

UC Santa Cruz

UC Santa Cruz Previously Published Works

Title

Chapter Three Coupling FACS and Genomic Methods for the Characterization of Uncultivated Symbionts

Permalink

<https://escholarship.org/uc/item/9j05b9kw>

Authors

Thompson, Anne
Bench, Shellie
Carter, Brandon
et al.

Publication Date

2013

DOI

10.1016/b978-0-12-407863-5.00003-4

Peer reviewed

Title.

Coupling FACS and genomic methods for the characterization of uncultivated symbionts

Author names and affiliations.

Anne Thompson. University of California, Santa Cruz. Santa Cruz, CA.

Shellie Bench. Stanford University, Stanford, CA

Brandon Carter. University of California, Santa Cruz. Santa Cruz, CA.

Jonathan Zehr¹. University of California, Santa Cruz. Santa Cruz, CA.

¹Corresponding author

University of California, Santa Cruz

Department of Ocean Sciences

1156 High Street

Santa Cruz, CA 95060.

Abstract

Symbioses between microbes are likely widespread and functionally relevant in diverse biological systems however they are difficult to discover. Most microbes remain uncultivated, symbioses can be relatively rare or dynamic, and intercellular connections can be delicate. Thus traditional methods such as microscopy are inadequate for efficient discovery and precise characterization of novel interactions, their metabolic basis, and the species involved. High-throughput metagenomic sequencing of entire microbial communities has revolutionized the field of microbial ecology, however genomic signals from symbionts can get buried in sequences from abundant organisms and evidence for direct links between microbial species cannot be gained from bulk samples. Thus, a specialized approach to the characterization of symbioses between naturally-occurring microbes is required. This chapter presents methods for combining fluorescence-activated cell sorting (FACS) to isolate and separate uncultivated symbionts with molecular biology techniques for DNA amplification in order to characterize uncultivated symbionts through genomic and metagenomic techniques.

Keywords: Symbiosis, flow cytometry, cell sorting, multiple displacement amplification, metagenomics.

1. Introduction

Symbioses occur between microorganisms in diverse biological systems. These associations range from delicate connections at cell surfaces to integration of one symbiont into the cell of another, similar to an organelle. The metabolic basis for the symbioses, the types of cellular structures that link cells, their dynamics, and the relative autonomy of the symbiotic partners, are likely as diverse and ecologically relevant as microbes themselves.

The majority of microbes remain uncultivated, so study of complex natural microbial communities is required to discover novel and characterize symbioses. Coupling of microscopy to fluorescence *in situ* hybridization (FISH) has been a successful approach to studying uncultivated symbioses (Orphan, et al., 2001). However these traditional methods are limited for several reasons. Morphological features of the symbioses can be damaged during sample handling. This is relevant to epiphytic symbioses or symbioses between delicate cells. Use of untreated samples could reduce artifacts and preserve intercellular connections, but garnering enough precise measurements of the symbionts' cellular parameters for statistical power is tedious. To characterize metabolisms of symbionts, studies have applied stable isotope analysis along with FISH (Orphan, et al., 2001). However, assessing the full metabolisms and evolutionary histories of symbionts is limited without genomic or metagenomic analysis and it would be difficult to recover intact cells from FISH preparations for whole-genome sequencing. An ideal approach to characterizing uncultivated symbionts will combine high-throughput methods that yield data on cellular parameters, phylogeny, and metabolism.

Fluorescence-activated cell sorting (FACS) is a widely-chosen method for separating and concentrating cells from complex samples prior to genomic analysis. Other methods such as serial dilution (Zhang, et al., 2006), micromanipulation (Kvist, et al., 2007; Hongoh, et al.,

2008), laser capture microdissection (Navin, et al., 2011), Raman tweezers (Brehm-Stecher & Johnson, 2004; Huang, et al., 2009) and microfluidics (Marcy, et al., 2007; Blainey, et al., 2011) have also been used effectively to separate cells prior to genetic or genomic analysis. However, for characterizing uncultivated symbionts these methods fall short of FACS in a number of ways.

FACS machines deposit single events (individual cells or multiple cells in association) into a variety of vessels quickly and accurately (Ibrahim & van den Engh, 2003, 2007). Precise assessment of pigment, cell size, membrane features, and genetic composition (cell cycle analysis) can be obtained in minutes for thousands of events (Ibrahim & van den Engh, 2007). Contaminating DNA is limited because only a very small droplet of liquid from the source sample and sheath fluid accompany each sorted event (Stepanuskas & Sieracki, 2007). Finally, FACS is a relatively gentle. The laminar flow fluidics of FACS prevents disruption of cells during sorting and this critical in the study of potentially delicate symbioses (Ibrahim & van den Engh, 2007).

Genomic analysis can dramatically increase understanding of the basis for cellular interactions and the ecosystem function of symbionts beyond measuring cellular parameters (Tripp, et al., 2010). However, current next-generation sequencing applications require more nucleic acids than can be obtained from thousands of sorted microbes. Thus, coupling FACS to nucleic acid amplification methods is essential.

MDA is a major advance over using PCR for Whole-Genome Amplification (WGA) (Telenius, et al., 1992; L. Zhang, et al., 1992; Cheung & Nelson, 1996; Dean, et al., 2001). MDA uses the Phi29 DNA polymerase to generate high quantities of double-stranded DNA that can be longer than 10 kb, making it an excellent starting material for sequencing applications and genomic analyses (Figure 1).

Besides MDA, other nucleic acid amplification techniques can also be useful in characterizing microbial symbioses (Figure 1). Especially in delicate associations, one symbiotic partner may remain unknown. In these cases, FACS can be coupled to nested PCR, which is applied to sorted single-events (host-symbiont complexes), to amplify phylogenetically relevant genes for sequencing. This is a relatively low cost method of identifying symbiotic partners before further sequencing efforts (Thompson, et al., 2012) (Figure 1).

Together FACS, MDA and nested PCR can provide researchers with information on the species specificity of symbiotic associations, metabolic potential, ecological function, and evolutionary history of uncultivated microbial symbionts. This chapter focuses on how Fluorescence-Activated Cell Sorting (FACS) can be coupled with genomic analysis to identify and characterize uncultivated symbionts (Figure 1).

2. Fluorescence-Activated Cell Sorting (FACS)

2.1 Flow cytometer features

While several high-throughput flow sorters are available, the BD Biosciences Influx™ Cell Sorter has been the FACS of choice for several recent studies on uncultivated symbionts (Zehr, et al., 2008; Cuvelier, et al., 2010; Tripp, et al., 2010; Vaultot, et al., 2012). The Influx™ can be equipped with ten laser paths supporting collection of 24 data parameters, with the choice of lasers dependent on target cell types. A 488 nm laser (Sapphire Coherent) is commonly used to study phytoplankton as it excites chlorophyll *a* (measured with a band-pass optical filter of 692-40 nm) and other auto-fluorescent pigments such as phycoerythrin (572-27 nm). Unpigmented cells can be stained with SYTO-13 (Invitrogen, Carlsbad, CA) (Del Giorgio, et al., 1996; Stepanauskas & Sieracki, 2007) and 4',6-diamidino-2-phenylindole (DAPI) and detected

using the 488 nm or UV laser (355 nm) (JDS Uniphase Xcyte Laser), respectively. The multiple lasers of the Influx™ are especially useful when cells are hybridized to synthetic or antibody-based probes. Functional fluorophores are available across the light spectrum and are another means of distinguishing between cell populations via FACS (Orcutt, et al., 2008).

2.2 Preparation for sorting

To minimize contamination ultraclean conditions are required for cell sorting prior to DNA amplification. Sample collection tubes and the sheath fluid reservoir are treated with UV for 2 hours before filling (Rodrigue, et al., 2009). Sheath fluid may be as simple as a 1% NaCl solution prepared with UV-treated ddH₂O and heat combusted NaCl (450°C for 4 hours), or a commercially available formula such as BioSure (BioSure, Grass Valley, CA USA). Sheath fluid and sample lines are cleaned by running warm water followed by bleach solution (5-10%) and extensive flushing with UV-treated, DNA-free ddH₂O (Stepanauskas & Sieracki, 2007; Rodrigue, et al., 2009).

Nozzle diameter sizes from 70 µm to 200 µm are available for the Influx™. A 70 µm nozzle is commonly used and reliably sorts cells less than 10 µm in width. Larger nozzles provide more reliable sorting of large particles (up to 25 µm for the 140 µm nozzle) and blockages are less likely when processing samples containing large particles. However, sort droplet size increases with nozzle size, which increases chances of contamination from the sheath fluid. Lower sorting efficiency due to coincident drop occupancy is also a disadvantage of larger nozzle diameters. Choosing the smallest diameter nozzle that still allows sorting of the largest target cells is recommended.

2.3 Sorting

The Influx™ default for triggering data collection and sorting is light scatter in the forward direction (FSC). After preparing the fluidics, the laser alignment and focus are optimized using standard fluorescent beads (Spherotech, Lake Forest, IL). The Influx™ is equipped with FACST™ Software (BD Biosciences) providing users a flexible interface to optimize analysis settings, create sort gates, and collect data. Graphical interpretations and data analysis are conducted in FlowJo (Tree Star Inc, Ashland, USA).

Environmental samples may contain organisms or particles that are larger than the nozzle diameter. Pre-filtration of samples with a pore-size smaller than the nozzle diameter (Partec Celltrics, Swedesboro, NJ, USA) will reduce risk of clogging the sample line or nozzle opening. The maximum sort rate of the Influx™ is 40,000 events per second (Ibrahim & van den Engh, 2007). For environmental samples, diluting the sample to an event rate of less than 1,000 events per second will reduce coincident drop occupancy and enhance sorting efficiency. When sorting single cells, additional dilutions or sorting target cell populations twice can help to guarantee pure single-event sorts (Rodrigue, et al., 2009). Setting the sort mode drop count (1.0 Drop, 2.0 Drop, or 1.5 Drop) and sort mode attributes (Enrich, Single, Pure, Yield, or Recovery) allow choices in tradeoffs between sorting efficiency and purity (refer to the Influx™ user manual).

2.4 Row by column sorting to locate the target symbiont population

Most environmental samples contain numerous populations of microbes in addition to the target symbiotic organisms. In order to proceed efficiently with DNA amplification and sequencing after sorting, it is necessary to identify a cell population that is enriched in the target symbionts. Identification of the best sort region is accomplished by screening sorted populations

of unknown cells with a quantitative assay specific for one or both of the symbionts (Zehr, et al., 2008). The sort region determined to contain the target organism(s) is a source of template for MDA or nested PCR.

Sort gates are created as a matrix of overlapping rows and columns that encompass all cell populations in the sample (Figure 2). Cell populations known not to contain the target symbionts may be excluded from the sort gates to enrich sorts of target symbionts. For example, the strong phycoerythrin signal of the marine cyanobacterium *Synechococcus* can be used to exclude these cells from sorts.

Efficient and quantitative screening of row and column sorts is best carried out by a phylogenetically-selective assay such as quantitative real-time PCR (qPCR) targeting one or both symbionts. The qPCR assay is applied to whole cells so sample losses that accompany DNA extraction are prevented. Preparation of all reagents, tubes, and plates is conducted in a UV-treated PCR workstation to prevent contamination. Before sorting, fill wells of qPCR tubes or plates with 10 μ l of nuclease-free water. Initially, sort populations of 500 cells per gate into each well, in replicate, and screen alongside standards and no template controls (NTC) in a real-time thermal cycler. The area of overlap between positive (amplification of the target gene by qPCR) rows and columns pinpoints the sort region that contains the most target symbiont (Figure 2).

3. MDA and metagenomic sequencing of uncultured symbiont populations

Genomic analysis can be instrumental in determining the metabolism of uncultivated symbionts (Tripp, et al., 2010). However, genome sizes of symbiotic partners can be vastly different. Especially in symbioses between bacteria (smaller genomes) and microbial eukaryotes, whose genomes can be much larger or present in multiple copy numbers, it is advantageous to

isolate the one partner in an enriched sample to ensure adequate genome coverage following sequencing. Whole-genome amplification via MDA is effective at amplifying even very small amounts of DNA from sorted cells for next-generation sequencing.

The enzyme Phi29 and kits for MDA are commercially available through several companies. Previous work for sequencing symbiont populations used the GenomiPhi V2 DNA amplification kit (GE Healthcare Life Sciences) (Zehr, et al., 2008) and more recent studies performing MDA on single cells or cell populations have used the Qiagen RepliG Mini or Midi kit (Reyes-Prieto, et al., 2010; Yoon, et al., 2011; Bench, et al., *In press*) or RepliPhi Phi29 DNA polymerase (Epicentre Biotechnologies) (Rodrigue, et al., 2009) with success. Studies comparing a variety of WGA methods found that RepliG produced the highest yields of DNA with the least amplification bias across the genomes (Pinard, et al., 2006; Pan, et al., 2008). In MDA reactions with very little template DNA a significant concern is the production of template-independent products (TIPs), which can contaminate sequencing processes and confound bioinformatic analyses. To avoid TIP contamination start with larger numbers of cells (e.g. >5,000 cells per reaction) when possible and limit the MDA reaction incubation time (Pan, et al., 2008). The number of sorted cells will depend on their abundance in the sample, the expected diversity of the target population, difficulty of cell lysis, or numbers of non-target contaminating cells in the sort gate. Amplification of genomic DNA from very few (1 to 100) *Prochlorococcus* cells reached a maximum product yield within 5 hours of incubation, with evidence of TIP production after 6 hours (Rodrigue, et al., 2009). Based on those results, the protocol below keeps MDA incubation times at 5.75 hours.

3.1 Sampling to separate symbiotic partners

Separation of symbiotic partners depends on the physical strength of their attachment. For delicate marine associations symbiotic partners are separated by filtration to concentrate the sample, vortex mixing, and a freeze/thaw cycle. 100-fold concentrated samples are created by filtering 1 L of seawater onto Type GV Durapore filters (pore size depends on expected size of target symbionts) with 10 psi vacuum pressure. The damp filter is immediately placed in a 15 mL polypropylene tube with 10 mL of sterile (0.22 μm filtered) seawater with the cell-covered side submerged and facing the center of the tube. Vortexing resuspends the cells from the filter. The concentrated sample is frozen in liquid nitrogen and stored at -80°C .

Prior to FACS analysis, the cell concentrate is thawed on ice then mixed again by vortexing. The freeze-thaw cycle should disrupt any delicate intercellular connections and leave the one symbiont intact as in a study on symbiotic cyanobacteria (Zehr, et al., 2008; Thompson, et al., 2012). Once thawed, the concentrated sample is used as soon as possible, and certainly within a few hours.

3.2 Performing MDA on sorted cells

From the concentrated sample, sorts of the target cell population are directed into 100 μl aliquots of 1X Tris-EDTA buffer (10mM Tris, 1 mM EDTA, pH 8.0), in replicate, in sterile microfuge tubes. Briefly spin the tubes to collect any cells that may have landed on the tube walls and store at -80°C until MDA.

A modified version of a user-developed protocol for MDA using the Qiagen RepliG Midi DNA amplification kit is appropriate for starting with whole cells (Bench, et al., *In press*) and is as follows:

The 1.5mL tubes used to receive sorted cells are marked where the pellet will collect then centrifuged at 14,000 rpm (20,800 x g) for 40 minutes. Following centrifugation, use a fine pipette tip to remove supernatant, while avoiding the pellet, which may not be visible. Vortex or flick tubes vigorously to resuspend the cell pellet in any remaining liquid.

To each sample add 2.5 µl of PBS and mix well. Then add 3.5 µl of Qiagen Buffer D2 (2.5 µl 1mM DTT and 27.5 µl Buffer DLB) and mix well. Incubate the samples at 65 °C for 5 minutes then place the samples on ice and add 3.5 µl of Qiagen Stop buffer. Next, prepare the reaction master mix in sterile UV-treated tubes. For one reaction, mix 10 µl RT-PCR grade H₂O (Ambion, Life Technologies), 1 µl RepliG Midi DNA Polymerase and 29 µl RepliG Midi Reaction Buffer. Combine one aliquot of the master mix (40 µl) with the cell lysis mix in a sterile PCR tube and incubate at 30 °C for 5-16 hours (5.75 hours is recommended, see above) in a thermal cycler with a heated lid. Following isothermal amplification, heat samples to 65 °C for 5 minutes, cool to 4 °C, then store at -20°C. Positive and negative control reactions are amplified alongside sorted cells. Prior to submission for next-generation sequencing, amplified genomic DNA should be quantified using Pico Green (Invitrogen Corporation, Carlsbad, CA) and checked for quality and fragment size distribution using the Agilent 2100 Bioanalyzer (Agilent Technologies, Inc. Santa Clara, CA) DNA 7500 or 12000 chip.

3.4 Bioinformatic analysis of symbiont genomes from FACS and MDA

A wide variety of bioinformatic tools are available for assembly and annotation of genomic data generated using next-generation sequence data. These tools are continually evolving and the choice of tool or pipeline is often determined by the user's experience and preferences. For sequence data from MDA product it is important to screen out TIP or

contaminant sequences. However, analyzing all the individual reads produced by current technologies requires significant computational time and processor power. Assembly can condense millions of reads into a manageable number of contigs. Pyrosequencing (i.e. 454) data can be assembled using the GS Data Analysis Software package (<http://www.454.com/products/analysis-software/>), but the best program for *de novo* assembly of Illumina data is debated (for detailed discussion of assemblers see (Paszkiewicz & Studholme, 2010; W. Zhang, et al., 2011)). The contigs generated by assembly can be screened using algorithms that identify open reading frames (ORFs) and by sequence similarity searches (e.g. BLAST). TIP data will be apparent as random sequences that do not have ORFs, and will usually have no similarity to sequences in public databases. If DNA was amplified from a sample that had contaminating cells (e.g. if other species are present in the sort gate containing the target symbionts), contigs can be further screened by comparing them to genomes of closely related species. Automated identification of ORFs and corresponding functional and taxonomic annotation for microbial genomes (and metagenomes) can be obtained by submitting contigs to online services. These include RAST (Aziz, et al., 2008) and the JGI IMG system (Markowitz, et al., 2012), both of which were useful for annotating the genome of a cyanobacterial symbiont from FACS and MDA (Zehr, et al., 2008; Tripp, et al., 2010). Once a genome has been assembled and annotated, the subsequent analyses will often include detailed phylogenetic comparisons and determination of metabolic capabilities. Pathways that are novel or unexpectedly lacking can be particularly meaningful for microbes suspected of existing in symbioses.

4. Nested PCR approach to analyze 18S and 16S rRNA genes from single cells

Analysis of individual symbiotic associations is critical to unequivocally link the identity of an uncultivated symbiont to its host or symbiotic partner. Furthermore, relatively high-throughput analysis of single associations is essential to explore the species specificity of the association and this can be accomplished with nested PCR using universal primers for phylogenetically-informative genes.

The physical connection between cells engaged in symbiosis can vary in its tolerance to sample handling. Preventing disruption of intercellular connections by turbulence, freezing, or by chemical treatment is essential to exploring intact associations. For open ocean studies, the InfluxTM can be equipped for the field (Thompson, et al., 2012; Worden, et al., 2012), which is required for analyzing untreated open ocean samples. The target symbiont can be located among cell populations from the untreated samples with the same sort strategy described in Section 3.

4.1 Sorting single cells and screening by qPCR

Important considerations in applying nested PCR with universal primers to single cells include: 1) one event is sorted into each well, 2) contaminating DNA is minimized, 3) the sample matrix (i.e. seawater) does not inhibit the PCR, and 4) target cells readily lyse following freeze/thaw cycles or during the first few cycles of PCR.

Aliquots of 10 μ l (sorted events and NTCs) or 8 μ l (standards) of nuclease-free water are added to wells as in Figure 3. Single events, either single cells or multiple cells in association (i.e. the symbiont-host complex), are sorted individually into 72 wells of a 96-well PCR plate. No events are sorted into wells for standards or NTCs (Figure 3). After sorting, cover plates with sterile sealing foil, briefly centrifuge, then store at -80 °C or -20 °C until screening by qPCR.

For qPCR, thaw plates and briefly centrifuge. Prepare master mix for qPCR and aliquot

15 µl to all wells. Carefully add 2 µl of standards, in replicate, to designated wells. Perform 45 cycles of amplification and analyze the results according to standard qPCR procedures. Move the entire volume (25 µl) of positive (amplification of approximately one gene copy) reactions to new sterile PCR tubes and store at -20 °C before nested PCR.

4.2 Amplification of 16S and 18S rRNA by PCR from positive qPCRs using universal primers

Nested PCR using universal primers for 18S and 16S rRNA genes is applied to the positive reactions from the qPCR assay to determine the phylogeny of the symbiotic partners. The nested PCR is useful for amplifying genes present in low abundance. Product from one round of PCR using "outer primers" to amplify a large fragment of the rRNA gene is used as template in a second round of PCR that targets a smaller region of the amplicon using "inner primers".

For nested PCR, use a high performance polymerase mixture such as TaKaRa Ex Taq (Takara Bio, Inc.) to ensure amplification if targets are difficult to amplify. For the first round of nested PCR use the outer primers EukA/B (Medlin, et al., 1988) and Eub27F/Eub1492R (Weisburg, et al., 1991) for amplification of 18S and 16S rRNA genes, respectively. 75 µl of PCR master mix should be added directly to the saved reaction from the qPCR assay (25 µl) and amplified for 35 cycles alongside positive and negative controls. Store completed outer primer reactions at -20 °C or immediately use 1 µl as template in 25 µl reactions for the second round of nested PCR with inner primers. Use the internal primers Euk18S-555F/1269R (López-García, et al., 2003) and 358F/907R (Lane, 1991) for the 18S and 16S rRNA reactions, respectively.

4.3 Sequencing rRNA genes from nested PCR and phylogenetic analysis

PCR products from the inner primers are analyzed by gel electrophoresis, any bands are cut from the gel, extracted (QIAquick Gel Extraction Kit), and cloned using kits such as pGEM-T (Promega, Madison, WI) or TOPO-TA (Life Technologies). The clones can be prepared by plasmid mini-prep and sequenced. Quality control and generation of contigs for identical or very similar sequences (>99%) can be generated in Sequencher (Gene Codes Corporation, Ann Arbor, MI USA) or similar analysis software. Phylogenetic analysis and alignment of the sequences can be conducted through BLASTn to GenBank (Benson, et al., 2013) or through other tools such as the Ribosomal Database Project (Wang, et al., 2007), SILVA (Quast, et al., 2013), or SINA (Pruesse, et al., 2012). Negative controls often contain amplification products after the qPCR and nested PCR procedures, so it is important to work under as clean conditions as possible for highest efficiency of sequencing nested PCR products (Thompson, et al., 2012).

Figure legends.

Figure 1. Strategy to characterize uncultivated microbial symbionts using FACS coupled to MDA for genomic analysis and to nested PCR for identification of unknown partner cells to suspected symbionts. Dashed line represents strategies for symbiont characterization that be effective following identification of all symbiotic partners.

Figure 2. Flow cytometry density plot of 47,000 events from a coastal seawater sample showing multiple populations of unknown phytoplankton. Darker shading indicates higher numbers of cells. Sort gates of rows (R) and columns (C) are positioned to cover all phytoplankton populations. The sorts from each row and column are screened with a qPCR assay specific to one symbiont. In this example, the qPCR screen was positive (bold type) for cells from gates C3 and

R1, meaning the target population is present where these gates overlap (large black box). With this information, additional sorts of cells from two smaller gates (r1 and r2) are conducted and the target symbiont population is pinpointed in r2. Standard 3 μm diameter beads (b) are included for internal reference.

Figure 3. Strategy for application of nested PCR, with phylogenetically-informative universal primers, to single-event sorts from untreated samples to determine the identity of uncultivated symbionts.

References.

- Aziz, R., Bartels, D., Best, A., DeJongh, M., Disz, T., Edwards, R., et al. (2008). The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics*, 9(1), 75.
- Bench, S. R., Heller, P., Frank, I., Arciniega, M., Shilova, I. N., & Zehr, J. P. (*In press*). Whole genome comparison of six *Crocospaera watsonii* strains with differing phenotypes. *Journal of Phycology*.
- Benson, D., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D., Ostell, J., et al. (2013). GenBank. *Nucleic Acids Res.*, 41(Database issue), D26-42.
- Blainey, P. C., Mosier, A. C., Potanina, A., Francis, C. A., & Quake, S. R. (2011). Genome of a Low-Salinity Ammonia-Oxidizing Archaeon Determined by Single-Cell and Metagenomic Analysis. [doi:10.1371/journal.pone.0016626]. *PLoS ONE*, 6(2), e16626.
- Brehm-Stecher, B. F., & Johnson, E. A. (2004). Single-Cell Microbiology: Tools, Technologies, and Applications. *Microbiology and Molecular Biology Reviews*, 68(3), 538-559.
- Cheung, V. G., & Nelson, S. F. (1996). Whole genome amplification using a degenerate oligonucleotide primer allows hundreds of genotypes to be performed on less than one nanogram of genomic DNA. *Proceedings of the National Academy of Sciences*, 93(25), 14676-14679.
- Cuvelier, M. L., Allen, A. E., Monier, A., McCrow, J. P., Messié, M., Tringe, S. G., et al. (2010). Targeted metagenomics and ecology of globally important uncultured eukaryotic phytoplankton. *Proceedings of the National Academy of Sciences*, 107(33), 14679-14684.
- Dean, F., Nelson, J., Giesler, T., & Lasken, R. (2001). Rapid amplification of plasmid and phage DNA using Phi29 polymerase and a multiply-pimed rolling circle amplification. *Genome Research*, 11, 1095 - 1099.
- Del Giorgio, P. A., Bird, D. F., Prairie, Y. T., & Planas, D. (1996). *Flow cytometric determination of bacterial abundance in lake plankton with the green nucleic acid stain SYTO 13* (Vol. 41). Waco, TX, ETATS-UNIS: American Society of Limnology and Oceanography.
- Hongoh, Y., Sharma, V. K., Prakash, T., Noda, S., Taylor, T. D., Kudo, T., et al. (2008). Complete genome of the uncultured Termite Group 1 bacteria in a single host protist cell. *Proceedings of the National Academy of Sciences*, 105(14), 5555-5560.
- Huang, W. E., Ward, A. D., & Whiteley, A. S. (2009). Raman tweezers sorting of single microbial cells. *Environmental Microbiology Reports*, 1(1), 44-49.

- Ibrahim, S. F., & van den Engh, G. (2003). High-speed cell sorting: fundamentals and recent advances. *Current Opinion in Biotechnology*, *14*(1), 5-12.
- Ibrahim, S. F., & van den Engh, G. (2007). Flow Cytometry and Cell Sorting. *Adv Biochem Engin/Biotechnol*, *106*, 19-39.
- Kvist, T., Ahring, B., Lasken, R., & Westermann, P. (2007). Specific single-cell isolation and genomic amplification of uncultured microorganisms. *Applied Microbiology and Biotechnology*, *74*(4), 926-935.
- Lane, D. J. (1991). 16S/23S rRNA sequencing. In Stackebrandt & Goodfellow (Eds.), *Nucleic Acid Techniques in Bacterial Systematics* (pp. 115-175). New York: John Wiley and Sons.
- López-García, P., Philippe, H., Gail, F., & Moreira, D. (2003). Autochthonous eukaryotic diversity in hydrothermal sediment and experimental microcolonizers at the Mid-Atlantic Ridge. *Proceedings of the National Academy of Sciences*, *100*(2), 697-702.
- Marcy, Y., Ouverney, C., Bik, E. M., Lösekann, T., Ivanova, N., Martin, H. G., et al. (2007). Dissecting biological “dark matter” with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. *Proceedings of the National Academy of Sciences*, *104*(29), 11889-11894.
- Markowitz, V. M., Chen, I.-M. A., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., et al. (2012). IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic Acids Research*, *40*(D1), D115-D122.
- Medlin, L., Elwood, H. J., Stickel, S., & Sogin, M. L. (1988). The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. [doi: 10.1016/0378-1119(88)90066-2]. *Gene*, *71*(2), 491-499.
- Navin, N., Kendall, J., Troge, J., Andrews, P., Rodgers, L., McIndoo, J., et al. (2011). Tumour evolution inferred by single-cell sequencing. [10.1038/nature09807]. *Nature*, *472*(7341), 90-94.
- Orcutt, K. M., Gundersen, K., Wells, M. L., Poulton, N. J., Sieracki, M. E., & Smith, G. J. (2008). Lighting up phytoplankton cells with quantum dots. *Limnol. Oceanogr. Methods*, 653-658.
- Orphan, V. J., House, C. H., Hinrichs, K.-U., McKeegan, K. D., & DeLong, E. F. (2001). Methane-Consuming Archaea Revealed by Directly Coupled Isotopic and Phylogenetic Analysis. *Science*, *293*(5529), 484-487.
- Pan, X., Urban, A. E., Palejev, D., Schulz, V., Grubert, F., Hu, Y., et al. (2008). A procedure for highly specific, sensitive, and unbiased whole-genome amplification. *Proceedings of the National Academy of Sciences*, *105*(40), 15499-15504.
- Paszkiwicz, K., & Studholme, D. J. (2010). De novo assembly of short sequence reads. *Briefings in Bioinformatics*, *11*(5), 457-472.
- Pinard, R., de Winter, A., Sarkis, G., Gerstein, M., Tartaro, K., Plant, R., et al. (2006). Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing. *BMC Genomics*, *7*(1), 216.
- Pruesse, E., Peplies, J., & Glöckner, F. O. (2012). SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics*, *28*(14), 1823-1829.
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, *41*(D1), D590-D596.
- Reyes-Prieto, A., Yoon, H. S., Moustafa, A., Yang, E. C., Andersen, R. A., Boo, S. M., et al. (2010). Differential Gene Retention in Plastids of Common Recent Origin. *Molecular Biology and Evolution*, *27*(7), 1530-1537.
- Rodrigue, S., Malmstrom, R. R., Berlin, A. M., Birren, B. W., Henn, M. R., & Chisholm, S. W. (2009). Whole genome amplification and *de novo* assembly of single bacterial cells. [doi:10.1371/journal.pone.0006864]. *PLoS ONE*, *4*(9), e6864.
- Stepanuskas, R., & Sieracki, M. E. (2007). Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. *Proceedings of the National Academy of Sciences*, *104*(21), 9052-9057.

- Telenius, H. k., Carter, N. P., Bebb, C. E., Nordenskjöld, M., Ponder, B. A. J., & Tunnacliffe, A. (1992). Degenerate oligonucleotide-primed PCR: General amplification of target DNA by a single degenerate primer. *Genomics*, *13*(3), 718-725.
- Thompson, A. W., Foster, R. A., Krupke, A., Carter, B. J., Musat, N., Vaultot, D., et al. (2012). Unicellular Cyanobacterium Symbiotic with a Single-Celled Eukaryotic Alga. *Science*, *337*(6101), 1546-1550.
- Tripp, H. J., Bench, S. R., Turk, K. A., Foster, R. A., Desany, B. A., Niazi, F., et al. (2010). Metabolic streamlining in an open-ocean nitrogen-fixing cyanobacterium. [10.1038/nature08786]. *Nature*, *464*(7285), 90-94.
- Vaultot, D., Lepère, C., Toulza, E., Iglesia, R. D. I., Poulain, J., Gaboyer, F., et al. (2012). Metagenomes of the picoalga *Bathycoccus* from the Chile coastal upwelling. *PLoS ONE*, *7*(6), e39648.
- Wang, Q., Garrity, G., Tiedje, J., & Cole, J. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new. *Appl Environ Microbiol.*, *73*(16), 5261-5267.
- Weisburg, W. G., Barns, S. M., Pelletier, D. A., & Lane, D. J. (1991). 16S ribosomal DNA amplification for phylogenetic study. *Journal of Bacteriology*, *173*(2), 697-703.
- Worden, A. Z., Janouskovec, J., McRose, D., Engman, A., Welsh, R. M., Malfatti, S., et al. (2012). Global distribution of a wild alga revealed by targeted metagenomics. *Current Biology*, *22*(17), R675-R677.
- Yoon, H. S., Price, D. C., Stepanauskas, R., Rajah, V. D., Sieracki, M. E., Wilson, W. H., et al. (2011). Single-Cell Genomics Reveals Organismal Interactions in Uncultivated Marine Protists. *Science*, *332*(6030), 714-717.
- Zehr, J. P., Bench, S. R., Carter, B. J., Hewson, I., Niazi, F., Shi, T., et al. (2008). Globally Distributed Uncultivated Oceanic N₂-Fixing Cyanobacteria Lack Oxygenic Photosystem II. *Science*, *322*(5904), 1110-1112.
- Zhang, K., Martiny, A. C., Reppas, N. B., Barry, K. W., Malek, J., Chisholm, S. W., et al. (2006). Sequencing genomes from single cells by polymerase cloning. [10.1038/nbt1214]. *Nat Biotech*, *24*(6), 680-686.
- Zhang, L., Cui, X., Schmitt, K., Hubert, R., Navidi, W., & Arnheim, N. (1992). Whole genome amplification from a single cell: implications for genetic analysis. *Proceedings of the National Academy of Sciences*, *89*(13), 5847-5851.
- Zhang, W., Chen, J., Yang, Y., Tang, Y., Shang, J., & Shen, B. (2011). A Practical Comparison of *De Novo* Genome Assembly Software Tools for Next-Generation Sequencing Technologies. [doi:10.1371/journal.pone.0017915]. *PLoS ONE*, *6*(3), e17915.

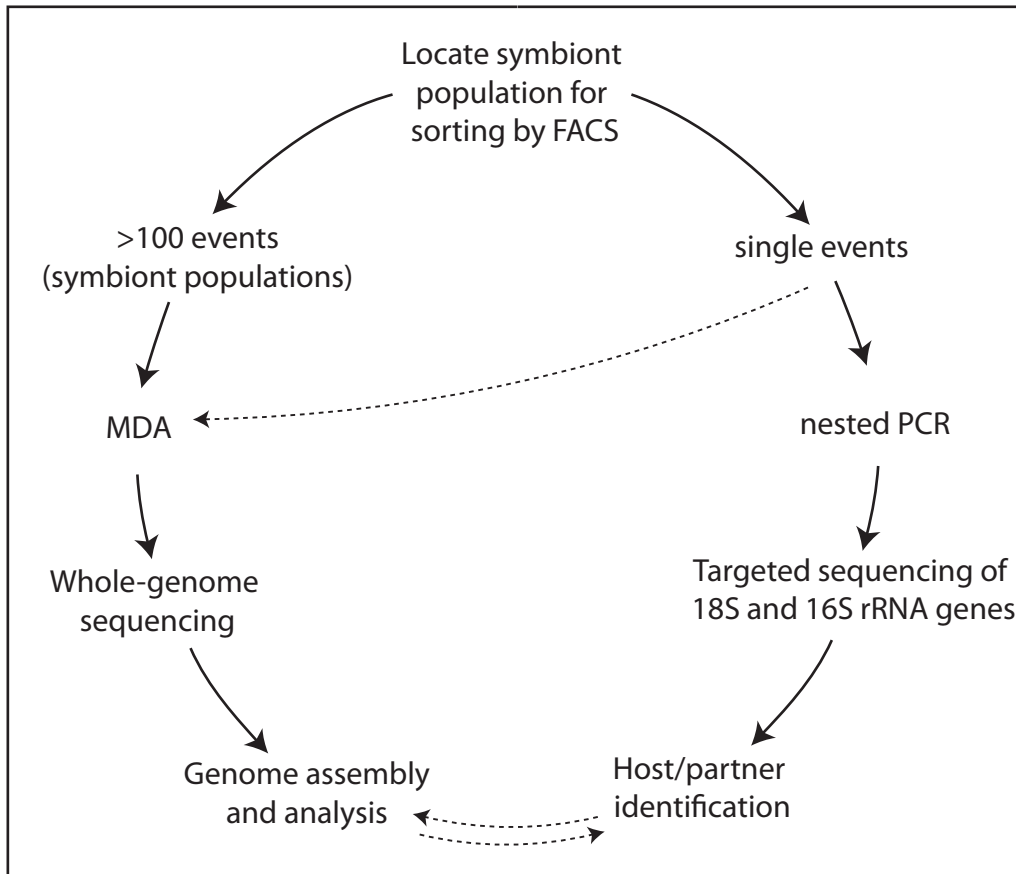


Figure 1. Strategy to characterize uncultivated microbial symbionts using FACS coupled to MDA for genomic analysis and to nested PCR for identification of unknown partner cells to suspected symbionts. Dashed line represents strategies for symbiont characterization that would be effective following identification of all symbiotic partners.

Figure 2

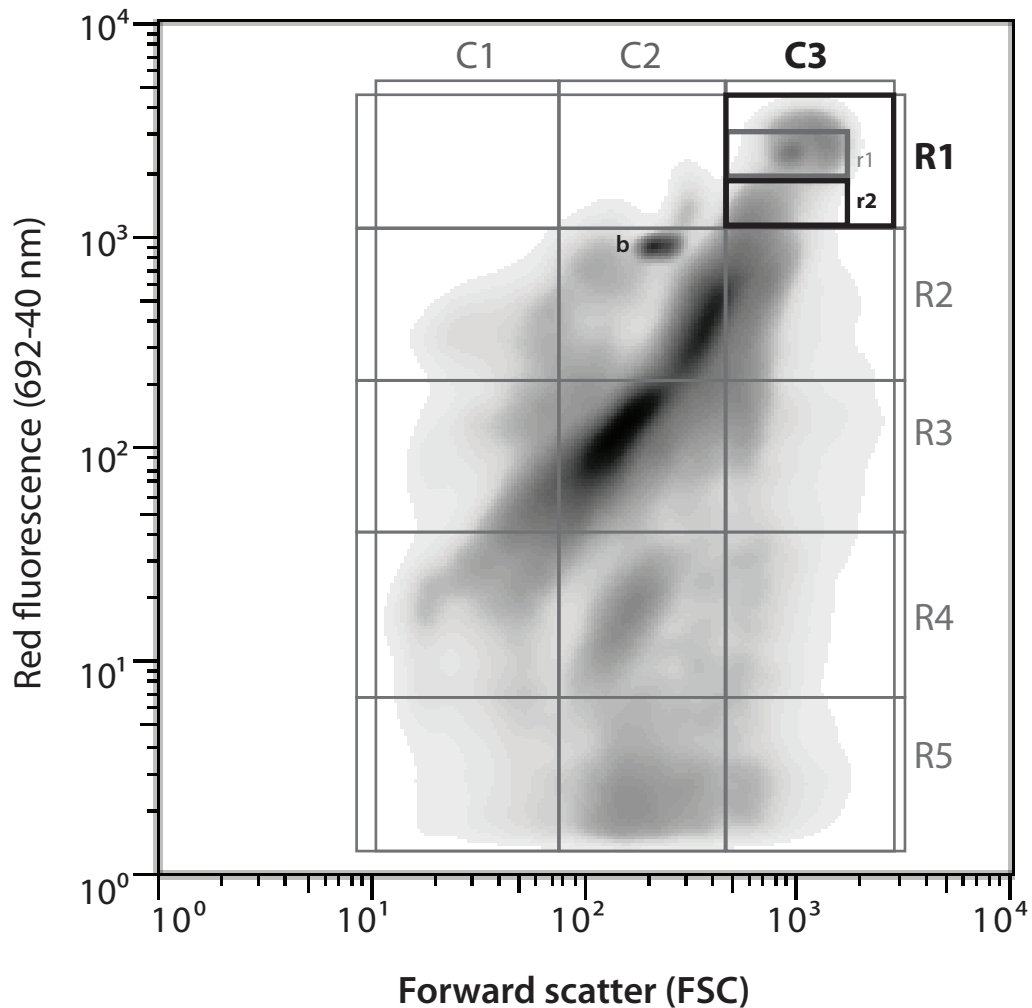


Figure 2. Flow cytometry density plot of 47,000 events from a coastal seawater sample showing multiple populations of unknown phytoplankton. Darker shading indicates higher numbers of cells. Sort gates of rows (R) and columns (C) are positioned to cover all phytoplankton populations. The sorts from each row and column are screened with a qPCR assay specific to one symbiont. In this example, the qPCR screen was positive (bold type) for cells from gates C3 and R1, meaning the target population is present where these gates overlap (large black box). With this information, additional sorts of cells from two smaller gates (r1 and r2) are conducted and the target symbiont population is pinpointed in r2. Standard 3 μm diameter beads (B) are included for internal reference.

Figure 3

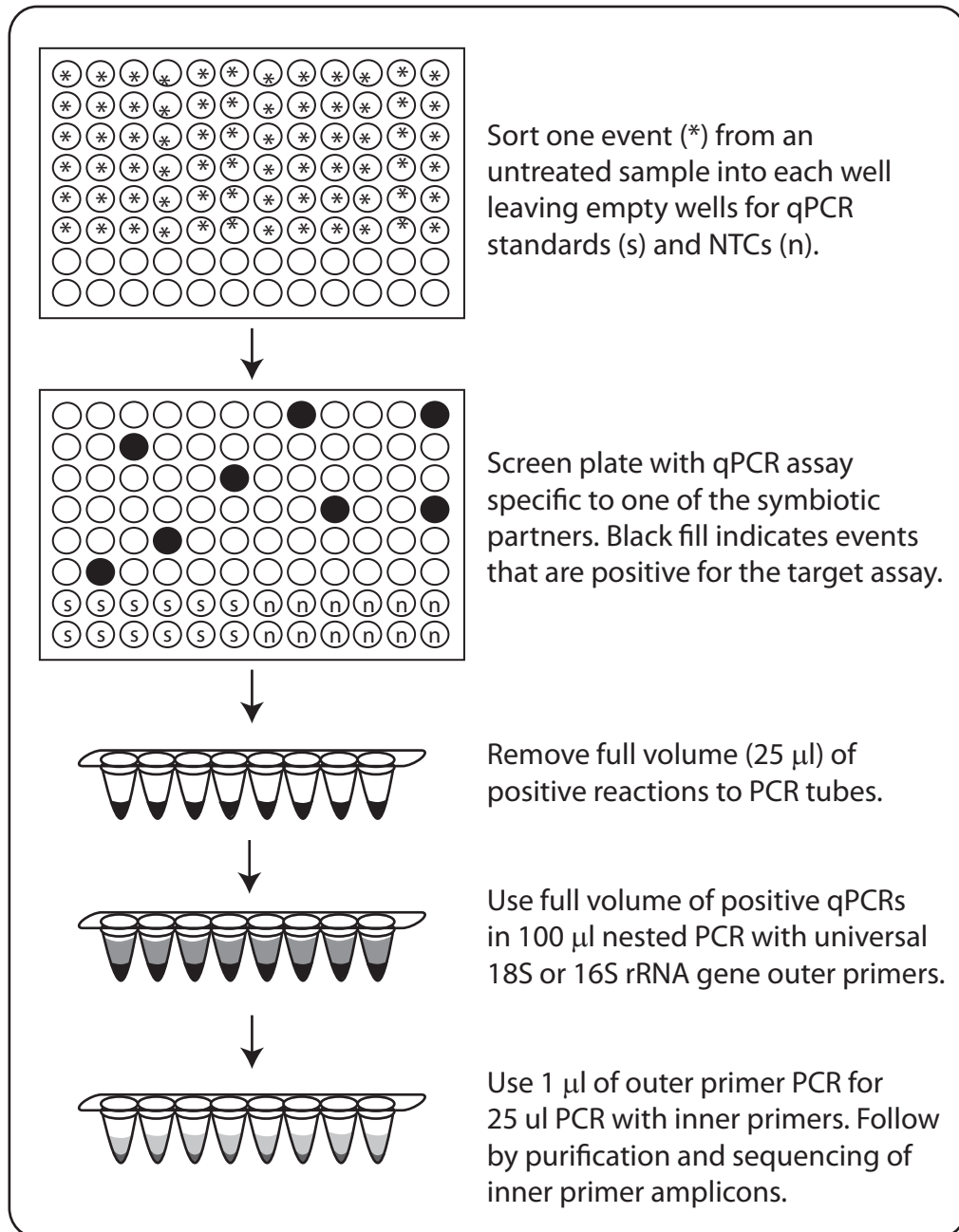


Figure 3. Strategy for application of nested PCR with phylogenetically-informative primers to single-event sorts of untreated sample to determine the identity of uncultivated symbionts.