Algal Research 4 (2014) 41–51

Contents lists available at ScienceDirect

Algal Research

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

Molecular diagnostics for monitoring contaminants in algal cultivation

Scott P. Fulbright [a,b], M. Kristen Dean [a], Greg Wardle [c], Peter J. Lammers [c,d], Stephen Chisholm [a,b,*]

[a] Bioagricultural Sciences and Pest Management, Colorado State University, Fort Collins, CO 80523-1177, United States
[b] Cell and Molecular Biology Graduate Program, Colorado State University, Fort Collins, CO 80523-1005, United States
[c] Solix Biosystems, Fort Collins, CO, 80524, United States
[d] Energy Research Laboratory, New Mexico State University, Las Cruces, NM 88003-8001, United States

ARTICLE INFO

Article history:
Received 31 May 2013
Received in revised form 24 October 2013
Accepted 6 November 2013
Available online 3 January 2014

Keywords:
Algae
Contamination
Diagnostics

ABSTRACT

There is currently great interest in mass cultivation of microalgae for production of fuels and other high value products. Since algae have not previously been grown at the scales and with the precision required for these endeavors, sensitive methods are needed for enumeration of elite algal varieties relative to "weedy" invader strains that are ubiquitous in the environment and a common issue with culture management. The ideal monitoring strategy would be inexpensive and identify weedy algae long before they become prominent in cultures of elite varieties. Herein, multiple polymerase chain reaction (PCR)-based tools for monitoring contaminants are presented. These include resources to identify unknown strains, to routinely monitor dominant constituents in cultures, and to detect contaminants constituting as little as one in 10^8 cells in a culture. Quantitative PCR was shown to be 10^4 times more sensitive for detecting weeds than flow cytometry. During characterization of these tools, it was demonstrated that contamination is a common phenomenon and that early detection is necessary for informed decision making during culture selection for subculturing or scale-up. Thus, implementation of strategies for monitoring contaminants in algal cultivation is a critical component of culture management for optimal productivity.

1. Introduction

Microalgae (herein, "algae") comprise a highly diverse set of photosynthetic eukaryotes that arose via independent endosymbiotic events [1,2]. Because strains from divergent taxa produce oils appropriate for use in production of renewable biofuel, general interest in algae has increased significantly [3]. Oil productivity in some algal varieties is significantly greater than even the most robust oil-producing traditional crops [4], and genetic modification is now common in multiple relevant algal strains and thus may be used to further enhance high-oil-productivity strains [5–8]. Following agricultural convention, these desired algal varieties with high oil productivity and other inherent or engineered qualities that make them suitable crops for commercial production may be generically referred to as "elite" lines. Algae have not historically been cultivated at the scales nor with the technical precision required for affordable, reliable mass cultivation and quality-controlled fuel production. Major barriers that currently limit the potential of algal biofuels include proven, stable, large-scale (>1000 ha) cultivation methods for appropriate high-oil-content algal strains and an understanding of culture maintenance and pest management strategies [9].

Because algae are ubiquitous in the environment, there are constant opportunities for low oil content algae to contaminate cultures and compete with elite strains for sunlight and nutrients. Such contaminants are appropriately referred to as "weeds" and must be managed as such to minimize their impact on crop productivity and resulting fuel quality. Because lipids are more reduced than carbohydrates and proteins, high-oil elite algae require more photosynthetically derived reductant per unit biomass than weedy strains containing less oil. Thus, weedy algae may grow faster than elite strains and have the potential to become abundant or dominant in a culture [10]. Both open pond and closed photobioreactor systems are known to be invaded by weedy species, grazers and pathogens [11–13], so such invasions must be expected regardless of the cultivation system. Clearly, algal culture monitoring methods will be needed along with pest management programs for algae-based biofuel production, and culture monitoring is equally important for production facilities, research laboratories and culture collections [14]. To be included as part of a routine culture monitoring regime, these tools and related protocols should be of low or moderate cost, versatile for adaptation to various algal communities, able to be implemented immediately, require only limited technical expertise, and be informative.

Current culture monitoring methods vary in throughput, instrumentation, degree of experience required and cost. Growers may use microscopy to manually observe cultures and identify algae based on morphology and pigmentation. This methodology is low throughput

* Corresponding author at: Bioagricultural Sciences and Pest Management, Colorado State University, Fort Collins, Colorado 80523-1177, United States.
E-mail address: chisholm@colostate.edu (S. Chisholm).

2211-9264/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.algal.2013.11.008

and requires considerable expertise distinguishing strains. Microalgae are small (1–100 μm in diameter), and distinct genera may have nearly identical overall appearances [15,16]. Furthermore, algal strains of the same species may be morphologically indistinguishable, yet harbor cryptic genetic diversity that affects crop value [17]. In addition to standard microscopy, it is common to use flow cytometry and imaging flow cytometry to group cells based on phenotypes such as size and chlorophyll content [18]. Despite its increased throughput, flow cytometry has limited ability to identify algae with certainty or to distinguish strains with similar phenotypes.

Nucleic-acid-based methods may be used to unambiguously identify algae, for example by sequencing or otherwise characterizing a portion of algal genomes. Genes encoding RNA subunits of prokaryotic or eukaryotic ribosomes are commonly characterized for taxonomic and phylogeny purposes. Relevant to the work presented herein, there are evolutionarily constrained regions of rRNA genes ideal for design of PCR primers of broad specificity or for comparison of distantly related organisms, as well as interspersed variable regions that may be used to distinguish more closely related organisms [19–21]. Additionally, there are millions of rRNA sequences deposited in general nucleotide databases (i.e., Genbank, http://www.ncbi.nlm.nih.gov/genbank/) and specialized rRNA databases (i.e., SILVA, http://www.arb-silva.de).

In this work, molecular tools were developed for routine monitoring of elite and weedy algae in laboratory and production cultures. The various tools and procedures involved characterization of 18S rRNA genes. In the analyses presented, the polymorphism among algal 18S rRNA genes was sufficient to distinguish different genera, species of the same genus, and geographic isolates seemingly of a single species. Specifically, PCR primers were designed to amplify an approximately 1500 nt region of 18S rRNA genes from three classes of algae: *Bacillariophyceae*, *Eustigmatophyceae*, and *Chlorophyceae* (herein referred to as "BEC"). These amplicons can be sequenced for definitive identification of strains, or they can be digested with a restriction enzyme to generate allele-specific fragmentation patterns for rapid, inexpensive characterization of strains and cultures (Fig. 1, left panel). Two strategies for culture monitoring based on quantitative PCR (QPCR) were also compared for their ability to detect weedy algae at low abundance in elite cultures (i.e., allele-specific QPCR probes and allele-specific QPCR primers; Fig. 1, middle and right panels, respectively). We chose the more promising allele-specific QPCR primer method and compared its sensitivity and specificity to that of flow cytometry for detecting weedy algae at low abundance in cultures. In addition to clarifying the utility and limitations of these tools, we demonstrate the importance of sensitive and accurate weed detection during selection of potential innocula for scale-up or subculturing.

2. Materials and methods

2.1. Sampling

Samples were collected (Solix Biosystems; [11]). Approximately 1.5 mL of culture was sampled from cultures ranging in biomass density between 0.5 and 5 g(dry weight)/L, equivalent to 9×10^7 and 1×10^9 cells/mL, respectively. Other samples came from agar plates where single colonies or numerous colonies were picked using a pipette tip and placed into F/2 media. The samples were centrifuged at 6000 ×*g* for 10 min at room temperature and the supernatant was decanted. Cell pellets were less than 100 mg and were stored at −20 °C until DNA extraction.

2.2. Flow cytometry

Samples were analyzed using a guava easyCyte HT + flow cytometer (EMD Millipore) equipped with an argon laser (488 nm) and 680/30 nm bandpass filter. For each sample, 20,000 events (i.e., cells) were scored for red fluorescence to identify chlorophyll-positive cells and for low-angle forward scatter to determine approximate diameter. Algal cells were identified as chlorophyll-positive events, and populations of algal genera were distinguished by size.

2.3. DNA extraction

Total DNA was isolated from frozen cell pellets. Cells were disrupted by grinding in liquid nitrogen with a mortar and pestle for 5 min or by mechanical disruption using a bead beater (BioSpec Products) or paint shaker (Fluid Management). Frozen cell pellets in microcentrifuge tubes were shaken 3 × 1 min in the presence of 0.5 mm zirconia/silica beads (BioSpec Products Inc.). Prior to and between each round of shaking, the biomass was flash frozen in liquid nitrogen. Following

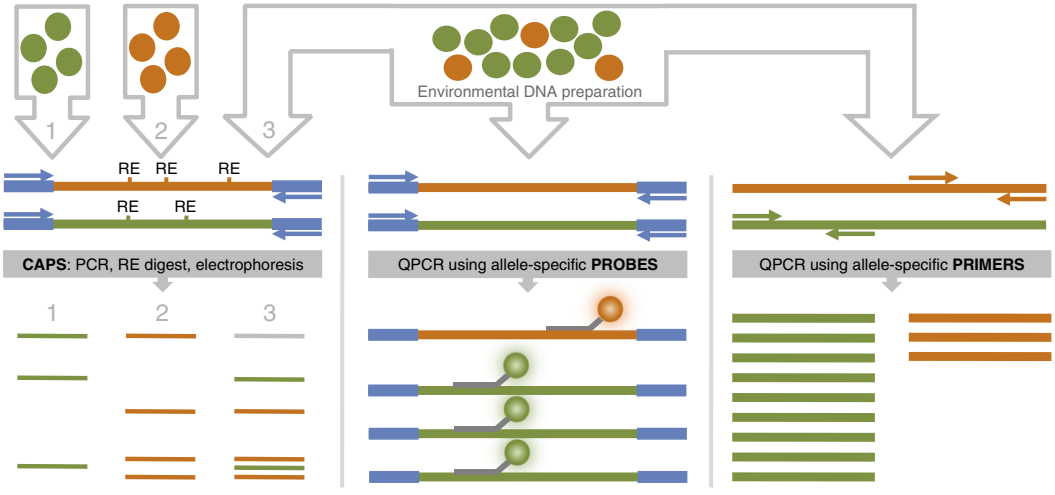

Fig. 1. Nucleic acid-based diagnostics for monitoring algal cultures. Schematic overview contrasting three strategies for monitoring algal cultures. Left panel: Using cleaved amplified polymorphic sequences (CAPS), a portion of the 18S rRNA gene is amplified from different algae (represented as green or orange cells) using a single set of primers with broad specificity (blue arrows). Amplicons are digested with an appropriate restriction enzyme (RE, restriction enzyme cut sites) and restriction fragments are resolved by electrophoresis. Allele-specific fragmentation patterns may be used to identify algae in unialgal cultures (e.g., inputs 1 & 2) or mixed cultures (input 3). Some restriction fragments may be shared by multiple organisms and are not useful for diagnostic purposes (e.g., gray fragment in restriction pattern 3). Middle panel: fluorescent probes in QPCR reactions detect allele-specific polymorphisms within 18S rRNA amplicons produced using primers with broad specificity (blue arrows). Relative fluorescence intensity from multiplexed probes with distinct fluorophores may be used to estimate relative abundances of organisms in cultures. Right panel: allele-specific QPCR primers amplify 18S rRNA gene regions from specific organisms in a culture and estimate their relative abundances.

cell disruption, genomic DNA extraction was done using the Easy-DNA kit (Invitrogen) or DNeasy Plant Mini kit (Qiagen), according to manufacturer's instructions. Isolated DNA concentration was determined using a spectrophotometer (ND-1000 Thermo Scientific).

2.4. 18S rRNA sequence alignments and BEC 18S primer design

The 18S rRNA gene sequences of representative members of algal classes *Bacillariophyceae*, *Eustigmatophyceae* and *Chlorophyceae* were retrieved from GenBank. A total of 117 unique sequences (42 *Bacillariophyceae*, 19 *Eustigmatophyceae*, 56 *Chlorophyceae*) larger than 1000 nt were aligned using ClustalW (Supplemental File 1). Primers (BEC 18S Forward & Reverse; Table 1) were designed to anneal to highly conserved regions (Supplemental Fig. 1) and to generate amplicons of approximately 1500 bp.

2.5. BEC 18S PCR

The PCR of 18S rRNA genes was done using 50 μL reactions containing a final concentration of 10 ng template DNA, 0.5 μM each BEC 18S Forward and Reverse primers (Integrated DNA Technologies; Table 1), 1 U High Fidelity Phusion DNA Polymerase (New England Biolabs), 1× HF buffer and 0.2 mM dNTPs (Fisher Scientific), Thermal cycling consisted of initial denaturation at 98 °C for 2 min; 40 cycles of denaturation at 98 °C for 30 s, annealing at 55 °C for 30 s and extension at 72 °C for 1 min; and a final extension at 72 °C for 10 min. Amplicons were resolved using agarose gel electrophoresis and visualized following ethidium bromide staining.

2.6. Cloning

As needed, PCR products were either cloned directly or following gel purification using a Geneclean (Qbiogene) or QIAquick PCR purification kit (Qiagen). Amplicons were ligated into the pSC-B vector and transformed into *Escherichia coli* cells using StrataClone Blunt PCR cloning kit (Agilent) according to manufacturer's instructions. Transformations were plated on selective LB agar and incubated overnight at 37 °C. Colonies of putative transformants were isolated, used to inoculate 5 mL of selective LB media and grown overnight at 37 °C with agitation. Plasmids were extracted using a Gene JET Plasmid Miniprep Kit (Fermentes) following manufacturer's instructions. Plasmids were eluted in 200 μL water.

2.7. Sequencing and analysis

Purified PCR products or plasmids were sequenced using ABI BigDye Terminator v3.1 chemistry and an ABI 3130xL Genetic Analyzer at the Colorado State University Proteomics and Metabolomics Facility. Primers for sequencing included BEC 18S primers (Table 1) and standard M13 primers. To determine algal strain identity, DNA sequences were queried against GenBank using BLASTn.

2.8. Cleaved Amplified Polymorphic Sequences (CAPS)

BEC 18S amplicon sequences were aligned to identify polymorphisms within *Hae*III restriction enzyme cut sites. PCR products were generated with BEC 18S primers. Restriction digests were done using 20 μL reaction volumes containing 10 μL PCR product, 1 U *Hae*III (New England Biolabs), 1 × BSA, 1 × NEB Buffer 4. Reactions were incubated overnight at 37 °C followed by inactivation at 80 °C for 20 min. Digest products were resolved by gel electrophoresis using 1%–2.5% agarose or 4% metaphor agarose (Lonza) and visualized following ethidium bromide staining. For high resolution, 1 μL of restriction digest was prepared with the DNA 1000 kit (Agilent) and assayed using a Bioanalyzer 2100 (Agilent) according to manufacturer's instructions.

2.9. Real-time QPCR and probe threshold cycle analysis

Real-time QPCR assays used a CFX96 Real-Time System (BioRad). Threshold cycles (C_t) were identified using single threshold determination and baseline-subtracted analysis. For all QCPR amplicons, amplification efficiencies (Table 1) were calculated using C_t values from a series of reactions in which templates were serial dilutions of linearized plasmid DNA containing the target sequence [19]. To establish specificity, primers were used in QPCR reactions in which template was linearized plasmid DNA containing a non-target 18S rRNA gene sequence.

Table 1
PCR amplicons, primers and probes.

	Amplicon length (nt)	Amplification efficiency[1]	Primer/Probe	SNPs[2]	Sequence[3]
BEC 18S amplicon					
BEC 18S	~1500[4]	n/a	BEC 18S Forward	n/a	CCGTAGTAATTCTAGAGCTAATAC
			BEC 18S Reverse	n/a	CGGTGTGTACAAAGGGCAGGGACGTAATC
Allele-specific QPCR **probes**					
Nannochloropsis v. *Tetraselmis*	*Nannochloropsis*: 181	*N. salina*: 94.7%	NT Forward	n/a	GGATGTTTTCATTAATCAAGAAC
			NT Reverse	n/a	GTTTCAGCCTTGCGACCATA
	Tetraselmis: 183	*T. striata*: 93.3%	*Nannochloropsis* probe	7	FAM-cgggcgTGCCGACTAGGGATCGGTGGGTGCATcgcccg-BHQ
			Tetraselmis probe	16	HEX-ccggccCAGACGTTTTTTTGATGACTCTGCCAGCAggccgg-BHQ
N. salina v. *N. oculata*	135	*N. salina*: 93%	NsNo Forward	n/a	GCTCGTAGTTGGATTTCTGG
			NsNo Reverse	n/a	CACAGTAAAAGATAGGGATCC
		N. oculata:103%	*salina* probe	9	FAM-ccggggTCGGTTCCGATAAGGGGCCGTACTGTTGccccgg-BHQ
			oculata probe	4	HEX-cccggcGCGGCTCTTACATTAAGTTGTCGGCGTgccggg-IBQ
Allele-specific QPCR **primers**					
Nannochloropsis diagnostic	104	*N. salina*: 92.6%	Nanno Forward	14	CGTCGGGATCCCTATCTTTT
			Nanno Reverse	10	AGACCACCAAGGTCGTATCTTA
Tetraselmis diagnostic	140	*T. striata*: 92%	Tetra Forward	13	TAGCTCCTGGGCTTCACTGT
			Tetra Reverse	9	CCAACAAGATAAGCCAGAGTCC

[1] For primers designed to amplify multiple alleles, amplification efficiency for each allele is given.
[2] Relative to non-target allele.
[3] Uppercase: target sequence; lowercase, self-complementary stem sequences of probes; FAM, 6-fluorescein amidite; HEX, hexachloro fluorescein; BHQ, black hole quencher; IBQ, Iowa black quencher.
[4] Precise amplicon sizes vary by organism.

2.10. Allele-specific fluorescent QPCR probes

A total of 81 gene sequences of the 18S rRNA gene from the genera *Tetraselmis* and *Nannochloropsis* were retrieved from GenBank, aligned using ClustalW or ClustalOmega (Supplemental File 2), and viewed with JalView. Primers (NT Forward & Reverse; Table 1) were designed in conserved regions. Probes (*Nannochloropsis* & *Tetraselmis* probes; Table 1 and Supplemental Fig. 2) were designed as "molecular beacons" [20] that are stem-loop structures with 5′ fluorophores and 3′ quenchers. Primers were used in QPCR with a reaction volume of 10 μL, containing a final concentration of 10 ng template DNA, 300 nM each NT Forward and Reverse primer, 150 nM probe and 1 × SsoFast Probes Supermix (BioRad). Multiplexed reactions contained two probes with unique fluorophores each at 150 μM. Thermal cycling consisted of initial denaturation at 95 °C for 2 min; 40 cycles of denaturation at 95 °C for 5 s, annealing and extension at 60 °C for 4 s; followed by melt curve analysis from 65 °C to 95 °C in 0.5 °C increments. Fluorescence was measured using the HEX and FAM channels during annealing/extension steps and the melt curve.

The same methodology was used to design primers (NsNo Forward & Reverse; Table 1) and probes (*salina* & *oculata* probes; Table 1) for discrimination of QPCR amplicons produced from 18S rRNA genes of *Nannochloropsis salina* (GenBank accession AF045048.1) and *Nannochloropsis oculata* (GenBank accession U38902.1) (Supplemental Fig. 3).

2.11. Allele-specific QPCR primers

The alignment of 81 *Tetraselmis* and *Nannochloropsis* 18S rRNA gene records (see above; Supplemental File 2) was used to identify regions highly conserved within each individual genus but polymorphic between the genera (Supplemental Fig. 4). Primers (Table 1) were designed to specifically amplify *Nannochloropsis* (Nanno Forward & Reverse) or *Tetraselmis* (Tetra Forward & Reverse) strains (Supplemental Fig. 4). Primers were used in 10 μL QPCR reactions containing a final concentration of 10 ng template DNA, 300 nM each forward and reverse primer, and 1 × SsoAdvanced SYBR Green Supermix (BioRad). Thermal cycling consisted of initial denaturation at 98 °C for 2 min; 40 cycles of denaturation at 98 °C for 30 s, annealing and extension at 67 °C for 30 s; followed by melt curve analysis from 65 °C to 95 °C in 0.5 °C increments. Fluorescence was measured using the SYBR channel during annealing/extension steps and the melt curve.

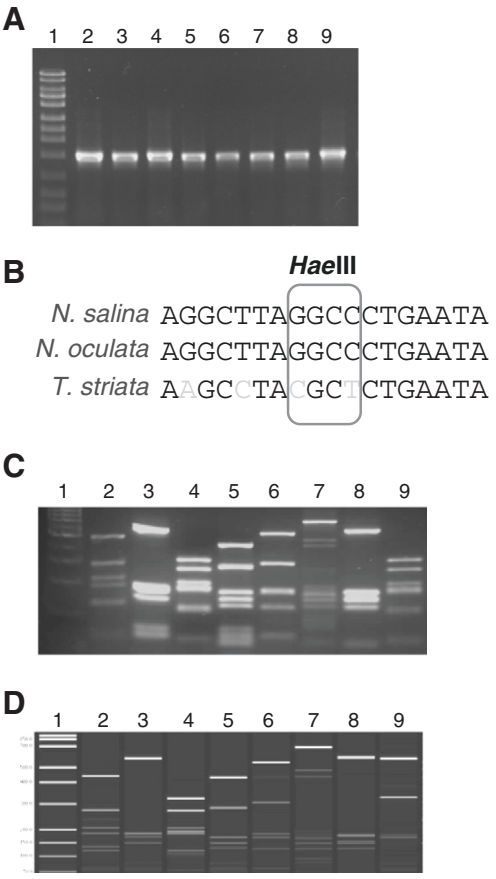

Fig. 2. BEC amplicon production and CAPS analysis. A) BEC amplicons were produced from gDNA templates and resolved by electrophoresis. Lanes: 1. size standard, 2. & 3. *N. salina*, 4. *N. oculata*, 5. & 6. *T. striata* (Ute isolate), 7. *Prymnesium parvum*, 8. *Phaeodactylum tricornutum*, 9. *Chlorella vulgaris*. B) A representative 18 nt region within the BEC amplicon containing a *Hae*III restriction site polymorphism. In the example shown, *Hae*III should cut the *N. salina* and *N. oculata* BEC amplicons at this position, but not the *T. striata* amplicon. C) *Hae*III-digested BEC 18S amplicons were resolved using a 2.5% agarose gel. Lanes: 1. size standard, 2. *N. salina*, 3. *T. striata* (Ute isolate), 4. *N. oculata*, 5. *Chlorella vulgaris*, 6. *Prymnesium parvum*, 7. *Pavlova lutheri*, 8. *Phaeodactylum tricornutum*, 9. *N. oculata*. D) *Hae*III-degested BCEP 18S amplicons were resolved using an Experion automated electrophoresis system. Lanes: 1. size standard, 2. *N. salina*, 3. *T. striata* (Ute isolate), 4. *N. oculata*, 5. *Chlorella vulgaris*, 6. *Prymnesium parvum*, 7. *Pavlova lutheri*, 8. *Phaeodactylum tricornutum*, 9. *T. striata* (Poudre isolate).

3. Results and discussion

3.1. BEC primers for identification of algae

To facilitate identification of algae in laboratory and production cultures, we designed primers to amplify the 18S rRNA gene of three major algal classes. The 18S rRNA gene sequences of algae from the BEC classes were retrieved from GenBank. A total of 117 unique 18S rRNA gene records containing sequences larger than 1000 nt were aligned using ClustalW or ClustalOmega (Supplemental File 1) and regions highly conserved among all sequences were identified (Supplemental Fig. 1). Within such conserved regions, primers were designed to generate amplicons of approximately 1500 bp (Table 1, BEC 18S Forward & Reverse), including multiple variable 18S rRNA gene regions. This primer set produced specific amplicons ("BEC 18S amplicons") from representatives of BEC classes (Fig. 2A). These amplicons may be cloned using standard procedures and subsequently sequenced to definitively identify algae by querying 18S rRNA databases using BLAST. Nucleotide sequencing of cloned amplicons provided information sufficient to unambiguously identify the corresponding algal species, even for related species such as *N. salina* and *N.oculata*. Compared to using microscopy or flow cytometry for algal identification, there are numerous advantages to these amplicons. Researchers without expertise in algal morphology may identify algae they have not previously encountered and they may have a high degree of confidence in the identification. Furthermore, sequence comparisons may discriminate algae such as *N. salina* and *N. oculata* that are morphologically indistinguishable even to experienced researchers. This method of strain identification is only limited by the length of accurate sequence recovered and the population of properly annotated 18S rRNA gene sequences in GenBank and other databases.

3.2. CAPS analysis discriminates algal strains

We anticipate growers will commonly work with a small number of elite algae and – for any particular location and production environment – will encounter a finite set of weeds. Using sequencing to routinely characterize algal populations of limited diversity would be inefficient. Therefore, we developed a cleaved amplified polymorphic sequences (CAPS) assay [21] for discrimination of strains based on nucleotide polymorphisms in restriction enzyme recognition sequences within BEC 18S amplicons (Fig. 1, left panel).

Based on alignments of BEC 18S amplicon sequences from representative algal strains, there are numerous nucleotide polymorphisms among these amplicons. Some of these polymorphisms produce or eliminate restriction endonuclease recognition sequences and are specific to particular genera, species or strains (e.g., Fig. 2B). As a result, strain-specific restriction fragmentation patterns are produced after BEC 18S amplicons are digested with an appropriate restriction enzyme. These unique fragmentation patterns may be readily distinguished using gel electrophoresis and used to putatively identify organisms. For example, the BEC 18S amplicons of *N. salina*, *N. oculata* and *Tetraselmis striata* (a prevalent weed in saline cultures) contain 8, 8 and 6 predicted *Hae*III sites, respectively. The predicted restriction fragmentation patterns should be unique to each organism and therefore useful for identification purposes. Specifically, *N. salina*, *N. oculata* and *T. striata* digest products are predicted to include unique fragments of 449, 333 and 600 nt, respectively.

To confirm these and other algae can be distinguished using restriction fragmentation patterns, genomic DNA was extracted from presumed unialgal cultures, BEC 18S amplicons were produced by PCR, amplicons were digested with *Hae*III, and the resulting restriction fragments were resolved using electrophoresis. Indeed, fragmentation patterns for representative BEC algal strains could be differentiated following this procedure (Fig. 2C). The assay even discriminated the related species *N. salina* and *N. oculata* that have only 31 nucleotide polymorphisms between their full-length (1790 nt) 18S rRNA genes and are indistinguishable by microscopy or flow cytometry. It is evident the CAPS procedure is an effective tool for rapid and inexpensive routine characterization of cultures. To compare separation and visualization technologies, restriction digest products were resolved using 2.5% agarose (Fig. 2C) and an Experion automated electrophoresis system (Bio-Rad) (Fig. 2D). Electrophoretic separation of restriction fragments using a 2.5% agarose gel commonly resolves fragments between 200–1000 nucleotides and therefore provides resolution sufficient to discriminate the selected algae. For the Experion capillary system, Bio-Rad 1K LabChips, which separate DNA fragments between 15–1500 nucleotides, were used. As seen in Fig. 2C & 2D, both technologies sufficiently resolved *Hae*III-digested BEC amplicons to allow discrimination of algae based on restriction fragmentation patterns.

In Fig. 2D, lanes 3 and 9 contain previously unreported isolates of *T. striata* ("Ute" and "Poudre" isolates) recovered from southwestern and northern central Colorado, respectively. Following observation of their distinct restriction fragmentation patterns, the BEC amplicons from each isolate were sequenced. The isolates have unique 18S sequences and are therefore distinct. Queries of GenBank revealed the 18S sequences from both isolates are most similar to *T. striata* strain SAG 41.85 (GenBank record JN904000.1). In Fig. 2C, it was shown that the CAPS analysis was able to distinguish species of the same genus, *N. salina* and *N. oculata*. As shown in Fig. 2D, even these two geographical isolates of the same weedy algal species have distinct fragmentation patterns resulting from our standard CAPS analysis, with the Poudre isolate having a distinguishing restriction fragment of approximately 320 nt. In addition to demonstrating the versatility of the CAPS procedure, this result indicates that different geographic locations will have genetically distinct populations of weeds, even though those weeds may belong to the same genus or species and further demonstrates CAPS analysis may distinguish algae that appear identical when observed by microscopy.

Another sample examined by CAPS analysis was a stock culture of *Dunaliella salina*. The observed fragmentation pattern (Fig. 2C, lane 9) for this culture did not match the fragmentation pattern predicted based on sequence of the *D. salina* BEC amplicon. The observed fragmentation pattern lacked a 994 nt fragment predicted based on the *D. salina* 18S rDNA sequence and was seemingly identical to that of *N. oculata* (Fig. 2C, lane 4), with a distinguishing band of approximately 333 nt. Sequencing of cloned BEC 18S amplicon from this sample confirmed the algae to be *N. oculata*, indicating the supposed *D. salina* culture had at some point been mislabeled or contaminated with – and eventually dominated by – *N. oculata*. This demonstrates the practical use of this simple and rapid CAPS procedure for monitoring dominant algae in cultures.

3.3. CAPS analysis distinguishes abundant species in polyalgal cultures

We next determined whether the CAPS procedure could recognize two distinct algal strains in a culture if those algae are present at similar levels. To mimic a mixed culture but allow more precise control of DNA ratios, genomic DNA was extracted from *T. striata* and *N. salina* cultures, combined at different ratios, and used as templates in BEC CAPS analyses. As shown in Fig. 3A, restriction fragments indicative of both algal strains were visible in each of the three reactions using mixed templates. We did not determine the limits for detection of a less-abundant algal strain in a culture dominated by another algal strain. Even if BEC amplicons were produced in a quantitative fashion (amplicons of this size are not), the dynamic range of technologies used to observe fragmentation patterns following *Hae*III digestion and electrophoretic separation limit the potential to visualize the fragmentation pattern from a weed at low abundance in an elite culture. In the course of this work, we used CAPS analyses to identify numerous cultures as containing multiple algal strains, some of which are represented in Fig. 3B. Each of these cultures was intended to contain *N. salina*, and the known *N. salina* fragmentation pattern was present in all lanes. However in lanes 2, 3, 4, 9 and 10, additional

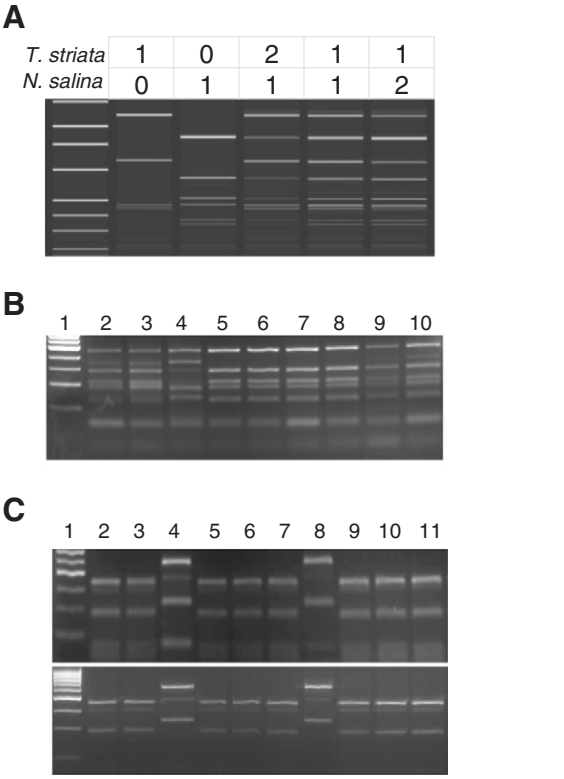

Fig. 3. CAPS analyses of polyalgal cultures. A) Genomic DNA from *T. striata* and *N. salina* were mixed at the indicated ratios and used as template for BEC 18S CAPS analysis. *Hae*III restriction fragments were resolved using Experion capillary electrophoresis. Leftmost lane: size standard. B) Nine *N. salina* cultures were analyzed by the *Hae*III CAPS procedure. Resulting restriction fragments were resolved using a 2.5% agarose gel. Lane 1, size standard; lanes 2–10, individual cultures. C) BEC amplicons from a contaminated culture were cloned and individual clones were analyzed using the BEC *Hae*III CAPS analysis. Resulting fragmentation patterns from 10 clones are shown following separation using a 2.5% agarose gel (top panel) or 4% Metaphor agarose (bottom panel). Lane 1, size standard; lanes 2–11, individual clones.

restriction fragments were visible, indicated unwanted algal strains contaminated these cultures.

It is easy to envision situations in which culture components may not be confidently identified based on fragmentation patterns. Nevertheless, if a novel algal strain begins to dominate a culture, its fragmentation pattern will likely be distinct from those of the targeted elite algae strain or contaminants encountered previously. Additionally, the presence of multiple organisms in a culture may result in a fragmentation pattern too complex to deconvolute with confidence. These two scenarios are readily distinguished. In the case of a single organism of unknown identity, the sum of the individual restriction fragments should total approximately 1500 nt (the size of BEC amplicons). If multiple organisms are present, the sum of the individual restriction fragments should clearly exceed 1500 nt. In either case, a simple solution is to clone an aliquot of the same BEC amplicons that were used as input for CAPS analysis. In the case of a single novel organism, sequencing a clone should be sufficient to identify the algal strain. In the case of a complex mixture of algae, multiple individual clones may be used as templates for CAPS analysis. Each clone will produce a fragmentation pattern corresponding to a single algal strain from the mixed culture.

Such an approach was used to confirm identities of algae present in a polyalgal culture. Initial CAPS analysis of the culture indicated at least two algal strains were present, but their respective identities could not be confidently determined based on the fragmentation pattern (data not shown). Subsequently, an aliquot of the relevant BEC 18S amplicons was cloned into a standard PCR-cloning plasmid, plasmid DNA was extracted individually from ten clones and used as template in PCR with BEC 18S primers. Products were digested with *Hae*III and restriction fragments were resolved using 2.5% agarose and 4% Metaphor agarose (Fig. 3C). Based on the fragmentation patterns, two clones were *T. striata* and the remaining eight were *Chlorella vulgaris* (Fig. 3C). Individual clones were subsequently sequenced to confirm the CAPS-based identification (data not shown). It is possible the ratio of *T. striata*:*C. vulgaris* clones approximated relative abundance of these algae in the initial culture. However, since this CAPS procedure used 1500 nt amplicons from standard PCR as inputs for the *Hae*III digest, the results should not be interpreted in a quantitative manner.

3.4. Allele-specific QPCR probes for monitoring low-abundance organisms

Given the potential impact of contamination by weedy algae, it would be optimal to detect weeds when they are at low levels relative to elite strains so that remediation strategies may be pursued to salvage the culture. The CAPS procedure is suitable for discriminating algal species and identifying dominant culture constituents, but is not practical for detection of a weed at low abundance in a culture dominated by an elite algal strain. We compared the capabilities of two QPCR-based procedures for early detection of weedy algae (Fig. 1, middle and right panels). Both QPCR assays amplify portions of the 18S rRNA gene and must be customized to detect algae of interest (i.e., elite strains and common weeds). One assay uses a single set of primers to amplify the same 18S region from different algae, and amplicons are then distinguished using multiplexed allele-specific fluorescent probes. The second assay uses allele-specific QPCR primers to produce amplicons from the 18S rRNA genes of different algae.

The fluorescent probe assay (Fig. 1, middle panel) uses “molecular beacons”, which are stem-loop structures in which the loop corresponds to the allele-specific sequence [20]. One end of the self-complimentary stem is fused to a fluorophore, the other to a quencher. When the probe is not bound to a target amplicon, it assumes the stem-loop conformation, bringing the quencher into proximity with the fluorophore, thereby suppressing probe fluorescence. During each cycle of QPCR, probes anneal to target amplicons and fluoresce; this fluorescence level is used to quantify amplicon production per cycle. Probes with distinct fluorophores may be multiplexed to detect multiple alleles in a single QPCR reaction.

In theory, these probes may be designed with enough specificity to distinguish amplicons differing by as little as a single nucleotide. To establish the ability of fluorescent QPCR probes to distinguish sequences with limited polymorphism, a probe set was designed to distinguish *N. salina* and *N. oculata*. We aligned 18S rRNA sequences from *N. salina* and *N. oculata* and identified conserved regions for QPCR primers (Supplemental Fig. 3). We designed primers (Table 1, NsNo Forward & Reverse) to produce a 135 nt amplicon spanning 15 polymorphic positions and developed allele-specific probes to distinguish amplicons derived from *N. salina* and *N. oculata* (Table 1, *salina* and *oculata* probes). To test whether the QPCR primers amplify *N. salina* and *N. oculata* templates with similar efficiencies and produce a single amplicon, these primers were analyzed in QPCR reactions in which the template was linearized plasmid containing cloned BEC 18S amplicons derived from *N. salina* or *N. oculata*. These QPCR primers amplified *N. salina* and *N. oculata* templates with efficiencies of 93.0% and 103%, respectively (Table 1; Supplemental Fig. 5).

To determine the specificity of probes for their intended targets, *N. salina* and *N. oculata* sequences were individually used as templates for QPCR in which the *salina* and *oculata* probes were both present. When *N. salina* DNA was used as template in a QPCR reaction containing *salina* and *oculata* probes, only the *salina* probe effectively detected amplicons (Fig. 4A). Similarly, when *N. oculata* DNA was used as template with both probes, only the *oculata* probe effectively detected amplicons (Fig. 4A). Additionally, the *salina* and *oculata* probes each produced specific signal above background levels with as little as 0.001 pg template DNA. Similar results were observed when *N. salina* and *N. oculata* genomic DNA were used as template in the QPCR reactions (data not shown). In summary, the *salina* and *oculata* QPCR probes were efficient, specific and sensitive when used in QPCR reactions with template derived from a single organism.

Nannochloropsis and *Tetraselmis* were then used to demonstrate the ability of QPCR probes to discriminate elite and weedy algae. We aligned an approximately 650 nt region from 81 GenBank records of 18S rRNA genes from strains of the genera *Nannochloropsis* and *Tetraselmis* (Supplemental Fig. 2 and Supplemental File 2). We anticipated the diversity within and between these genera would present challenges in designing primers to produce a single amplicon with similar efficiencies from different strains. Nonetheless, we identified regions highly conserved among *Nannochloropsis* and *Tetraselmis* strains and designed QPCR primers (Table 1, NT Forward & Reverse) to anneal within these regions (Supplemental Fig. 2). Though the primer binding sites are conserved among *Nannochloropsis* and *Tetraselmis* 18S genes, there are 25 genus-discriminating positions within these amplicons and allele-specific probes were designed based on these polymorphisms (Supplemental Fig. 2 and Table 1, *Nannochloropsis* and *Tetraselmis* probes). As described above for the *salina* and *oculata* probes, the primers and probes designed to monitor *Nannochloropsis* and *Tetraselmis* were shown to efficiently produce amplicons from both templates (Table 1; Supplemental Fig. 6) and to be highly specific for their intended targets (Fig. 4B), respectively.

3.5. Loss of allele-specific probe signal with complex templates

Since the intended use of these probes was to detect unwanted algae at low abundance within a culture dominated by an elite strain, we determined the sensitivity of QPCR probes when the template was serial dilutions of DNA from one algae made in a background of DNA from another algae. To test the *salina* and *oculata* probes, template DNAs were linearized plasmids containing cloned *N. salina* or *N. oculata* 18S BEC amplicons. The concentration of *N. oculata* template was held constant

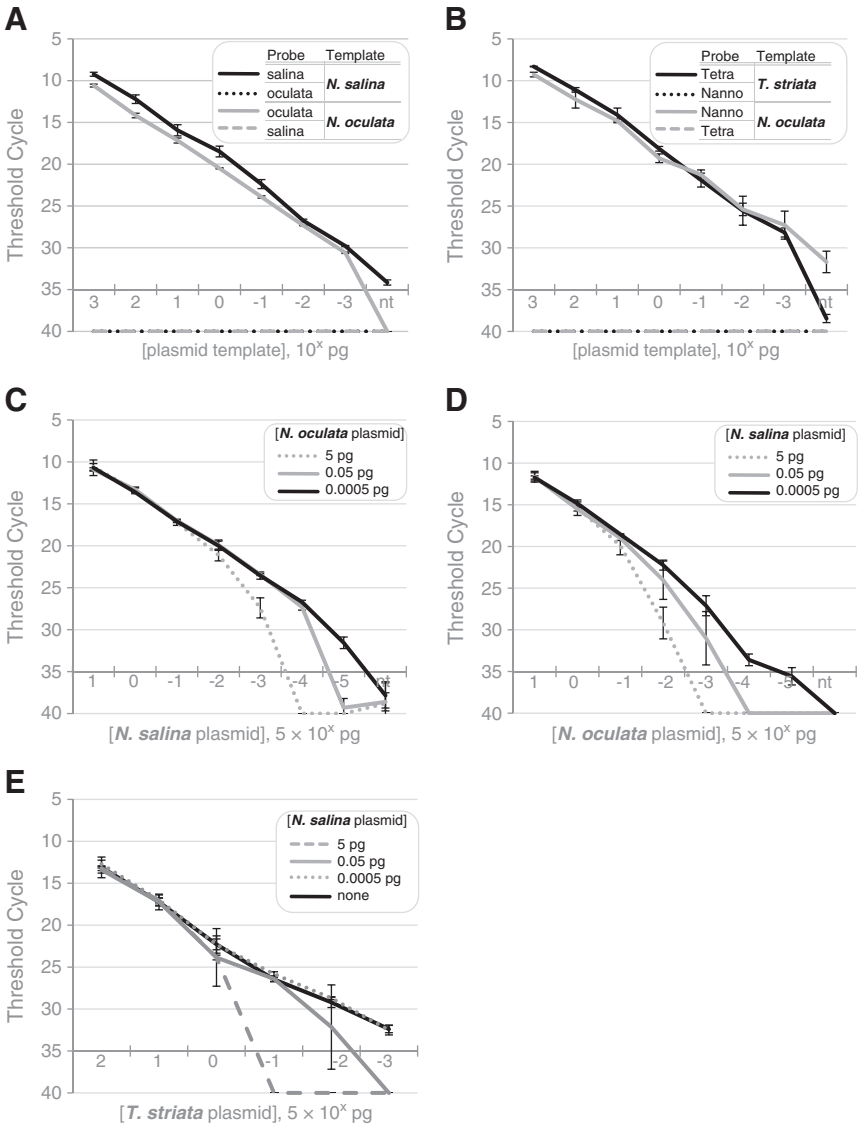

Fig. 4. Fluorescent probes are specific and sensitive, but do not accurately detect minority organisms in polyalgal samples. For all panels, serial dilutions of plasmid DNA templates containing cloned 18S BEC amplicons from the specific target were made alone or in the presence of the non-target templates indicated in inset graphical legends for individual panels. Forty or 45 cycles of QPCR were done for all experiments; if no fluorescence above background was detected, a C_t value of 40 was assigned. A) Fluorescent probes to discriminate *N. salina* (salina) and *N. oculata* (oculata) efficiently detect as little as 10^{-3} pg target template, but do not produce fluorescence above background levels in reactions containing only non-target templates. B) Fluorescent probes to discriminate *Nannochloropsis* (Nanno) and *Tetraselmis* (Tetra) efficiently detect as little as 10^{-3} pg target template, but do not produce fluorescence above background levels in reactions containing only non-target templates. C) Fluorescent probes inefficiently detect specific targets diluted in an excess of non-target template. Plasmid templates containing cloned *N. salina* 18S BEC amplicon was serially diluted in 5 pg, 0.05 pg or 0.0005 pg plasmid containing cloned *N. oculata* BEC 18S amplicon. When the ratio of *N. salina*:*N. oculata* was 1:1000 or greater, the salina probe did not efficiently detect *N. salina*, as indicated by higher C_t values than when the *N. salina* is more concentrated than or approximately equal in concentration to *N. oculata*. D) Fluorescent probes inefficiently detect specific targets diluted in an excess of non-target template. Plasmid templates containing cloned *N. oculata* 18S BEC amplicon was serially diluted in 5 pg, 0.05 pg or 0.0005 pg plasmid containing cloned *N. salina* BEC 18S amplicon. When the ratio of *N. oculata*:*N. salina* was 1:100 or greater, the oculata probe did not efficiently detect *N. oculata*, as indicated by higher C_t values than when the *N. oculata* is more concentrated than or approximately equal in concentration to *N. salina*. E) Fluorescent probes inefficiently detect specific targets diluted in an excess of non-target template. Plasmid templates containing cloned *N. oculata* 18S BEC amplicon was serially diluted in 100 ng, 1 ng or 10 pg *N. salina* genomic DNA. When the ratio of *N. oculata* plasmid:*N. salina* gDNA was 1:20,000 or greater, the oculata probe did not efficiently detect *N. oculata*. F) Fluorescent probes inefficiently detect specific targets diluted in an excess of non-target template. Plasmid templates containing cloned *T. striata* 18S BEC amplicon was serially diluted in 5 pg, 0.05 pg, 0.0005 pg or no plasmid containing cloned *N. salina* BEC 18S amplicon. When the ratio of *T. striata*:*N. salina* was 1:10 or greater, the Tetraselmis probe did not efficiently detect *T. striata*, as indicated by higher C_t values than observed in the absence of *N. salina* or when the *T. striata* is more concentrated than or approximately equal in concentration to *N. salina*.

in the reactions at 5 pg, 0.05 pg or 0.0005 pg, while the *N. salina* template ranged from 50 pg to 0.00005 pg per reaction. The *salina* probe performed as expected when the *N. oculata* template was less abundant or nearly equal in concentration to the *N. salina* template. However, when the ratio of *N. oculata*:*N. salina* template was 1000:1 or greater, there was strong interference with detection of the *N. salina* allele (Fig. 4C). A similar and more extreme interference was observed in reciprocal experiments in which serial dilutions of *N. oculata* DNA were made in a background of *N. salina* DNA (Fig. 4D). In this case, there was strong interference with and high variability of the *oculata* probe signal when the ratio of *N. salina*:*N. oculata* is 10:1.

This interference phenomenon occurs when using the corresponding primers and probes in QPCR reactions with mixed *Tetraselmis* and *Nannochloropsis* templates. As shown in Fig. 4E, there was interference with detection of the weed *Tetraselmis* when the ratio of *Nannochloropsis*:*Tetraselmis* template was 10:1 or greater. Varying QPCR primer or $MgCl_2$ concentrations did not alleviate this interference (data not shown). We did not determine the molecular basis for the observed interference with probe signal when multiple 18S

rDNA alleles were present, though a similar phenomenon was previously reported in experiments using QPCR to detect multiple organisms with concentration differences greater than three orders of magnitude [22]. Fluorescent QPCR probes are extremely sensitive and specific to their targets; however, due to the signal interference phenomenon they are not appropriate for detecting a small amount of weedy algae in a culture dominated by an elite strain.

3.6. Allele-specific primers detect minority algae in polyalgal cultures

As an alternative to multiplexed probes, we designed allele-specific primers to detect either *Tetraselmis* or *Nannochloropsis* in polyalgal cultures (Fig. 1, right panel). Using the same alignment of 81 *Tetraselmis* and *Nannochloropsis* 18S rRNA gene records from above (Section 3.4 and Supplemental File 2), regions polymorphic between genera but conserved within each genus were identified (Supplemental Fig. 4) and used to design allele-specific QPCR primers (Table 1, *Nannochloropsis* and *Tetraselmis* diagnostics). In this assay, accumulation of QPCR products is monitored by incorporation of a fluorescent dsDNA-binding dye that is not specific for any particular amplicon. To determine whether the genus-discriminating primers would amplify the non-target genus, *Tetraselmis* primers were included in QPCR reactions with 10 ng purified plasmid DNA containing a cloned fragment of the *N. oculata* 18S rRNA gene. In reactions containing 10 ng *N. oculata* plasmid but lacking *Tetraselmis*, there was not significant accumulation of product within 40 QPCR cycles (Fig. 5A), indicating that the *Tetraselmis* primers do not efficiently amplify *Nannochloropsis* 18S rRNA gene templates. Similarly, in the absence of *N. oculata* template, the *Nannochloropsis* primers did not effectively amplify *T. striata* plasmid template within 35 QPCR cycles (Fig. 5B). Therefore, both sets of genus-discriminating primers efficiently amplified their intended targets, but not the non-target template. Throughout the course of this work, negative controls using these primers sets produced results similar to those described here, such that C_t values greater than 35 were considered non-specific amplification and C_t values less than 35 were regarded as specific amplification.

The sensitivity of primer-based QPCR assays was determined using reactions in which the template was serial dilutions of DNA from one algae made in a background of DNA from another algae. Dilutions of *T. striata* plasmid (ranging from 50 pg to 0.000005 pg) were made in a background of 10 ng *N. oculata* gDNA, and these mixed DNAs were used as templates in QPCR reactions with the *Tetraselmis* primers. Even when the ratio of *N. oculata*:*T. striata* template was 2×10^8:1, C_t values were nearly identical to those observed when no *Nannochloropsis* template was included (Fig. 5A). In similar experiments, *N. oculata* plasmid was serially diluted in 10 ng *T. striata* template, and the mixed templates were included in QPCR reactions with the *Nannochloropsis* primers. The *Nannochloropsis* primers effectively detected *N. oculata* even when the ratio of *T. striata*:*N. oculata* was 2×10^7:1 (Fig. 5B).

The allele-specific primer assay is very effective at detecting a weed at an early point, however the assay is not without limitations. Firstly, it is challenging to design allele-specific primers that eliminate the potential of amplifying targets other than the organism of interest. For this reason, we refer to these assays as "allele-specific" rather than "organism-specific". For example, the allele-specific primers we use to detect *Tetraselmis* also amplify the 18S rRNA gene of *Chlorella vulgaris* (data not shown). While *C. vulgaris* is often considered an unwanted weed and it is useful to have tools to detect it a low abundance, if amplicons are produced and detected with these primers a researcher would not immediately be sure whether the contaminant was *Tetraselmis* or *Chlorella*, though sequencing the QPCR amplicons would reconcile this. Additionally, there is limited potential for multiplexing of these reactions, so a separate QPCR reaction would likely be needed to assay for each organism of interest.

3.7. QPCR outperforms flow cytometry for early detection of weeds

During scale-up, biomass from smaller cultures is used to inoculate larger cultures. Even at small scales, algal cultures are often maintained by subculturing. Thus, many cultures – regardless of scale or cell density – are old and many generations removed from starter material of confirmed identity [10]. A logical point for weed monitoring is during selection of inocula for subculturing or scale-up. By analyzing potential sources of inoculum in parallel, researchers may identify the sample with the least relative contamination and select this for use.

Presently, flow cytometry is commonly used to monitor algal cultures. To compare the ability of allele-specific QPCR primers to detect weeds in a mixed cell population with that of a flow cytometer, known ratios of *T. striata* and *N. salina* cells were analyzed using

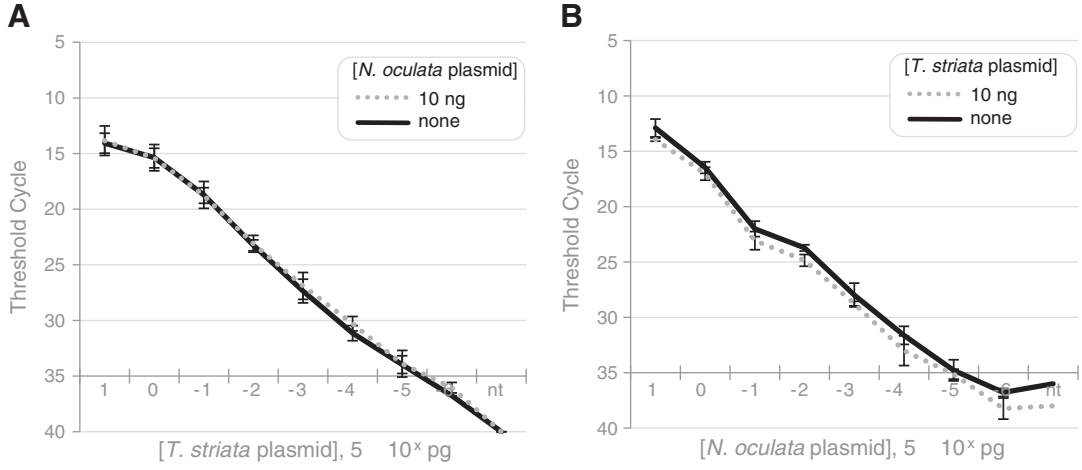

Fig. 5. Allele-specific QPCR primers are specific, sensitive and detect minority organisms in polyalgal samples. A) Allele-specific primers were used to amplify *Tetraselmis* templates alone (solid line) or diluted in a background of 10 ng *Nannochloropsis* DNA (dotted line). Data from three replicates is shown and standard deviations indicated. The primers detect *Tetraselmis* with nearly identical efficiencies in the absence or presence of an excess of *Nannochloropsis* template, indicating the primers do not amplify *Nannochloropsis* template and that the presence of a massive excess of *Nannochloropsis* template does not interfere with the primers' ability to detect *Tetraselmis*. nt, no template. B) A distinct set of allele-specific primers were used to amplify *Nannochloropsis* templates alone (solid line) or diluted in a background of 10 ng *Tetraselmis* DNA (dotted line). Data from three replicates is shown and standard deviations indicated. The primers detect *Nannochloropsis* with nearly identical efficiencies in the absence or presence of an excess of *Tetraselmis* template, indicating the primers do not amplify *Tetraselmis* template and that the presence of a massive excess of *Tetraselmis* template does not interfere with the primers' ability to detect *Nannochloropsis*. nt, no template.

both technologies. Cell densities of unialgal cultures were determined and serial dilutions of *T. striata* cells were made in a background of 1×10^8 *N. salina* cells mL^{-1}, such that the final concentration of *T. striata* ranged from 10% to 0.000001% of cells. For each aliquot of the dilution series, *T. striata* and *N. salina* cells were counted using flow cytometry.

Both expected and observed (counted in triplicate) flow cytometry results are shown in Fig. 6A (left panel, dashed gray and solid black lines, respectively). Flow cytometry accurately detected *T. striata* cells when they were 10% — 0.01% of the population. In the four dilutions for which *T. striata* comprised less than 0.01% of the population, flow cytometry overestimated abundance. Not only are the flow cytometry results inaccurate estimations of *T. striata* abundance, these four samples that varied in *T. striata* abundance by a factor of 10^4 were scored as having similar levels of weedy cells (0.001%–0.008%). If these samples represented potential sources of inocula for subculturing or scale-up, a grower using this data would be unaware of the differences in *Tetraselmis* concentration among the cultures and therefore would likely make suboptimal culture management decisions.

The inaccuracy of flow cytometry for detecting *Tetraselmis* cells when they are at low abundance in populations may be due to factors such as sampling error and technical limitations of the instrument. For example, differences in buoyant densities across algal species can result in settling artifacts. Even with conscientious mixing, this can preclude injection of a representative portion of cells of rapidly settling organisms like *Tetraselmis* into the flow cytometer. In addition to potential false negatives, flow cytometry results may include false positives, in that debris or aggregates of smaller cells may be scored as a single event within the size range of (in this case) *Tetraselmis* cells.

Following analysis by flow cytometry, the remaining cells in each serial dilution were pelleted and gDNA was extracted for use as template in QPCR with the *Tetraselmis* allele-specific primers. In Fig. 6A (right panel), the expected and observed data are shown (dashed gray and solid black lines, respectively). The *Tetraselmis* allele-specific primers effectively detected *Tetraselmis* cells at all dilutions tested. Importantly, C_t values correlated strongly with relative abundance of *T. striata* cells across the dilution series ($R^2 = 0.9982$). In QPCR reactions of 100% efficiency, the C_t values for 10-fold dilutions should differ by 3.32 [19]. Thus, the expected curve was plotted by extending a line with a slope of −3.32 from the observed data point for the most concentrated sample (10% *T. striata* cells; $C_t = 14.41 \pm 0.178$). As

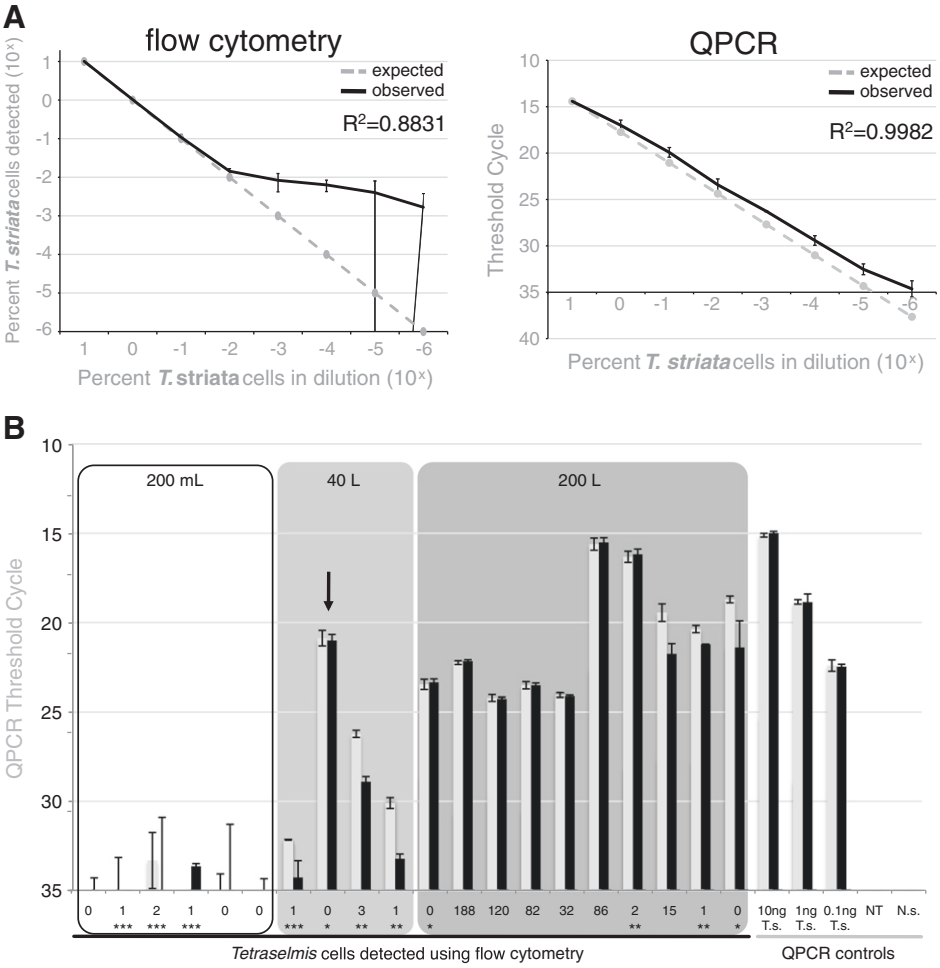

Fig. 6. QPCR has greater sensitivity, accuracy and dynamic range than flow cytometry. A) Comparison of flow cytometry and allele-specific QPCR primers for detection of *T. striata* cells serially diluted in *N. salina* cells. Left panel: Flow cytometry was used to score 20,000 events, and percent *Tetraselmis* cells detected and standard deviations across three technical replicates are shown (black line). The expected curve (dashed gray line) indicates actual percentages of *T. striata* cells in the dilution series. Right panel: Triplicate QPCR reactions were done using DNA extracted from the serial dilutions. Observed data and standard deviations for triplicate reactions are shown (black line). The expected curve (dashed gray line) was plotted by extending a line with a slope of −3.32 starting from the observed data point for the most concentrated sample (see accompanying text). For both panels, standard deviations across three replicates are shown, and the correlation of determination (R^2) between the observed data and the amount of *T. striata* cells in each dilution is given. B) Comparison of allele-specific QPCR primers and flow cytometry for detection of *T. striata* cells in 20 *N. salina* cultures between 200 mL and 200 L. Flow cytometry was used to score 20,000 events. The number of events identified as *Tetraselmis* is indicated on the x-axis for each sample. Allele-specific QPCR primers were used to detect *Tetraselmis*. Threshold cycles and standard deviations from triplicate reactions in repeated experiments (gray and black columns) are plotted on the y-axis. Black arrow, a 40 L culture flow cytometry indicated was free of *Tetraselmis* was shown using QPCR to be heavily contaminated with *Tetraselmis*. T.s., *T. striata* plasmid; NT, no template; *N.s.*, *N. salina* plasmid.

seen in Fig. 6A, the observed data approaches the expected values across all dilutions tested.

The C_t values from the QPCR assay correspond with relative abundance of (in this case) *Tetraselmis* cells in samples, though not necessarily to an absolute number of cells. By comparing C_t values from samples analyzed in parallel, researchers may establish the relative levels of weeds in cultures and make informed decisions regarding which cultures to use as inocula for scale-up or subculturing. Thus, the greater dynamic range of the QPCR assay provides a measure of certainty that flow cytometry does not.

3.8. Early weed detection enables informed culture management decision-making

To demonstrate the utility of the QPCR primer assay for culture monitoring when selecting inoculum for scale-up or subculturing, 20 *N. salina* cultures (ranging from 200 mL to 200 L) were analyzed using QPCR and flow cytometry. Flow cytometry was used to characterize each sample, and the number of *Tetraselmis*-like events identified is indicated in Fig. 6B. The remaining cells in each aliquot were used for DNA extraction and analysis by QPCR. To confirm reproducibility of the results, QPCR was done in triplicate on two occasions (Fig. 6B, gray and black bars). The QPCR assay resulted in strong *Tetraselmis* signals in three samples for which flow cytometry detected no *Tetraselmis* (Fig. 6B, single asterisks). Further, strong QPCR signals indicating *Tetraselmis* resulted from analysis of 4 additional samples for which flow cytometry identified between only 1 and 3 *Tetraselmis* cells (Fig. 6B, double asterisks). For four samples in which one or two *Tetraselmis*-like events were detected by flow cytometry, the QPCR assay produced no signal or inconsistent weak signals across triplicates from the two repetitions of the experiment (Fig. 6B, triple asterisks).

The quantity of *Tetraselmis* cells in these cultures was not determined using an independently validated method, so it is not feasible to definitively conclude which technology more accurately quantified *Tetraselmis* in the samples. Among other potential sources of error, all PCR-based procedures are susceptible to false positives resulting from contamination of DNA preparations or other reagents with target sequences, or false negatives resulting inefficient primer binding to potential targets. Nonetheless, results from experimental controls (Fig. 6B), the relative accuracy of the technologies in quantifying the *Tetraselmis* dilution series (Fig. 6A), as well as quality control experiments in preceding sections, all indicate the QPCR assay effectively detects *T. striata* DNA at 0.0000005% of a mixed template (Fig. 5A) or *Tetraselmis* cells when they constitute as little as 0.000001% of a culture (Fig. 6A), yet does not give positive signal in the presence of *Nannochloropsis* cells or DNA, or in the absence of template (e.g., Fig. 6B, QPCR controls). Furthermore, the data in Fig. 6A (left panel) demonstrate flow cytometry is inaccurate for quantification of *T. striata* at low levels.

Data from the 20 samples analyzed above illustrate that culture contamination becomes more prominent in aging cultures during the scale-up process (Fig. 6B) and that informed decisions regarding culture selection may minimize this. The six samples from 200 mL cultures had little (if any) *Tetraselmis*, as detected using QPCR diagnostics. Samples from four 40 L cultures varied greatly in the amount of *Tetraselmis*, and all samples from 200 L cultures had significant quantities of *Tetraslemis* detected by QPCR.

In the absence of additional culture remediation strategies, it would be practical to discard significantly contaminated cultures as early as possible in the scale-up process and to preferentially use non- or less-contaminated cultures as inoculum. The QPCR assay allows determination of relative amounts of *Tetraselmis* in samples analyzed in parallel. With respect to inoculum selection from the 40 L cultures in Fig. 6B, the QPCR data is more informative than flow cytometry. Flow cytometry indicated there were between zero and three *Tetraselmis* cells in each of these cultures. The QPCR assay established relative levels of contamination among these samples, clarifying which 40 L culture would be the most appropriate for use as inoculum. In fact, the culture that QPCR data suggested was the most contaminated with *Tetraselmis* was the culture flow cytometry data indicated was free of *Tetraselmis* (Fig. 6B, black arrow). If growers relied on flow cytometry data for culture characterization and sample selection, they would have likely chosen to use the most contaminated 40 L culture as inoculum for scale-up or subculturing. Therefore, the superior accuracy and sensitivity of the QPCR assay for the detection of weeds at low abundance provide critical information for culture management and selection of inoculum.

3.9. Technology overview

Given the ubiquitous nature of weedy algae, routine validation of algal cultures is an essential element of weed management and quality control. Depending on the growth characteristics of elite strains, it may take many months to scale up from a small maintenance culture to production cultures covering hundreds to thousands of acres. In laboratory settings, cultures may be maintained for years by subculturing. If weeds or other undesired algae grow undetected during this time, they may easily render cultures unusable.

The unique characteristics of 18S rRNA genes make them particularly well suited for diagnostic purposes. The presence of both highly conserved and hypervariable regions within these genes allow the production of various PCR-based monitoring tools. As demonstrated throughout this report, 18S rRNA genes contained polymorphisms sufficient to discriminate algal genera and species. The BEC primers are useful for initial characterization of algal strains, in that they amplify templates from a broad range of algae, providing enough sequence information to unambiguously identify strains. Once BEC sequences are known for strains, restriction fragmentation patterns may be predicted for development of CAPS assays. In all cases herein, *Hae*III digestion of BEC 18S amplicons distinguished strains and was useful for rapid routine monitoring of major culture constituents. The BEC 18S CAPS procedure – coupled with sequencing of BEC 18S amplicons, as needed – is a practical qualitative strategy for monitoring algae that are abundant in cultures.

In the results presented, even two geographic isolates seemingly of the same species were discriminated using CAPS analysis of BEC amplicons. Nonetheless, it is anticipated that *Hae*III digestion of BEC 18S amplicons will not discriminate all organisms. In that case, researchers may wish to digest BEC 18S amplicons using alternate restriction enzymes or to develop similar resources to examine additional loci. Loci encoding large ribosomal subunits (eg, 28S rRNA), internal transcribed spacers between adjacent rRNA genes, and the chloroplast-encoded large subunit of RuBisCo (*rbcL*) are commonly used for taxonomic studies and such sequences from numerous organisms are contained in standard online databases [23].

Relative to the qualitative CAPS analyses, quantitative PCR approaches provide vastly increased sensitivity for early detection of unwanted organisms. Though the allele-specific probes described herein did not effectively detect minority organisms, allele-specific primers accurately detected and established relative concentrations of weedy cells comprising as little as 0.000001% of a culture. In contrast, flow cytometry may not always be able to distinguish weedy and elite algae, and it gives inaccurate results when weedy cells constitute less than 0.01% of a population. Thus, incorporating an allele-specific primer assay into current weed monitoring practices will allow growers to identify weeds at low abundance and to make informed decisions regarding culture management and inoculum selection.

Recently, taxonomic surveys of microbial environments – in particular, 16S rRNA profiling of prokaryotic communities – have taken advantage of high throughput sequencing technologies [24]. This is

particularly useful for characterization of complex communities that may contain hundreds to thousands of distinct taxonomic units of bacteria with relative abundances that vary by orders of magnitude. Such technologies seem well suited for identification of all algae in a culture population, and to confidently detect weeds at extremely low abundance. At present, several factors preclude implementation of these technologies as part of a standard culture monitoring regime. There are high costs associated with instrumentation, sample preparation, and sequencing. The resulting datasets are massive and require considerable time and expertise to properly analyze and interpret. Most current technologies balance sequence read length, accuracy, and throughput. To distinguish closely related organisms, it would be ideal to have long sequences of high accuracy; to detect an organism of low abundance (e.g., a weed), it would be ideal to maximize throughput. It is likely at some point the associated direct costs will decrease, data analysis will require less expertise, and the timeframe for sequencing and analysis will be reduced, such that eukaryotic community profiling by advanced sequencing will be a practical component of routine culture monitoring.

Throughout this work it was demonstrated that culture contamination is common and likely increases throughout the culture scale-up process. Therefore, monitoring cultures for contaminants is essential for efficient cultivation of elite algal strains. We anticipate advances in high-throughput sequencing, flow cytometry and additional technologies will eventually provide efficient and cost effective alternatives to the PCR-based monitoring described in this report. However, the technologies presented may be implemented immediately at little to moderate cost and involve procedures accessible to most researchers possessing a general familiarity with PCR. We urge industrial and academic growers of algae to implement such monitoring strategies as part of a their standard quality control procedures.

Acknowledgments

The authors thank Joel Butler, Jillian Lang, Ron Parsons, Tracy Yates, Mike DeSarro, John Walden, Evan Piland and Mark Clark of Solix BioSystems for their assistance with the sampling and analyses. The described work was supported with funds from Solix Biosystems, the Colorado State University Clean Energy Supercluster, the US Department of Energy under contract DEEE0003046 awarded to the National Alliance for Advanced Biofuels and Bioproducts, and the US National Science Foundation under contract IIA-1301346 awarded to PJL. SPF is supported by a National Science Foundation Graduate Research Fellowship.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.algal.2013.11.008.

References

[1] J.M. Archibald, The evolution of algae by secondary and tertiary endosymbiosis, in: G. Piganeau (Ed.), Genomic Insights into the Biology of Algae, Vol. 64, Academic Press Ltd-Elsevier Science Ltd, London, 2012, pp. 87–118.
[2] O. De Clerck, K.A. Bogaert, F. Leliaert, Diversity and evolution of algae: primary endosymbiosis, in: G. Piganeau (Ed.), Genomic Insights into the Biology of Algae, Vol. 64, Academic Press Ltd-Elsevier Science Ltd, London, 2012, pp. 55–86.
[3] Q. Hu, M. Sommerfeld, E. Jarvis, M. Ghirardi, M. Posewitz, M. Seibert, et al., Microalgal triacyglycerols as feedstocks for biofuel production: perspectives and advances, Plant J. 54 (2008) 621–639.
[4] Y. Chisti, Biodiesel from microalgae, Biotechnol. Adv. 25 (2007) 294–306.
[5] W.J. Henley, R.W. Litaker, L. Novoveská, C.S. Duke, H.D. Quemada, R.T. Sayre, Initial risk assessment of genetically modified (GM) microalgae for commodity-scale biofuel cultivation, Algal Res. 2 (2013) 66–77.
[6] Z. Perrine, S. Negi, R.T. Sayre, Optimization of photosynthetic light energy utilization by microalgae, Algal Res. 1 (2012) 134–142.
[7] D.R. Ort, X. Zhu, A. Melis, Optimizing antenna size to maximize photosynthetic efficiency, Plant Physiol. 155 (2011) 79–85.
[8] J.N. Rosenberg, G.A. Oyler, L. Wilkinson, M.J. Betenbaugh, A green light for engineered algae: redirecting metabolism to fuel a biotechnology revolution, Curr. Opin. Biotechnol. 19 (2008) 430–436.
[9] J.B. Shurin, R.L. Abbott, M.S. Deal, G.T. Kwan, E. Litchman, R. McBride, et al., Industrial-strength ecology: trade-offs and opportunities in algal biofuel production, Ecol. Lett. (2013), http://dx.doi.org/10.1111/ele.12176.
[10] J. Bull, S. Collins, Algae for biofuel: will the evolution of weeds limit the enterprise? Evolution, 2012.
[11] J. Quinn, T. Yates, N. Douglass, K. Weyer, J. Butler, T.H. Bradley, et al., Nannochloropsis production metrics in a scalable outdoor photobioreactor for commercial applications, Bioresour. Technol. 117 (2012) 164–171.
[12] C.M.M. Gachon, M. Strittmatter, D.G. Muller, J. Kleinteich, F.C. Kupper, Detection of differential host susceptibility to the marine oomycete pathogen *Eurychasma dicksonii* by real-time PCR: not all algae are equal, Appl. Environ. Microbiol. 75 (2009) 322–328.
[13] Y.T. Li, D.X. Han, K.S. Yoon, M. Sommerfeld, Q. Hu, An example of algal culture contamination and collapse in an open pond, J. Phycol. 47 (2011) S53–S54.
[14] M.B.J. Moniz, I. Kaczmarska, Barcoding of diatoms: nuclear encoded ITS revisited, Protist 161 (2010) 7–34.
[15] R. Rippka, J. Deruelles, J.B. Waterbury, M. Herdman, R.Y. Stanier, Generic assignments, strain histories and properties of pure cultures of cyanobacteria, J. Gen. Microbiol. 111 (1979) 1–61.
[16] J.D. Hall, K. Fucikova, C. Lo, L.A. Lewis, K.G. Karol, An assessment of proposed DNA barcodes in freshwater green algae, Cryptogam. Algol. 31 (2010) 529–555.
[17] M.B.J. Moniz, F. Rindi, P.M. Novis, P.A. Broady, M.D. Guiry, Molecular phylogeny of antarctic prasiola (prasiolales, trebouxiophyceae) reveals extensive cryptic diversity, J. Phycol. 48 (2012) 940–955.
[18] M. Garmendia, M. Revilla, L. Zarauz, Testing the usefulness of a simple automatic method for particles abundance and size determination to derive cost-effective biological indicators in large monitoring networks, Hydrobiologia 704 (2013) 231–252.
[19] T. Nolan, R.E. Hands, S.A. Bustin, Quantification of mRNA using real-time RT-PCR, Nat. Protoc. 1 (2006) 1559–1582.
[20] K. Abravaya, J. Huff, R. Marshall, B. Merchant, C. Mullen, G. Schneider, et al., Molecular beacons as diagnostic tools: technology and applications, Clin. Chem. Lab. Med. 41 (2003) 468–474.
[21] A. Konieczny, F.M. Ausubel, A procedure for mapping *Arabidopsis* mutations using codominant ecotype-specific PCR-based markers, Plant J. 4 (1993) 403–410.
[22] S.M. Handy, D.A. Hutchins, S.C. Cary, K.J. Coyne, Simultaneous enumeration of multiple raphidophyte species by quantitative real-time PCR: capabilities and limitations, Limnol. Oceanogr. Methods 4 (2006) 193–204.
[23] K.P. Fawley, M.W. Fawley, Observations on the diversity and ecology of freshwater *Nannochloropsis* (Eustigmatophyceae), with descriptions of new taxa, Protist 158 (2007) 325–336.
[24] J.G. Caporaso, J. Kuczynski, J. Stombaugh, K. Bittinger, F.D. Bushman, E.K. Costello, et al., QIIME allows analysis of high-throughput community sequencing data, Nat. Methods 7 (2010) 335–336.