence of a concrete river bottom has been known to influence the oxygen content of freshwater and this factor is expected to be one of the key factors which would influence communities in the concrete bottom condition. Nevertheless, if River organisms such as oxygenic autotrophs generate oxygen to the aboveground environment, it would help to offset such a concern by performing a beneficial function.

Since bacteria play a huge role in breakdown of wastes, nitrogen cycling, plant growth promotion, and pathogenicity, differences due to a bottom-up effect warranted a closer look. Lack of oxygen in the underwater environment was expected to be one of the key factors which would influence communities in the concrete bottom condition. Furthermore, concrete paved rivers contribute to the urban heat island effect, which involves increased light intensity and heat [13]. Urban rivers generally have a cooling effect on the metropolis by virtue of the water that flows from them and the green spaces they support [16].

In the absence of rain, the LA River has been fed by water from three water reclamation plants. Ackerman et al found in 2003 that there were higher ratios of ammonia to nitrate near the water reclamation plants [10]. The benefits of using reclaimed water are obvious in terms of ecosystem services as a river fed by recycled water would be expected to provide more habitat than a dry river bed. The year-round supply of water has the potential to support wildlife and vegetation. The water sources have been shown to increase the NO^{3-} concentration near the treatment plant sources, but it also would be expected to dilute the concentration of other pollutants such as hydrocarbons from households and industry pollutants such as heavy metals. The proximity to a water reclamation plant could influence the diversity of bacterial sequences recovered from different sampling locations. A potential abundance of nitrate from water treatment plants was considered to be a priority at sites nearby to Glendale [10], however the flow of water to wildlife through the river conduit would be expected to promote the diversity and abundance of organisms. On balance, the river would otherwise be a dry ravine during most of the year due to the Mediterranean climate, if it were not for the releases from the water treatment plants.

In this study we sought to investigate the impact of urbanization, presence of concrete river bottom, and nutrient pollution on microbial communities along the L.A. River. This was done by meta-barcoding and community analysis of environmental DNA (eDNA). Organisms that perform beneficial functions in the LA River Ecosystem were identified and quantified among samples taken along the river [17]. This study focused on both eukaryotic and prokaryotic microbes including archaea, bacteria, cyanobacteria, fungi, plants, and eukaryotic algae. Differential abundance of these organism types was measured and analyzed computationally. This work contributes to a better understanding of the microbial ecology of the LA River ecosystem and helps identify urbanization impacts on microbial communities.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 March 2023 doi:10.20944/preprints202303.0337.v1

2. Materials and Methods

2.1. Sample Collection

The original data was generated as part of a BioBlitz program by University of California CALeDNA. CALeDNA is a collaboration of scientists creating a baseline of data for the biodiversity in California. Samples were collected by the UC CALeDNA team led by Miroslava Ramos, project manager. 90 replicated samples were collected from sediment over a 51-mile span of the channelized portion of the LA River and its tributaries. Three subsamples were taken from each sampling location and bulked after DNA extraction to capture a capture a picture of the diversity within a 1-foot radius. In total, there were 180 subsamples.

Table 1 lists the sampling sites by their GPS coordinates for reference. The sampling sites were spread throughout the LA River Watershed. Tillman WRP is near Sepulveda Dam. Note that Verdugo Wash flowed to Glendale Narrows, and Glendale also received water from the intermediary Glendale Water Reclamation Plant. Also depicted is Arroyo Seco, a naturalized area that flows into the industrialized area of Maywood, providing contrast.

Table 1. Tabulation of the types of genomic data that were available for the LA River [94].

Marker	Description	Target Organisms	Forward Primer	Reverse Primer	Reference
FITS	Fungal rRNA Internal Transcribed Spacer	Fungi	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG	Miya et al. 2015
16S	Prokaryotic rRNA small subunit	Bacteria, archaea	GTGYCAGCMGCCGCGGTAA	GGACTACNVGGGTWTCTAAT	F: 515F and R: 806R, see Caporaso et al., 2012
18S	Eukaryotic rRNA small subunit	Fungi, algae, protists	GTACACACCGCCCGTC	TGATCCTTCTGCAGGTTCACCTAC	Amaral-Zettler et al. 2009; Euk_1391f and EukBr
CO1	Mitochondrial cytochrome oxidase subunit I	Animals	ATGCGATACTTGGTGTGAAT	GACGCTTCTCCAGACTACAAT	Gu et al. 2013
12S	Mitochondrial rRNA small subunit	Fish, birds, snakes, insects	GGWACWGGWTGAACWGTWTAYCCYCC	TANACYTCnGGRTGNCCRAARAAYCA	Leray et al. 2013
PITS	Plant rRNA Internal Transcribed Spacer	Plants	GGAAGTAAAAGTCGTAACAAGG	CAAGAGATCCGTTGTTGAAAGTT	F: ITS5, White et al., 1990; R: 5.8S, Epp et al. 2012

2.2. DNA Isolation and Amplification

DNA was extracted using the Qiagen DNEasy PowerSoil Kit. Six molecular markers specific to different kingdoms of life were amplified from the eDNA for amplicon sequencing. Amplicon libraries from each sample type with Illumina barcode adapters were sequenced on the MiSeq platform at 35,000 paired reads each. Quality control was performed in QIIME [18]; Cutadapt was used to remove Illumina adaptor sequences, DADA2 was used for quality score trimming

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 March 2023 doi:10.20944/preprints202303.0337.v1

and identification of unique ASVs. Taxonomies were assigned to Amplicon Sequence Variants with an 80% likelihood cutoff from the CRUX database. A GreenGenes classifier was used. Each marker dataset was outputted into an ASV (Amplicon Sequence Variant) table for downstream analysis using the Anacapa toolkit [19].

For this differential abundance analysis, computation focused on the bacteria and fungi. However, the results of the differential abundance analysis may also include algae and nematodes, for example. Table 3 shows the covariates that were contrasted in DESeq2.

Table 3. List of Covariates that were tested for association with differential abundance of bacterial and fungal taxa.

Marker	Covariate	Factor Levels Tested
16S	LA River Site	Glendale Narrows, Verdugo Wash
16S	River Condition	Soft-Bottom, Concrete
16S	Habitat	Frequently Submerged, Fully Submerged
FITS	Habitat	Frequently Submerged, Fully Submerged
FITS	LA River Site	Maywood, Arroyo Seco

2.3. Statistical Approach

The goal of this project was to examine sample diversity using a variety of methods using a Euclidean distance matrix [20]. The Euclidean distance is given by:

$$d(j1, j2) = [(X1j_1 - X1j_2)^2 + \cdots + (Xnj_1 - Xnj_2)^2]^{1/2} \qquad [21]$$

The methods utilizing the Euclidean dissimilarity measure will include Neighbor joining of samples [22], UPGMA of samples [22], Heatmap visualization using Chi-square standardization of samples, and PAM (partitioning around medoids) clustering applied to PCA. Ranacapa [23] was used to perform a PERMANOVA beta diversity test and visualize with Principal Coordinates Analysis (PCoA) to help with hypothesis development.

PAM clustering was applied to PCA to investigate whether samples cluster by location in an unsupervised model, and if the PCA reflected a spatial relationship inherent in the genetic distances. The PAM function from the cluster package was used [24]. First, K representative medoids are arbitrarily selected, then swapping cost C_{ih} to swap medoid h and non-medoid i is calculated. If the resulting value is negative, then the medoid and non-medoid are swapped. The process is repeated until there is no change. Principal components analysis reveals population stratification and PAM is used for classification of samples.

Classification of samples was expected based on the taxonomic composition of samples; that is, if there were differentially abundant taxa between groupings then separation into different PAM clusters would be expected. To select the optimal number of clusters K, the PAM model with the highest average silhouette value was selected. Factor analysis of the most important taxon features in the PCA for each marker dataset gave some preliminary evidence about which particular taxa may be differentially abundant. Relative abundance was compared for important plant taxa using a pivot table in Excel.

2.4. Chi Square Test of Proportions for the 18S Marker

The data published originally as Table 2, *Richness of Main Taxonomic Groups of Fungi in Freshwater Ecosystems* from a study that has counts for the main taxonomic groups of fungi in freshwater ecosystems has been used for the comparison [25]. The information captures data from 22 publicly available datasets from around the world. Initial exploration of the data revealed that there were few *Cryptomycota* and Chytridiomycota identified in the pooled LA River samples. The Chi-square test tested whether the proportion of Ascomycota: Basidiomycota in the LA River differed significantly from freshwater and river environments in the published data. The hypotheses that were tested for this analysis are contained in supplemental materials.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 March 2023 doi:10.20944/preprints202303.0337.v1

Table 2. The table of metadata for the LA River sites show the distribution of the samples across the site features.

Kit_Name	LA River Site	Latitude	Longitude	Habitat	River Condition
K0585_T9	Arroyo Seco	34.203154	-118.166402	Frequently submerged, intertidal, marsh	soft
K0593_C3	Arroyo Seco	34.203274	-118.166417	Terrestrial, not submerged	soft
K0594_E4	Arroyo Seco	34.202987	-118.166335	Terrestrial, not submerged	soft
K0595_B2	Arroyo Seco	34.203593	-118.166448	Terrestrial, not submerged	soft
K0595_L7	Arroyo Seco	34.203567	-118.166415	Terrestrial, not submerged	soft
K0595_T9	Arroyo Seco	34.204139	-118.166314	Terrestrial, not submerged	soft
K0597_M8	Arroyo Seco	34.20375	-118.166481	Terrestrial, not submerged	soft
K0599_L7	Arroyo Seco	34.20331	-118.166408	Frequently submerged, intertidal, marsh	soft
K0526_B2	Bowtie Parcel	34.108161	-118.246186	Fully submerged	soft
K0529_L7	Bowtie Parcel	34.108149	-118.246176	Fully submerged	soft
K0672_C3	Bowtie Parcel	34.108433	-118.246959	Fully submerged	soft
K0672_G5	Bowtie Parcel	34.108278	-118.246926	Fully submerged	soft
K0674_E4	Bowtie Parcel	34.108186	-118.246584	Fully submerged	soft
K0678_E4	Bowtie Parcel	34.108131	-118.246003	Fully submerged	soft
K0679_B2	Bowtie Parcel	34.108278	-118.246341	Fully submerged	soft
K0679_M8	Bowtie Parcel	34.108374	-118.246774	Fully submerged	soft
K0528_A1	Bull Creek	34.181558	-118.497717	Frequently submerged, intertidal, marsh	soft
K0528_E4	Bull Creek	34.182029	-118.49771	Frequently submerged, intertidal, marsh	soft
K0528_K6	Bull Creek	34.181975	-118.497849	Frequently submerged, intertidal, marsh	soft
K0529_K6	Bull Creek	34.181652	-118.497718	Frequently submerged, intertidal, marsh	soft
K0529_T9	Bull Creek	34.181651	-118.497716	Fully submerged	soft
K0530_A1	Bull Creek	34.181419	-118.497763	Frequently submerged, intertidal, marsh	soft
K0530_B2	Bull Creek	34.181342	-118.497657	Frequently submerged, intertidal, marsh	soft
K0530_E4	Bull Creek	34.1814	-118.497865	Frequently submerged, intertidal, marsh	soft
K0528_G5	Compton Creek	33.843656	-118.206466	Frequently submerged, intertidal, marsh	soft
K0528_L7	Compton Creek	33.843055	-118.205667	Fully submerged	soft
K0528_T9	Compton Creek	33.843328	-118.2061	Frequently submerged, intertidal, marsh	soft
K0529_A1	Compton Creek	33.843196	-118.205854	Frequently submerged, intertidal, marsh	soft
K0530_C3	Compton Creek	33.843311	-118.206092	Frequently submerged, intertidal, marsh	soft
K0530_K6	Compton Creek	33.842877	-118.205544	Frequently submerged, intertidal, marsh	soft
K0530_L7	Compton Creek	33.842749	-118.205402	Fully submerged	soft
K0530_M8	Compton Creek	33.843196	-118.205854	Frequently submerged, intertidal, marsh	soft
K0529_C3	Elysian Valley	34.083829	-118.228152	Fully submerged	concrete
K0672_T9	Elysian Valley	34.084621	-118.228071	Frequently submerged, intertidal, marsh	concrete
K0673_A1	Elysian Valley	34.084217	-118.228066	Frequently submerged, intertidal, marsh	concrete
K0673_G5	Elysian Valley	34.084227	-118.228048	Fully submerged	concrete
K0674_G5	Elysian Valley	34.08455	-118.228053	Fully submerged	concrete
K0676_B2	Elysian Valley	34.08449	-118.228157	Fully submerged	concrete
K0676_T9	Elysian Valley	34.084721	-118.228145	Fully submerged	concrete
K0677_A1	Elysian Valley	34.084482	-118.228157	Frequently submerged, intertidal, marsh	concrete
K0593_T9	Glendale	34.155282	-118.275211	Fully submerged	concrete
K0594_L7	Glendale	34.15459	-118.276618	Fully submerged	concrete
K0596_C3	Glendale	34.155107	-118.275459	Fully submerged	concrete

 doi:10.20944/preprints202303.0337.v1

K0596_E4	Glendale	34.154774	-118.27637	Frequently submerged, intertidal, mars	concrete
K0596_L7	Glendale	34.154918	-118.276231	Fully submerged	concrete
K0596_T9	Glendale	34.154973	-118.275799	Fully submerged	concrete
K0597_K6	Glendale	34.154997	-118.275944	Fully submerged	concrete
K0597_L7	Glendale	34.155157	-118.27542	Fully submerged	concrete
K0526_C3	Glendale Narrows	34.102813	-118.242742	Fully submerged	concrete
K0526_G5	Glendale Narrows	34.103427	-118.242642	Fully submerged	concrete
K0529_B2	Glendale Narrows	34.103109	-118.242634	Fully submerged	soft
K0529_G5	Glendale Narrows	34.103652	-118.242686	Fully submerged	concrete
K0529_M8	Glendale Narrows	34.103251	-118.242645	Fully submerged	concrete
K0672_B2	Glendale Narrows	34.10274	-118.242669	Fully submerged	concrete
K0678_B2	Glendale Narrows	34.103274	-118.242544	Fully submerged	concrete
K0678_K6	Glendale Narrows	34.103437	-118.24275	Fully submerged	concrete
K0672_A1	Long Beach	33.762909	-118.202355	Fully submerged	soft
K0674_M8	Long Beach	33.762738	-118.202271	Fully submerged	concrete
K0676_M8	Long Beach	33.762683	-118.202126	Fully submerged	concrete
K0677_B2	Long Beach	33.762833	-118.202418	Fully submerged	concrete
K0677_E4	Long Beach	33.762907	-118.202298	Fully submerged	concrete
K0677_L7	Long Beach	33.762841	-118.20235	Fully submerged	concrete
K0678_L7	Long Beach	33.762906	-118.202305	Fully submerged	soft
K0701_C3	Long Beach	33.76269	-118.202303	Fully submerged	concrete
K0527_A1	Maywood	33.986755	-118.171412	Frequently submerged, intertidal, marsh	concrete
K0527_C3	Maywood	33.988033	-118.172607	Fully submerged	concrete
K0527_E4	Maywood	33.987023	-118.171842	Fully submerged	concrete
K0527_K6	Maywood	33.986686	-118.171342	Fully submerged	concrete
K0527_L7	Maywood	33.987668	-118.172288	Fully submerged	concrete
K0527_T9	Maywood	33.986617	-118.171324	Fully submerged	concrete
K0539_L7	Maywood	33.986776	-118.17165	Fully submerged	concrete
K0593_G5	Sepulveda Dam	34.168961	-118.475292	Fully submerged	soft
K0594_A1	Sepulveda Dam	34.168698	-118.475195	Fully submerged	soft
K0594_T9	Sepulveda Dam	34.168961	-118.475292	Fully submerged	soft
K0595_G5	Sepulveda Dam	34.168941	-118.47461	Terrestrial, not submerged	soft
K0597_T9	Sepulveda Dam	34.1688	-118.475049	Fully submerged	soft
K0599_G5	Sepulveda Dam	34.16868	-118.474846	Frequently submerged, intertidal, marsh	soft
K0599_K6	Sepulveda Dam	34.168906	-118.475125	Fully submerged	soft
K0599_T9	Sepulveda Dam	34.168758	-118.474733	Rarely submerged, wetland, arroyo	soft
K0593_A1	Tujunga Wash	34.258032	-118.386781	Fully submerged	concrete
K0593_E4	Tujunga Wash	34.258403	-118.386614	Fully submerged	concrete
K0595_M8	Tujunga Wash	34.257481	-118.386845	Fully submerged	concrete

 doi:10.20944/preprints202303.0337.v1

K0596_B2	Tujunga Wash	34.258667	-118.386473	Fully submerged	concrete
K0597_E4	Tujunga Wash	34.258716	-118.386376	Fully submerged	concrete
K0599_A1	Tujunga Wash	34.258424	-118.386387	Fully submerged	concrete
K0599_E4	Tujunga Wash	34.258395	-118.386592	Fully submerged	concrete
K0599_M8	Tujunga Wash	34.258016	-118.386744	Fully submerged	concrete
K0593_L7	Verdugo Wash	34.203216	-118.237654	Fully submerged	soft
K0595_A1	Verdugo Wash	34.202985	-118.237755	Fully submerged	soft
K0596_G5	Verdugo Wash	34.202611	-118.237615	Fully submerged	soft

Overdispersion is common in taxonomic count data for environmental samples. The model that was implemented in DESeq2 to answer these research questions was a negative binomial model. In this data, zero-inflation is also suspected. The way that DESeq2 dealt with overinflation in this analysis was to analyze only positive counts. Exploratory plots for dispersion in the fungi dataset were generated to further investigate the appropriateness of the model (see supplemental materials).

Differential Abundance Analysis

For differential abundance analysis, DESeq2 was employed [26]. The DESeq2 package has handled RNA-seq or ChIP-seq, metabarcoding ASV tables, and any similar genomic data that consisted of counts. The goal was to correct some problems associated with using Chi-square test and the Poisson distribution for this type of data, which may not effectively control Type I error [26].

It was assumed that the number of reads in sample *j* assigned to gene or taxon $i=Kij\sim NB(\mu ij, \sigma^2)$ follows a negative binomial distribution (NB), which is commonly used for modelling of data in the presence of overdispersion [26].

The following further assumptions were made:

1. The mean parameter is the expectation value for *Kij* and is proportional to the actual number of sequence counts for gene *i* under the experimental condition ϱ. The size factor is also accounted for, which is essentially the coverage or sequencing depth of the genetic library for each sample.
2. The variance σ^2 is the sum of the shot noise and the raw variance.
3. The model uses a pooled variance from genes (or taxa) with similar count values to estimate the per gene raw variance.

Kij follows a Poisson distribution. If the rate that fragments are assigned to known sequences depends on a random variable *Rij=rij*, and the size factor, *sij*, then when *Rij* is modeled by the gamma distribution, $Kij\sim NB(\mu ij, \sigma^2)$, the cycle has been completed.

In terms of fitting the model, data exists in a *n* x *m* table of *Kij* counts; *i=1…n* genes in *j=1…m* samples. The parameters used were:

1. *m* size factors, including 1 for each sample.
2. *n* expression strength parameters *qip* for each condition ϱ. In other words, the expectation values for the abundance of counts for gene or taxon *i* are proportional to *qip*.
3. The pooled variance parameter simulates the dependence of *Vip* on the expectation value for the mean, *qip*, for each condition ϱ.

The size factor *sij* allows comparisons between samples with different sequencing depths. Size factors are estimated by the median of observed count ratios [26]. *qip* is estimated by a transformation of the average counts from *j* samples on condition ϱ. The fit can be applied to small numbers of replicates using local regression to estimate the raw variance. The method is a gamma family GLM for local regression that implements R locfit.

A hypothesis rejection in DESeq would mean that the difference in counts between two samples was larger than would be expected if the samples were replicates from the same individual or tissue [26]; the rejection does not indicate what is responsible for the difference. A rejection shows a taxon, protein, or gene count was differentially abundant between two samples. However, a hypothesis rejection would not reveal if it was more different than what would typically be seen if two separate locations along the same river were sampled. It would also not reveal if the difference would have a greater magnitude than if one compared the differential abundance of that taxa between two different

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 March 2023 doi:10.20944/preprints202303.0337.v1

rivers. It empowers the user to detect differences, while controlling Type I error. Volcano plots were subsequently visualized in SystemPipeR [27] and Enhanced Volcano [90].

3. Results

The Unweighted Unifrac distance method coupled with PERMANOVA, visualized by PrinCoA was the most sensitive for detection of differences between groups based on sampling site, habitat or depth. The Chi squared standardized heatmap was not sensitive. PCA alone was not sensitive, although the factor loadings were useful for revealing the few important taxa that differed between samples. PAM coupled with PCA was more useful for identifying highly similar groups of samples, and elucidating community structure. PCA with PAM gave a better visualization than the hierarchical clustering methods for this sample size, although overall the PAM and UPGMA results were very similar.

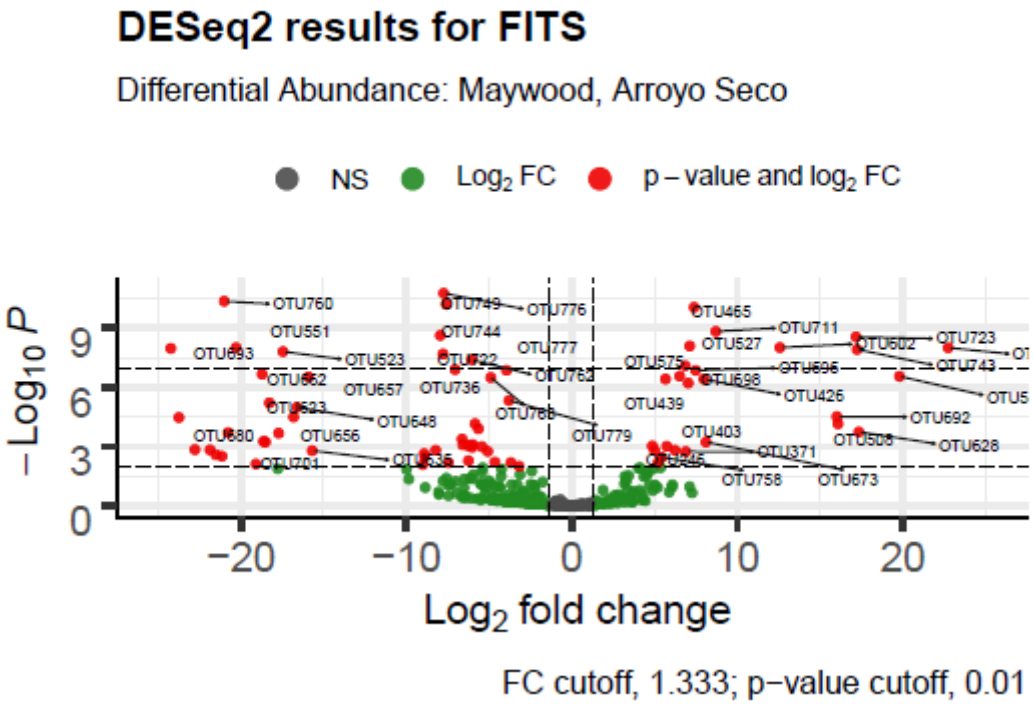
Table 3 shows the medians and ranges for taxon abundance and sequences per sample. The FITS marker had a median number of sequences per sample of 18,157. Table 3 displays the Summary Statistics resulting from the NJ (Neighbor Joining) and UPGMA (Unweighted Pair Group Method with Arithmetic Mean) Tree analyses in R phyloseq. As shown in Table 3, the Branch Length means were similar but the variance is higher for Neighbor Joining, with respect to the FITS marker. The higher variance for Neighbor Joining would be expected.

Table 3. Abundance of ASVs and Assigned Sequences per sample across the LA River sites.

LA RIVER	Taxon Abundance			Assigned Seqs/ Sample		
Marker	**Min**	**Med**	**Max**	**Min**	**Med**	**Max**
FITS	17	211	183,729	369	18,157	40,447
16S	29	181	109,927	1	15,178	44,190
18S	30	168	299,045	386	24,799	56,966
COI	30	208	153,574	14	18,555	41,257
12S	30	713	31,898	0	953	30,699
PITS	0	265	238,793	133	9,642	24,730

Depicted in Figure S3 is the PCA for the fungal ITS sequences that were recovered from the LA River sediment samples. The first two principal components capture about 37% of the variation in the data. Fungi samples separate high on PC 2 based on abundance of *Penicillium,* which may be important to the decomposition of leaf litter along the river, and *Cladosporium* sequences, which produce the antibiotic and antimalarial metabolite Cladosporin [28]. Low on PC 2, the separation is based on abundances of *Desmodesmus armatus* and *Desmodesmus* sp. variants of algae, especially in Maywood, Glendale Narrows, Glendale, and Elysian Valley. These genera have been known to break down radioactive materials.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 March 2023 doi:10.20944/preprints202303.0337.v1



Figure

Other results from the DESeq2 analysis showed that in frequently submerged river condition samples, there was a trend toward a differential abundance of fungi and a decrease in the abundance of bacteria, when compared with submerged samples. In frequently submerged sediment samples, *Capniodales sp.* were differentially abundant (p<1*10^-13), as well as *Penicillium sp.* (p<0.0005). Notably, *Tricladium angulatum* (p<1.5*10^-46), *Monocillium tenue* (p<2.5*10^-39), *Acremonium nepalense* (p<5*10^-30), and *Peziza badia* (p=9.5*10^-15) also had differentially higher abundance in frequently submerged samples.

As shown in Table 3, the 16S assay had a strong median number of sequences per sample at 15,178. This shows that the sample had good sequencing depth. As shown in Table 4, the Branch Length means are similar but the variance is about 50,000 units higher for Neighbor Joining, with respect to the 16S marker. In terms of the number of clusters reflected by the rooted and unrooted trees, both trees point to k=5 for the number of clusters in terms of bacteria.

Table 4. Summary statistics from the Neighbor Joining and UPGMA Trees for each marker. The trees were generated from the Euclidean Distance Matrix. Tree topological distances have been provided in the far-right column.

LA RIVER	Branch Length NJ		Branch Length UPGMA		NJ vs. UPGMA
Marker	**Mean**	**Variance**	**Mean**	**Variance**	**Tree Distance**
FITS	1,657	5,419,114	1,585	4,124,851	8,195
16S	620	460,349	609	417,224	2,473
18S	2,018	5,534,355	1,978	4,278,736	10,919
COI	2,312	8,746,132	2,114	6,010,691	9,697
12S	634	4,710,694	1,585	4,124,851	12,130
PITS	1,457	6,728,373	1,351	4,241,554	8,516

Figure S5 shows the PCA for the bacterial 16S DNA sequences that were recovered from the LA River sediment samples. The first two principal components capture about

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 March 2023 doi:10.20944/preprints202303.0337.v1

42% of the variation in the data. Bacteria DNA samples separate by numerous important taxa factor loadings such as abundance of *Erythrobacteracea*, *Proteobacteria*, and *Oscillatoriales cyanboacterium*.

Among others, samples from Maywood and Glendale scored low on PC 2, in the direction of high cyanobacteria abundance. Figure 3 shows the PCA plot for the 16S samples color coded by the best PAM clustering. The best PAM clustering in this case was k=4 with the highest average silhouette width. The samples in the second cluster, colored red, are from Glendale Narrows. The third cluster, colored green, is mostly made up of sediment samples from Maywood and Glendale.

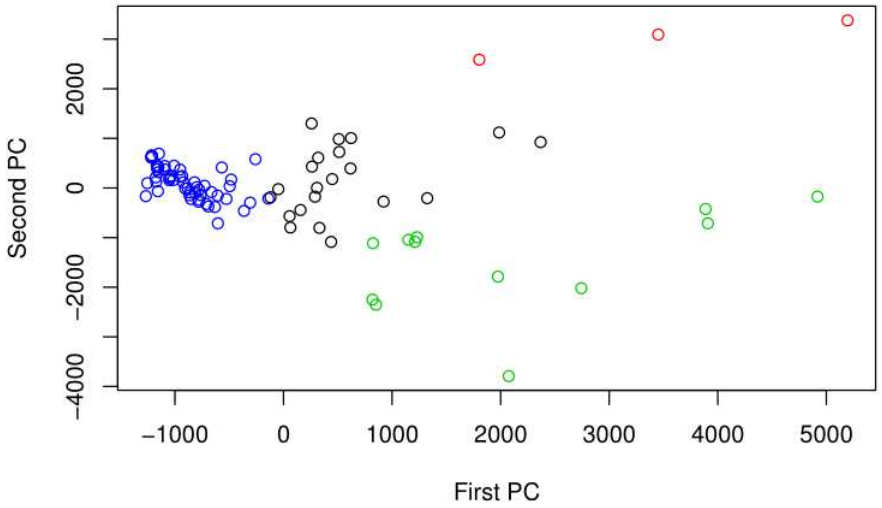


Figure 3. PCA for Bacterial identified sequences from the 16S marker by sample, color coded by the best PAM clustering. Note that there is evidence of overdispersion, in particularly high on PC1.

Among the bacteria with differentially higher abundance of 16S sequences in Glendale Narrows, *Cyanobacteria microcystis* (p<1.5*10^-7) and *Oscillatoriales cyanobacterium* (p<3*10^-14). *Verrucomicrobia* were also differentially more abundant in Glendale Narrows (p<4*10^-23). On the other hand, the alphaproteobacteria *Devosia* from *Rhizobiales* had differentially higher counts of sequences in samples from Verdugo Wash.

Table 5. The results of the differential abundance analysis for Glendale vs. Verdugo Wash. Positive log fold change results represent sequences that were differentially abundant at the Glendale site. Negative log fold changes represent sequences that were differentially abundant at the Verdugo Wash site.

Table 6. Positive log fold change results represent sequences that were differentially abundant at the soft bottom sites. Negative log fold changes represent sequences that were differentially abundant at the concrete sites.

log2FoldChange	padj	Taxon	Notes
22.09927	3.71E-23	*Prosthecobacter* sp.	possible pathogen, anaerobic, tubulin like genes, low nutrient environments
34.31956	1.53E-41	*Dechloromonas* sp.	may oxidize benzene
-22.258	5.73E-05	*Devosia* sp.	Nitrogen fixer
-25.3115	1.67E-05	*Bacillus* sp.	many beneficial species
23.78784	1.22E-06	Chromatiaceae (unclassified)	purple sulfur bacteria, use sulfide to fix carbon and generate oxygen
-30.519	0.009416	*Sandaracinobacter* sp.	metabolism of sulfide to cysteine (or from serine)

 doi:10.20944/preprints202303.0337.v1

25.68591	0.000938	*Chloroflexaceae* (unclassified)	green non-sulfur bacteria, many heat-loving anoxygenic photoheterotrophs [29, 30]
-22.3636	0.00014	endosymbiont of *Ridgeia piscesae*	Gammaproteobacterium, symbiont of a tubeworm
-6.85917	4.08E-06		anaerobic bacterium MO-CFX2 *Chloroflexi*
17.1087	4.15E-08	Rhodocyclales (unclassified)	nitrogen fixing or nitrogen reducing
33.82601	2.58E-14	*Phormidium setchellianum*	Potential cause of gastroenteritis, concentrates caused neuro- and hepato-toxicity in mice [31]
20.18264	0.000268	*Cytophaga xylanolytica*	xylan degrading, does well in sulfogenic and methanogenic environments, anaerobic and gliding
-23.4117	0.002659	*Synechococcus* sp.	Photolysis of sulfide or water, produces neurotoxins [32]
11.0032	0.000123	*Scenedesmaceae* (unclassified)	Green algae, may degrade radioactive materials
8.245038	0.000199	*Flavobacterium* sp.	Often associated with plant resistance to pathogens
7.271474	0.005122	*Oscillatoriales* cyanobacterium HF1	Cyanobacterium which may cause illness or death in humans and animals
10.11933	0.001645	*Tetradesmus obliquus*	Produces valuable saturated and unsaturated esters, extract has anticancer and antimicrobial effects [33, 34]
28.7773	1.03E-07	*Microcystis* sp.	Cyanobacterium which is toxic to humans [35]
28.91261	5.24E-05	*Rhodocyclaceae* bacterium enrichment culture clone Y62	nitrogen fixing or nitrogen reducing

log2Fold Change	**padj**	**Taxon**	**Notes**
-25.207183	3.06E-23	Oscillatoriales cyanobacterium YACCYB599	Cyanobacteria which may cause illness or death in humans and animals
-24.66764915	4.55E-23	*Chroococcus subviolaceus*	Freshwater or high salinity environments, Cyanobacteria which can survive with low O2 [36]
-24.50212313	4.55E-23	*Haliea* sp.	Marine gamma proteobacterium which tolerates up to12% salinity [37, 38]
24.49667323	3.81E-31	*Halomonas* sp.	chloride and saline tolerance
24.12963073	1.43E-27	*Marmoricola* sp.	Denitrifying bacteria [39]
10.00393321	8.21E-09	alpha proteobacterium LS7-MT	Methanol oxidizer, lives in high temperatures [40]
9.188395232	2.37E-18	*Nitrosarchaeum koreense*	Aerobic ammonia-oxidizing archaea [41]
-8.382519826	0.001244	Microcystaceae (unclassified)	Common Eutrophic Bloomer, toxin-producing Cyanobacterium
7.849119335	3.12E-07	*Acidobacterium* sp. SCGC AAA007-P13	Potential saprobe
-7.732408042	4.32E-08	Oscillatoriales cyanobacterium IRH12	Cyanobacterium which may cause illness or death in humans and animals

 doi:10.20944/preprints202303.0337.v1

-7.389766623	0.000539	*Roseisolibacter agri*	Grows in low oxygen environments [42]
-7.310779292	1.03E-07	*Pleurocapsa concharum*	Ostracod-dependent Cyanobacterium [43]
7.242636088	5.51E-07	*Devosia* sp.	Nitrogen-fixing bacteria
6.970043209	0.001616	*Nitrospira* sp. enrichment culture clone LD3	Nitrifying bacteria nitrite oxidizing bacteria
6.533527317	1.83E-13	gamma proteobacterium SCGC AAA007-P21	Uncultivated bacterioplankton
6.503508981	0.001529	alpha proteobacterium Schreyahn_AOB_Aster_Kultur_5	Cultured alphaproteobacterium
-6.479686479	0.000178	Chlamydomonadales (unclassified)	Green algae [44]
-6.382235759	0.000425	*Chloronema giganteum*	Photoautotrophic, anoxygenic green non-sulfur bacteria [91]
-6.230017507	0.002384	*Chamaesiphon* sp.	Widely distributed Cyanobacterium [45]
6.02052523	0.007591	*Altererythrobacter* sp.	Alkaline or salt tolerant aerobic phototroph, anoxygenic [46, 47, 48]
5.990283542	0.000524	Mycobacteriaceae (unclassified)	Potential human and animal pathogens
5.737312813	2.78E-06	Acidobacteriaceae (unclassified)	Likely saprobe of plant organic matter
-5.72085055	0.009826	Candidatus *Viridilinea mediisalina*	Anaerobic phototroph, salt-tolerant and prefers alkaline environments [49]
-5.56037325	2.59E-05	Veillonellaceae bacterium 6-15	bacterial vaginosis
-5.548460876	0.000699	*Phormidium setchellianum*	Cyanobacterium with possible antitumor agents, neuro and hepatotoxicity
-5.531306605	0.003193	*Calothrix* sp. UAM 374	Cyanobacterium which grows on plants and hard substrates [89]
5.344610141	0.0001	Candidatus *Nitrosocosmicus sp.*	Aerobic ammonia-oxidizing archaea
-5.019693824	0.003193	*Treponema stenostreptum*	syphilis relative
-4.952937198	0.001067	Leptolyngbyaceae (unclassified)	Thermophilic and potentially iron-loving Cyanobacterium [50]
-4.934291389	0.000964	Holophagaceae (unclassified)	Anaerobic dweller of freshwater sediments [51]

 doi:10.20944/preprints202303.0337.v1

4.926495832	0.009823	unidentified eubacterium RB01 (Verrucomicrobia)	
-4.711954167	0.002384	Xanthomonadaceae bacterium	Potential phytopathogens
-4.711366069	0.005914	*Leptolyngbya geysericola*	Alkaline tolerant non-heteroctic Cyanobacterium, produces calcite on microplastics [52]
4.50039412	4.71E-06	Caldilineales bacterium	Thermophilic and anaerobic [53]
-4.35065315	0.009823	*Fusibacter* sp. enrichment culture	Thiosulfate reducing, potentially halotolerant
-4.16646108	0.002439	*Desulfomicrobium* sp.	oxidizes sulfide and arsenate in the presence of CO2 and acetate [54], reduces nitrate to ammonium [55]
-3.874861377	0.005914	Oscillochloridaceae (unclassified)	anoxygenic phototrophic bacteria [29, 56]
-3.695598612	0.009826	Pleurocapsales (unclassified)	Cyanobacterium from calcareous environments
3.602101991	0.002384	*Vicinamibacter silvestris*	Polyphosphate accumulating organisms
2.378738101	0.004923	Firmicutes (unclassified)	High abundance in suburban rivers, negatively correlated with ammonia concentration
2.253024076	0.008829	*Stenotrophobacter terrae*	opportunistic pathogen
2.126473277	0.00044	Vicinamibacteraceae (unclassified)	Degrades chitin [57]
2.033767588	0.003193	Actinobacteria (unclassified)	Many denitrifying bacteria [58, 59]

The soft-bottom river condition was associated with a differentially higher abundance of *Alphaproteobacteria* and a decreased abundance of *Cyanobacteria Pleurocaps* (p<1*10^-6) and *Phormidium* (p<0.0007), *Oscillatoria* (p<3*10^-23), and *Chroococci* (p<5*10^-23) when contrasted with concrete sites. Notably, *Devosia* was more abundant in soft bottom (p<6*10^-7) whereas *Desulfomicrobium* (p<0.003) was more abundant under concrete conditions. On the other hand, *Verrucomicrobia* and *Haliaceae* family *Proteobacteria* were differentially abundant in the soft bottom condition (p<5*10^-23, p=0.01 respectively).

Most of the bacteria that were differentially expressed in the concrete sites were Cyanobacteria and autotrophs. There was also a trend toward a differentially high abundance of DNA sequences from potential human and plant pathogens, including the potential plant pathogen *Xanthomonas,* Clostridia, and bacteria related to the agents that cause reproductive infections. Nevertheless, the soft bottom sites also had differentially high abundances of *Norcardiaceae* and *Verrucomicrobia,* which are also potential pathogens. For the concrete sites, there was a less clear picture of the nitrogen cycle when considering the bacteria alone. There was a clear picture of the nitrogen cycle for the soft bottom sites, as well as a candidate species for phosphate accumulation.

The highest number of assigned sequences per sample was for the 18S marker, as shown in Table 2. This suggests that the highest overall sequencing depth was for the 18S assay. As shown in Table 4, the Branch Length means were both near 2,000 but the variance is about 125,000 units higher for Neighbor Joining, with respect to the 18S marker. For both tree topologies, k=4 is apparent for the number of clusters in terms of 18S sequences identified by the assay.

In Figure S6, the PCA for the 18S DNA sequences that were recovered from the LA River sediment samples is shown. The first two principal components capture about 46% of the variation in the data. The PCA by sample for 18S validates the FITS results, because the samples scored low on PC 2 based on factor loadings for *Desmodesmus* and other *Scenedesmaceae* taxa of algae. Further, samples scored high on PC 2 based on *Podocopida* and *Cypridida* high relative sequence abundance. *Podocopida* is a crustacean that has freshwater and brine-dwelling groups [60]. The *Cyprididae*

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 March 2023 doi:10.20944/preprints202303.0337.v1

are a group of freshwater Ostracods [61]. Figure S7 shows the 18S PCA color coded by the best PAM clustering, which was k=5, with the highest average silhouette width. The red samples in cluster 2 were all from Glendale. Cluster 5, in light blue, corresponds to the Long Beach sediment samples. Considering the spatial heterogeneity displayed by the samples, there is a sense that the genetic material is funneling into Long Beach, reflecting the physical landscape. The fourth cluster, in dark blue, is comprised of Sepulveda Dam, Tujunga Wash, and Arroyo Seco.

The observed alpha diversity for fungi sequences based on the 18S marker is shown in Figure 5. Los Angeles River Proportions of Ascomycota to Basidiomycota were compared to Freshwater and River habitats worldwide. The equality of these proportions were tested on a Chi square distribution. The results showed that the proportion of Ascomycota vs. Basidiomycota in the LA River differed significantly from freshwater and river environments worldwide, based on published 18S data [25].

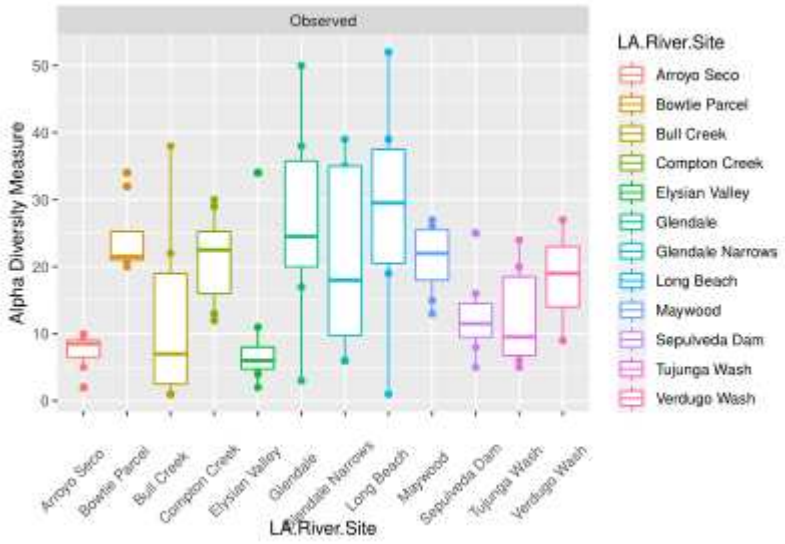
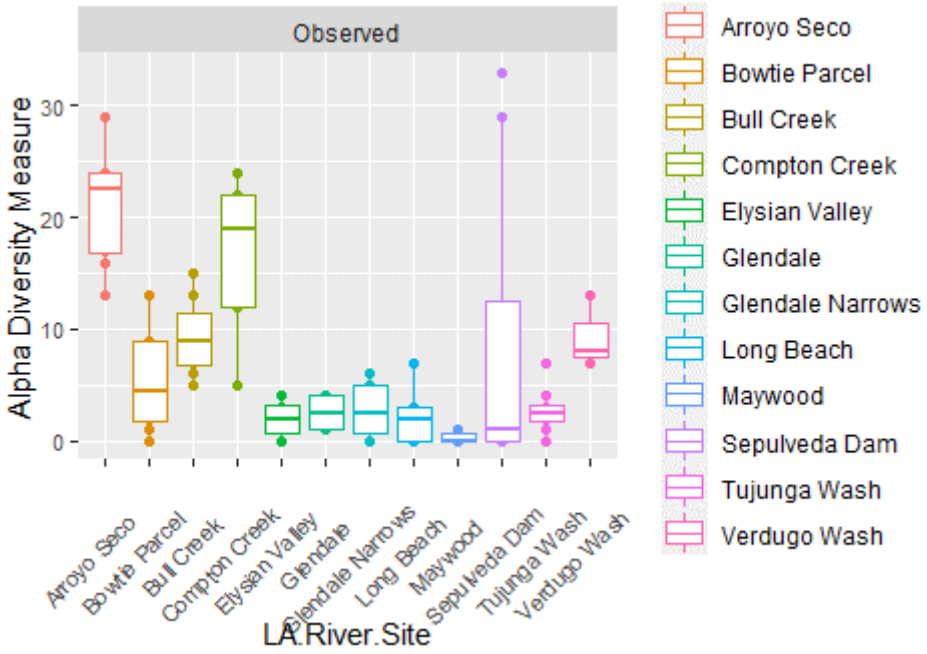


Figure 5. The observed alpha diversity of plant species is depicted in boxplots. This figure answers the question: Which site had the highest number of plant species detected overall? Note that the highest Observed Alpha Diversity tended to be in Glendale, Glendale Narrows, and Long Beach. Again, there is evidence of overdispersion, especially for the Bull Creek, Glendale, and Long Beach samples.



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 March 2023 doi:10.20944/preprints202303.0337.v1

Figure 3. The boxplot of Observed alpha diversity shows that the species richness for Ascomycota is the highest for Arroyo Seco, Bull Creek, Compton Creek, and Maywood.

The data that was used for this portion of the analysis was publicly available [25] amplicon sequence variants tables, also known as ASVs or OTUs. OTU stands for operational taxonomic unit. Essentially, these tables have counts of sequences that were identified from organisms in the environment. The goal is to compare the proportions of different Divisions of fungi in the LA River to other environments.

The results of the Chi square test for equality of proportions shows that the proportion of Ascomycota to Basidiomycota for the LA River is not equal to the proportion of Ascomycota to Basidiomycota for Freshwater Habitats ($p<0.0005$) nor River Habitats ($p<10^{-11}$) described by Lepère's analysis of worldwide freshwater data. In terms of the River Habitats, the proportion of Ascomycota to Basidiomycota is 21.5%-39.2% higher in the LA River. Furthermore, for the freshwater habitat comparison, the proportion of Ascomycota to Basidiomycota is between 7.3%-25.74% higher for the LA River, based on the 95% confidence intervals. When comparing the mosaic plots in Figure S9 and Figure 4, the gap between the proportions of Ascomycota to Basidiomycota appears smaller for the LA River compared to Freshwater Habitats in Lepère et al's study [25], compared with river environments.

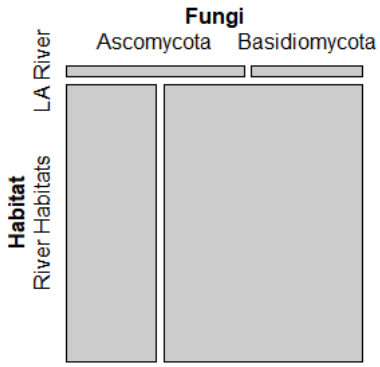


Figure 4. The mosaic plot shows that there is a difference in the proportion of ascomycetes to basidiomycetes in the LA River compared to River Habitats worldwide [25]. This gap was larger than the gap shown in Figure S9 for freshwater habitats.

The alpha diversity analysis for ascomycetes is plotted in Figure 3. The mosaic plot shows that the sites that had the most Ascomycota species were detected at Arroyo Seco, Bull Creek, Compton Creek, and Maywood. Maywood had much variability; two points were outliers with high counts >25, whereas most values were near zero. It is also interesting to note that more than 50 taxa of Ascomycota were identified only to the Family level, and some of these may represent heretofore uncharacterized ascomycetes. Based on these results, an interesting junction of the LA River to investigate ascomycete sequences to a higher depth, would be Arroyo Seco and Maywood, which were geographically connected.

The plot of alpha diversity for all fungi, given in Figure 3, shows which sites had the most different types of fungi in any Division. Overall, there were 132 taxa of fungi identified. Arroyo Seco, Bull Creek, Compton Creek, Maywood, and Verdugo Wash accumulated the most taxa. An interesting aspect of this, is that out of the 132 taxa of fungi, over 30% were ascomycetes identified only to the family level.

The COI marker performed well in terms of median sequences per sample, which was 18,555. As shown in Table 3, the Branch Length mean is about 200 units longer for NJ? and the variance is about 275,000 units higher for Neighbor Joining, with respect to the COI marker. For both tree topologies, k=3 is apparent for the number of clusters in terms of COI sequences identified. This seems to reflect that the animal diversity detected by the assay has less breadth than the biodiversity captured by 16S or FITS in this instance.

In Figure S10, the PCA for the COI DNA sequences that were recovered from the LA River sediment samples is shown. The first two principal components capture about 33% of the variation in the data. The COI assay captured a picture of lower diversity for the sequences. Samples score low on PC 2 based on relative abundance of *Dicrotendipes* species, non-biting bloodworms [62] . Also low on PC 2 were samples with high relative abundance of *Eucypris virens*, a cyprididine ostracod [63].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 March 2023 doi:10.20944/preprints202303.0337.v1

The PCA plot for the COI samples color coded by the best PAM clustering is shown in Figure S11. The best PAM clustering in this case was k=3, which exhibited the highest average silhouette width. For the COI sequences, 73 of the samples fall into the first cluster shown in black, ranging from Bowtie Parcel to Verdugo Wash. The second cluster, in red, is comprised of Glendale and Sepulveda sediment samples. The third cluster, shown in green, is made up of only 2 samples from Tujunga Wash and Glendale. This supports the observation that samples were similar for this marker.

The abundance of sequences per taxon was lower than the other markers assayed for 12S, at only 31,898, maximum. Furthermore, the median number of sequences per sample was 953. As shown in Table 3, the Branch Length means differ for NJ and UPGMA. The UPGMA mean branch length is 1585 whereas the NJ branch length is about 600. The variance is higher for Neighbor Joining, for the 12S marker, consistent with the other markers. For the NJ tree topology, k=2 appears to be the number of clusters, whereas for UPGMA, k=3 is apparent for the number of clusters in terms of 12S sequences identified.

In Figure S12, the PCA for the 12S DNA sequences that were recovered from the LA River sediment samples is given. The first two principal components capture about 63% of the variation in the data. Samples appeared similar in this assay, except for the sample which is high on PC 2 from Elysian Valley that contained a high relative abundance of salmon sequences, which appeared to be an error. In that case, since the taxon is too rare among samples it could be excluded from the analysis because it might be an error, or unlikely to be relevant to many individuals in the population. Figure S13 shows the PCA plot for the 12S samples color coded by the best PAM clustering, which was k=5, with the highest average silhouette width. 79 out of 90 samples fall into the first cluster, shown in black. The second cluster is mostly made up of Sepulveda Dam sediment DNA samples. The first and third clusters were similar to one another. The fifth cluster, in light blue, is made up of a single sample from Long Beach.

The median number of assigned sequences per sample was relatively low for the plant ITS assay at 9,642, although it was not the lowest of all markers. Nevertheless, the number of sequences per taxa had a high maximum at 238,793. As shown in Table 3, the Branch Length means were similar for NJ and UPGMA, and the variance is about 250,000 units higher for Neighbor Joining, with respect to the PITS marker. For both tree topologies, k=4 is reflected for the number of clusters in terms of plant sequences identified.

In Figure S14, it is possible to view the PCA for the plant DNA sequences that were recovered from the LA River sediment samples. The first two principal components capture about 34% of the variation in the data. One of the samples from Elysian Valley separates out on PC4 due to a high abundance of *Paspalum distictum* sequences. This is a knotgrass found in most of the Southern US and Pacific Northwest, where it is native but can become weedy [64]. It plays a role in wetland restoration since it tolerates waterlogged and saline environments, as well as providing food for deer [64]. Samples from Arroyo Seco separate out high on PC 3 based on differential abundance of *Alnus rhombifolia* sequences. Interestingly, most of the Alnus sequences came from a Tujunga Wash sample. White alders are native to streamside habitats in the US [65]. Alders have been shown to be key to nitrogen cycling in riparian environments since they form an association with *Frankia* bacteria. For that reason, they are better at colonizing disturbed habitats [65].

Various samples separate out on PC1 due to abundance of reads that were assigned only to the phylum level. The main factor that separates samples on PC 2 is the abundance of willow species, especially in Bull Creek, Bowtie Parcel, and Arroyo Seco. Most of the Salix sequences came from two samples from Arroyo Seco. Figure 6 shows the PCA for the plant sequences, color coded by the best PAM clustering. The best PAM clustering for the FITS markers was k=4. The model with four clusters had the highest average silhouette width. The second cluster, shown in red, is composed of Arroyo Seco and Bull Creek. The third cluster consists of sediment samples from Compton Creek, Sepulveda Dam, and Glendale adjacent sites. The fourth cluster, in blue, is made up of Arroyo Seco samples. The first cluster is made up of a mixture of all other samples, which were similar to one another, shown in black.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 March 2023 doi:10.20944/preprints202303.0337.v1

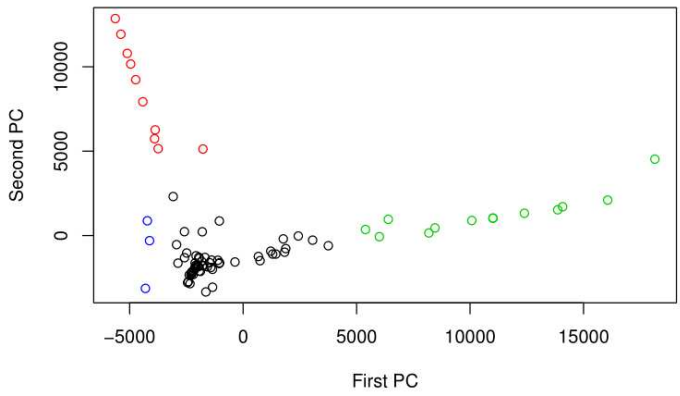


Figure 6. PCA for identified plant sequences from the PITS marker by sample is presented, color coded by the best PAM clustering.

4. Discussion

This study has investigated the associations between microorganisms and environmental conditions including soft-bottom versus concrete, degree of urbanization, and proximity to a water treatment plant. The physical distance between samples appears to be mirrored by the genetic distance, based on the evidence from PCA with PAM clustering for the 18S markers. Matsuoka et al found similar results along a river network in Japan in 2019, where they found that fungal DNA assemblages had a spatial structure and samples which were closer to one another tended to be more similar. Overall, our results agree with numerous studies of urban, eutrophic, and brackish freshwater bodies since Proteobacteria, Bacteroidetes, Firmicutes, Cyanobacteria, Chloroflexi, Actinobacteria, Acidobacteria were all well-represented [66–69]. The elevated presence of *Verrucomicrobia* and *Gammaproteobacteria* aligned more with the brackish metagenome [69]. The ostracods detected in high abundance are not known indicator species for heavy metal contamination [70].

In Glendale Narrows, downstream from water reclamation plants, there were differential abundance of cyanobacteria and algae. According to Garcia et al, the greatest social costs associated with irrigating with reclaimed water are the costs to recreation and the risks to human health due to the potential for the presence of hazardous substances [71]. Eutrophication can lead to hypoxic conditions; since hypoxia can be fatal to fish, this may partly explain the low 12S diversity. However, at Glendale Narrows, indicator species for both low nutrient environments and ammonia-abundance were also present. A potential explanation for this is the high abundance of plant species at Glendale Narrows, which assimilate nitrogen. Microbes with nutrient cycling capabilities such as Nitrogen reduction or Nitrogen fixation have been known to be associated with plant growth promotion, or may be associated with toxicity. Nevertheless, our results do not agree with Francis et al 2012, where plant species diversity was expected to decrease in urban environments compared to rural environments [72].

Eukaryotic microbes in the rootzone such as *Basidiomycota* and *Ascomycota* may help plants with phosphorus solubilization, but may be pathogenic to plants or humans. Organisms such as these fungi which promote phosphorus mineralization have received less attention over the years [73], although they play important roles in nutrient cycling. Fungi such as *Pleurotus* have been shown to mycoremediate contamination with *E. coli* [74]. The results indicate that LA River biome is rich with Ascomycota beyond the expected proportion for freshwater bodies, including rivers. *Penicillium* sp. are known to bioaccumulate arsenic and cadmium, and are thus mycoremediators of metals [75].

Nitrogen cycling was explained by differential abundance of ammonia oxidizing archaea, the complete ammonia oxidizers *Nitrospira* sp., nitrate reducing bacteria *Marmoricola* sp., and nitrogen fixing bacteria *Devosia* sp. were differentially abundant at soft-bottom sites (p adj < 0.002). The proposed nitrogen cycle for the soft bottom condition is shown Fgiure 7 Ammonia oxidizing archaea were represented by more species. This result partly disagrees with Cai et al's findings [76] since ammonia oxidizing archaea were more represented. However, some *Nitrospira* bacteria are complete ammonia oxidizers, so they may be equally important. Interestingly, the results from a recent study indicated that nitrogen pollution in river sediments also contributed to bacteria community shifts [67]. In contrast, differential abundance of several *Cyanobacteria* and other anoxygenic phototrophs was associated with the concrete bottom

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 March 2023 doi:10.20944/preprints202303.0337.v1

sites, which suggested the accumulation of excess nitrogen. *Desulfomicrobium* may play a part in nitrate reduction in concrete environments, but conserve more nutrition [55], and is sulfate-dependent [54]. Since denitrification generally requires substrate that is made under aerobic conditions [77], it makes sense that denitrifying bacteria were not as abundant in the concrete environments. Clostridia are indicator species for fecal contamination and sewage [78]. In regards to the reproductive pathogens, as Hervé et al noted street gutters are important in dispersal of putative pathogens from anthropogenic waste [79] and bioremediating species.

The diversity of cyanobacterial species observed indicated health within the cyanobacteria community. As Stal noted in 2007, *Cyanobacteria* have involvement in two essential biogeochemical processes on Earth, since they capture both CO_2 and N_2 [80]. *Cyanobacteria* have been known to colonize hostile environments [80] and to produce toxins that bring health risks to the public such as liver damage, eye irritation, vomiting, and death [81]. However, only 1-2 species of algae were highly represented, which is not an indicator of health for the ecosystem. In Wang et al's freshwater study, elevated cyanobacteria were associated with bacterioplankton, whereas algae were associated with zooplankton [82]. Heterogeneity and diversity of algae is tied to ecosystem services [83]. According to the Southern California Coastal Water Research Project, *Cladophora* algae supports the habitat of wading shorebirds [84]. Treating the underlying anaerobic conditions could promote algal and fish diversity.

The soft bottom sites tended to be represented by differential abundance of aerobes, whereas the concrete-associated species tended to be alkaliphilic, saliniphilic, calciphilic, sulfate dependent, and anaerobic. The presence of halophiles is a good indicator of salinity problems. Differential abundance of *Proteobacteria* was associated with soft bottom sites, and there was an apparent balance in the abundance of organisms responsible for nitrogen cycling.

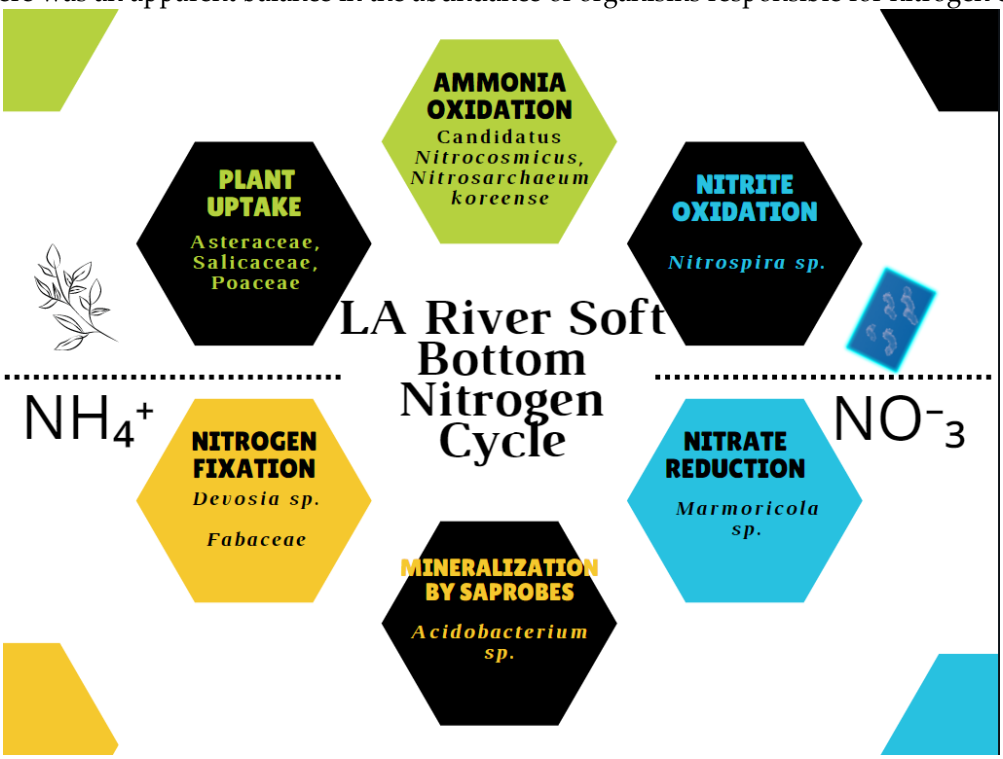


Figure 7. Proposed Nitrogen Cycle for the LA River Soft bottom condition.

In recent years, the city of Los Angeles has been reluctant to move toward a soft bottom channel restoration, since it would necessitate widening of the channel, which would potentially affect landowners and other infrastructure. Furthermore, although some activists have favored riparian plantings, this also has the potential to slow the flow of water. As the River was channelized in order to decrease flooding risk and efficiently carry away water, the introduction of a vegetative buffer would likely require a widening of the river, and possibly the river's overall footprint. As Levi et al pointed out, channel restoration benefits appear to be smaller when spread across a larger area [85]. Therefore, this type of effort may be most impactful when applied to the urban stretches that would benefit most from the intervention.

Based on the Plant Diversity Analysis, it was indicated that Maywood had high sequence abundances of weeds such as Datura, Atriplex, Oxalis, and Chenopodium, as well as high abundance of toxic Cyanobacteria based on the factor analysis, Maywood could benefit from the planting of perennial foliage that can also remediate air pollution [86]. According to Liu et al, air pollutants including particulate matter, nitrous oxide, and carbon monoxide also influence microbial and fungal communities [87]. Indications tended to suggest that sonicating devices at Maywood and Glendale Narrows for the control of *Cyanobacteria* should be considered, as well as perennial vegetative buffers in Maywood to combat noxious *Datura* plant species and toxic *Cyanobacteria* blooms. Interestingly, Maywood samples had differentially abundant *Tetradesmus* sp., including T. *obliquus*, which is a phosphorus accumulator and produces valuable lipids for biodiesel [33]. *T. obliquus* may also be used for animal feed; it is known to be rich in amino acids, including the essential amino acid leucine, with a low bioaccumulation of metals [34].

A surprising result is that some sites along the LA River were more diverse with plant life than rural Arroyo Seco, especially Bowtie Parcel, Glendale, Long Beach, and Maywood, based on observed alpha-diversity. This is most likely due to landscape plantings of exotic species near Glendale, coastal species at Long beach, and a diverse panel of weed sequences that were identified at Maywood. Plants prevent erosion and create habitat for birds, mammals, invertebrates, amphibians, and reptiles. Plants also help balance nitrogen cycling and can provide a buffer by absorbing some of the nutrients involved in eutrophication.

5. Conclusions

Further research should consider the efficacy of sonicating devices at Maywood and Glendale Narrows for the control of *Cyanobacteria* [88]. There were poorly characterized microbes and arthropods identified in this study that may present an opportunity for further investigation. These include a possible new species of *Capniodales* sooty mold in the submerged samples, little known *Chironomidae* lake flies in the Glendale Narrows sample, *Desulfomicrobia* in concrete environments, elusive *Eustigmatophyaceae* in Maywood, and unstudied *Verrucomicrobia* and *Flavobacter* in Glendale Narrows. Arroyo Seco and Maywood, which are geographically connected, present an interesting junction of the LA River to investigate ascomycetes and sequence to a higher depth. This is one of the first attempts to characterize the metagenome of the LA River. The diversity and interaction of the bacterial communities with plants and other organisms warrants more attention. The outcomes appear to involve interactions between environmental factors. Further research should consider functional analysis of similar associations.

Author Contributions: Conceptualization, S.S.; methodology, S.S.; formal analysis, S.S.; investigation, S.S., D.M., R.S.; writing—original draft preparation, S.S.; writing—review and editing, A.E.T., D.M., G.P., S.B., R.S., J.F.; visualization, S.S.; supervision, G.P., A.E.T., S.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Research data for the L.A. River Round 1 Project is available from CaleDNA at: https://data.ucedna.com/research_projects/los-angeles-river-round-1/pages/introduction.

Acknowledgments: The authors acknowledge support from the University of California CaleDNA program who kindly provided data to us prior to publication.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. St. Louis Federal Reserve. Federal Reserve Economic Data. https://fred.stlouisfed.org/
2. United States Census Bureau. Quick Facts: Los Angeles city, California. https://www.census.gov/quickfacts/losangelescitycalifornia
3. California Environmental Protection Board. Los Angeles River Watershed impaired waters. State Water Resources Control Board - Los Angeles. Retrieved July 6, 2022, from https://www.waterboards.ca.gov/rwqcb4/water_issues/programs/regional_program/Water_Quality_and_Watersheds/los_angeles_river_watershed/303.shtml
4. Protecting Our River. Accessed 10/17/2021. Protectingourriver.org
5. Qiu, Han, Gu, Likun, Sun, Bo, Zhang, Jianyun, Zhang, Miao, He, Shanshan, An, Shuqing, and Leng, Xin. "Metagenomic Analysis Revealed That the Terrestrial Pollutants Override the Effects of Seasonal Variation on Microbiome in River Sediments." Bulletin of Environmental Contamination and Toxicology 105.6 (2020): 892-98. Web.
6. Wei, Feifei, Sakata, Kenji, Asakura, Taiga, Date, Yasuhiro, and Kikuchi, Jun. "Systemic Homeostasis in Metabolome, Ionome, and Microbiome of Wild Yellowfin Goby in Estuarine Ecosystem." Scientific Reports 8.1 (2018): 3478-12. Web.

7. Garba, Fatima et al. “Deteriorating Water Quality State on the Structural Assemblage of Aquatic Insects in a North-Western Nigerian River.” Water Science 36.1 (2022): 22–31.
8. Qiusheng Yuan, Peifang Wang, Chao Wang, Juan Chen, Xun Wang, Sheng Liu, Indicator species and co-occurrence pattern of sediment bacterial community in relation to alkaline copper mine drainage contamination, Ecological Indicators, Volume 120, 2021, 106884, ISSN 1470-160X, https://doi.org/10.1016/j.ecolind.2020.106884.
9. Khalilova, E A et al. “Halophilic bacteria of salt lakes and saline soils of the Peri-Caspian lowland (Republic of Daghestan) and their biotechnological potential.” *Vavilovskii zhurnal genetiki i selektsii* vol. 25,2 (2021): 224-233. doi:10.18699/VJ21.026
10. Ackerman, D., et al. "Characterization of water quality in the LA River." (2003).
11. Desfor, Gene, and Keil, Roger. "Every River Tells a Story: The Don River (Toronto) and the Los Angeles River (Los Angeles) as Articulating Landscapes." Journal of Environmental Policy & Planning 2.1 (2000): 5-23. Web.
12. Gandy, Matthew. "Riparian Anomie: Reflections on the Los Angeles River." Landscape Research 31.2 (2006): 135-45. Web.
13. Seth J. Wenger, Allison H. Roy, C. Rhett Jackson, Emily S. Bernhardt, Timothy L. Carter, Solange Filoso,et al. "Twenty-six key research questions in urban stream ecology: an assessment of the state of the science," Journal of the North American Benthological Society, 28(4), 1080-1098, (27 October 2009)
14. Rachael E. Antwis, Sarah M. Griffiths, Xavier A. Harrison, Paz Aranega-Bou, Andres Arce, Aimee S. Bettridge, Francesca L. Brailsford, et al. Fifty important research questions in microbial ecology, FEMS Microbiology Ecology, Volume 93, Issue 5, May 2017, fix044, https://doi.org/10.1093/femsec/fix044
15. Orsi, Jared. Hazardous Metropolis: Flooding and Urban Ecology in Los Angeles, University of California Press, 2004. ProQuest Ebook Central, https://ebookcentral.proquest.com/lib/osu/detail.action?docID=227297.
16. Halecki, Wiktor, and Tomasz Stachura. “Evaluation of Soil Hydrophysical Parameters Along a Semiurban Small River; Soil Ecosystem Services for Enhancing Water Retention in Urban and Suburban Green Areas.” Catena (Giessen) 196 (2021): 104910–. Web. Accessed June 28, 2022.
17. Behera, Bijay Kumar, Patra, Biswanath, Chakraborty, Hirak Jyoti, Sahu, Parameswar, Rout, Ajaya Kumar, Sarkar, Dhruba Jyoti, Parida, Pranaya Kumar, Raman, Rohan Kumar, Rao, Atmakuri Ramakrishna, Rai, Anil, Das, Basanta Kumar, Jena, Joykrushna, Mohapatra, Trilochan, and Dikhit, Manas Ranjan. "Metagenome Analysis from the Sediment of River Ganga and Yamuna: In Search of Beneficial Microbiome." PloS One 15.10 (2020): E0239594. Web.
18. 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.; et al. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods **2010**, 7, 335–336.
19. Curd, EE, Gold, Z, Kandlikar, GS, et al. Anacapa Toolkit: An environmental DNA toolkit for processing multilocus metabarcode datasets. Methods Ecol Evol. 2019; 10: 1469– 1475. https://doi.org/10.1111/2041-210X.13214
20. Borg, I. and Groenen, P. (1997) Modern Multidimensional Scaling. Theory and Applications. Springer.
21. Hastie, Tibshirani, and Friedman (2009). The Elements of Statistical Learning (ESL) (2nd edition).
22. McMurdie and Holmes (2013) phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS ONE. 8(4):e61217
23. Kandlikar, G.S.; Gold, Z.; Cowen, M.C.; Meyer, R.S.; Freise, A.C.; Kraft, N.J.; Moberg-Parker, J.; Sprague, J.; Kushner, D.J.; Curd, E.E. ranacapa: An R package and Shiny web app to explore environmental DNA data with exploratory statistics and interactive visualizations. F1000Research 2018, 7, 1734. https://doi.org/10.12688/f1000research.16680.1.
24. Schubert, Erich, and Peter J. Rousseeuw. "Fast and eager k-medoids clustering: O (k) runtime improvement of the PAM, CLARA, and CLARANS algorithms." Information Systems (2021): 101804.
25. Lepère C, Domaizon I, Humbert J, Jardillier L, Hugoni M, Debroas D. 2019. Diversity, spatial distribution and activity of fungi in freshwater ecosystems. PeerJ 7:e6247 https://doi.org/10.7717/peerj.6247S
26. Anders, S., Huber, W. Differential expression analysis for sequence count data. Genome Biol 11, R106 (2010). https://doi.org/10.1186/gb-2010-11-10-r106
27. Backman, T.W.H and Girke, T. (2016). systemPipeR: NGS Workflow and Report Generation Environment. BMC Bioinformatics, 17: 388. 10.1186/s12859-016-1241-0.
28. Maria Lukarska, Andrés Palencia, Chapter Eleven - Aminoacyl-tRNA synthetases as drug targets, Editor(s): Lluís Ribas de Pouplana, Laurie S. Kaguni, The Enzymes, Academic Press, Volume 48, 2020, Pages 321-350, ISSN 1874-6047, ISBN 9780128202609, https://doi.org/10.1016/bs.enz.2020.07.001.
29. Hanada, S. (2014). The Phylum Chloroflexi, the Family Chloroflexaceae, and the Related Phototrophic Families Oscillochloridaceae and Roseiflexaceae . In: Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E., Thompson, F. (eds) The Prokaryotes. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-38954-2_165
30. John C. Willison, Jean-Pierre Magnin, Chapter Eight - Role and Evolution of Endogenous Plasmids in Photosynthetic Bacteria, Editor(s): J. Thomas Beatty, Advances in Botanical Research, Academic Press, Volume 66, 2013, Pages 227-265, ISSN 0065-2296, ISBN 9780123979230, https://doi.org/10.1016/B978-0-12-397923-0.00008-4.
31. Teneva, Ivanka et al. “Toxic potential of five freshwater Phormidium species (Cyanoprokaryota).” Toxicon : official journal of the International Society on Toxinology vol. 45,6 (2005): 711-25. doi:10.1016/j.toxicon.2005.01.018
32. “Synechococcus.” Microbewiki, Kenyon College, https://microbewiki.kenyon.edu/index.php/Synechococcus.

33. Oliveira, C.Y.B., Oliveira, C.D.L., Prasad, R., Ong, H.C., Araujo, E.S., Shabnam, N. and Gálvez, A.O. (2021), A multidisciplinary review of Tetradesmus obliquus: a microalga suitable for large-scale biomass production and emerging environmental applications. Rev. Aquacult., 13: 1594-1618. https://doi.org/10.1111/raq.12536
34. Luisa Gouveia, Jelena Molnar Jazić, Alice Ferreira, Snežana Maletić, Dragoljub Cvetković, Senka Vidović, Jelena Vladić,
35. Montclair State University. Microcystis. New Jersey Center For Water Science And Technology. Retrieved July 3, 2022, from https://www.montclair.edu/water-science/freshwater-cyanobacteria-of-new-jersey/visual-guide-to-cyanobacteria-in-new-jersey/coccoid/colonial/microcystis/
36. "Chroococcus." Microbewiki, Kenyon College, https://microbewiki.kenyon.edu/index.php/Chroococcus#Cell_Structure_and_Metabolism
37. Lucena, Teresa et al. "Haliea mediterranea sp. nov., a marine gammaproteobacterium." International journal of systematic and evolutionary microbiology vol. 60,Pt 8 (2010): 1844-1848. doi:10.1099/ijs.0.017061-0
38. Yang, L., Tan, Z., Wang, D. et al. Species identification through mitochondrial rRNA genetic analysis. Sci Rep 4, 4089 (2014). https://doi.org/10.1038/srep04089
39. Kasozi N, Kaiser H, Wilhelmi B. Metabarcoding Analysis of Bacterial Communities Associated with Media Grow Bed Zones in an Aquaponic System. Int J Microbiol. 2020 Oct 1;2020:8884070. doi: 10.1155/2020/8884070. PMID: 33061984; PMCID: PMC7547338.
40. Islam T, Hernández M, Gessesse A, Murrell JC, Øvreås L. A Novel Moderately Thermophilic Facultative Methylotroph within the Class Alphaproteobacteria. Microorganisms. 2021 Feb 25;9(3):477. doi: 10.3390/microorganisms9030477. PMID: 33668875; PMCID: PMC7996495.
41. Jung, Man-Young et al. "Nitrosarchaeum koreense gen. nov., sp. nov., an aerobic and mesophilic, ammonia-oxidizing archaeon member of the phylum Thaumarchaeota isolated from agricultural soil." International journal of systematic and evolutionary microbiology vol. 68,10 (2018): 3084-3095. doi:10.1099/ijsem.0.002926
42. Pascual, Javier et al. "Roseisolibacter agri gen. nov., sp. nov., a novel slow-growing member of the under-represented phylum Gemmatimonadetes." International journal of systematic and evolutionary microbiology vol. 68,4 (2018): 1028-1036. doi:10.1099/ijsem.0.002619
43. "Pleurocapsa." Microbewiki, Kenyon College, microbewiki.kenyon.edu/index.php/Pleurocapsa.
44. Hu Y, Xing W, Song H, Zhu H, Liu G and Hu Z (2019) Evolutionary Analysis of Unicellular Species in Chlamydomonadales Through Chloroplast Genome Comparison With the Colonial Volvocine Algae. Front. Microbiol. 10:1351. doi: 10.3389/fmicb.2019.01351
45. Kurmayer, R., Christiansen, G., Holzinger, A. et al. Single colony genetic analysis of epilithic stream algae of the genus Chamaesiphon spp.. Hydrobiologia 811, 61–75 (2018). https://doi.org/10.1007/s10750-017-3295-z
46. Dahal, Ram Hari, and Jaisoo Kim. "Altererythrobacter Fulvus Sp. Nov., a Novel Alkalitolerant Alphaproteobacterium Isolated from Forest Soil." International journal of systematic and evolutionary microbiology 68.5 (2018): 1502–1508. Web.
47. Li, Hui-Ping et al. "Altererythrobacter rhizovicinus sp. nov., isolated from rhizosphere soil of Haloxylon ammodendron." International journal of systematic and evolutionary microbiology vol. 70,1 (2020): 680-686. doi:10.1099/ijsem.0.003817
48. Fidalgo, Cátia et al. "Altererythrobacter halimionae sp. nov. and Altererythrobacter endophyticus sp. nov., two endophytes from the salt marsh plant Halimione portulacoides." International journal of systematic and evolutionary microbiology vol. 67,8 (2017): 3057-3062. doi:10.1099/ijsem.0.002079
49. Gaisin, Vasil A et al. "'Candidatus Viridilinea mediisalina', a novel phototrophic Chloroflexi bacterium from a Siberian soda lake." *FEMS microbiology letters* vol. 366,5 (2019): fnz043. doi:10.1093/femsle/fnz043
50. Brown, Igor, Tringe, Susannah G., Ivanova, Natalia, Goodwin, Lynne, Shapiro, Nicole, Alcorta, Jaime, Pan, Donald, Chistoserdov, Andrei, Sarkisova, Svetlana, Woyke, Tanja, and Maresca, ed., Julia A. High-Quality Draft Genome Sequence of the Siderophilic and Thermophilic Leptolyngbyaceae Cyanobacterium JSC-12. United States: 2021. Web. doi:10.1128/MRA.00495-21.
51. Fukunaga, Y., Ichikawa, N. (2014). The Class Holophagaceae . In: Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E., Thompson, F. (eds) The Prokaryotes. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-38954-2_161
52. Mateos Cárdenas, Alicia et al. "Microplastics in the Freshwater Environment." Reference Module in Earth Systems and Environmental Sciences 2021. Web.
53. Yamada, Takeshi, et al. "Anaerolinea Thermolimosa Sp. Nov., Levilinea Saccharolytica Gen. Nov., Sp. Nov. and Leptolinea Tardivitalis Gen. Nov., Sp. Nov., Novel Filamentous Anaerobes, and Description of the New Classes Anaerolineae Classis Nov. and CALDILINEAE Classis Nov. in the Bacterial Phylum Chloroflexi." *International Journal of Systematic and Evolutionary Microbiology*, vol. 56, no. 6, 2006, pp. 1331–1340., doi:10.1099/ijs.0.64169-0.
54. Rabus, Ralf et al. "A Post-Genomic View of the Ecophysiology, Catabolism and Biotechnological Relevance of Sulphate-Reducing Prokaryotes." Advances in microbial physiology 66 (2015): 55–. Print.
55. Zhang, Tong et al. "Investigation of Dissimilatory Nitrate Reduction to Ammonium (DNRA) in Urban River Network Along the Huangpu River, China: Rates, Abundances, and Microbial Communities." Environmental science and pollution research international 29.16 (2021): 23823–23833.
56. Keppen, O I et al. "Proposal of Oscillochloridaceae fam. nov. on the basis of a phylogenetic analysis of the filamentous anoxygenic phototrophic bacteria, and emended description of Oscillochloris and Oscillochloris trichoides in comparison with further new isolates." International journal of systematic and evolutionary microbiology vol. 50 Pt 4 (2000): 1529-1537. doi:10.1099/00207713-50-4-1529

57. Huber, Katharina J, and Jörg Overmann. "Vicinamibacteraceae Fam. Nov., the First Described Family Within the Subdivision 6 Acidobacteria." International journal of systematic and evolutionary microbiology 68.7 (2018): 2331–2334. Web.
58. Zhang, Haihan et al. "Nitrate reduction by the aerobic denitrifying actinomycete Streptomyces sp. XD-11-6-2: Performance, metabolic activity, and micro-polluted water treatment." Bioresource technology vol. 326 (2021): 124779. doi:10.1016/j.biortech.2021.124779
59. St. Clair, Savanah, et al. "Analysis of the soil microbiome of a Los Angeles urban farm." *Applied and Environmental Soil Science* 2020 (2020): 1-16.
60. WoRMS Editorial Board (2021). World Register of Marine Species. Available from http://www.marinespecies.org at VLIZ. Accessed 2021-04-20. doi:10.14284/170
61. ITIS Standard Report: Cyprididae. Retrieved Mar 2021, from the Integrated Taxonomic Information System on-line database, http://www.itis.gov.
62. Myers, P., R. Espinosa, C. S. Parr, T. Jones, G. S. Hammond, and T. A. Dewey. 2021. The Animal Diversity Web (online). Accessed at https://animaldiversity.org, 6/1/2021.
63. Smith R.J., Martens K. (2000) The ontogeny of the cypridid ostracod Eucypris virens (Jurine, 1820) (Crustacea, Ostracoda). In: Horne D.J., Martens K. (eds) Evolutionary Biology and Ecology of Ostracoda. Developments in Hydrobiology, vol 148. Springer, Dordrecht. https://doi.org/10.1007/978-94-017-1508-9_3
64. USDA, NRCS. 2021. The PLANTS Database (http://plants.usda.gov, 20 April 2021). National Plant Data Team, Greensboro, NC 27401-4901 USA.
65. White Alder (U.S. National PARK SERVICE). Last updated Jan 31, 2020. https://www.nps.gov/articles/white-alder.htm.
66. Gu, Qiyan et al. "Pollution Profile of Waterborne Bacterial and Fungal Community in Urban Rivers of Pearl River Estuary: Microbial Safety Assessment." Journal of freshwater ecology 36.1 (2021): 305–322.
67. Lin, Xianbiao et al. "Bacterial Community Shifts Driven by Nitrogen Pollution in River Sediments of a Highly Urbanized City." International journal of environmental research and public health 16.20 (2019): 3794–.
68. Wang, Peng et al. "Characteristics of Aquatic Bacterial Community and the Influencing Factors in an Urban River." The Science of the total environment 569-570 (2016): 382–389.
69. Luisa W Hugerth et al. "Metagenome-Assembled Genomes Uncover a Global Brackish Microbiome." bioRxiv (2015).
70. Iepure, Sanda et al. "Response of Microcrustacean Communities from the Surface—groundwater Interface to Water Contamination in Urban River System of the Jarama Basin (central Spain)." Environmental science and pollution research international 20.8 (2013): 5813–5826.
71. Garcia, X, and D Pargament. "Reusing Wastewater to Cope with Water Scarcity: Economic, Social and Environmental Considerations for Decision-Making." Resources, conservation and recycling 101 (2015): 154–166.
72. Francis, Robert A. "Positioning Urban Rivers Within Urban Ecology." Urban ecosystems 15.2 (2012): 285–291.
73. Li, Kexin et al. "Microbial Abundance and Diversity Investigations Along Rivers: Current Knowledge and Future Directions." Wiley interdisciplinary reviews. Water 8.5 (2021): e1547.
74. Pini, Andrea K, and Pamela Geddes. "Fungi Are Capable of Mycoremediation of River Water Contaminated by E. Coli." Water, air, and soil pollution 231.2 (2020).
75. Cecchi, Grazia et al. "From Waste to Resource: Mycoremediation of Contaminated Marine Sediments in the SEDITERRA Project." Journal of soils and sediments 20.6 (2019): 2653–2663.
76. Cai, Xianlei et al. "Properties of Ammonia-oxidising Bacteria and Archaea in a Hypereutrophic Urban River Network." Freshwater biology 67.2 (2022): 250–262. Web.
77. J.P. Megonigal, M.E. Hines, P.T. Visscher, 8.08 - Anaerobic Metabolism: Linkages to Trace Gases and Aerobic Processes, Editor(s): Heinrich D. Holland, Karl K. Turekian, Treatise on Geochemistry, Pergamon, 2003, Pages 317-424, ISBN 9780080437514, https://doi.org/10.1016/B0-08-043751-6/08132-9.
78. Wen, X.; Chen, F.; Lin, Y.; Zhu, H.; Yuan, F.; Kuang, D.; Jia, Z.; Yuan, Z. Microbial Indicators and Their Use for Monitoring Drinking Water Quality—A Review. *Sustainability* **2020**, *12*, 2249. https://doi.org/10.3390/su12062249
79. Herve, Vincent et al. "Aquatic Urban Ecology at the Scale of a Capital: Community Structure and Interactions in Street Gutters." The ISME Journal 12.1 (2018): 253–266.
80. Stal L.J. (2007) Cyanobacteria. In: Seckbach J. (eds) Algae and Cyanobacteria in Extreme Environments. Cellular Origin, Life in Extreme Habitats and Astrobiology, vol 11. Springer, Dordrecht. https://doi.org/10.1007/978-1-4020-6112-7_36
81. Bláha, Luděk et al. "Toxins produced in cyanobacterial water blooms - toxicity and risks." Interdisciplinary toxicology vol. 2,2 (2009): 36-41. doi:10.2478/v10102-009-0006-2
82. Wang, Kai, Mandy Razzano, and Xiaozhen Mou. "Cyanobacterial Blooms Alter the Relative Importance of Neutral and Selective Processes in Assembling Freshwater Bacterioplankton Community." The Science of the total environment 706 (2020): 135724–135724.
83. Caro-Borrero, Angela, and Javier Carmona-Jiménez. "The Use of Macroinvertebrates and Algae as Indicators of Riparian Ecosystem Services in the Mexican Basin: a Morpho-Functional Approach." Urban ecosystems 22.6 (2019): 1187–1200.
84. "Assessment of Aquatic Life Use Needs for the Los Angeles River." *Los Angeles River Environmental Flows Project*, Southern California Coastal Water Research Project, ftp.sccwrp.org/pub/download/DOCUMENTS/TechnicalReports/1154_LARiverAquaticLifeUses.pdf.

85. Levi, Peter S., and Peter B. McIntyre. "Ecosystem Responses to Channel Restoration Decline with Stream Size in Urban River Networks." Ecological applications 30.5 (2020): e02107–n/a.
86. North Dakota State University Extension. Prevent Cyanobacteria Blooms. Aug 2020. https://www.ag.ndsu.edu/news/news-releases/2020/aug-10-2020/prevent-cyanobacteria-blooms
87. Liu, Huan et al. "The Distribution Variance of Airborne Microorganisms in Urban and Rural Environments." Environmental pollution (1987) 247 (2019): 898–906.
88. US EPA. Control Measures for Cyanobacterial HABs in Surface Water. https://www.epa.gov/cyanohabs/control-measures-cyanobacterial-habs-surface-water
89. "Calothrix (Cyanobacteria)." *Phycokey - Calothrix*, University of New Hampshire Center for Freshwater Biology, cfb.unh.edu/phycokey/Choices/Cyanobacteria/cyano_filaments/cyano_unbranched_fil/tapered_filaments/CALOTHRIX/Calothrix_key.htm#:~:text=Habitat%3A,in%20the%20marine%20intertidal%20zone.
90. Blighe K, Rana S, Lewis M (2021). EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling. R package version 1.10.0, https://github.com/kevinblighe/EnhancedVolcano.
91. Gorlenko, Vladimir & Pierson, Beverly. (2015). Chloronema. 10.1002/9781118960608.gbm00381.
92. Matsuoka S, Sugiyama Y, Sato H, Katano I, Harada K, Doi H (2019) Spatial structure of fungal DNA assemblages revealed with eDNA metabarcoding in a forest river network in western Japan. Metabarcoding and Metagenomics 3: e36335. https://doi.org/10.3897/mbmg.3.36335
93. D. Baragaño, G. Ratié, C. Sierra, V. Chrastný, M. Komárek, J.R. Gallego, Multiple pollution sources unravelled by environmental forensics techniques and multivariate statistics, Journal of Hazardous Materials,Volume 424, Part B, 2022, 127413, ISSN 0304-3894, https://doi.org/10.1016/j.jhazmat.2021.127413.
94. UC CalEDNA DNA DATA SERVICE — CALeDNA (ucedna.com)
95. Badotti, F., de Oliveira, F.S., Garcia, C.F. et al. Effectiveness of ITS and sub-regions as DNA barcode markers for the identification of Basidiomycota (Fungi). BMC Microbiol 17, 42 (2017). https://doi.org/10.1186/s12866-017-0958-x
96. Gilbert J. (2015) Metagenomics, Metadata, and Meta-analysis. In: Nelson K.E. (eds) Encyclopedia of Metagenomics. Springer, Boston, MA. https://doi-org.ezproxy.proxy.library.oregonstate.edu/10.1007/978-1-4899-7478-5_17
97. Tran, Q., Pham, DT. & Phan, V. Using 16S rRNA gene as marker to detect unknown bacteria in microbial communities. BMC Bioinformatics 18, 499 (2017). https://doi.org/10.1186/s12859-017-1901-8
98. Kisand, Veljo, Talas, Liisi, Kisand, Anu, Stivrins, Normunds, Reitalu, Triin, Alliksaar, Tiiu, Vassiljev, Jüri, Liiv, Merlin, Heinsalu, Atko, Seppä, Heikki, and Veski, Siim. "From Microbial Eukaryotes to Metazoan Vertebrates: Wide Spectrum Paleo-diversity in Sedimentary Ancient DNA over the Last ~14,500 years." Geobiology 16.6 (2018): 628-39. Web.
99. Rach, Jessica, Bergmann, Tjard, Paknia, Omid, DeSalle, Rob, Schierwater, Bernd, and Hadrys, Heike. "The Marker Choice: Unexpected Resolving Power of an Unexplored CO1 Region for Layered DNA Barcoding Approaches." PloS One 12.4 (2017): E0174842. Web.
100. Pumariño L, Alomar O, Agustí N. Development of specific ITS markers for plant DNA identification within herbivorous insects. Bull Entomol Res. 2011 Jun;101(3):271-6. doi: 10.1017/S0007485310000465. Epub 2010 Nov 24. PMID: 21092379.
101. Green approach for the valorization of microalgae Tetradesmus obliquus, Sustainable Chemistry and Pharmacy, Volume 24, 2021, 100556, ISSN 2352-5541, https://doi.org/10.1016/j.scp.2021.100556.
102. Shoun H, Kano M, Baba I, Takaya N, Matsuo M. Denitrification by actinomycetes and purification of dissimilatory nitrite reductase and azurin from Streptomyces thioluteus. J Bacteriol. 1998 Sep;180(17):4413-5. doi: 10.1128/JB.180.17.4413-4415.1998. PMID: 9721277; PMCID: PMC107449.
103. Amaral-Zettler, L.A., McCliment, E.A., Ducklow, H.W. and Huse, S.M., 2009. A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. PloS one, 4: p.e6372.
104. Caporaso J. G., Lauber C. L., Walters W. A., Berg-Lyons D., Huntley J., et al. - Ultrahigh-throughput microbial community analysis on the Illumina HiSeq and MiSeqplatforms. The ISME Journal 6 (2012) 1621–1624. doi: 10.1038/ismej.2012.8 PMID: 22402401."
105. Epp, L.S., Boessenkool, S., Bellemain, E.P., Haile, J., Esposito, A., Riaz, T., Erseus, C., Gusarov, V.I., Edwards, M.E., Johnsen, A. and Stenøien, H.K., 2012. New environmental metabarcodes for analysing soil DNA: potential for studying past and present ecosystems. Molecular Ecology, 21: pp.1821-1833.
106. Gu, W., Song, J., Cao, Y., Sun, Q., Yao, H., Wu, Q., Chao, J., Zhou, J., Xue, W. and Duan, J., 2013. Application of the ITS2 region for barcoding medicinal plants of Selaginellaceae in Pteridophyta. PloS one, 8: p.e67818.
107. Leray, M., Yang, J.Y., Meyer, C.P., Mills, S.C., Agudelo, N., Ranwez, V., Boehm, J.T. and Machida, R.J., 2013. A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. Frontiers in Zoology, 10 p.34.