Algal Research 17 (2016) 341–347

Contents lists available at ScienceDirect

Algal Research

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

Pond Crash Forensics: Presumptive identification of pond crash agents by next generation sequencing in replicate raceway mass cultures of *Nannochloropsis salina*

Laura T. Carney [a,1], Joshua S. Wilkenfeld [b,1], Pam D. Lane [a], Owen D. Solberg [a], Zachary B. Fuqua [b], Nina G. Cornelius [b], Shaunette Gillespie [b], Kelly P. Williams [a], Tzachi M. Samocha [b], Todd W. Lane [a,*]

[a] Systems Biology, Sandia National Laboratories, PO Box 969, MS 9671, Livermore, CA 94551, USA
[b] Texas A&M AgriLife Research Mariculture Laboratory at Flour Bluff, 4301 Waldron Road, Corpus Christi, TX 78418, USA

ARTICLE INFO

Article history:
Received 22 January 2016
Received in revised form 3 May 2016
Accepted 11 May 2016
Available online xxxx

Keywords:
Algae mass culture
Biocontamination
Biomass algal crash
Second generation sequencing
Microbiome analysis

ABSTRACT

Productivity of algal mass culture can be severely reduced by contaminating organisms. It is, therefore, important to identify contaminants, determine their effect on productivity and, ultimately, develop countermeasures against such contamination. In the present study we utilized microbiome analysis by second-generation sequencing of small subunit rRNA genes to characterize the predator and pathogen burden of open raceway cultures of *Nannochloropsis salina*. Samples were analyzed from replicate raceways before and after crashes. In one culture cycle, we identified two algivorous species, the rotifer *Brachionus* and gastrotrich *Chaetonotus*, the presence of which may have contributed to the loss of algal biomass. In the second culture cycle, the raceways were treated with hypochlorite in an unsuccessful attempt to interdict the crash. Our analyses were shown to be an effective strategy for the identification of the biological contaminants and the characterization of intervention strategies.

1. Introduction

One of the major challenges to achieving high rates of long-term production in microalgal mass cultures, is the ability to maintain culture stability. Lost days of production due to pond crashes can significantly lower annualized production rates [26,31]. Methods exist for the tracking of many of the abiotic parameters, such as temperature, pH, and nutrient depletion, which can contribute to pond instability, however understanding of the contribution of the biological agents has been difficult to achieve. Aside from a few well-known causative agents, the diversity of algal pathogens, parasites, and predators has not been well characterized. What is understood, however, is that the economic impact of biological agents on the mass culture of algae can be substantial. Algal pathogens, primarily bacteria [6,11] and fungi [9,16], and zooplankton grazers (e.g., [29]) can cause severe algal biomass loss. Such decline can occur rapidly; for example, metazoan rotifers can ingest 200 algal cells min^{-1} $rotifer^{-1}$ [15] while doubling their density in 24 h [37].

In addition to being useful in community characterization (e.g., [3,23]), second generation sequencing (SGS) has brought about a revolution in detection and identification of new etiological agents in diverse systems. SGS of molecular barcode regions, such as ribosomal RNA genes and mitochondrial cytochrome oxidase, can identify agents to various taxonomic levels, providing measures of richness and diversity for bacterial (e.g., [2]) and eukaryotic (e.g., [7]) communities. Sequencing of specific single variable regions of the small subunit ribosomal RNA (SSU rRNA) gene can, generally, provide genus level identification [35]. Species level identification can be obtained with sequence data from multiple variable regions or additional molecular barcodes such as the internal transcribed spacer (ITS) region of the ribosomal RNA precursor transcript [1]. In terms of algal mass culture operations, it may not be necessary to identify an agent beyond the genus level.

There are several sources of bias, including differential nucleic acid extraction and amplification efficiencies that prevent sequencing datasets from being a quantitative measure of the relative abundances of species present in pond samples. Thus sequence data by itself is not sufficient to unambiguously identify the etiological agent of a crash, but rather, provides a presumptive or conjectural, identification of potential etiological agents. Confirmatory identification requires isolation of the agent, recapitulation of the crash, and re-isolation of the microorganism from the second infection thus fulfilling Koch's postulates [26]. When isolation of the etiological agent is desired, sequence data can

* Corresponding author at: Systems Biology, Sandia National Laboratories, PO Box 969, MS 9292, Livermore, CA 94551, USA.
E-mail address: twlane@sandia.gov (T.W. Lane).
[1] Heliae Development, LLC, 578 E. Germann Road, Gilbert, AZ 85297.

http://dx.doi.org/10.1016/j.algal.2016.05.011
2211-9264/

be used to generate genetic probes to track the agent and to facilitate the purification process. The increases in sequencing throughput has enabled multiplexing of libraries and driven down the per-sample costs of analysis and made SGS a practical solution for the identification of unknown deleterious species in algae culture samples.

Although some expertise in the nucleic acid manipulation and analysis is required, SGS does not require expertise in taxonomic identification via microscopy and is more effective at identifying morphological variants or low abundance species within a sample. SGS also has advantages over oligonucleotide-mediated methods such as PCR, in that it is not dependent on *a priori* knowledge of the organisms present. Microbiome analysis can also inform the development of specific probes, primers or biomarkers for both beneficial and deleterious species, which will allow rapid near real time tracking of organisms of interest. However, given the current cost and time required, SGS is unlikely to supplant either of these alternative methods for routine monitoring.

The marine alga, *Nannochloropsis salina*, shows promise as a candidate species for potential conversion of algae biomass to biofuel [5,33]. Here we report the use of SGS for the presumptive identification of etiological agents likely responsible for the *N. salina* culture instabilities in outdoor open raceways at the Texas A&M AgriLife Research Mariculture Laboratory, Corpus Christi, Texas.

2. Materials and methods

2.1. Raceway culture conditions and monitoring

Algae culture experiments were conducted in outdoor algae tanks at the Texas A&M AgriLife Research Mariculture Laboratory at Flour Bluff, Corpus Christi, Texas. Six fiberglass raceways were used for this study (3 m × 1 m; surface area of 2.78 m^2; 550 L with 20 cm culture depth). Raceways were equipped with a paddle wheel operating at 9 rpm, resulting in a flow rate of 50–60 cm sec^{-1}. CO_2 was supplied from compressed cylinders at a flow rate of 0.125 L min^{-1} for 8 h d^{-1}, during daylight hours.

Three replicate algae raceways (ARWs) for each cycle were inoculated on July 16, 2011 (cycle 7B) and on September 17, 2011 (cycle 8B) to a final *N. salina* cell density of 3.7×10^6 and 12.3×10^6 cells mL^{-1}, respectively. Nutrient concentrations are shown in Table 1 and physical parameters during the two cycles are given in Table 2. Contamination was monitored while making daily cell counts by hemocytometer and light microscopy observations. For cycle 7B, raceway culture samples (250–1000 mL) were collected on the day of inoculation and again after conditions in two of the raceways had deteriorated significantly. For cycle 8B, samples were taken from the stock culture used as the inoculum because it was observed to be contaminated, and from three replicate raceways after contamination was observed, and then on day 17 after replicates were treated on days 10, 12 and 14, post inoculation with sodium hypochlorite (see Fig. 1b) in an attempt to control the observed contamination.

Samples were shipped overnight on ice to Sandia National Laboratories for SGS analysis. Biomass was harvested from each sample by centrifugation (10,000 × *g* for 10 min) and filtration of the supernatant (0.2 μm). Biomass, harvested by these two sequential methods, was combined and stored at −80 °C.

2.2. Library preparation and second generation sequencing

Genomic DNA was purified from 200 mg of pelleted biomass from each sample (ZR Fungal/Bacterial DNA MiniPrep™, Zymo Research, Irvine, CA USA). For each sample, three replicate sequencing libraries were prepared for the hyper-variable V4 region of the eukaryotic SSU rRNA gene and the hyper-variable V6 region of the bacterial SSU rRNA using conditions and primers described in Carney et al. [4]. Multiplexed V4 and V6 libraries containing equimolar concentrations of PCR product from each sample were sequenced on Illumina GAII, HiSeq or MiSeq platforms, according to manufacturer's instructions (Illumina, San Diego CA),

2.3. Second generation sequencing data analysis

SGS data processing and Silva based analysis and classification was carried out according to the methods described in Carney et al. [4]. In an attempt to characterize the read pairs that remained unaligned after our Silva-based classification phase, we adopted a "join, cluster, and BLAST" approach. First, in order to join overlapping reads, read pairs were processed with FLASH [24] using default settings. Next, reads were processed with CD-HIT [22] and summarized at the sequence cluster level. Finally, we used BLAST to query these clustered sequences against the full Silva 111 and Genebank nr databases. Significant alignments were manually inspected and reported at whatever taxonomic level was possible given the available annotations associated with the BLAST hits.

Correlation analysis was performed to determine the reproducibility of microbiome analysis. For each sample/timepoint, the observed raw taxonomic count prior to cluster analysis was used to calculate the Pearson correlation coefficient between technical replicates. The calculations were performed at all taxonomic levels.

3. Results and discussion

We carried out microbiome analysis by SGS on samples taken from two sets of open raceway pond growth cycles referred to as cycle 7B and 8B. In the analysis of the eukaryotic microbiome, of >160 million paired reads that passed initial quality filters and were analyzed by the pipeline, 92% were mapped to genus-level in our edited version of the Silva 111 rRNA database (Table 3). In the analysis of the bacterial microbiome of >140 million paired reads that passed initial quality filters and were analyzed by the pipeline, 73% hit to the Silva 111 rRNA database, but only 17% on average were mapped to genus-level (Table 3). Replicate datasets 2 and 3 generated for each variable region were highly correlated ($r > 99.9\%$), thus only correlation analyses between rep 1 and 2 are reported. Replicates for eukaryotic variable region V4 were highly correlated at the genus level and above (mean correlation genus [±1SD]: $r = 99.1 \pm 1.7\%$). However, the V6 datasets were only moderately correlated at the phylum and genus level (mean correlation for phylum [±1SD]: $r = 78.8 \pm 24.6\%$; for genus: $r = 68.0 \pm 34.6\%$). The weaker correlation for V6 may be attributed to the fact that, for some of the samples, a majority of reads were not aligned to a sequence in the Silva database. In other words, the greater correlation for V4 versus V6 is explainable by the fact that Silva release 111 provides a much better survey of eukaryotic diversity than it does prokaryotic diversity.

Table 1
Nutrient formulations used in relevant algae raceways (ARWs), cycles 7B and 8B.

Nutrient	Chemical formula	Units	Nutrient L^{-1} of culture media	
			7B (ARW #s 6, 8 & 11)	8B (ARW #s 3, 7 & 12)
Ammonium sulfate	$(NH_4)_2SO_4$	g	0.2640	0.1320
Phosphoric acid (85%)	H_3PO_4	mL	0.0078	0.0240
Ferrous sulfate	$FeSO_4 \cdot 7H_2O$	g	0.0200	0.0100