Title
Sequencing of Genomes from Environmental Single Cells.

Authors
Bowers, Robert M
Lee, Janey
Woyke, Tanja

Publication Date
2018

DOI
10.1007/978-1-4939-7514-3_8

Peer reviewed

Chapter 8

Sequencing of Genomes from Environmental Single Cells

Robert M. Bowers, Janey Lee, and Tanja Woyke

Abstract

Sequencing of single bacterial and archaeal cells is an important methodology that provides access to the genetic makeup of uncultivated microorganisms. We here describe the high-throughput fluorescence-activated cell sorting-based isolation of single cells from the environment, their lysis and strand displacement-mediated whole genome amplification. We further outline 16S rRNA gene sequence-based screening of single-cell amplification products, their preparation for Illumina sequencing libraries, and finally propose computational methods for read and contig level quality control of the resulting sequence data.

Key words Single amplified genomes, Fluorescence-activated cell sorting, Multiple displacement amplification, Tagmentation, Illumina genome sequencing

1 Introduction

While most of the diversity on our planet is microbial, studies have largely been limited to those microbes amenable to cultivation, greatly skewing our view of the microbial world [1]. However, the advent of cultivation-independent molecular tools, chiefly utilizing DNA sequencing, has provided a window into the phylogenetic diversity and genetic makeup of this underexplored majority. 16S rRNA gene PCR-based molecular surveys have revolutionized our view and understanding of the breadth of environmental microorganisms, and as sequencing quality and throughput have improved, so too has the genetic toolkit used to analyze microbial ecosystems. Single-cell genomics and metagenomics are now providing access to microbial genomes directly from bulk environmental samples without the constraints of cultivation. Recent studies have begun to take advantage of both approaches, producing high-quality microbial genomes while simultaneously expanding our view of the tree of life [2–4].

Single-cell genomics can be viewed as highly complementary to shotgun metagenomics, in that metagenomic datasets cast a wide net, producing a snapshot of the whole community [5], while single-cell sequencing approaches act as a scalpel used to dissect

Steven R. Head et al. (eds.), *Next Generation Sequencing: Methods and Protocols*, Methods in Molecular Biology, vol. 1712, https://doi.org/10.1007/978-1-4939-7514-3_8,

the community and extract the genomes of specific interest [6]. However, single-cell sequencing can be challenging, as there are a number of critical steps, each of which is dependent on the successful completion of the previous step. Given the rather extensive process of going from a bulk sample to the sequence data of a single amplified genome (SAG), we outline some of the most common approaches for each of the individual steps and provide a detailed protocol.

The single-cell production workflow involves the following steps: sample preservation and preparation, single-cell isolation, cell lysis, whole genome amplification (WGA), phylogenetic screening of WGA products, library preparation, sequencing, and quality control of both reads and assembled contigs. Sample preservation and preparation are often sample specific, but critical to downstream genome quality as reported previously [7]. Single-cell isolation begins with a cell suspension step followed by isolation of individual cells using any number of different techniques including microfluidics [8–10], fluorescence-activated cell sorting [3, 11, 12], micropipetting [13], and optical tweezers [10, 14], among others. Cell lysis and whole genome amplification are the next steps in the process. The cell lysis step must lyse/disrupt as many cell types as possible while maintaining the integrity of the genomic DNA without introducing contaminants [6], and the WGA is used to produce sufficient quantities of the target genome for downstream sequencing. Many options for WGA are currently available [13, 15, 16], however isothermal multiple displacement amplification (MDA) is by far the most popular [3, 12, 17]. While significant coverage biases remain with MDA-based WGA [18], and chimeras are known to form along hyper-branched MDA DNA structures [19], other WGA methods include similar biases while increasing hands on time in the lab, ultimately increasing the potential for contamination. Following the WGA, MDA products are typically screened using 16S rRNA gene primers and Sanger sequencing to determine the identity of the isolated cells and to select a subset of interest for genome sequencing. Single-cell sequencing libraries are then prepared from the MDA product. The Illumina Nextera XT library kits have been particularly useful in the preparation of single-cell libraries as the transposase-based libraries limit hands on time by combining the fragmentation and adapter ligation steps into a single reaction. Finally, as single-cell MDA products average one chimeric junction per 20 kb of MDA sequence [19, 20], short-read technologies such as Illumina are well suited for SAG sequencing.

Single-cell genomics pipelines are inherently susceptible to contamination, as even trace amounts of contaminating DNA may be enriched during the WGA step, severely impacting the quality of the resulting SAGs. Contamination may appear from multiple sources including the samples themselves, the lab environment,

and even vendor-supplied reagents [11, 21, 22]. Furthermore, as Illumina sequencing platforms have increased in throughput, SAGs are now multiplexed within a single-sequencing run, which increases the likelihood of well to well cross contamination. Both reads and contigs (i.e., assembled datasets) should thus be assessed for contamination prior to biological data interpretation. Typically, read level decontamination is performed by screening the reads against a contaminant database composed of common contaminant sequences. There are a number of tools available to assist with read decontamination including DeconSeq [23] and various modules from the BBtools package (https://sourceforge.net/projects/bbmap). When SAGs are multiplexed on a single sequencing run, cross contamination can be assessed by mapping the reads of a given library/well to all assemblies in the multiplexed sequencing run, although this method requires that the multiplexed SAGs are not derived from highly similar taxa or that the taxonomically similar SAGs are removed from such cross contamination analysis. Following assembly, contig-level quality control should also be performed. Quality assurance and decontamination of assembled SAGs has traditionally been semi-manual where SAGs are screened for non-target 16S rRNA genes, abnormal k-mer frequencies, and/or variable GC content [20]; however, ProDeGe, a tool that combines composition (k-mer frequencies) and a BLAST-based screen to identify and remove questionable contigs, has enabled automated contaminant removal [24]. Finally, genome completeness can be estimated with either a universal set of single-copy marker genes or with automated software such as CheckM that calculates a set of optimal markers based on the query genome's position in a reference tree [25]. Once quality assurance has been completed, the resulting genomes are ready for phylogenetic analysis, metabolic reconstruction, and other comparative genomic analyses [26] and can be deposited into the public databases.

With the current protocol, we describe a method for obtaining genomes from individual bacteria and archaea using FACS-based cell isolation, MDA amplification, and 16S rRNA gene-based phylogenetic screening comparable to Rinke et al. [27], followed by library preparation and shotgun sequencing using the Illumina platform. We also briefly outline the bioinformatic quality assurance of SAGs from read and contig-level decontamination to genome completion estimates.

2 Materials

2.1 Sample Preparation and Preservation

1. Sample (for example, sediment or biofilm).
2. Sterile, filtered buffer solution, for example, 1× PBS (ThermoFisher Scientific, Waltham, MA, USA).
3. TE, 100×, pH 8.0.

4. Milli-Q water (EMD Millipore, Temecula, CA, USA).
5. Molecular-grade glycerol.
6. Sterile UV-treated seawater (Sigma-Aldrich, St. Louis, MO, USA).
7. Cryovials, 2 mL.
8. Microcentrifuge.
9. Vortex.
10. Centrifuge.
11. Standard light microscope.
12. Sterile cotton swabs.
13. Ultrasonic water bath.
14. Falcon tube, 50 mL.
15. Filter, 0.2 μm.

2.2 Single-Cell Collection Via Fluorescence-Activated Cell Sorting (FACS)

1. Ultrapure water, such as Milli-Q water, or filtered molecular biology-grade water.
2. Bleach (5% solution of sodium hypochlorite).
3. PBS liquid concentrate, 10×, sterile.
4. SYBR Green fluorescent nucleic acid stain (ThermoFisher Scientific).
5. Cell sorter.
6. PCR hood with UV light for decontamination.
7. Two 2-L quartz flasks for UV treatment of sheath fluid.
8. Two stir plates, stir bars for UV treatment of sheath fluid.
9. BD Falcon 40 μm nylon cell strainer.
10. Polypropylene round-bottom tubes, 5 mL.
11. Pall Acrodisc, 32-mm syringe filter with 0.1 μm Supor membrane.
12. BD Luer-Lok tip disposable syringe, 10 mL.
13. Optical microtiter plate, e.g., LightCycler Multiwell Plate 384 (Roche Diagnostics, Indianapolis, IN, USA).

2.3 Single-Cell Lysis and Whole Genome Amplification by MDA

1. Qiagen REPLI-g Single Cell Kit (Qiagen, Valencia, CA, USA).
2. SYTO 13, 5 mM (ThermoFisher Scientific).
3. Bleach (5% solution of sodium hypochlorite).
4. Spectraline XL-1500 UV Cross-linker.
5. PCR hood with UV light for decontamination.
6. Plate reader with temperature control, or a real-time thermocycler, e.g., LightCycler 480 (Roche Diagnostics, Indianapolis, IN, USA).
7. Eppendorf Safe-Lock microcentrifuge tubes, 1.5 mL.

2.4 Phylogenetic Screening

1. Ultrapure water, such as Milli-Q water, or filtered molecular biology-grade water.
2. SsoAdvanced SYBR Green Supermix (Biorad, Pleasanton, CA, USA).
3. Primer set of choice, for example 16S rRNA gene universal primer set (926wF/1392R primer, 10 μM each).
4. ExoSAP-IT (Affymetrix, Santa Clara, CA, USA).
5. Standard thermocycler, or a real-time thermocycler.
6. Plate shaker.
7. Optical microtiter plate.

2.5 Sequencing of Single-Amplified Genomes (SAGs)

1. Nextera XT DNA Library Preparation Kit (Illumina, San Diego, CA, USA).
2. Nextera XT Index Kit (Illumina).
3. Agencourt AMPure XP (Beckman Coulter, Carlsbad, CA, USA).
4. Nuclease-free water.
5. Ethanol, 200 proof, molecular biology-grade.
6. Bioanalyzer High Sensitivity DNA Kit (Agilent Technologies, Santa Clara, CA, USA).
7. Standard thermocycler.
8. Agilent Bioanalyzer 2100 (Agilent Technologies).

3 Methods

3.1 Sample Preparation

Sample processing and preservation differ depending on the type of sample. With the current protocol, we provide a recommended workflow (Fig. 1) for the production of SAGs from two commonly surveyed sample types: a soil sample (Subheading 3.1.1) and a biofilm sample (Subheading 3.1.2).

3.1.1 Sediment or Soil Sample

1. For a sediment or soil sample, thoroughly mix 5 g of sediment or soil with 10–30 mL sterile buffer in a 50 mL falcon tube. For soils and freshwater sediments, use 1× PBS as buffer. For marine sediments, use sterile-filtered seawater as buffer.
2. Vortex at 14,000 rpm for 30 s to dislodge microbes from soil.
3. Centrifuge at 2000 × *g* for 30 s to remove large particles.
4. Collect the supernatant and proceed to Subheading 3.2.

3.1.2 Biofilm Sample

1. Using a sterile cotton swab, collect a sample of biomass into microcentrifuge tubes already containing a sterile-filtered buffer solution (e.g., 1× PBS or seawater).
2. Sonicate the sample in the tube in an ultrasonic water bath for 10 min.

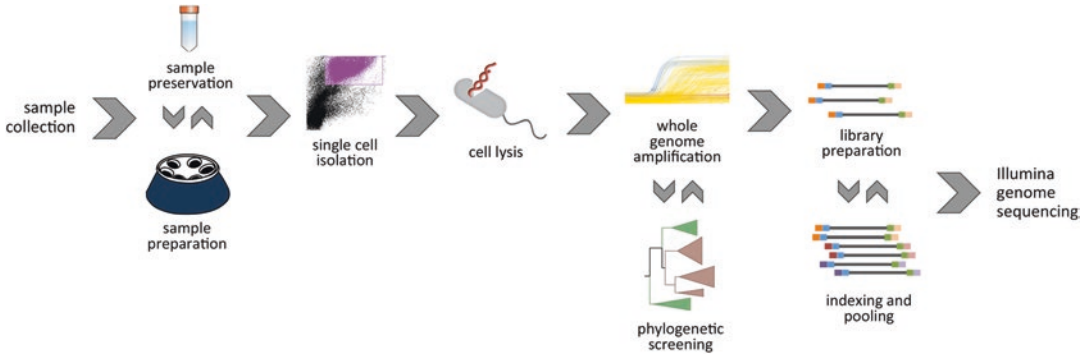

Fig. 1 Workflow for the preparation of sequencing libraries from environmental single cells. Environmental samples are typically preprocessed and cryopreserved. Single cells are then isolated using high throughput FACS, followed by lysis and MDA based whole genome amplification. MDA generated products are used as template in 16S rRNA gene based phylogenetic screening and identification. Finally, based on the results of the phylogenetic screen, a subset of single-cell MDA products is chosen for genome sequencing, following Illumina's tagmentation based Nextera XT library preparation protocol. Individual libraries may be sequenced on the MiSeq platform or pools of many barcoded libraries on higher throughput platforms such as NextSeq

3. Further shake the tube by hand for an additional 5 min.
4. Examine the sample under a microscope to ensure sufficient cell separation. If necessary, repeat **steps 2** and **3** above.

3.2 Sample Preservation

For best results and to minimize cell lysis during freeze-thawing, we recommend to cryo-protect the environmental sample and store at −80 °C.

1. Cryopreservation of cells should be done with sterile-filtered glycerol. Make Gly-TE stock by mixing 20 mL of 100× TE (pH 8.0), 60 mL of DIW (deionized water), and 100 mL of molecular-grade glycerol (*see* **Note 1**). Mix solution by vortexing thoroughly and pass through a 0.2 μm filter.
2. Transfer 100 μL of prepared GlyTE stock and 1 mL of sample solution to a sterile cryovial. Mix well by inverting.
3. Prepare several replicate vials for each sample. Store in liquid nitrogen or at −80 °C.

3.3 Preparation of FACS for Sterile Sort

As single-cell sorting and MDA are highly susceptible to exogenous contamination, it is essential to have a clean cell sorter with sterile sheath fluid. Briefly, fluidic lines in the cell sorter can be decontaminated by running bleach through the lines, and sheath fluid can be sterilized by UV treatment. This decontamination process is sufficient for a clean sorting process, provided it is performed before every run.

1. Within a clean hood, prepare 4 L of 1× PBS in two 2 L quartz flasks. Add magnetic stir bars to the flasks and place over plates

to begin stirring. Also include the empty sheath fluid tank and lid and arrange such that UV light will reach the inner surfaces. Begin the overnight UV exposure (at least 16 h). After UV exposure is complete, transfer the sterile sheath fluid to the sterile tank within the clean hood (*see* **Note 2**). Reserve at least 10 mL of clean sheath fluid for later use while sorting.

2. Prepare a second sheath tank with 1 L of 10% bleach (0.5% sodium hypochlorite final concentration) and run through the cell sorter for 2 h to decontaminate fluidic lines.
3. Dispose of any remaining bleach and rinse the sheath tank with sterile water. Run 1 L sterile water through the cell sorter for 30 min to rinse fluidic lines.
4. Install the dedicated clean tank with the sterile sheath fluid and begin running.
5. Prior to sorting, sterilize microtiter plates by UV treatment for 10 min. Do not seal or cover plates during UV treatment. To each well, add 2 μL of UV-treated, 1× PBS (*see* **Note 3**). Cover with optical seal and perform an additional UV treatment for 10 min.

3.4 Cell Separation by Flow Cytometry

Due to the large diversity of types of environmental samples and variation in technical operation of cell sorters, we here outline a general protocol for sorting single cells from environmental samples.

1. To avoid clogging the nozzle of the cell sorter, filter each environmental sample through an appropriately sized filter, relative to the nozzle size (e.g., use a 20 μm filter when working with a 70 μm nozzle).
2. Stain the cells in the sample with 1× SYBR green at 4 °C for 15 min in the dark (*see* **Note 4**).
3. Run the stained sample through the cell sorter and target the selected population with a sort gate (*see* **Note 5**).
4. Sort the targeted cell population into the UV-treated optical microtiter plates containing 2 μL 1× PBS per well (*see* **step 5** in Subheading 3.3). Any exogenous extracellular DNA that may be sorted with a single cell could present as contamination in downstream WGA. A two-step sort may be utilized to dilute this DNA (*see* **Note 6**). We recommend sorting into 96-well or 384-well plates and including positive and negative controls (*see* **Note 7**).
5. Seal plates with sterile foil and store at −80 °C.

3.5 Single-Cell Lysis

The Qiagen REPLI-g Single Cell Kit is used for both single-cell lysis and WGA steps. The WGA is based on the Phi29 catalyzed MDA reaction. Refer to the manufacturer's instructions for a

detailed protocol. Note that the recommended total reaction volume of 50 μL can be reduced by up to one-fifth without any adverse effects. As an example, we here outline a 25 μL reaction.

1. Before performing any work, wipe down and clean hood surfaces, pipettes, and equipment with 10% bleach. UV the clean hood for 60 min with equipment inside (*see* **Note 8**).
2. Prepare lysis Buffers DLB and D2 according to the manufacturer's instructions. To each well containing a sorted single cell in 2 μL 1× PBS, add 1.5 μL Buffer D2. Mix thoroughly by tapping and spin down at 1000 × *g* for 1 min. Incubate for 10 min at 65 °C.
3. Add 1.5 μL of Stop Solution to each well. Mix thoroughly by tapping and spin down at 1000 × *g* for 1 min. Store lysed cells (total reaction volume of 5 μL) briefly at 4 °C while preparing the WGA amplification master mix.

3.6 Whole Genome Amplification

For single-cell genome amplification, we recommend using the Qiagen REPLI-g Single Cell Amplification Kit in conjunction with SYTO staining for real-time amplification monitoring for the purpose of quality control (Fig. 2). Further, the recently developed WGA-X method is recommended for high GC templates [28].

1. Thaw the Qiagen REPLI-g Single Cell Amplification Kit reagents according to the manufacturer's instructions. Briefly, thaw the Phi29 DNA Polymerase on ice. Thaw all other reagents at room temperature.

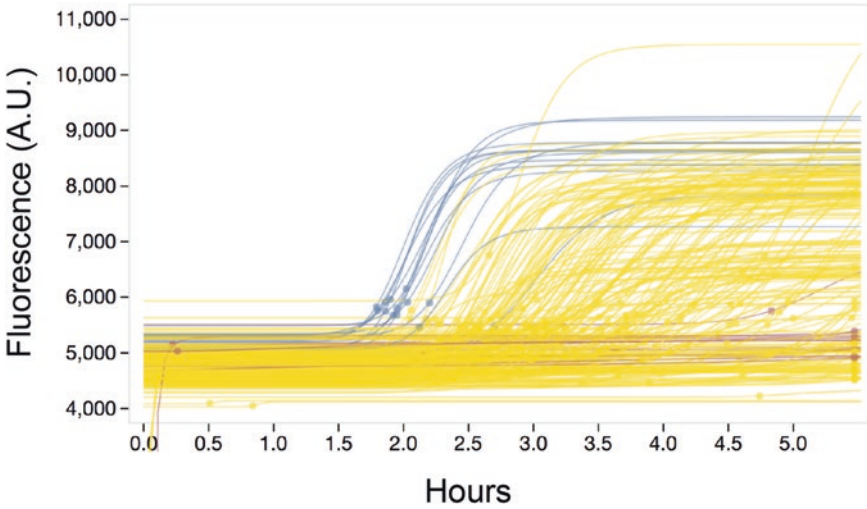

Fig. 2 Real-time kinetics for the whole genome amplification of single cells and controls in a 384-well plate. *Blue lines* show the amplification kinetics of the positive controls (100 cells sorted into a single well), *yellow* lines show the amplification kinetics of sorted single cells and *red lines* represent negative controls (no cells sorted)

2. Prepare enough master mix for all the reactions, adding Phi29 DNA Polymerase at the end. Each reaction should contain: 4.5 μL water (provided in the Qiagen kit, specifically for use in single-cell genomics), 14.5 μL Reaction Buffer, 1 μL Phi29 DNA Polymerase. Combine the water and Reaction Buffer first and mix thoroughly by vortexing before the addition of the polymerase last. Mix by vortexing and spin down (*see* **Note 9**).
3. To track amplification of single cells in real time, add SYTO 13 to the master mix for a final concentration of 0.5 μM (*see* **Note 10**). Mix by vortexing and spin down.
4. To each well containing 5 μL of lysed cells, add 20 μL of MDA master mix, for a total reaction volume of 25 μL. Cover the plate with an optical seal and centrifuge at 1000 × *g* for 1 min (*see* **Note 11**).
5. Incubate the plate at 30 °C for 2–5 h in a real-time thermocycler or plate reader, e.g., the Roche LightCycler 480 (*see* **Note 12**).
6. Heat-inactivate the Phi29 DNA Polymerase by incubating at 65 °C for 10 min. MDA products can be stored for multiple months at −80 °C (*see* **Note 13**).

3.7 Phylogenetic Screening

Phylogenetic screening is not mandatory, but useful for the accurate identification of potential target cells for genome sequencing. We here outline the use of universal 16S rRNA gene primers, but primers for specific functional genes may be used as well.

1. Make a 1:100 dilution of the MDA product using nuclease-free water. Mix dilution extremely thoroughly (*see* **Note 14**). Hand-pipetting is most effective, or alternatively, use a plate-shaker for 15 min at the maximum setting.
2. Transfer 2 μL of diluted MDA product as template to an optical microtiter plate.
3. Thaw the PCR reagents on ice: SsoAdvanced SYBR Green Supermix, 10 μM 926wF primer, 10 μM 1392R primer (*see* **Note 15**).
4. Prepare enough master mix for all PCR reactions. Each reaction should contain: 2.6 μL nuclease-free water, 5 μL SsoAdvance SYBR Green Supermix (2×), 0.2 μL 926wF primer (10 μM), 0.2 μL 1392R primer (10 μM). Mix by vortexing and spin down.
5. To each well containing 2 μL of diluted MDA product as template, add 8 μL of PCR master mix, for a total reaction volume of 10 μL. Cover the plate with an optical seal and centrifuge at 1000 × *g* for 1 min.
6. Refer to the manufacturer's protocol when selecting a cycling program. Adding a melt curve step to the cycling program will aid in analyzing and choosing PCR products for downstream

sequencing. Begin amplification of PCR products in a real-time thermocycler.

7. Purify and sequence PCR amplicons to identify WGA products. Prior to Sanger sequencing, PCR products are purified using ExoSAP-IT according to the manufacturer's instructions. Sanger sequencing is then performed on the clean PCR products, followed by sequence analysis.

3.8 Library Preparation and Sequencing of SAGs

The Illumina Nextera XT DNA Sample Preparation Kit is used to prepare indexed paired-end libraries from single-cell MDA products for next-generation sequencing. Briefly, libraries are prepared using a single-step reaction where the transposase simultaneously fragments and ligates Illumina sequencing adapters in a single 5 min reaction. This library preparation is cost-effective and Illumina is currently the most prevalent next-generation sequencing platform.

1. Before performing any work, wipe down and clean hood surfaces, pipettes, and equipment with 10% bleach (*see* **Note 16**).
2. Thaw reagents for tagmentation on ice. Mix by inverting gently and spin down. To sterile microcentrifuge tubes, add: 10 μL of Tagment DNA buffer, 5 μL of input DNA (1 ng total), and 5 μL of Amplicon Tagment Mix.
3. Mix by pipetting and spin down. Incubate on a thermocycler at 55 °C for 5 min and hold at 10 °C.
4. Once the thermocycler reaches 10 °C, immediately neutralize the reaction by adding 5 μL of Neutralize Tagement Buffer. Mix by pipetting and spin down. Incubate at room temperature for 5 min.
5. Thaw the appropriate indexes and the Nextera PCR master mix. To each reaction, add 15 μL of PCR master mix, 5 μL of Index 1, and 5 μL of Index 2 (*see* **Note 17**). Mix by pipetting and spin down. PCR amplify according to the manufacturer's protocol.
6. Perform cleanup of PCR product by using a 1.8× ratio of AMPure XP beads (*see* **Note 18**). For example, add 90 μL of beads to each 50 μL PCR reaction. Mix by vortexing and quick spin.
7. Shake samples at 1800 rpm for 2 min, then incubate at room temperature for 5 min.
8. Place tubes on a magnetic stand to pellet the beads (*see* **Note 19**). Carefully remove and discard the supernatant once the solution has become clear. Make sure not to remove any beads.
9. Without removing tubes from the magnetic stand, add 200 μL of freshly prepared 80% ethanol to each sample. Let samples sit for 30 s. Remove and discard ethanol. Repeat, for a total of two 80% ethanol washes (*see* **Note 20**).

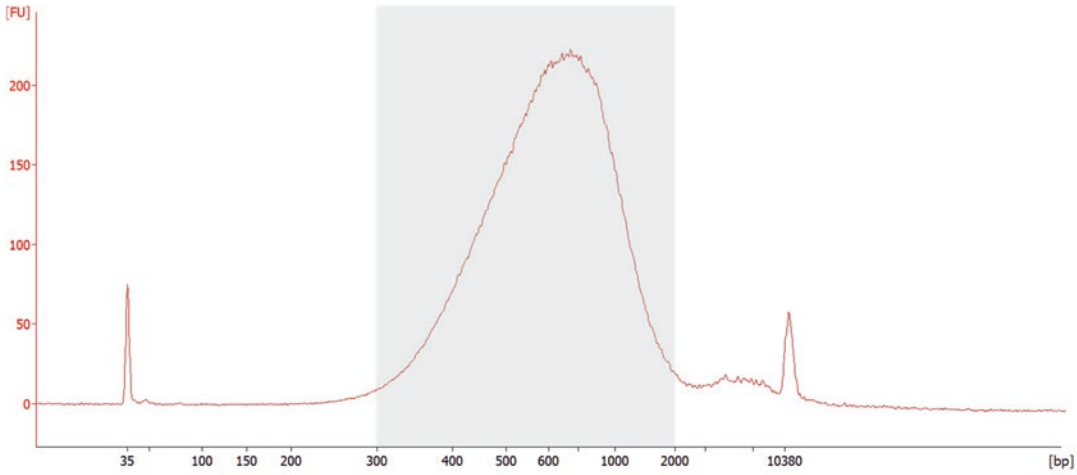

Fig. 3 Library QC of a pool of barcoded Nextera XT single-cell libraries on Agilent HS Bioanalyzer. 35 bp and 10,380 bp peaks represent markers used as internal standards. When using the protocol, library insert size distributions typically ranges from 300 to 2000 bp

10. Leave tube-caps open and let the samples air-dry on the magnetic stand for 15 min until residual 80% ethanol has evaporated.
11. Add 52 μL of Resuspension Buffer to each tube. Resuspend the beads and mix thoroughly.
12. Shake the samples at 1800 rpm for 2 min, then incubate at room temperature for 2 min.
13. Place the tubes on the magnetic stand to pellet the beads. Once the liquid is clear, carefully transfer 50 μL of eluted PCR product to a new tube.
14. Perform quality control on the final library. Run 1 μL of library on the Agilent 2100 Bioanalyzer using the High Sensitivity Kit for size and quantification (Fig. 3) (*see* **Note 21**).
15. Barcoded libraries can be pooled in equimolar ratios according to Illumina's pooling protocols and procedures for sequencing on the Illumina platform. For library concentrations, refer to the Illumina platform-specific protocol.
16. Sequence individual libraries or library pools using the Illumina platform, such as MiSeq or NextSeq, according to the manufacturer's protocol. NextSeq is specifically recommended for library pools [28].

3.9 Assembly of SAG Sequences and Data Quality Assurance

To ensure high-quality genomes, a few steps need to be taken before data are useable for downstream biological inference. Contaminant sequences can be identified and removed at both the read and contig (post-assembly) levels. At the read level, data are

typically removed if they match sequences from a lab specific contaminant database. Once the reads have been screened for contaminants, the decontaminated reads are used as input for a SPAdes assembly [29]. SPAdes has been specifically developed to deal with the highly uneven coverage produced during the MDA step, as SPAdes uses multiple coverage cutoffs, resulting in a larger fraction of useable data, ultimately yielding more complete assemblies. After the assembly, an additional round of contaminant screening should be performed at the contig level. Once potential contaminants have been removed, the resulting SAGs should be analyzed for genome completeness, and each of the abovementioned steps should be documented as metadata during public database submission.

1. Perform quality trimming and adapter trimming using any one of the currently available tools such as cutadapt (http://code.google.com/hosting/moved?project=cutadapt), Trimmomatic [30], BBDuk.sh from the BBMap (http://sourceforge.net/projects/bbtools) package package, among others.
2. Screen reads for contaminants against either premade contaminant databases such as the one employed in DeconSeq [23] or assemble an in-house contaminant database and filter out contaminant reads using BBDuk.sh from the BBMap package (http://sourceforge.net/projects/bbtools).
3. Use cleaned reads as input to SPAdes [29] for denovo genome assembly.
4. Check assembled contigs for additional contaminants either manually (https://img.jgi.doe.gov/er/doc/SingleCellDataDecontamination.pdf) or by taking an automated approach using a tool like ProDeGe, which takes a feature-based (k-mer) approach combined with BLAST-based screening to identify suspect contigs in an assembly [24].
5. Check resulting QC'd and largely contaminant free genomes for estimated completeness and a final screen for contaminants using either a universal set of single copy markers or by implementing an automated strategy as recently outlined in CheckM [25] (*see* **Note 19**).

4 Notes

1. Glycerol is extremely viscous and is most accurately transferred via syringe. Store Gly-TE stock at −20 °C.
2. We suggest dedicating one tank to clean sheath fluid for single-cell genomics use.
3. The amount of 1× PBS added to the plate prior to sorting is determined by the total MDA reaction volume chosen (Subheading 3.5).

4. Minimize exposure of SYBR green to direct light.
5. For a wide variety of environmental samples, it can be difficult to define the target population. For example, sediment samples generally show a high background signal that may interfere with locating a distinct cell population. Other, low-biomass environments may not have a high enough cell density to easily identify a population. Some possible solutions include utilizing different nucleic acid stains or applying different sample preparation methods.
6. Begin by sorting at least 10,000 target cells into 1 mL of UV-treated 1× PBS. Backflush the sample line for 1 min, run 10% bleach solution through the sample line for 5 min, and backflush for an additional 5 min. Sort this "pre-sorted" population of cells into the UV-treated plates containing 2 μL 1× PBS per well.
7. Positive controls are rows or columns where 10–100 cells are sorted in each well, and negative controls are rows or columns in which no cells are sorted into the corresponding wells (Fig. 2).
8. Because single-cell genomics processes are so highly susceptible to outside contamination, it is important to reserve a dedicated space for single-cell work, with separate equipment and to be maintained as sterile as possible.
9. Do not let the MDA master mix warm to a temperature above 30 °C. The Phi29 enzyme becomes inactivated at temperatures above 30 °C.
10. Before adding to the master mix, minimize exposure of SYTO13 to direct light.
11. Centrifuge the plate thoroughly to ensure that no bubbles remain in the wells. The presence of bubbles may interfere with the real-time readings measured by the instrument.
12. If directly monitoring the amplification reaction in real-time, incubation can be cut short when the negative controls begin to amplify.
13. Due to the hyper-branched nature of MDA products, MDA DNA is difficult to homogenize after a freeze-thaw cycle. If possible, proceed directly to phylogenetic screening.
14. MDA product is extremely viscous. It is important to thoroughly mix both the MDA product (pre-dilution) and the MDA dilution as well.
15. Following the manufacturer's instructions, minimize exposure of SsoAdvance SYBR Green Supermix to direct light.
16. Prevent PCR contamination of pre-PCR processes by physically separating lab spaces where pre and post-PCR processes are performed.
17. If pooling multiple libraries together for a sequencing run, refer to the manufacturer's protocol to select the correct

indexes for each sample, dependent on the number of libraries being pooled.

18. In preparation for bead cleanup, follow the manufacturer's instructions on best handling practices: let the beads equilibrate to room temperature before use, always vortex the beads thoroughly each time before use, and prepare fresh 80% ethanol solution before each use.
19. The pelleting process may take up to 5 min.
20. Be sure to remove all residual 80% ethanol from each sample. The tubes should remain on the magnetic stand throughout washing.
21. A typical library will have an insert size distribution of ~300–2000 bp (Fig. 3).
22. Data can be annotated and further analyzed in the IMG system (http://img.jgi.doe.gov/) [31].

Acknowledgment

The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported under Contract No. DE-AC02-05CH11231. We would like to thank Bill Andreopoulos for his assistance in preparing Fig. 2.

References

1. Amann RI, Ludwig W, Schleifer KH (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol Rev 59:143–169
2. Eloe-Fadrosh EA, Ivanova NN, Woyke T, Kyrpides NC (2016) Metagenomics uncovers gaps in amplicon-based detection of microbial diversity. Nat Microbiol 1:15032
3. Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F, Darling A, Malfatti S, Swan BK, Gies EA, Dodsworth JA, Hedlund BP, Tsiamis G, Sievert SM, Liu W-T, Eisen JA, Hallam SJ, Kyrpides NC, Stepanauskas R, Rubin EM, Hugenholtz P, Woyke T (2013) Insights into the phylogeny and coding potential of microbial dark matter. Nature 499:431–437
4. McLean JS, Lombardo M-J, Badger JH, Edlund A, Novotny M, Yee-Greenbaum J, Vyahhi N, Hall AP, Yang Y, Dupont CL, Ziegler MG, Chitsaz H, Allen AE, Yooseph S, Tesler G, Pevzner PA, Friedman RM, Nealson KH, Venter JC, Lasken RS (2013) Candidate phylum TM6 genome recovered from a hospital sink biofilm provides genomic insights into this uncultivated phylum. Proc Natl Acad Sci U S A 110:E2390–E2399
5. Thomas T, Gilbert J, Meyer F (2012) Metagenomics – a guide from sampling to data analysis. Microb Inform Exp 2:3
6. Stepanauskas R (2012) Single cell genomics: an individual look at microbes. Curr Opin Microbiol 15:613–620
7. Clingenpeel S, Schwientek P, Hugenholtz P, Woyke T (2014) Effects of sample treatments on genome recovery via single-cell genomics. Isme J:1–4. https://doi.org/10.1038/ismej.2014.92
8. Marcy Y, Ouverney C, Bik EM, Lösekann T, Ivanova NN, Martin HG, Szeto E, Platt D, Hugenholtz P, Relman DA, Quake SR (2007) Dissecting biological 'dark matter' with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. Proc Natl Acad Sci U S A 104:11889–11894
9. Gawad C, Koh W, Quake SR (2016) Single-cell genome sequencing: current state of the science. Nat Rev Genet 17:175–188
10. Lo S-J, Yao D-J (2015) Get to understand more from single-cells: current studies of microfluidic-based techniques for single-cell analysis. Int J Mol Sci 16:16763–16777
11. Stepanauskas R, Sieracki ME (2007) Matching phylogeny and metabolism in the uncultured

marine bacteria, one cell at a time. Proc Natl Acad Sci U S A 104:9052–9057

12. Swan BK, Tupper B, Sczyrba A, Lauro FM, Martinez-Garcia M, González JM, Luo H, Wright JJ, Landry ZC, Hanson NW, Thompson BP, Poulton NJ, Schwientek P, Acinas SG, Giovannoni SJ, Moran MA, Hallam SJ, Cavicchioli R, Woyke T, Stepanauskas R (2013) Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. Proc Natl Acad Sci U S A 110: 11463–11468
13. Zong C, Lu S, Chapman AR, Xie XS (2012) Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. Science 338:1622–1626
14. Landry ZC, Giovanonni SJ, Quake SR, Blainey PC (2013) Optofluidic cell selection from complex microbial communities for single-genome analysis. Methods Enzymol 531:61–90
15. Hoeijmakers W, Bártfai R, Françoijs K-J, Stunnenberg HG (2011) Linear amplification for deep sequencing. Nat Protoc 6:1026–1036
16. Duhaime MB, Deng L, Poulos BT, Sullivan MB (2012) Towards quantitative metagenomics of wild viruses and other ultra-low concentration DNA samples: a rigorous assessment and optimization of the linker amplification method. Environ Microbiol 14:2526–2537
17. Kashtan N, Roggensack SE, Rodrigue S, Thompson JW, Biller SJ, Coe A, Ding H, Marttinen P, Malmstrom RR, Stocker R, Follows MJ, Stepanauskas R, Chisholm SW (2014) Single-cell genomics reveals hundreds of coexisting subpopulations in wild Prochlorococcus. Science 344:416–420
18. Marine R, McCarren C, Vorrasane V, Nasko D, Crowgey E, Polson SW, Wommack KE (2014) Caught in the middle with multiple displacement amplification: the myth of pooling for avoiding multiple displacement amplification bias in a metagenome. Microbiome 2:3
19. Lasken RS, Stockwell TB (2007) Mechanism of chimera formation during the multiple displacement amplification reaction. BMC Biotechnol 7:19
20. Woyke T, Xie G, Copeland A, González JM, Han C, Kiss H, Saw JH, Senin P, Yang C, Chatterji S, Cheng J-F, Eisen JA, Sieracki ME, Stepanauskas R (2009) Assembling the marine metagenome, one cell at a time. PLoS One 4:e5299
21. Blainey PC (2013) The future is now: single-cell genomics of bacteria and archaea. FEMS Microbiol Rev 37:407–427
22. Woyke T, Xie G, Copeland A, González JM, Han C, Kiss H, Saw JH, Senin P, Yang C, Chatterji S, Cheng J-F, Eisen JA, Sieracki ME, Stepanauskas R (2011) Decontamination of MDA reagents for single cell whole genome amplification. PLoS One 6:e26161
23. Schmieder R, Edwards R (2011) Fast identification and removal of sequence contamination from genomic and metagenomic datasets. PLoS One 6:e17288
24. Tennessen K, Andersen E, Clingenpeel S, Rinke C, Lundberg DS, Han J, Dangl JL, Ivanova NN, Woyke T, Kyrpides N, Pati A (2015) ProDeGe: a computational protocol for fully automated decontamination of genomes. ISME J. https://doi.org/10.1038/ismej.2015.100
25. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. https://doi.org/10.7287/peerj.preprints.554v2
26. Lasken RS, McLean JS (2014) Recent advances in genomic DNA sequencing of microbial species from single cells. Nat Rev Genet 15:577–584
27. Rinke C, Lee J, Nath N, Goudeau D, Thompson B, Poulton N, Dmitrieff E, Malmstrom R, Stepanauskas R, Woyke T (2014) Obtaining genomes from uncultivated environmental microorganisms using FACS-based single-cell genomics. Nat Protoc 9:1038–1048
28. Stepanauskas R, Fergusson EA, Brown J, Poulton NJ, Tupper B, Labonté JM, Becraft ED, Brown JM, Pachiadaki MG, Povilaitis T, Thompson BP, Mascena CJ, Bellows WK, Lubys A (2017) Improved genome recovery and integrated cell-size analyses of individual uncultured microbial cells and viral particles. Nat Commun 8:84. https://doi.org/10.1038/s41467-017-00128-z
29. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477
30. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170
31. Markowitz VM, Chen I-MA, Chu K, Szeto E, Palaniappan K, Pillay M, Ratner A, Huang J, Pagani I, Tringe S, Huntemann M, Billis K, Varghese N, Tennessen K, Mavromatis K, Pati A, Ivanova NN, Kyrpides NC (2014) IMG/M 4 version of the integrated metagenome comparative analysis system. Nucleic Acids Res 42:D568–D573