Challenges of metagenomics and single-cell genomics approaches for exploring cyanobacterial diversity

Michelle Davison · Eric Hall · Richard Zare · Devaki Bhaya

Received: 27 June 2014 / Accepted: 10 December 2014
© Springer Science+Business Media Dordrecht 2014

Abstract Cyanobacteria have played a crucial role in the history of early earth and continue to be instrumental in shaping our planet, yet applications of cutting edge technology have not yet been widely used to explore cyanobacterial diversity. To provide adequate background, we briefly review current sequencing technologies and their innovative uses in genomics and metagenomics. Next, we focus on current cell capture technologies and the challenges of using them with cyanobacteria. We illustrate the utility in coupling breakthroughs in DNA amplification with cell capture platforms, with an example of microfluidic isolation and subsequent targeted amplicon sequencing from individual terrestrial thermophilic cyanobacteria. Single cells of thermophilic, unicellular *Synechococcus* sp. JA-2-3-B’a(2-13) (*Syn* OS-B’) were sorted in a microfluidic device, lysed, and subjected to whole genome amplification by multiple displacement amplification. We amplified regions from specific CRISPR spacer arrays, which are known to be highly diverse, contain semi-palindromic repeats which form secondary structure, and can be difficult to amplify. Cell capture, lysis, and genome amplification on a microfluidic device have been optimized, setting a stage for further investigations of individual cyanobacterial cells isolated directly from natural populations.

Keywords Multiple displacement amplification (MDA) · Whole genome amplification (WGA) · Single cell · Microfluidics · Cyanobacteria · CRISPR

Background

Cyanobacteria represent an ancient, diverse, and ecologically important phylum. They are ubiquitous in both terrestrial and marine environments (Coelho et al. 2013; Shih et al. 2013) and are of significant interest from a variety of viewpoints. Evolutionarily, cyanobacteria played a crucial role in oxygenation of the early earth atmosphere (Hoehler et al. 2001), and were the first pioneers in primary endosymbiosis, giving rise to modern day plastids (Larkum et al. 2007). Ecologically, cyanobacteria are primary producers, performing oxygenic photosynthesis in a wide range of associations and symbiotic relationships (Paerl et al. 2000; Lesser et al. 2004; Bergman et al. 2007), and play critical roles in global nitrogen and carbon cycles (Karl et al. 2012; Bar-Zeev et al. 2013). Environmental nutrient imbalances lead to toxic blooms of cyanobacteria and freshwater eutrophication (Oliver 2012; McMahon and Read 2013). More recently, cyanobacteria have been examined for their potential in waste water remediation (Martins et al. 2011; Olguín 2012), and as an efficient chassis in biotechnology (Ducat et al. 2011; Hess 2011; Berla 2013). Although the ecology, physiology, and molecular biology of cyanobacteria have all been extensively studied for many decades (Gupta and Carr 1981; Fay 1992; Robinson et al. 1995), it is only in the last few years...
that new genome sequencing technologies and single-cell capture technologies have made a major impact in our understanding of cyanobacterial diversity. Many of these technologies have been optimized for the study of model organisms such as E. coli or yeast, so there are specific challenges in their use with cyanobacteria and microbial populations. To fully appreciate these issues, we begin with a mini-review of the major strategies that are currently used to capture genomic diversity.

In the second section, we describe in greater detail the methods available for single-cell capture and subsequent sequencing strategies. Because such methods have not been widely used with cyanobacteria, we describe the optimization of a protocol by which single cyanobacterial cells were sorted in a microfluidic (“chip”) device followed by lysis and whole genome amplification by on-chip multiple displacement amplification (MDA). We focused on amplification from specific regions of the genome including regions from specific CRISPR spacer arrays, which are highly diverse and are part of the recently identified adaptive immune response. By doing so, we provide a proof of concept study that suggests that this pipeline has the potential to be used for the study of natural diversity in cyanobacterial populations.

There are currently three primary strategies for acquiring sequence data to capture cyanobacterial diversity:

(i) classical isolation methods, by which axenic strains are first isolated from environmental samples, followed by DNA extraction and sequencing to get complete genome sequences

(ii) metagenomics, a culture-independent means by which total DNA is directly extracted from environmental samples and sequenced

(iii) single-cell methods, by which individual cells are isolated (either directly from the environment or from axenic populations) and DNA is extracted for amplification and subsequent sequencing.

Initial sequencing projects focused on acquiring the genomes of axenic microbes from available culture collections, with Haemophilus influenzae being the first bacterial genome ever to be sequenced (Fleischmann et al. 1995) and the unicellular cyanobacterium Synechocystis sp. PCC6803 being the second (Kaneko et al. 1996). However, with the advent of next generation sequencing, the choice of sequencing strategies and platforms has begun to have a significant impact on the sorts of questions that can be addressed. As there are advantages and problems associated with each method (Fig. 1), it is also recommended to use these methods in combination to mediate the shortcomings of each individual sequencing strategy. For example, the use of fully sequenced genomes as “anchors” or “references” in conjunction with population metagenomics revealed the existence of specialized cyanobacterial populations containing genes for specific metabolic functions (Bhaya et al. 2007; Kashtan et al. 2014).

Sequencing cyanobacterial isolates

The focus on culturing of axenic strains and genome sequencing is widespread (Laloui et al. 2002; Parkhill and Wren 2011). However, the ability to isolate a strain in the laboratory can often skew the distribution of sequenced individuals. Within the phylum of Cyanobacteria, almost 40 % of the currently sequenced cyanobacterial genomes cluster within the marine Synechococcus/Prochlorococcus subclade isolated from various oceanic locations and depths. This has provided a rich source of genomic-based experimental data which have emerged from a number of laboratories (Scanlan et al. 2009; Flombaum et al. 2013; Mackey et al. 2013; Thompson et al. 2013; Axmann et al. 2014; Kashtan et al. 2014). On the other hand, until recently, there was a conspicuous lack of genome sequences from the diverse morphologies that represent the deeply branched cyanobacterial lineages. To address this, a major collective effort was made to sequence the full genomes of a greater range of cyanobacterial lineages to capture phylogenetic and phenotypic diversity (Shih et al. 2013). Fifty-four strains covering all five morphological subsections, in addition to a range of lifestyles and metabolisms, were sequenced using isolates primarily from the Pasteur collection (Shih et al. 2013). Based on this study, over 21,000 novel proteins, with no homology to known proteins, were discovered, as well as the longest collection of CRISPR spacer-repeat units, 650, ever characterized in cyanobacteria. By sampling across a wide phylogenetic distribution, the groundwork has been laid for cyanobacteria to emerge as a powerful comparative genomic model system (Shih et al. 2013).

Sequencing metagenomes

In an alternate approach, metagenomic surveys have uncovered cyanobacterial signatures across many habitats, including many biofilm assemblages such as arctic mats and sub-zero sediments (Varin et al. 2012; Lay et al. 2013), glacial streams (Wilhelm et al. 2013), thermophilic hot springs (Bhaya et al. 2007; Heidelberg et al. 2009; Klatt et al. 2011 #, Ionescu et al. 2010; Mackenzie et al. 2013), non-thermophilic mats (Germán Bonilla-Rosso 2012; Kirk Harris et al. 2013; Lindemann 2013; Mobberley et al. 2013) coastal mats (Balskus et al. 2011; Burrow et al. 2013), as well as free-living populations, such as those in oligotrophic ocean waters (Shi et al. 2011; Malmstrom et al. 2013), the Red Sea (Thompson et al. 2013), soil crusts
Analyses of deep metagenomic sequencing data have yielded insights into cyanobacterial diversity, exposing a wealth of novel sequences (Bejà et al. 2012; Malmstrom et al. 2013). This has led to an understanding of both specific and general mechanisms of environmental adaptation and survival strategies to cope with physiochemical stresses (Thompson et al. 2013). It has allowed for reconstructions of metabolic interactions in complex assemblages (Burow et al. 2013). It has also yielded insights into cyanophage predator/prey interactions, and phage diversity (Sharon et al. 2009; Mizuno et al. 2013; Deng et al. 2014), and characteristics that have the potential to be harnessed for biofuel production (Kennedy et al. 2008; Rittmann et al. 2008; Zhou and Li 2010; Hess 2011). All of these examples indicate the power of metagenomic methodologies that rely on a culture-independent way of examining diversity in populations. It is however important to underscore that each sequence read is representative of DNA acquired from a single individual (or cell). As a consequence of read-length limits of NextGen sequencing platforms (up to 300 bp for Illumina, 1 Kb for 454 Ti, and approx. 20 Kb for PacBio) linkage information or the complete genomic make up of an entire cell or species cannot be acquired.

Capture and sequencing of single cells

At the other end of the spectrum, the ability to carry out genome sequencing of single cells can reveal “individuality” (Germán Bonilla-Rosso 2012; Kashtan et al. 2014). In this case, particular collections of SNPs may define an individual, however, this information is lost in a consensus genome, which represents the overall average of a population of cells. Interactions that effect the entire population often occur at the single-cell level (Balázsí et al. 2011; LeRoux et al. 2012). This enables links to be established between observed ecology and genomics and may shed light on how variants with seemingly identical genomes are able to co-exist within the same environment (Woyke et al. 2009). In as few as ten sequenced single cells, even from the highly represented Prochlorococcus clade, 394 new genes were discovered (Malmstrom et al. 2013). Despite significant metagenomic sequencing of this species, these genes remained undiscovered in previous surveys (Inskeep et al. 2010).
differing amplification efficiencies of individual genes. Sequences with secondary structure, or repeated regions notoriously difficult to amplify (Pinard et al. 2006; Linnarsson 2010), may be hard to identify in broad-scale metagenomic surveys, yet are more easily discovered in the genome of a single cell (Malmstrom et al. 2013). While single-cell genomic methods neatly sidestep the great unsolved problem of isolation under laboratory conditions, the serious shortcoming is that in discovering particular variants of interest, they are unavoidably destroyed in the capture and sequencing process (Wade 2002; Martinez-Garcia et al. 2012; Rusch et al. 2013). It is technically very challenging to get a complete or “close to complete” genome sequence from a single cell; furthermore, it requires robust methods to capture a single cell. Here we describe the single-cell capture methods that have been used for bacterial cells.

Single-cell capture methods

The earliest and simplest foray into single-cell capture was the dilution to extinction method, wherein a population of cells is serially diluted until only a single cell remains, from which an axenic culture can then be cultivated (Button et al. 1993). A relatively inexpensive method, dilution can also easily be adapted to be high throughput by use of microtitre plates and/or robotics (Connon and Giovannoni 2002). Variants most abundant in the population are the most successfully captured, rather than “weeds”, i.e., low abundance population variants, which thrive in an abundance of nutrients (Hugenholtz 2002). The slow-growing, ubiquitous SAR11 marine bacterioplankton is an excellent example of a bacterium successfully isolated by serial dilution (Connon and Giovannoni 2002; Rappe et al. 2002). However, serial dilution is very unlikely to capture rare variants, nor does it guarantee a final isolation of one individual cell (Ishii et al. 2010).

Another straightforward method for single-cell isolation is flow cytometry and fluorescence-activated cell sorting (FACs). FACs is not only high throughput, but also allows for selection of cells based on multiple metrics, such as fluorescence levels and cell morphology (Davey and Kell 1996; Konokhova et al. 2013). Flow cytometry has even been adapted to monitor internal levels of metabolites, for time and cost-efficient screening of a mutant library (Binder et al. 2012). A single-cell capture workflow, complete with robotics for automatic liquid handling, can be assembled from “out-of-the-box” instruments (Kalisky et al. 2011). However, the drawbacks of FACs include costly large reaction volumes in downstream applications, and aggregated input cells may pose an especial challenge (Kalisky et al. 2011).

Micromanipulation methods (including mechanical manipulations and “optical tweezers”) are well suited to remove cells from dense biofilm matrices, or close symbiotic interactions (Ashkin and Dziedzic 1987; Hongoh et al. 2008; Sato et al. 2009; Blainey et al. 2011). Mechanical manipulations utilize physical contact with micropipette tubes, while optical methods require no direct contact, but depend on near-infrared lasers (Ishoy et al. 2006). However, in using lasers, careful calculation is required not to cause substantial photodamage to cells (Emmert-Buck et al. 1996). Additionally, micromanipulation methods are very time consuming, and unlikely to be adaptable to high-throughput pipelines (Blainey et al. 2011).

Another means to separate individual cells is through the use of microfluidic devices. An intricate design of valves and channels allows cells and precise reagent amounts to flow into distinct chambers (Wheeler et al. 2003; Zare and Kim 2010; Blainey and Quake 2014). Of note, microfluidics is the only method in which analysis (such as PCR, Whole Genome Amplification (WGA), protein purification, and cell growth) can also be performed on the same device (Chueh et al. 2011). By restricting reactions to very small volumes, diffusion distances are decreased, enhancing molecular interactions, in particular for amplifying an entire genome from femtograms of template DNA in a typical bacterium (Rodrique et al. 2009; Blainey 2013). Although microfluidics is not as high throughput as flow cytometry or FACs, it can be multiplexed (unlike micromanipulation methods) (Melin and Quake 2007; Kalisky et al. 2011). Microfluidics fills a distinct void left by other methods, in that single-cell experiment can be performed.

Microfluidic techniques can also be uniquely tailored to address specific problems. To separate a mixture of cells, with different morphologies and or membrane proteins, axenic strains of cyanobacteria were used to create an imprint on a polymeric film in a microfluidic device. This imprint was shown to preferentially capture cells that matched the original imprint from a mixture of cells with an 80–90 % efficiency rate (Schirhagl et al. 2012). This technology may hold promise for capturing single cyanobacterial cells with a wide variety of morphologies, such as branching or filamentous types, which have been difficult to capture by other methods (Grindberg et al. 2011).

Harnessing single-cell technologies for robust use with cyanobacteria present a unique set of challenges. Cyanobacterial species exhibit many morphologies, such as spherical, rod, and spiral, as well as filamentous and branching species and groups of filamentous cells surrounded by a thick sheath (Singh and Montgomery 2011). Lysis of cyanobacterial cells is notoriously difficult, as they typically have cell wall characteristics of both gram-
positive and gram-negative bacteria (Hoi czyk and Hansel 2000). In addition, many cyanobacteria produce copious amounts of ill-defined extracellular polymeric substances, which provide protection from viral lysis and dehydration, but can form hydrogen bonds with proteins, lipids, and DNA and may interfere with downstream molecular techniques (Pereira et al. 2009; Philippis et al. 2011).

One model system readily adapted for environmental single-cell exploration is the microbial mats that grow in the alkaline siliceous hot springs of Yellowstone National Park (YNP) (Ward et al. 2006). These dense biofilms are inhabited by a stable thermophilic community of moderate complexity, which include cyanobacteria, (Synechococcus sp.), Chloroflexi, and other less well-characterized heterotrophs (van der Meer et al. 2010; Klatt et al. 2011). Two isolates have been fully sequenced: Synechococcus sp. JA-3-Ab (Syn OS-A) isolated from 58 to 65 °C gradient and Synechococcus sp. JA-2-3-B’a(2-13) (Syn OS-B’) from 51 to 61 °C gradient in Octopus Spring, YNP (Bhaya et al. 2007; Melendrez et al. 2011). Comparative genomics revealed several interesting features about these genomes. For instance, large-scale genomic architecture, also known as synteny, is not well conserved, between the genomes of Syn OS-A and Syn OS-B’, (Bhaya et al. 2007).

The lack of synteny between the two genomes suggests that a high degree of recombination may be occurring in these organisms. Although the underlying cause of this recombination remains unclear, there is evidence of transposon activity in these communities (Nelson et al. 2012). Finally, it was observed that one of the most variable regions between Syn OS-A and Syn OS-B’ were the CRISPR spacer arrays, suggesting that viral–host interactions may be a powerful driving force that shapes the microbial mat communities (Heidelberg et al. 2009). In brief, the CRISPR–Cas adaptive immune system allows the host to incorporate foreign plasmid or phage DNA (termed “protospacers”) into CRISPR loci within its own genome (Sorek et al. 2013; Barrangou and Marraffini 2014). These “spacers” are subsequently expressed as small RNAs and are used to identify and target highly complementary invading DNA for degradation (Heidelberg et al. 2009). Thus, CRISPR arrays are a promising target to investigate diversity at the single-cell level. CRISPR spacer data from individual cells may shed light on how CRISPR spacers are acquired in natural systems, overall population turn-over rates, and the level of diversity in these regions between individuals. We describe a proof of concept study in which individual axenic Syn OS-B’ cells are sorted, captured, and visualized via a microfluidic device monitored by microscopy. We optimized lysis procedures, carried out Whole Genome Amplification (on-chip and off-chip), followed by amplification of loci of interest, including specific CRISPR arrays.

**Materials and methods**

**Growth and strain conditions**

An axenic culture of Synechococcus sp. JA-2-3-B’a(2-13) (Syn OS-B’) was grown in liquid DH10 media under 75 μmol photon m−2 s−1 light with continuous shaking (Adams et al. 2008). Fresh 1 mL aliquots were taken from cultures in exponential phase, centrifuged at 6,000×g for 5 min to gently pellet cells, 900 L of the supernatant was aspirated, and the concentrated cells were washed extensively with DH10 media (D media supplemented with 10 mM HEPES and adjusted to pH 8.2 (Castenholz 1969; Gomez-Garcia et al. 2011)). The pellet was washed up to five times, and stored no longer than 24 h at 4 °C before cell capture on the microfluidic device.

**Microfluidic device design and fabrication**

Individual cells were captured on a microfluidic device originally designed by (Marcy et al. 2007). The chip is fabricated with polydimethylsiloxane (PDMS), a silicon-based polymer commonly used in microfluidic devices due to its non-reactivity and flow properties, and mounted onto a glass cover slip for added stability (Ng et al. 2002; Friend and Yeo 2010). This design was slightly modified and fabricated by Eric Hall (Fig. 2) (Hall 2012) such that it is

---

**Fig. 2 Schematic of a Microfluidic device design modified by Marcy et al. (2007).** The inner working area of the microfluidic device consists of 2 layers: an upper layer containing 10-μm channels (shown in **black**) through which cells and reagents can flow, in addition to the 25-μm chambers (shown in **blue**) for specific chemical reactions. The lower layer contains the valves (shown in **red**) by which cell and reagent movement through channels in the upper layer can be controlled, and hydration lines (shown in **green**) to prevent dehydration of chamber volumes (Hall 2012). Individual cells are captured at (1, **purple oval**) and moved to chamber (2, **blue square**) to begin lysis with lysis buffer. The lysate is then pushed into chamber (3, **blue square**) with the addition of commercially available alkaline denaturation and lysis (DLB) solution (Qiagen, USA) to complete the cell lysis and denaturing released DNA. Stop buffer is added, and the lysate is pushed to chamber (4, **blue square**) where the pH is neutralized. A PCR mastermix solution is added to chamber (5, **large blue square**) where an isothermic MDA reaction occurs at 33 °C for 16 h. Amplified DNA is extracted at (6) with approximately 50 μL of TE buffer.
Movement of the chip itself was achieved by placement on a 638 nm wavelength (CrystaLaser; Reno, NV, USA). Visualized by excitation with a laser beam with 25 mW at 488 nm, injected into the inlet (Fig. 2). Single cells were moved via negative pressure, and isolated in the intersection (Fig. 2, purple circle) with the use of pneumatic valves (Unger et al. 2000). A cross flow of lysis buffer was used to “push” the isolated cell into the 3.0 nL lysis chamber (Fig. 2, chamber 2) as well as fill it. Six parallel cell captures, as well as two negative controls in which no cell was captured, were included in every chip run. The relative position of the negative controls was varied with each run.

Visualization of single cells within the microfluidic device

Movement and fluorescence of cells within the chip were observed via a Nikon Eclipse TE2000-U microscope with real-time imaging recorded by an EM-CCD camera (Andor iXon). As Syn OS-B’ are naturally fluorescent, cells were visualized by excitation with a laser beam with 25 mW at 638 nm wavelength (CrystaLaser; Reno, NV, USA). Movement of the chip itself was achieved by placement on a moveable stage (Listep Marzhäuser) that was remotely controlled by the user.

Captured single-cell lysis, multiple displacement amplification, and recovery

Lysis of captured cells was performed with the following optimized protocol: (1) A 2-h lysis incubation at 37.1 °C with Ready-Lyse (Epicenter, WI) in DH10 media (Fig. 2, chamber two), (2) a 90-min denaturation and lysis buffer (Repli-G) incubation at 69 °C (Fig. 2, chamber three), and (3) a 60-min stop incubation at room temperature (Fig. 2, chamber four). Lysis was followed by an on-chip MDA reaction with Phi29 polymerase for 16 h (Genomiphi, GE Healthcare, PA, USA) at 33 °C. Amplified DNA was recovered in μL volumes, and 1 μL was subjected to a second 50 μL off-chip MDA (as per kit instructions) to allow for a workable amount of amplified single-cell DNA.

Lysozyme inhibition of Phi29 amplification

Whole genome amplifications were performed in the conditions described in Fig. 3. Varying amounts of starting DNA template were used, either a “high” concentration at 11 ng/mL or a “low” concentration at 1.1 ng/mL. MDA reactions were performed as per the manual specifications (Qiagen, USA), with all kit solutions passed through a 0.2-μm filter and exposed to UV radiation for 1 h prior to use. To determine if either lysozyme was inhibiting the Phi29 MDA amplification, we subjected all samples to quantification via digital MDA (dMDA) using a Fluidigm chip with the manual specifications with EvaGreen as the fluorescent marker (Fluidigm, CA, USA).

Optimization of MDA reaction yield from single-cell input DNA

Previous studies have shown that MDA kit reagents can contain contaminating DNA that do not interfere with experiments in which there are many cells, and can inhibit successful amplification from single cells (Kalisky et al. 2011; Woyke et al. 2011). Contamination was detected in Qiagen Repli-G kit reagents (Fig. 3). To reduce contamination, Phi29 enzyme was expressed and purified by the method established by Blainey and Quake (Blainey 2010). All MDA kit solutions were filtered with a 0.2-μm filter and exposed to UV radiation for 1 h prior to use. UV radiation is commonly used to sterilize reagents, plasticware, and surfaces to prevent DNA contamination from unwanted sources during very sensitive applications (such as PCR or MDA) by creating pyrimidine dimers that prevent the DNA contaminant from acting as an effective template. (Ou 1991; Tamariz et al. 2006).

Amplification of genes of interest

To amplify the T1B CRISPR region from Syn OS-B’, located at the genome positions 1428062–1429246 (Heidelberg et al. 2009), primers were designed to exploit the unique repeat characteristic of the CRISPR array (Fig. 4a, b). The forward primer was placed outside the CRISPR loci, while the reverse primer was placed within the repeat regions of the CRISPR array, resulting in a “ladder-like” PCR product (Fig. 4). This design allowed the capture of unique and/unknown spacer sequences, while retaining spacer order information, from which recent and historical viral assaults can be determined. Primers were tested on DNA extracted from axenic cultures of Syn OS-B’ and Syn OS-A as well as total mat DNA extract (Fig. 4). In addition, variant general primers previously generated to three other loci of interest such as A0014/B0276 (a GUN4-like family protein), A0191/B1231 (a ParA homolog), and A0949/B2479 (a pentapeptide repeat protein), and previously published primers for universal bacterial 16S RNA sequences V1–V3 region were used as a positive control (Table 1) (Sundquist et al. 2007). Invitrogen recombinant Taq polymerase and mastermix were used for all amplifications. The following thermocycler program was used: (1) 94 °C for 2 min initial denaturation, (2) 94 °C for 60 s
denaturation, (3) 55 °C for 45 s annealing, (4) 72 °C for 2 min extension, (5) repeat from step 2 (29×), (6) 72 °C for 6 min extension/repair, (7) hold at room temperature. PCR reactions were visualized in 0.8 % agarose gels with ethidium bromide and run at 100 mV for 20 min. Primers were then used on amplified DNA recovered from single cells (Fig. 5).

Results and discussion

We have shown that unicellular thermophilic cyanobacteria can be captured, lysed, and subjected to whole genome amplification by Phi29 in a microfluidic device, and specific genes of interest can subsequently be amplified. However, we observed a vast difference in the rate of success for individual primer pairs. This was also noted in an earlier report, where capture and whole genome amplification of approximately ten *Synechococcus elongatus* PCC 7942 cells resulted in only 70–80 % coverage of the genome (Chueh et al. 2011). We successfully designed primers to amplify the T1B CRISPR loci from *Syn OS-B* (Fig. 4). We observed a wide range of successful amplifications for three control loci, but nearly 100 % success with the 16S RNA V1–V3 primer pair. *Syn OS-B* contains two copies of the 16S loci,

### Results and discussion

We have shown that unicellular thermophilic cyanobacteria can be captured, lysed, and subjected to whole genome amplification by Phi29 in a microfluidic device, and specific genes of interest can subsequently be amplified. However, we observed a vast difference in the rate of success for individual primer pairs. This was also noted in an earlier report, where capture and whole genome amplification of approximately ten *Synechococcus elongatus* PCC 7942 cells resulted in only 70–80 % coverage of the genome (Chueh et al. 2011). We successfully designed primers to amplify the T1B CRISPR loci from *Syn OS-B* (Fig. 4). We observed a wide range of successful amplifications for three control loci, but nearly 100 % success with the 16S RNA V1–V3 primer pair. *Syn OS-B* contains two copies of the 16S loci,
which may have contributed to the high success rate we observed. For the T1B CRISPR region, we consistently observed the lowest number of successful PCRs. This could potentially be due to the secondary structure caused by the palindromic repeats present in the CRISPR array, which are known to cause amplification biases using Phi29 (Ballantyne et al. 2007). Recently advances have been made to improve amplification efficiency of the Phi29 polymerase by adding helix-turn-helix domains to increase DNA-binding capability, and the use of random hexamers optimized for the organism in question (Alsmadi et al. 2009; de Vega et al. 2010). Use of these chimeric Phi29 polymerases, or hexamers enriched for nucleotide motifs found throughout the genome may improve overall single-cell coverage.

We observed that both the choice and concentration of lysozyme used within the microfluidic device play a crucial role in the success or failure of single-cell whole genome amplification. We observed that lysozyme from different suppliers could inhibit the MDA amplification reaction. Specifically, the Sigma lysozyme, at the concentrations used to disrupt the cell wall (visualized as a quenching of autofluorescence in lysed cells), on the microfluidic device inhibited successful DNA amplification (Fig. 3). This is in part due to the DNA-binding capabilities of lysozyme, and in part due to the nature of the microfluidic device, in which reagents can be added to chambers, but not removed (Lin et al. 2009; Hall 2012). Both specific binding from DNA motifs at the C and N termini, in addition to non-specific electrostatic binding, may interfere with DNA replication, resulting in whole genome amplification failure (Fig. 3) (Lin et al. 2009).

We also found the high-temperature incubation (69 °C) required to lyse the cyanobacteria was at the technical limit of the PDMS microfluidic device, and often led to chamber collapse and DNA recovery failure (personal communication, Eric Hall). Use of a different material to build the chip may lend itself to more robust behavior at higher temperatures required for complete lysis. We are also developing the use of a cyanophage-encoded lysozyme (Heidelberg et al. 2009) which may be more efficient in lysing cyanobacterial cells.

In future studies, we will further investigate heterogeneous natural populations found in the hot spring microbial mats. Sequencing all six identified CRISPR arrays found in Syn OS-B’ would result in an unparalleled glimpse into CRISPR spacer acquisition and turnover rates in the environment, both of which remain uncharacterized as of now (Bhya et al. 2011). Additionally, obtaining CRISPR spacer data from individuals, combined with metagenomic host and viral sequence, allows for investigation into host-viral co-evolution in a naturally occurring population. Metagenomes provide an exhaustive depth of coverage, while single-cell sequence retains the full genomic context. In a dovetailed approach, deeply targeted amplicon sequencing, resulting in statistically significant coverage of regions of interest, such as CRISPR loci, could be leveraged to estimate overall population level diversity in CRISPR spacers (Rosen et al. 2012). Such multipronged approaches have already been used with great success in deep ocean communities (Cameron Thrash et al. 2014) and herald a second revolution in understanding the dynamics of cyanobacterial diversity.

### Table 1 Primers used in this study

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Gene Family</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>A0014F</td>
<td>GUN4-like family protein</td>
<td>rGCsAGGATCTCrATsAGCA</td>
</tr>
<tr>
<td>A0014R</td>
<td></td>
<td>GTACC GGCCyGAAAsCTGAG</td>
</tr>
<tr>
<td>A0191F</td>
<td>ParA homolog</td>
<td>GTAAACCAGrkTGGGAGGGT</td>
</tr>
<tr>
<td>A0191R</td>
<td></td>
<td>GTCGAmGAGCCGCTyAAACT</td>
</tr>
<tr>
<td>A0949F</td>
<td>Pentapeptide repeat family protein</td>
<td>AAGCTGGGCTTGCTCArGT</td>
</tr>
<tr>
<td>A0949R</td>
<td></td>
<td>ATsCTGCTrGskGCCmAykCT</td>
</tr>
<tr>
<td>V1F</td>
<td>16S</td>
<td>AGAGTTTGATCMTGGCCCTCAG</td>
</tr>
<tr>
<td>V3R</td>
<td></td>
<td>ATGCCCGGGCTGCTGG</td>
</tr>
<tr>
<td>T1Bfor</td>
<td>CRISPR T1B</td>
<td>GATCGGTGGGTGTGTTTCCCTACT</td>
</tr>
<tr>
<td>T1Brev</td>
<td></td>
<td>CGCAAGGGGACCGAAAC</td>
</tr>
</tbody>
</table>

Ambiguous bases are as follows: \( r \): A or G, \( s \): C or G, \( v \): A or C or G, \( k \): G or T, \( m \): A or C

Fig. 5 Targeted amplicons amplified from three biological replicate single cells. L: OGeneRuler DNA ladder Mix (Fermentas). Purple arrow indicates 3,000 bp, yellow arrow indicates 1,000 bp, and red arrow indicates 500 bp. Amplification success was observed. Amplification of the 16S primers was observed in all three captured single cells (lane 5). Amplification of the T1B CRISPR array was successful only from cell 2 (lane 4)
References


revealed through single-cell genomics and biogeographic analysis. ISME J 7(1):184–198


