Computational and Structural Biotechnology Journal 13 (2015) 18–23

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

Communications

Transcriptomics in the tropics: Total RNA-based profiling of Costa Rican bromeliad-associated communities

Shana K. Goffredi [a,*], Gene E. Jang [a], Mohamed F. Haroon [b]

[a] Occidental College, United States
[b] University of Queensland, Australia

ARTICLE INFO

Article history:
Received 16 October 2014
Received in revised form 1 December 2014
Accepted 3 December 2014
Available online 13 December 2014

Keywords:
Bromeliad
Metatranscriptome
Methanogen
Rainforest
Diversity

ABSTRACT

RNA-Seq was used to examine the microbial, eukaryotic, and viral communities in water catchments ('tanks') formed by tropical bromeliads from Costa Rica. In total, transcripts with taxonomic affiliation to a wide array of bacteria, archaea, and eukaryotes, were observed, as well as RNA-viruses that appeared related to the specific presence of eukaryotes. Bacteria from 25 phyla appeared to comprise the majority of transcripts in one tank (Wg24), compared to only 14 phyla in the other (Wg25). Conversely, eukaryotes from only 16 classes comprised the majority of transcripts in Wg24, compared to 24 classes in the Wg25, revealing a greater eukaryote diversity in the latter. Given that these bromeliads had tanks of similar size (i.e. vertical oxygen gradient), and were neighboring with presumed similar light regime and acquisition of leaf litter through-fall, it is possible that pH was the factor governing these differences in bacterial and eukaryotic communities (Wg24 had a tank pH of 3.6 and Wg25 had a tank pH of 6.2). Archaeal diversity was similar in both tanks, represented by 7 orders, with the exception of Methanocellales transcripts uniquely recovered from Wg25. Based on measures of FPKG (fragments mapped per kilobase of gene length), genes involved in methanogenesis, in addition to a spirochaete flagellin gene, were among those most highly expressed in Wg25. Conversely, aldehyde dehydrogenase and monosaccharide-binding protein were among genes most highly expressed in Wg24. The ability to observe specific presence of insect, plant, and fungi-associated RNA-viruses was unexpected. As with other techniques, there are inherent biases in the use of RNA-Seq, however, these data suggest the possibility of understanding the entire community, including ecological interactions, via simultaneous analysis of microbial, eukaryotic, and viral transcripts.

1. Introduction

In the Neotropics, plants within the family Bromeliaceae are important epiphytes due to their relative abundance and to their compact foliar-based catchments (or 'tanks') capable of retaining water. Because epiphytic bromeliads virtually lack absorptive roots, these tanks provide the plants with water and nutrients released by heterotrophic decomposition of impounded debris [2,46]. Tank bromeliads can contain as much as 50,000 l of suspended water ha^{-1} of canopy [8], a habitat that has potential significance with regard to local biodiversity and community structure. It has been shown that increased environmental complexity results in increased species diversity for both bacteria and animals [21,43]. This may be particularly true for bromeliads, which not only contribute to the heterogeneity of the forest, but themselves are complex structures with compartments of different sizes and unique physical–chemical characteristics [1,24]. Faunal abundance and diversity are enhanced in bromeliad tanks, which house up to 100× greater animal densities than surrounding forest floor habitats [34,39,40]. A vast array of aquatic taxa take advantage of these catchments, including numerous insect families, ostracods, copepods, rotifers, and amphibians, as well as protozoans, algae, fungi, bacteria, and archaea [3,9,11,12,14,28,38].

Due to their small size, defined boundaries, and relatively simple faunal communities, bromeliad catchments are ideal for the investigation of whole ecosystem complexity, a notoriously difficult task for lakes and other large, more complex freshwater habitats [24]. As a consequence, bromeliad habitats have been used for the study of carbon and nitrogen cycling, food web dynamics, and even the effects of warming on trophic cascades [20,27,30,32,33]. Similarly, bromeliad tanks have been investigated for their unique bacterial and archaeal communities, as well as the influence of various 'micro-limnological' parameters, such as pH, oxygen, dissolved organic carbon, and light, on the microbial communities [11,12,15,30]. pH conditions, for example, can have a major impact on the structure of freshwater microbes [17], and in a study of Costa Rican bromeliads it was discovered that Alphaproteobacteria, Acidobacteria, Planctomycetes, and Bacteroidetes dominated tanks of low pH (<5.1),

* Corresponding author at: Biology Dept., Occidental College, 1600 Campus Rd, Los Angeles, CA 90041, United States. Tel.: +1 323 259 1470.
E-mail address: sgoffredi@oxy.edu (S.K. Goffredi).

http://dx.doi.org/10.1016/j.csbj.2014.12.001

while Betaproteobacteria, Firmicutes, and Bacteroidetes dominated tanks of high pH (>5.3), based on 16S rRNA gene surveys [11]. The influence of pH was not observed for archaea [12] and is generally less resolved for animal communities [4,7]. Petrin and colleagues determined that the ability of stream invertebrates to adapt to naturally low pH conditions was variable among different taxonomic groups and that total species richness and abundance did not change with pH [35].

With a new appreciation for the intertwined worlds of metazoans and microbes, it is an exciting prospect to explore the possibility of examining the entire community (from viruses to eukaryotes) with a single technique. The current state of knowledge with regard to plant-associated aquatic habitats is based on the counting of animals or protozoa or, in the case of bacteria and archaea, based on 16S rRNA ribosomal surveys (ex. [11,12,14,28,38,42]). There have been no papers published on viral communities within these abundant and important ecosystems. Each of these studies, like most, was conducted with a specific organismal domain in mind. Despite decades of investigation, total community dynamics has not been measured for any given bromeliad tank community. Additionally, DNA-based analysis (ex. 16S rRNA gene surveys), although a technique that has yielded much new information about the living world, does little to inform our understanding of functional microbial status and activity in situ. Metatranscriptomic analysis (the study of the RNA from the entire community of organisms), however, allows for detection of mRNA transcripts from all domains, rather than specific taxa. Advances in high-throughput sequencing and bulk extraction of mRNA allow for RNA-based studies of organismal functioning in complex environments. Despite certain technological challenges, microbial metatranscriptomics has already helped elucidate microbial responses to oil spills and the resulting deep-sea hydrocarbon plumes, differences between anoxic and oxic paddy soils, and metabolic activity of methane-producing microbes, to name a few [31,41,47]. Similarly, metatranscriptomic approaches have been used to investigate the functional diversity of the eukaryotic microorganisms within the rumen of muskoxen [37] and in forest soils [6,44]. Notably, Urich et al. obtained information on the structure and function of organisms from all three domains of life, in a single RNA-centered metatranscriptomic experiment with forest soil [44]. While no single 'omics' approach is able to elucidate the truly complex nature of entire communities, the current results suggest the possibility of using similar RNA-based analyses to profile total community dynamics, including eukaryotes, archaea, bacteria, and possibly viruses. We provide here preliminary data on the diversity of total communities associated with two bromeliad tanks from La Selva Biological Station in a lowland tropical rainforest in Costa Rica.

2. Methods

2.1. Sample collection

Bromeliad tank debris and water were collected from two *Werauhia gladioliflora* specimens in La Selva Biological station, a wet (4 m annual rainfall) lowland neotropical reserve located in northeastern Costa Rica (10°25′52″ N, 84°00′12″ W). Samples were collected in June 2012, from two plants, neither of which was in bloom, growing at heights of ~2.0 m from the ground. The bromeliads had tanks of ~10 cm depth (i.e. similar vertical oxygen gradients), and were neighboring with presumed similar light regime and acquisition of leaf litter through-fall. Since previous studies revealed strong correlations between bacterial diversity and bromeliad tank pH, which can range from pH 3 to 7 [11], two tanks at the pH extremes were chosen for comparison of gene expression from all 3 domains of life, in addition to viruses. At the time of sample collection, 'Wg24' had a tank pH of 3.56, whereas 'Wg25' had a pH of 6.21. The entire tank contents were collected from the deepest horizons within the central axil of each plant, and put into clean, RNase-free 15 ml conical tubes for transport to the lab. Tank samples were centrifuged at 15,000 rpm for 10 min, water was removed, and the remaining debris was resuspended in 3× LifeGuard Soil Preservation Solution (MoBio Laboratories), and stored at −80 °C for further analysis. Samples were at room temperature in the LifeGuard solution for ~15 h during air transport back to Occidental College.

2.2. Total nucleic acid extraction

Nucleic acid extraction was performed at Occidental College, using RNase-free materials (tubes and pipette tips) on workstations treated with RNase*Zap* Wipes (Ambion). The RNA PowerSoil Total RNA Isolation Kit (MoBio Laboratories) was used for bulk extraction of RNA, with slight protocol modifications to extract sufficient quantities of nucleic acid for metatranscriptomic analysis (~90–120 ng μl^{-1}). Modifications included the use of freshly prepared low-pH phenol:chloroform:isoamyl alcohol (25:24:1 ratio, respectively) for nucleic acid extraction (PowerSoil Step 5), a room temperature incubation for secondary debris precipitation (PowerSoil Step 9), gentle encouragement of fluid flow through the RNA capture column using a syringe (PowerSoil Steps 16–19), and an overnight nucleic acid precipitation step (PowerSoil Step 20). Successful nucleic acid extraction was verified by NanoDrop UV spectrophotometry, ensuring sufficient RNA concentrations and purity. RNA extraction was immediately followed by removal of contaminant DNA from the extracted RNA, using TURBO DNA-free DNase Treatment and Removal Reagents (Ambion). A treatment of 4 Units TURBO DNase per 100 μl RNA was applied, with a 20 min 37 °C incubation for sufficient DNA contaminant removal.

2.3. RNA preparation and sequencing

Total extracted RNA was submitted to Otogenetics Corporation (Norcross, GA USA) for RNA-Seq analysis. Briefly, 9–12 μg of total RNA was subjected to rRNA depletion using the RiboZero Meta-Bacteria kit (Epicentre Biotechnologies) and cDNA was generated from the depleted RNA using the NEBNext mRNA Sample Prep kit (New England Biolabs). Illumina libraries were created from the cDNA using NEBNext reagents (New England Biolabs). The quality, quantity and size distribution of the Illumina library products were determined using an Agilent Bioanalyzer 2100. The libraries were then submitted for Illumina HiSeq2000 sequencing according to standard operations.

2.4. Bioinformatic analysis

Paired-end 100 nucleotide (nt) reads were generated and checked for data quality using FASTQC (Babraham Institute). For removal of ribosomal RNA sequences (5S, 16S, 23S and the eukaryote rRNA equivalents) that were not subtracted via RiboZero depletion, SortMeRNA software [25] was used to filter by comparison against the Silva rRNA archaea, bacteria and eukaryotic rRNA databases [36]. For Wg24, 13,562,864 remained after rRNA removal (78%), out of 17,258,552 initial reads. For Wg25, 10,761,162 remained after rRNA removal (67%), out of 15,998,012 initial reads. These ~10–14 million 100 nt reads were assembled into larger contigs (500–5000 nt in length) using CLC Genomics Workbench v.6.0.1 (2533 and 1938 contigs were generated for Wg24 and Wg25, respectively). Sequencing reads were quality filtered and Illumina adapters were trimmed using the CLC Genomics Workbench with a quality threshold of 0.01. Trimmed reads were assembled de novo using the CLC Genomics Workbench using an automatic kmer size of 22 and a bubble size of 50. To assign function to the assembled metatranscriptomic contigs, Prodigal v.2.60 was used to call open reading frames (ORFs; [22]), which were annotated functionally and taxonomically by blastp, using an E-value of $<10^{-5}$ against the Integrated Microbial Genomes (IMG) system v.4.0 [29] which includes archaeal, bacterial and eukaryotic full and draft genomes, as well as viral and environmental sourced genes. Of the 2533 contigs within library Wg24, 434 had functional gene annotations without a taxonomic affiliation, while an additional 166 had no homologs in public databases. Similarly, of the 1938 contigs within library Wg25, 128 had functional gene annotations