

Single-Cell Technologies in Environmental Omics Studies

Rimantas Kodzius^{a,c,d,*}, Takashi Gojobori^{a,b,d}

^a Computational Bioscience Research Center (CBRC), Saudi Arabia

^b Biological and Environmental Sciences and Engineering Division (BESE), Saudi Arabia

^c Computer, Electrical and Mathematical Sciences and Engineering Division (CEMSE), Saudi Arabia

^d King Abdullah University of Science and Technology (KAUST), Saudi Arabia

* Corresponding author at: King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia.

E-mail addresses: Rimantas.Kodzius@kaust.edu.sa (R. Kodzius),

Takashi.Gojobori@kaust.edu.sa (T. Gojobori).

Abstract

Environmental studies are primarily done by culturing isolated microorganisms or by amplifying and sequencing conserved genes. Difficulties understanding the complexity of large numbers of various microorganisms in an environment led to the development of techniques to enrich specific microorganisms for upstream analysis, ultimately leading to single-cell isolation and analyses. We discuss the significance of single-cell technologies in omics studies with focus on metagenomics and metatranscriptomics. We propose that by reducing sample heterogeneity using single-cell genomics, metaomic studies can be simplified.

Keywords (max 6)

Metagenomics, Heterogeneity, Sequencing, Microdroplet, FACS

Introduction

Emerging technologies in metagenomics and single-cell genomics (SCG) are paving the way to new methods of analyzing and understanding our environment. Data obtained from sequencers are a useful source of information for the discovery of new species, pathways, and genes. Furthermore, with this new information, producer genomes can be modified by genetic engineering.

The emergence of metagenomic research

Early environmental studies of organism diversity required that samples be cultured to increase DNA quantity before genomic libraries could be constructed. At the time, sequencing of ribosomal RNA (rRNA), mainly 16S and 18S rRNA (Lane et al., 1985; Pace et al., 1986; Field et al., 1988), was the best option available for identifying environmental diversity. However, early on it became clear that uncultivable species constitute the majority of sample communities (Hugenholtz et al., 1998). Fortunately, polymerase chain reaction (PCR) of rRNA enables access to both cultivable and uncultivable sample diversity. Because PCR can be performed directly on environmental samples without cloning, it can be used to amplify genes of interest directly from the environment.

Now that researchers could extract nucleic acids from many organisms (cultivable and uncultivable) to construct and sequence their rRNA libraries, they began to quantify the variety of organisms in an environmental sample (Schmidt et al., 1991). Based on 16S rRNA sequencing, an organism's taxonomy was assessed and phylogenetic trees were constructed, allowing for the identification and comparison of microorganisms present in a given sample.

Interest in sequencing both noncoding rRNA and protein coding genes is growing, with specific interest in reconstructing bacterial genomes in bulk. Shotgun metagenomic sequencing allows for a comprehensive sampling of all genes from all organisms existing in a complex sample, making it possible to evaluate bacterial diversity and to detect the abundance of microbes in a specific environment (Stein et al., 1996). Furthermore, it became possible to reconstitute complete genomes of uncultivable microorganisms (Hugenholtz, 2002). In addition to the discovery of new species of archaea, bacteria, protozoa, algae, fungi, and eukaryotes, it became viable to reconstruct viral genomes. This was an important advancement because viruses lack a shared universal phylogenetic marker such as bacterial 16S rRNA or eukaryotic 18S rRNA (Breitbart et al., 2002; Venter et al., 2004).

With the arrival of high-throughput sequencing, DNA could be sequenced directly without cloning. Typically, the complexity of the sample is estimated by creating 16S/18S rRNA libraries, which provide information about how deep the libraries need to be sequenced. Generally, more complex samples contain a greater diversity of sequences and, therefore, require deeper sequencing.

Rather than using genomic DNA (gDNA), the metatranscriptome approach involves the collection of the entire RNA community, which is used to construct cDNA libraries that are sequenced. Although RNA is less stable than DNA, it gives a snapshot of the current state of the community by revealing which genes are active and expressed. To date, marine water studies represent some of the most important metatranscriptome studies (Poretzky et al., 2009; Shi et al., 2009; McCarren et al., 2010; Mason et al., 2012).

Single-cell genomics

Metagenomic sequencing is a helpful tool for understanding environmental communities. However, assembling gene catalogues and composite genomes can be challenging. Furthermore, it is difficult to distinguish between genes that originate from the same or different organisms in genomes that are already assembled (Fig. 1). Therefore, there is considerable interest in developing a system to understand the organization of discovered genes and pathways within genomes.

Scaling-up and deep-sequencing are two approaches that only partially solve the challenge of understanding sample heterogeneity. Different approaches are necessary to compare metagenomic samples collected from different environments and to understand the content and organization of microorganism genomes in the environment.

Some attempts have been made to enrich the culture of a specific organism of interest on the basis of their function in the environment. This kind of targeted-metagenomic approach allows for the development of composite microbial genomes (Hess et al., 2011). In addition to the uncultivability of many organisms, other problems arise when faster growing microorganisms outgrow slower growing microorganisms, subsequently introducing additional biases (Wu et al., 2009).

Therefore, instead of culturing, a complex sample can be enriched for a target cell population using fluorescence-activated cell sorting (FACS) (Stein et al., 1996; Hallam et al., 2006; Bergquist et al., 2009), which sorts cells based on their shape, size, and density, for example (sorting

of features depends on fluorophores used) (Wallner et al., 1997; Sekar et al., 2004). While this approach enables the streamlining of downstream biocomputational elements from composite genomes, it is not guaranteed that the phenotypic features of mixed cells will represent functional couplings (organisms that are integrated networks of functional interactions).

Alternatively, specific organisms of interest can be enriched by processing a small number of cells that are relevant to a specific function in the environment (Podar et al., 2007). With the advancement of whole genome amplification (WGA) techniques (e.g., multiple displacement amplification (MDA) or rolling circle amplification (RCA)), the gDNA of a limited number of cells can be amplified and sequenced. However, WGA techniques introduce potential bias because the abundance of the specific sequences is distorted—all sequences are not equally amplified. Therefore, gDNA amplification of isolated populations with a limited number of cells compromises the quantitative analysis of metagenomics (Yilmaz et al., 2010).

Genome sequencing from individual cells provides direct access to fine-scale heterogeneity of complex microorganism populations (Pamp et al., 2012). One of the easiest ways to perform microbial genome sequencing is from the isolate, when a single cell is grown in corresponding medium and large quantities of high-quality gDNA can be extracted and libraries prepared. However, this method is only accessible with cultivable microbes, has a low throughput (only a few cells can be processed in this way due to limited media formulations), and long processing times that introduce bias from slow growth and possible mutations.

Recent developments in single-genome amplification techniques facilitate SCG using physical cell separation (usually by FACS on 96-well plates), cell lysis, and WGA. An amplified genome from a single cell can then be sequenced. While metagenomics is performed by collecting the microorganism community and extracting and sequencing total DNA, SCG separates individual cells from the microorganism community. Often, the sample quality of amplified genomes is assessed by 16S/18S rRNA sequencing to ensure that the cells have different genomes of interest (Woyke et al., 2009). Data from sequencing provides quantitative information on genomic variability in microorganism populations—gene insertions, losses, duplications, and genome rearrangements can be analyzed on a single-cell level, and complex metabolic pathways can be analyzed for an individual cell. Data from metagenomics and SCG are complementary: Metagenomic data contribute to SCG sequence assembly (Blainey et al., 2011) and SCG uncovers genomic information from individual and often uncultivable cells. Therefore, together they provide information about the metabolic potential and evolution of specific microorganisms. At present, however, SCG remains limited by high sequencing costs and a low throughput (96-well plates).

While currently SCG is considered only as a complement to cultivation and metagenomics (providing genomic information from individual and uncultivable cells), we predict that SCG will emerge as a stand-alone tool, delivering information about the organization of genes within genomes and providing clues to metabolic potential and evolutionary history. The future of SCG is promising, with new fields developing such as single-cell transcriptomics (Tang et al., 2009), metabolomics (Amantonico et al., 2010), viral-particle sequencing (Allen et al., 2011), and matched bioinformatics.

Single-cell isolation and processing

The population level provides a general understanding of the genome and its regulation by analyzing millions of cells at once. However, population-based averaged measurements do not allow for a view of the heterogeneity in biological systems, nor do they represent the current state of the population. Beyond 16S/18S RNA sequencing and metagenomics, SCG is gaining in popularity for studying both cultivable and especially uncultivable microorganisms. The process to obtain, multiply, and sequence genetic material from either live or dead cells is relatively simple – the cell needs to be isolated (captured) and lysed and then the released genetic material is amplified and the genomic library can be prepared and sequenced.

Targeted approach for single-cell isolation

Previous reviews provide detailed descriptions of the two different methods of separating and isolating single cells (Stepanauskas, 2012; Blainey, 2013; Yaron et al., 2014). Micromanipulation is a targeted approach: targeted cells are identified, isolated, and inspected microscopically. Suitable cells are transferred using a micropipette (Zhang et al., 2012b), laser source (Dobes et al., 2013), optical tweezers (Lanigan et al., 2008), or by real-time microfluidic flow (Skafte-Pedersen et al., 2012). Using microfluidics a desired phenotype correlating to a specific genome can be selected and live-single-cell resolution, and fluorescence compatibility are possible. In addition, cells can be easily observed before capture (Tadmor et al., 2011; Leung et al., 2012). WGA is advantageous because reaction volumes are very low and reaction efficiency is higher than with nonmicrofluidic approaches, providing better genome amplification coverage (Marcy et al., 2007).

These types of targeted approaches are gentle in a way that cells remain viable for further analysis. The disadvantages of a targeted approach are low throughput, low-parallelization potential, and the requirement for specialized technical knowledge.

Random encapsulation to isolate single cells

Random encapsulation is the second option for isolating cells. Here the cells are randomly selected by serial dilution (the simplest method), which involves diluting the cell culture until one cell is contained in a specific volume (partition) of medium (Zhang et al., 2006). The diluted liquid can then be deposited into microwells or the cells can be encapsulated by microdroplets (Thorsen et al., 2001; Raghunathan et al., 2005; Stepanauskas and Sieracki, 2007; Agresti et al., 2010; Zeng et al., 2010). Fluidigm has developed commercial microfluidic devices for automated handling of single cells in a single-phase system using multiple valves. The Fluidigm C1 Single-Cell Auto Prep System is often coupled with the BioMark HD System for gene expression analysis (Citri et al., 2012). Using this system, individual cells can be isolated, processed, and profiled for genomic analysis. However, Fluidigm commercial solutions are only suitable for use with larger cells (e.g., mammalian), and cannot be used for analyses of microorganisms. Additionally, because the maximum number of compartments on a chip made by valves is limited ($\sim 4 \times 10^4$) so is the throughput (the number of cells that can be analyzed). Although cells in Fluidigm are captured randomly, those captured in the chip can be inspected before continuing with any additional treatments.

Fluorescence-activated cell sorting

Flow cytometry is the most popular method of random encapsulation. FACS allows the selection of cells matching specific criteria, such as the size, shape, color, or even presence of

specific nucleic acids or activities, using fluorescent dyes (Raghunathan et al., 2005; Stepanauskas and Sieracki, 2007; Rodrigue et al., 2009; Woyke et al., 2009; Hess et al., 2011; Swan et al., 2011; Yoon et al., 2011; Dupont et al., 2012; Rinke et al., 2014). Among the many advantages of FACS are its high throughput, high sorting speed, and ability to sort live cells. In addition, FACS can detect the presence of cells in a droplet, reject empty droplets, and direct the cells with desired properties into multiwell/microtiter plates (Fig. 2). From here, cells can be lysed and genomes can be amplified in single wells. On the down side, FACS equipment is expensive, specialized technical expertise is required, individual cells cannot be visualized during the process, cell throughput is only in the hundreds (deposition into microwells), and the processing volume is high (both sample and reagent). Furthermore, cells are exposed to high electric charge and high pressure during FACS, which compresses them and makes them less viable. Algae cells, for example, usually require a longer growth time in medium following FACS procedures. In addition, cells are mixed in a one-phase solution; therefore, FACS can only distinguish cells that differ in intracellular or surface-associated fluorescence. To separate cells that differ in protein secretion, a technique such as two-phase microdroplet technology should be used.

Microdroplet technology

Microdroplets are composed of an emulsion that forms when cells in the water phase are vigorously mixed with oil; oil droplets encapsulate cells in the water phase. Originally, PCR performed on an emulsion was thought to be much more specific than analysis of cells in the water phase alone, yielding greater amounts of reaction products and forming fewer chimeric byproducts (Williams et al., 2006). Moreover, the droplets formed contained a limited number of template molecules (ideally only one).

Advances in microfluidics allow for control of droplet size to the nanoliter or even the picoliter, providing aqueous uniform (monodisperse) droplets in oil (Teh et al., 2008). In this publication we use the term microdroplets to describe all droplets smaller than a milliliter (e.g., nanoliter and picoliter sized). Droplets can be manipulated in a variety of ways: splitting the droplets, adding new reagents, incubating on-chip or off-chip, and sorting or breaking the droplets to recover the content (even viable cells can be released). Microdroplets can be produced in large numbers at a speed of thousands of droplets per second (or millions per hour) for a single channel (Zeng et al., 2010; Casadevall i Solvas et al., 2011) and both droplet formation and cell encapsulation can be seen using high-speed cameras. Droplets are sorted by laser excitation and fluorescence measurements according to specific indications such as presence or absence of a cells, dead or viable cells (using specific stains as indicators), or those containing a cell with a specific cell fluorescence marker.

Microdroplet technology, a subset of microfluidics, offers other advantages such as a high potential for parallelization and automation (controlled by a computer program such as LabView), gentle cell handling, low potential for contamination, and low sample and reagent consumption. Physical phenomena such as diffusion and heat transfer take place quickly in small volumes, speeding up the whole process. A small droplet volume facilitates a high sample concentration in the droplet (Zhong et al., 2008), and more efficient PCR and MDA reactions (Marcy et al., 2007). Microdroplet technology makes it possible to reach high concentrations of cell-secreted molecules, enabling rapid detection of droplets that contain cells producing molecules of interest (Koster et al.,

2008; Brouzes et al., 2009). Furthermore, microdroplet technology allows transcription reaction to convert small amounts of mRNA to cDNA more efficiently, reaching 54% compared with conventional methods in tubes, which yield only 12% (Zhong et al., 2008). Finally, single beads with functionalized surface or attached PCR primers can be introduced to the droplets to facilitate various reactions inside the microdroplet (Kumaresan et al., 2008).

One of the main advantages of the two-phase system of microdroplets over FACS is the ability to continue to handle single cells as separated units. The droplet acts as an independent well or reaction chamber where cells can be grown and later screened for expressed products. Released gDNA can be efficiently amplified in microdroplets and analyzed by biochemical and genetic assays or used for sequencing-library preparation. Analyses show that genotype and phenotype are through compartmentalization such that the cells and their secreted molecules are contained within droplets, even during sorting and analyzing (Fig. 2). These encapsulated cells can be lysed and intracellular molecules can be analyzed. A droplet content readout can be performed using fluorescence (such as fluorescence polarization, fluorescence resonance energy transfer (FRET), or Raman or Mass spectroscopy).

Bacterial, yeast, plant, insect, and mammalian cells can be analyzed using microdroplet technology on agarose (Zhang et al., 2012a) or other microgels (Rossow et al., 2012). Even multicellular organisms can be encapsulated and grown in a droplet (Clausell-Tormos et al., 2008). Theoretically, it is possible to measure any secreted molecule for which a fluorescently labeled ligand exists such as insulin, growth factors, and cytokines. The activities of various cell proteins can be monitored – cell surface, intracellular, or secreted proteins. The very high-throughput nature of the system makes it well suited for performing high-throughput screening (Brouzes et al., 2009) as is required for screening for antibodies or other molecules that inhibit enzyme activity. Directed enzyme evolution is a key area for the future focus of microdroplet technology. Using fluorescent substrates, cells can be regrown and enzymes with improved function and stability can be expressed.

By combining the strengths of microdroplet technology—capturing single cells and cultivating them in gel microdroplets and then sorting the single microcolonies using FACS—the complete genome from a single bacterial cell can be assembled (Dichosa et al., 2014).

Since the establishment of the two-phase droplet-based microfluidic systems at the laboratories of Prof. Andrew D. Griffiths (Clausell-Tormos et al., 2008; Fallah-Araghi et al., 2012) and Prof. David A. Weitz (Clausell-Tormos et al., 2008; Agresti et al., 2010), many other academic and industrial laboratories are establishing similar technology. This is not surprising considering the advantages of microdroplets, especially in high-throughput screening. There are already commercial solutions using microdroplets for quantitative absolute real-time PCR (such as Bio-Rad's QX200 Droplet Digital PCR System, Life Technologies QuantStudio 3D Digital PCR System or RainDrop Digital PCR System), and the opportunities for microdroplet technology continue to expand. Although there is no complete commercial solution currently available for cell screening using microdroplets, a customized system can be built as demonstrated by Mazutis et al. (Mazutis et al., 2013). As microdroplet technology continues to advance, we can expect more and more sequenced single-cell genomes deposited into data banks. For example, onboard their boat, during their research expedition, (Lim et al., 2014) sequenced and assembled 26 marine microbial genomes and two

marine microbial metagenomes. Recent studies have already sequenced, analyzed, and stored thousands of single-cell genomes in data banks (pers. comm.).

Platforms for studying various cells

Living cells can roughly be classified into bacteria, archaea, and eukaryotes. Although viruses are not considered “living,” they constitute an important part of analyses. Various single-cell analysis platforms are suitable for handling different organisms with different efficiency. Some are even capable of handling multicellular organisms.

Viruses are present in virtually every environment; they are the most numerous and diverse biological entities (Edwards and Rohwer, 2005). It seems likely that to isolate single viruses and sequence viral genomes, a cultivable virus-host system must first be established. For example, eukaryotic algae is a host to an incredible diversity of viruses (Short, 2012). FACS has been used to sort various natural viral populations into microtiter plates; however, the majority of wells (98%) contained no viruses (Marie et al., 1999; Brussaard et al., 2000; Brussaard, 2004). Allen et al. (2011) successfully used FACS to sort SYBR Green (Invitrogen) stained viruses directly onto agarose beads; >75% of agarose beads contained one or more (one–five) viral particles (Allen et al., 2011). Therefore, using both FACS and single-droplet systems, viruses can potentially be isolated and sorted into single particles and without dependence on the viral-host system. We suspect that Fluidigm chip wells may be too large for use with viral particle isolation, but may be suitable for a group of viruses, when viruses could be co-isolated with the host.

Archaea and bacteria have a similar cell structure; only cell composition and organization set these two domains apart. Bacterial cells are ordinarily 0.5–5.0 µm in length, while archaea can be larger than 15 µm in length. Essentially, all methods reviewed here thus far can be used to separate and process single archaeal and bacterial cells. However, during FACS high currents may slow the growth of organisms, especially in algae. It is noteworthy to mention that Fluidigm chips are currently only suitable to handle organisms no smaller than 3–5 µm.

Eukaryotes are set apart from other life domains by their membrane-bound nucleus and membrane-bound organelles. Eukaryotic organisms may be multicellular or single-celled organisms; cells typically range from about 10–100 µm in size, in the order of ten-times larger than bacteria. Dilution, FACS, and droplet sorting are all suitable methods for processing eukaryotic cells; droplet size can easily be adjusted depending on cell size. Fluidigm chips are well suited for use on smaller eukaryotic cells as they are capable of handling cells up to 25 µm.

Multicellular organisms can theoretically also be sorted and processed by various techniques. FACS is an established technology that can easily sort multicellular organisms. If incubated in droplets for an adequate amount of time, single-celled organisms may express a protein or ligand and have the opportunity to multiply, potentially creating a multicellular structure. At present, Fluidigm chips cannot accommodate studies of multicellular organisms.

		Single-celled				
Technology	Virus	Archaea	Bacteria	Eukaryote	Multicellular	Note
Cultivation	Yes (need host)	Yes	Yes	Yes	Yes	Media requirement;

						Low throughput
Dilution	Possible	Yes	Yes	Yes	Yes	Low throughput
FACS	Yes	Yes	Yes	Yes	Possible	Slow growth after FACS; One-phase system
Droplet	Possible	Yes	Yes	Yes	Yes	High throughput; Incubation possible
Fluidigm	No	No	No	Yes	No	Chips handle single cells 3–25 μm

Table 1. Various technologies have strengths and weaknesses depending on the organism being analyzed.

Conclusion

Molecular studies of our environment and ecology facilitate the understanding of living communities. With the advancement of molecular tools, especially sequencing machines, it is possible to sequence whole communities from selected locations. Environmental sampling and 16S/18S rRNA sequencing allows us to understand which microorganisms are thriving. Genomic libraries of uncultivable microbes can now be prepared and sequenced, providing insight into the presence of populations in the environment and allowing for them to be quantified.

Sequencing of whole metagenomes or metatranscriptomes delivers more information than 16S/18S rRNA library sequencing; they provide information about the presence or absence of both coding and noncoding genes. However, the metagenomic approach uses short sequencing reads, which do not account for heterogeneity within a cell population. Culture enrichment and sample enrichment by FACS and WGA only partially resolves our understanding of the heterogeneity of complex samples.

Recent advancements in single-cell technologies are enabling the isolation and amplification of genomic material from single cells that can then be prepared into libraries and sequenced. Data obtained from SCG compliments data obtained by metagenomics. Isolating a single cell is a complex task. Popular approaches include targeted (specific cells are identified and selected) and random selection, which include a separation of cells into chambers. Currently, microdroplet technology, where a cell is encapsulated by an aqueous microdroplet surrounded by an oil phase, is gaining in popularity.

We anticipate the same trend for SCG as have observed with other sequencing projects. Advancements in sequencing technologies and single-cell handling in microdroplets is facilitating the laborious isolation of single cells and is allowing the sequencing of dozens of single cells simultaneously; in time, thousands and hundreds of thousands of cells will be sequenced this way. The unprecedented throughput and the linkage of a genome to a phenotype through compartmentalization, along with other benefits, makes microdroplet technology unique and

perfectly suitable for deciphering the pathways of complex environmental samples. Microdroplet technology can also be used to select the most effective producer cell, which can then be improved through genetic engineering, and high throughput permits the selection of the best performing enzyme or other molecules of interest.

Ultimately a combination of methods (such as microdroplet and FACS) is already delivering astonishing results (Dichosa et al., 2014). SCG technology and metagenomics go hand-in-hand, where data from both can be used to assemble single-cell genomes. Microdroplet technology has a broad scope of application in biology. Customized systems are being setup in various labs and commercial solutions are being developed so that more labs will be able to perform SCG studies on various environmental samples, including marine samples. Applying the advanced SCG and metagenomics tools, marine will be the ideal resource for the bioprospecting (Kodzius and Gojabori, 2015).

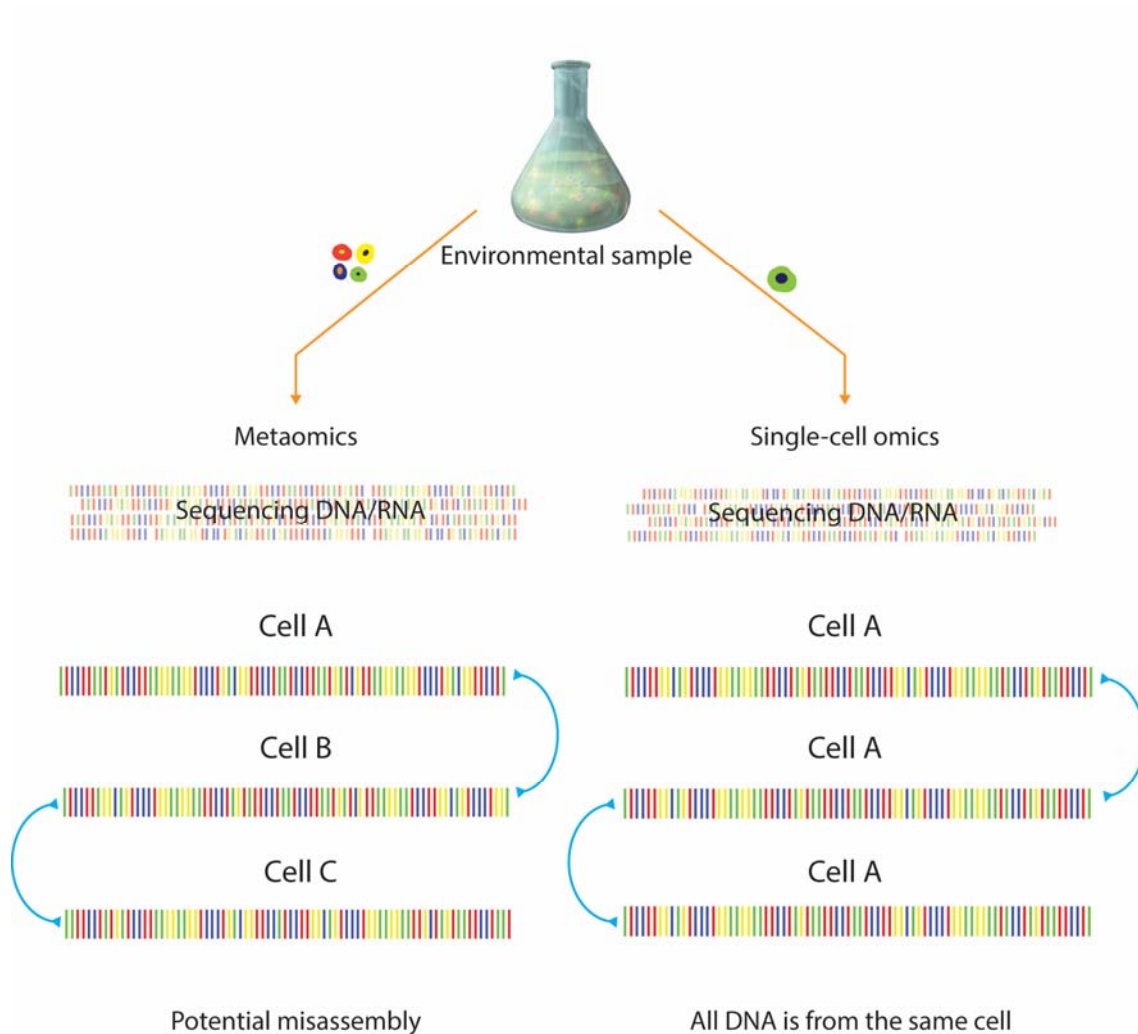


Fig. 1. Differences between metaomic and single-cell genomic methods

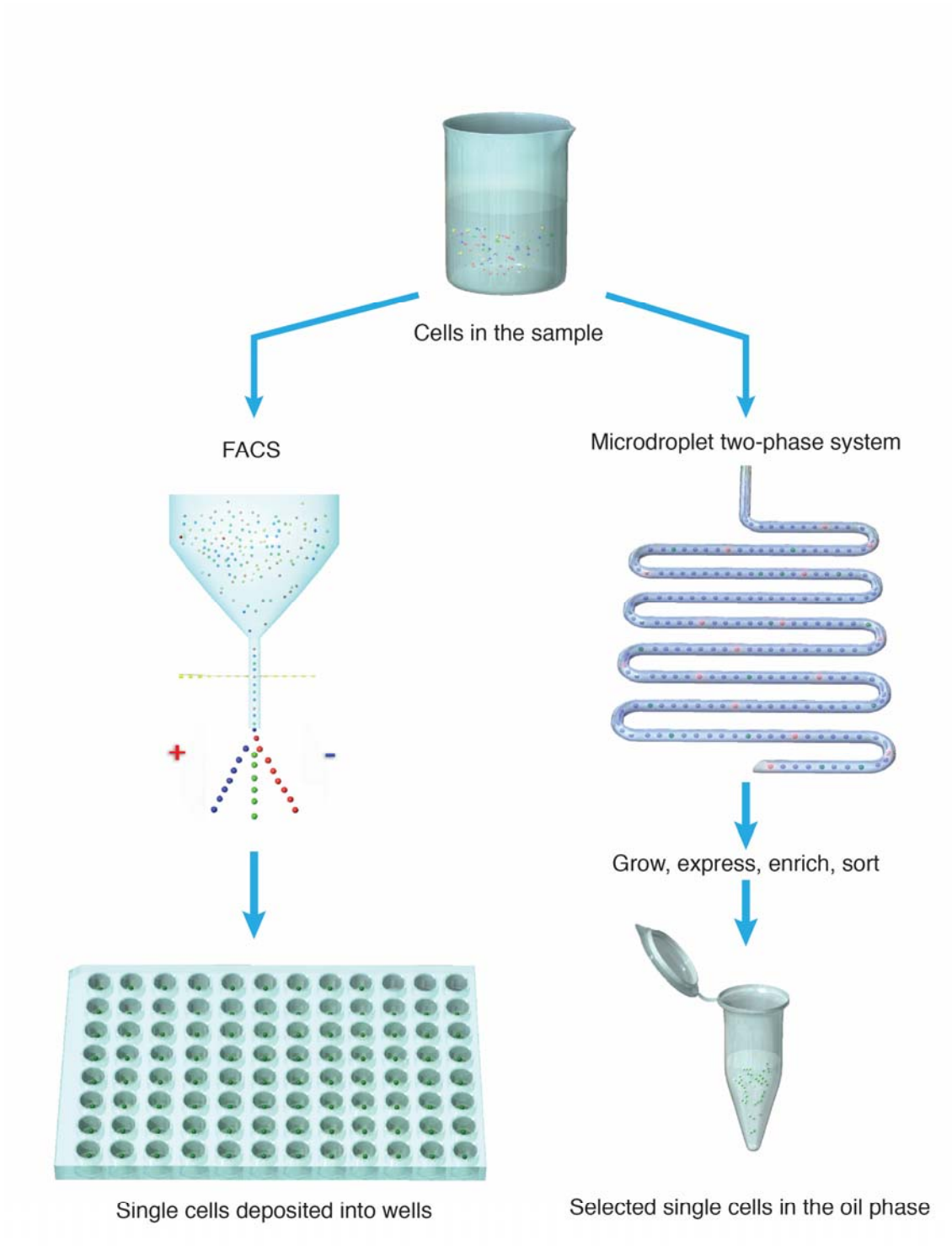


Fig. 2. Differences between FACS and the two-phase microdroplet system

References

- Agresti, J.J., Antipov, E., Abate, A.R., Ahn, K., Rowat, A.C., Baret, J.C., Marquez, M., Klibanov, A.M., Griffiths, A.D. and Weitz, D.A., 2010. Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. *Proc Natl Acad Sci U S A* 107, 4004-9.
- Allen, L.Z., Ishoey, T., Novotny, M.A., McLean, J.S., Lasken, R.S. and Williamson, S.J., 2011. Single virus genomics: a new tool for virus discovery. *PLoS One* 6, e17722.
- Amantonico, A., Urban, P.L. and Zenobi, R., 2010. Analytical techniques for single-cell metabolomics: state of the art and trends. *Anal Bioanal Chem* 398, 2493-504.
- Bergquist, P.L., Hardiman, E.M., Ferrari, B.C. and Winsley, T., 2009. Applications of flow cytometry in environmental microbiology and biotechnology. *Extremophiles* 13, 389-401.
- Blainey, P.C., 2013. The future is now: single-cell genomics of bacteria and archaea. *FEMS Microbiol Rev* 37, 407-27.
- Blainey, P.C., Mosier, A.C., Potanina, A., Francis, C.A. and Quake, S.R., 2011. Genome of a low-salinity ammonia-oxidizing archaeon determined by single-cell and metagenomic analysis. *PLoS One* 6, e16626.
- Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J.M., Segall, A.M., Mead, D., Azam, F. and Rohwer, F., 2002. Genomic analysis of uncultured marine viral communities. *Proc Natl Acad Sci U S A* 99, 14250-5.
- Brouzes, E., Medkova, M., Savenelli, N., Marran, D., Twardowski, M., Hutchison, J.B., Rothberg, J.M., Link, D.R., Perrimon, N. and Samuels, M.L., 2009. Droplet microfluidic technology for single-cell high-throughput screening. *Proc Natl Acad Sci U S A* 106, 14195-200.
- Brussaard, C.P., 2004. Optimization of procedures for counting viruses by flow cytometry. *Appl Environ Microbiol* 70, 1506-13.
- Brussaard, C.P., Marie, D. and Bratbak, G., 2000. Flow cytometric detection of viruses. *J Virol Methods* 85, 175-82.
- Casadevall i Solvas, X., Niu, X., Leeper, K., Cho, S., Chang, S.I., Edel, J.B. and deMello, A.J., 2011. Fluorescence detection methods for microfluidic droplet platforms. *J Vis Exp*.
- Citri, A., Pang, Z.P., Sudhof, T.C., Wernig, M. and Malenka, R.C., 2012. Comprehensive qPCR profiling of gene expression in single neuronal cells. *Nat Protoc* 7, 118-27.
- Clausell-Tormos, J., Lieber, D., Baret, J.C., El-Harrak, A., Miller, O.J., Frenz, L., Blouwolff, J., Humphry, K.J., Koster, S., Duan, H., Holtze, C., Weitz, D.A., Griffiths, A.D. and Merten, C.A., 2008. Droplet-based microfluidic platforms for the encapsulation and screening of Mammalian cells and multicellular organisms. *Chem Biol* 15, 427-37.
- Dichosa, A.E., Daughton, A.R., Reitenga, K.G., Fitzsimons, M.S. and Han, C.S., 2014. Capturing and cultivating single bacterial cells in gel microdroplets to obtain near-complete genomes. *Nat Protoc* 9, 608-21.
- Dobes, N.C., Dhopeswarkar, R., Henley, W.H., Ramsey, J.M., Sims, C.E. and Allbritton, N.L., 2013. Laser-based directed release of array elements for efficient collection into targeted microwells. *Analyst* 138, 831-8.
- Dupont, C.L., Rusch, D.B., Yooseph, S., Lombardo, M.J., Richter, R.A., Valas, R., Novotny, M., Yee-Greenbaum, J., Selengut, J.D., Haft, D.H., Halpern, A.L., Lasken, R.S., Neelson, K., Friedman, R. and Venter, J.C., 2012. Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. *ISME J* 6, 1186-99.
- Edwards, R.A. and Rohwer, F., 2005. Viral metagenomics. *Nat Rev Microbiol* 3, 504-10.
- Fallah-Araghi, A., Baret, J.C., Ryckelynck, M. and Griffiths, A.D., 2012. A completely in vitro ultrahigh-throughput droplet-based microfluidic screening system for protein engineering and directed evolution. *Lab Chip* 12, 882-91.
- Field, K.G., Olsen, G.J., Lane, D.J., Giovannoni, S.J., Ghiselin, M.T., Raff, E.C., Pace, N.R. and Raff, R.A., 1988. Molecular phylogeny of the animal kingdom. *Science* 239, 748-53.
- Hallam, S.J., Mincer, T.J., Schleper, C., Preston, C.M., Roberts, K., Richardson, P.M. and DeLong, E.F., 2006. Pathways of carbon assimilation and ammonia oxidation suggested by environmental genomic analyses of marine Crenarchaeota. *PLoS Biol* 4, e95.

- Hess, M., Sczyrba, A., Egan, R., Kim, T.W., Chokhawala, H., Schroth, G., Luo, S., Clark, D.S., Chen, F., Zhang, T., Mackie, R.I., Pennacchio, L.A., Tringe, S.G., Visel, A., Woyke, T., Wang, Z. and Rubin, E.M., 2011. Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science* 331, 463-7.
- Hugenholtz, P., 2002. Exploring prokaryotic diversity in the genomic era. *Genome Biol* 3, REVIEWS0003.
- Hugenholtz, P., Goebel, B.M. and Pace, N.R., 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J Bacteriol* 180, 4765-74.
- Kodzius, R. and Gojobori, T., 2015. Marine metagenomics as a source for bioprospecting. *Mar Genomics*.
- Koster, S., Angile, F.E., Duan, H., Agresti, J.J., Wintner, A., Schmitz, C., Rowat, A.C., Merten, C.A., Pisignano, D., Griffiths, A.D. and Weitz, D.A., 2008. Drop-based microfluidic devices for encapsulation of single cells. *Lab Chip* 8, 1110-5.
- Kumaresan, P., Yang, C.J., Cronier, S.A., Blazej, R.G. and Mathies, R.A., 2008. High-throughput single copy DNA amplification and cell analysis in engineered nanoliter droplets. *Anal Chem* 80, 3522-9.
- Lane, D.J., Pace, B., Olsen, G.J., Stahl, D.A., Sogin, M.L. and Pace, N.R., 1985. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc Natl Acad Sci U S A* 82, 6955-9.
- Lanigan, P.M., Chan, K., Ninkovic, T., Templer, R.H., French, P.M., de Mello, A.J., Willison, K.R., Parker, P.J., Neil, M.A., Ces, O. and Klug, D.R., 2008. Spatially selective sampling of single cells using optically trapped fusogenic emulsion droplets: a new single-cell proteomic tool. *J R Soc Interface* 5 Suppl 2, S161-8.
- Leung, K., Zahn, H., Leaver, T., Konwar, K.M., Hanson, N.W., Page, A.P., Lo, C.C., Chain, P.S., Hallam, S.J. and Hansen, C.L., 2012. A programmable droplet-based microfluidic device applied to multiparameter analysis of single microbes and microbial communities. *Proc Natl Acad Sci U S A* 109, 7665-70.
- Lim, Y.W., Cuevas, D.A., Silva, G.G.Z., Aguinaldo, K., Dinsdale, E.A., Haas, A.F., Hatay, M., Sanchez, S.E., Wegley-Kelly, L., Dutilh, B.E., Harkins, T.T., Lee, C.C., Tom, W., Sandin, S.A., Smith, J.E., Zgliczynski, B., Vermeij, M.J.A., Rohwer, F. and Edwards, R.A., 2014. Sequencing at sea: challenges and experiences in Ion Torrent PGM sequencing during the 2013 Southern Line Islands Research Expedition. *Peerj* 2, e433v1.
- Marcy, Y., Ishoey, T., Lasken, R.S., Stockwell, T.B., Walenz, B.P., Halpern, A.L., Beeson, K.Y., Goldberg, S.M. and Quake, S.R., 2007. Nanoliter reactors improve multiple displacement amplification of genomes from single cells. *PLoS Genet* 3, 1702-8.
- Marie, D., Brussaard, C.P.D., Thyrrhaug, R., Bratbak, G. and Vaulot, D., 1999. Enumeration of marine viruses in culture and natural samples by flow cytometry. *Appl Environ Microbiol* 65, 45-52.
- Mason, O.U., Hazen, T.C., Borglin, S., Chain, P.S., Dubinsky, E.A., Fortney, J.L., Han, J., Holman, H.Y., Hultman, J., Lamendella, R., Mackelprang, R., Malfatti, S., Tom, L.M., Tringe, S.G., Woyke, T., Zhou, J., Rubin, E.M. and Jansson, J.K., 2012. Metagenome, metatranscriptome and single-cell sequencing reveal microbial response to Deepwater Horizon oil spill. *ISME J* 6, 1715-27.
- Mazutis, L., Gilbert, J., Ung, W.L., Weitz, D.A., Griffiths, A.D. and Heyman, J.A., 2013. Single-cell analysis and sorting using droplet-based microfluidics. *Nature Protocols* 8, 870-891.
- McCarren, J., Becker, J.W., Repeta, D.J., Shi, Y., Young, C.R., Malmstrom, R.R., Chisholm, S.W. and DeLong, E.F., 2010. Microbial community transcriptomes reveal microbes and metabolic pathways associated with dissolved organic matter turnover in the sea. *Proc Natl Acad Sci U S A* 107, 16420-7.
- Pace, N., Stahl, D., Lane, D. and Olsen, G., 1986. The Analysis of Natural Microbial Populations by Ribosomal RNA Sequences, in: Marshall, K.C. (Ed.), *Advances in Microbial Ecology*. Springer US, pp. 1-55.

- Pamp, S.J., Harrington, E.D., Quake, S.R., Relman, D.A. and Blainey, P.C., 2012. Single-cell sequencing provides clues about the host interactions of segmented filamentous bacteria (SFB). *Genome Res* 22, 1107-19.
- Podar, M., Abulencia, C.B., Walcher, M., Hutchison, D., Zengler, K., Garcia, J.A., Holland, T., Cotton, D., Hauser, L. and Keller, M., 2007. Targeted access to the genomes of low-abundance organisms in complex microbial communities. *Appl Environ Microbiol* 73, 3205-14.
- Poretsky, R.S., Hewson, I., Sun, S., Allen, A.E., Zehr, J.P. and Moran, M.A., 2009. Comparative day/night metatranscriptomic analysis of microbial communities in the North Pacific subtropical gyre. *Environ Microbiol* 11, 1358-75.
- Raghunathan, A., Ferguson, H.R., Jr., Bornarth, C.J., Song, W., Driscoll, M. and Lasken, R.S., 2005. Genomic DNA amplification from a single bacterium. *Appl Environ Microbiol* 71, 3342-7.
- Rinke, C., Lee, J., Nath, N., Goudeau, D., Thompson, B., Poulton, N., Dmitrieff, E., Malmstrom, R., Stepanauskas, R. and Woyke, T., 2014. Obtaining genomes from uncultivated environmental microorganisms using FACS-based single-cell genomics. *Nat Protoc* 9, 1038-48.
- Rodrigue, S., Malmstrom, R.R., Berlin, A.M., Birren, B.W., Henn, M.R. and Chisholm, S.W., 2009. Whole genome amplification and de novo assembly of single bacterial cells. *PLoS One* 4, e6864.
- Rossow, T., Heyman, J.A., Ehrlicher, A.J., Langhoff, A., Weitz, D.A., Haag, R. and Seiffert, S., 2012. Controlled synthesis of cell-laden microgels by radical-free gelation in droplet microfluidics. *J Am Chem Soc* 134, 4983-9.
- Schmidt, T.M., DeLong, E.F. and Pace, N.R., 1991. Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *J Bacteriol* 173, 4371-8.
- Sekar, R., Fuchs, B.M., Amann, R. and Pernthaler, J., 2004. Flow sorting of marine bacterioplankton after fluorescence in situ hybridization. *Appl Environ Microbiol* 70, 6210-9.
- Shi, Y., Tyson, G.W. and DeLong, E.F., 2009. Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column. *Nature* 459, 266-9.
- Short, S.M., 2012. The ecology of viruses that infect eukaryotic algae. *Environ Microbiol* 14, 2253-71.
- Skafto-Pedersen, P., Hemmingsen, M., Sabourin, D., Blaga, F.S., Bruus, H. and Dufva, M., 2012. A self-contained, programmable microfluidic cell culture system with real-time microscopy access. *Biomed Microdevices* 14, 385-99.
- Stein, J.L., Marsh, T.L., Wu, K.Y., Shizuya, H. and DeLong, E.F., 1996. Characterization of uncultivated prokaryotes: Isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. *Journal of Bacteriology* 178, 591-599.
- Stepanauskas, R., 2012. Single cell genomics: an individual look at microbes. *Curr Opin Microbiol* 15, 613-20.
- Stepanauskas, R. and Sieracki, M.E., 2007. Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. *Proc Natl Acad Sci U S A* 104, 9052-7.
- Swan, B.K., Martinez-Garcia, M., Preston, C.M., Sczyrba, A., Woyke, T., Lamy, D., Reinthaler, T., Poulton, N.J., Masland, E.D., Gomez, M.L., Sieracki, M.E., DeLong, E.F., Herndl, G.J. and Stepanauskas, R., 2011. Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. *Science* 333, 1296-300.
- Tadmor, A.D., Ottesen, E.A., Leadbetter, J.R. and Phillips, R., 2011. Probing individual environmental bacteria for viruses by using microfluidic digital PCR. *Science* 333, 58-62.
- Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B.B., Siddiqui, A., Lao, K. and Surani, M.A., 2009. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods* 6, 377-82.
- Teh, S.Y., Lin, R., Hung, L.H. and Lee, A.P., 2008. Droplet microfluidics. *Lab Chip* 8, 198-220.
- Thorsen, T., Roberts, R.W., Arnold, F.H. and Quake, S.R., 2001. Dynamic pattern formation in a vesicle-generating microfluidic device. *Phys Rev Lett* 86, 4163-6.
- Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., Wu, D., Paulsen, I., Nelson, K.E., Nelson, W., Fouts, D.E., Levy, S., Knap, A.H., Lomas, M.W., Neelson, K., White,

- O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y.H. and Smith, H.O., 2004. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304, 66-74.
- Wallner, G., Fuchs, B., Spring, S., Beisker, W. and Amann, R., 1997. Flow sorting of microorganisms for molecular analysis. *Appl Environ Microbiol* 63, 4223-31.
- Williams, R., Peisajovich, S.G., Miller, O.J., Magdassi, S., Tawfik, D.S. and Griffiths, A.D., 2006. Amplification of complex gene libraries by emulsion PCR. *Nat Methods* 3, 545-50.
- Woyke, T., Xie, G., Copeland, A., Gonzalez, J.M., Han, C., Kiss, H., Saw, J.H., Senin, P., Yang, C., Chatterji, S., Cheng, J.F., Eisen, J.A., Sieracki, M.E. and Stepanauskas, R., 2009. Assembling the marine metagenome, one cell at a time. *PLoS One* 4, e5299.
- Wu, D., Hugenholtz, P., Mavromatis, K., Pukall, R., Dalin, E., Ivanova, N.N., Kunin, V., Goodwin, L., Wu, M., Tindall, B.J., Hooper, S.D., Pati, A., Lykidis, A., Spring, S., Anderson, I.J., D'Haeseleer, P., Zemla, A., Singer, M., Lapidus, A., Nolan, M., Copeland, A., Han, C., Chen, F., Cheng, J.F., Lucas, S., Kerfeld, C., Lang, E., Gronow, S., Chain, P., Bruce, D., Rubin, E.M., Kyrpides, N.C., Klenk, H.P. and Eisen, J.A., 2009. A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature* 462, 1056-60.
- Yaron, J.R., Ziegler, C.P., Tran, T.H., Glenn, H.L. and Meldrum, D.R., 2014. A convenient, optimized pipeline for isolation, fluorescence microscopy and molecular analysis of live single cells. *Biol Proced Online* 16, 9.
- Yilmaz, S., Allgaier, M. and Hugenholtz, P., 2010. Multiple displacement amplification compromises quantitative analysis of metagenomes. *Nat Methods* 7, 943-4.
- Yoon, H.S., Price, D.C., Stepanauskas, R., Rajah, V.D., Sieracki, M.E., Wilson, W.H., Yang, E.C., Duffy, S. and Bhattacharya, D., 2011. Single-cell genomics reveals organismal interactions in uncultivated marine protists. *Science* 332, 714-7.
- Zeng, Y., Novak, R., Shuga, J., Smith, M.T. and Mathies, R.A., 2010. High-performance single cell genetic analysis using microfluidic emulsion generator arrays. *Anal Chem* 82, 3183-90.
- Zhang, H., Jenkins, G., Zou, Y., Zhu, Z. and Yang, C.J., 2012a. Massively parallel single-molecule and single-cell emulsion reverse transcription polymerase chain reaction using agarose droplet microfluidics. *Anal Chem* 84, 3599-606.
- Zhang, K., Martiny, A.C., Reppas, N.B., Barry, K.W., Malek, J., Chisholm, S.W. and Church, G.M., 2006. Sequencing genomes from single cells by polymerase cloning. *Nat Biotechnol* 24, 680-6.
- Zhang, X., Leung, C., Lu, Z., Esfandiari, N., Casper, R.F. and Sun, Y., 2012b. Controlled aspiration and positioning of biological cells in a micropipette. *IEEE Trans Biomed Eng* 59, 1032-40.
- Zhong, J.F., Chen, Y., Marcus, J.S., Scherer, A., Quake, S.R., Taylor, C.R. and Weiner, L.P., 2008. A microfluidic processor for gene expression profiling of single human embryonic stem cells. *Lab Chip* 8, 68-74.