Published in final edited form as:
FEMS Microbiol Rev. 2013 May ; 37(3): . doi:10.1111/1574-6976.12015.

The future is now: single-cell genomics of bacteria and archaea

Paul C. Blainey[1,2]

[1]Broad Institute of Harvard and MIT, Cambridge, MA, USA

[2]Department of Biological Engineering, MIT, Cambridge, MA, USA

Abstract

Interest in the expanding catalog of uncultivated microorganisms, increasing recognition of heterogeneity among seemingly similar cells, and technological advances in whole-genome amplification and single-cell manipulation are driving considerable progress in single-cell genomics. Here, the spectrum of applications for single-cell genomics, key advances in the development of the field, and emerging methodology for single-cell genome sequencing are reviewed by example with attention to the diversity of approaches and their unique characteristics. Experimental strategies transcending specific methodologies are identified and organized as a road map for future studies in single-cell genomics of environmental microorganisms. Over the next decade, increasingly powerful tools for single-cell genome sequencing and analysis will play key roles in accessing the genomes of uncultivated organisms, determining the basis of microbial community functions, and fundamental aspects of microbial population biology.

Keywords

single-cell analysis; whole-genome amplification; multiple displacement amplification; microfluidics; microencapsulation; micromanipulation

Introduction

Networks of microorganisms constitute the chemical infrastructure of Earth's biosphere. Microbial communities varying in complexity and vigor entwine every ecosystem on the planet. Humans depend on these microbial systems for global primary production (Liu *et al*., 1997), ecosystem services (Brussaard, 1997; Matson *et al*., 2011), industrial processes (Prescott & Dunn, 1949; Bai *et al*., 2008), and most intimately in regard to the human microbiome (Proctor, 2011; Relman, 2011). As a result, there is tremendous interest in describing the physiology of these microorganisms, their relationships to one another, and their impact on human society. Presumptively, our ability to predict the response of microbial communities to perturbation (anthropogenic and otherwise) will improve in accord with the depth of our understanding. The standard of knowledge is higher yet for efforts to engineer the function of microbial systems.

The ongoing revolution in genomic science and sequencing technology has strongly impacted environmental genomics, providing increasingly comprehensive 'shotgun' (random) coverage of DNA in environmental samples (Berry *et al*., 2003; Tyson *et al*., 2004; Venter *et al*., 2004) and making the genome sequencing of bacterial and archaeal isolates routine (Fleischmann *et al*., 1995; Bult *et al*., 1996). Next-generation sequencing has

Correspondence: Paul C. Blainey, 7, Cambridge Center, Cambridge, MA 02142,, USA. Tel.: +1 650 714 7320;, pblainey@mit.edu.

been applied broadly in metagenomic sequencing studies and helped the field advance beyond single-gene PCR-based studies. While providing rapid access to the catalog of community genes and enabling comparisons among different communities, the analysis of metagenomic data has remained largely gene and pathway centric (notable exceptions are discussed below).

Single-gene studies, metagenomic assemblies, and the genome sequences of a limited number of cultured isolates are not a sufficient basis on which to accurately model the responses of natural microbial networks or engineer the function of artificial communities as we desire. For example, gene catalogs and composite genomes assembled from metagenomic data do not presently distinguish between genes that are tightly coupled within the context of the same organism and genes that are coupled across different organisms. This is a critical limitation, because only gene products encoded by the same organism can freely come into contact with one another to form complexes, drive signaling pathways, or carry out multi-step enzymatic transformations of diffusible substrates at maximum rates. A systems-level predictive understanding of microbial physiology absolutely demands the interpretation of genes and pathways in a full genomic context. Furthermore, individual organisms (indeed single cells) encoding full genomes are the basic replicating unit of biology and an important unit of evolutionary selection, factors that cannot be ignored in understanding the development of microbial networks in larger populations as a function of time.

The need for whole genomes from microbial communities is clear. Although bringing novel isolates into axenic culture remains important to enable functional studies of new microorganisms, the traditional isolate sequencing paradigm falls short in three respects as a general approach to genome sequencing. First, although data on failed attempts are not often reported, the yield of axenic cultures from randomly targeted environmental organisms is understood to be low. Second, the distribution of successful cases is strongly biased. While an effort bias explains the preponderance of database genomes from heterotrophic human pathogens, other intrinsic biases have been recognized that favor isolation of organisms similar to those already cultured, as well as toward faster-growing organisms and those that depend to a lesser extent on interactions with the community network (Wu *et al*., 2009). Finally, traditional isolation techniques are labor-intensive and slow, sometimes requiring years of effort due to the need for serial enrichment culture and the slow growth of organisms under suboptimal conditions, prompting the invention of automated systems (Connon & Giovannoni, 2002).

The demand for greater numbers of more diverse genomes than are being delivered through isolate sequencing can be satisfied by an emerging spectrum of cultivation-agnostic approaches for genome sequencing (Fig. 1). These methods, although not requiring culture-based isolation, can be carried out in parallel with culture-based studies if conditions for growth are known. The lack of a requirement for axenic culture allows very broad application, with some schemes leveraging DNA present in the sample at the time of collection exclusively, permitting limited fixation of samples. The span of culture-agnostic genomics techniques ranges from new ways of processing standard metagenomic data sets to single-cell sequencing. These methods have variable requirements and result in genomic data sets with differing properties; as such, particular communities and target organisms are best served by different combinations of approaches. Such ambitious study designs will become increasingly tractable as sample preparation procedures are streamlined and bespoke instrumentation is commercialized.

Composite genomes can be amassed from metagenomic contigs by classifying (or ‘binning’) reads according to the abundance of related reads and lineage-specific signatures such as

nucleotide content signatures (Tyson *et al*., 2004; Woyke *et al*., 2006; Dick *et al*., 2009; Hess *et al*., 2011; Luo *et al*., 2011; Tanaseichuk *et al*., 2011; Wang *et al*., 2012b; Fig. la). Although challenging in data sets from more complex microbial communities and for organisms with significant strain heterogeneity, this approach is expected to scale favorably with increased sequencing depth and advancements in assembly of metagenomic data (Mavromatis *et al*., 2007; Namiki *et al*., 2011; Peng *et al*., 2011; Treangen *et al*., 2011; Wrighton *et al*., 2012). One way to enhance the targeting of organisms associated with a particular community function is through the use of enrichment culture under a condition designed to bloom organisms associated with a function of interest and/or to select against other organisms prior to the collection of a 'targeted metagenomic' data set (Hess *et al*., 2011; Fig. lb). Another avenue is the processing of tiny consortia that exhibit lower diversity simply because they contain a limited number of cells. Although not yet explored to a great extent, analysis of such microconsortia is now accessible thanks to the advances in whole-genome amplification (WGA) technology, although relative abundances of sequences from different strains are likely to change during WGA (Yilmaz *et al*., 2010). When such physically aggregated groups of cells are selected, additional information about functional couplings can be accessed.

A third method, termed 'targeted enrichment' (Stein *et al*., 1996; Hallam *et al*., 2006a, b; Bergquist *et al*., 2009), seeks to segregate a single organism by physically enriching for a target cell population based on combinations of phenotypic characteristics such as size, shape, density, and the spectral characteristics of native and applied fluorophores (Wallner *et al*., 1997; Sekar *et al*., 2004; Fig. 1c). The selected cells are then the basis for sequencing and assembly of a composite genome. This approach faces the challenge that the cell types are not uniquely delineated by measured properties in more complex samples. Even so, limited enrichment may improve the results of efforts to 'bin' genomes from mixed-organism data sets. Low yield of select cells in targeted enrichment is now readily ameliorated by the application of modern WGA methods and high-yield sequence library preparation procedures (Podar *et al*., 2007).

Notwithstanding its limited applicability, isolate sequencing remains the gold standard method for microbial genome sequencing. In isolate sequencing, microgram quantities of high-quality DNA are purified from large-scale axenic cultures, which can be further propagated for functional studies of the microorganism. In many cases, isolate cultures originate from a single cell, facilitating construction of a clonal consensus genome, as all the cells sequenced are descendants of a single original cell (Fig. 1d). However, lengthy procedures for enrichment culture under artificial conditions may select for new variants that are not representative of natural populations present in the original sample. Unlike genomes produced by the other methods discussed, isolate genomes are fully verifiable in the sense that additional equivalent material can typically be produced for confirmatory analyses and have the further advantage that the genome sequences obtained are insensitive to the sequencing and informatics methodologies used.

Single-cell sequencing is unique among current genomic approaches in yielding access to the genomes of individual cells without the complications of culture or compositing data from multiple cells or strains. Formally, single-cell genomic data sets are free of uncertainty in the grouping of sequence reads according to the strain of origin and can resolve extremely fine strain structure, hyper-variable loci, and phase variation at the whole-genome level without ambiguity.

The ability to resolve fine-scale heterogeneity is important in sequencing the genomes of asexual microorganisms, which undergo recombination less frequently than they reproduce. Populations of such organisms have the potential to rapidly diversify in a variety of patterns,

as new mutations are not necessarily mixed through the population and there is no immediate imperative to maintain compatibility for recombine with the rest of the population (Koeppel *et al*., 2008; Caro-Quintero & Konstantinidis, 2012). The resulting mutational 'fuzziness', variation in the relative importance of mutation *vis a vis* recombination, and variation in the rate of genetic exchange between geographically or ecologically distinct populations complicate phylogenic analysis and efforts to circumscribe microbial species (Koonin *et al*., 2001; Gevers *et al*., 2005; Smillie *et al*., 2011; Denef & Banfield, 2012; Shapiro *et al*., 2012). Single-cell sequencing promises to provide direct access to fine-scale heterogeneity in complex microbial populations by resolving and linking 'fuzzy' diversity across whole genomes sequenced from individual cells (Pamp *et al*., 2012).

More generally, cells are the fundamental quanta of biology, representing the most granular level where biological entities can command the full spectrum of biochemical activities. Not coincidentally, cells are both biologically relevant units of living matter and containers that physically subdivide biological samples—this is a convenience to experimental biology that should be utilized to the best effect possible. Analyses targeted to individual cells are an ideal approach to capitalize on this opportunity in biological science. Today, technologies enabling single-cell analysis are progressing and diversifying rapidly, as well as powering discovery in many fields of biological science. Besides environmental microbiology, many groups are currently active in applying single-cell genomic methods including WGA and shotgun sequencing to human cells. Although such applications to human genetics are not a focus of this review, many of the methods employed in these studies are related to those used for microbial samples and will be referenced where appropriate.

Single-cell single-gene sequencing studies were the first single-cell sequencing experiments. These studies, first in human cells (Küppers *et al*., 1993; Sucher & Deitcher, 1995; Maryanski *et al*., 1996; Findlay, 1998; Dietmaier *et al*., 1999), then in microorganisms (Ruiz Sebastián & O'ryan, 2001), depended on the physical (often manual) isolation of individual cells and the use of these cells as templates for PCR amplification and sequencing of specific genomic loci. The advent of higher-throughput automated systems enabled the application of this approach to larger numbers of bacterial cells (Fig. 1e). Such techniques have been applied for multiplex PCR, product recovery, and sequencing of multiple loci per cell in uncultivated organisms to link phenotype-determining functional genes with phylogenetic markers, identifying phenotype–phylotype relationships (Ottesen *et al*., 2006). Alternatively, this approach can be taken to correlate phage and host marker genes to establish phage–host relationships (Tadmor *et al*., 2011). Nevertheless, such methods depend on targeting highly conserved or previously characterized genes with specific primers, which imposes strong biases and limits the scope of the approach for discovery.

Of all the genome-sequencing methods, single-cell whole-genome-sequencing workflows require the most demanding sample preparation: (1) single cells must be isolated with high confidence, (2) each cell's envelope must be compromised such that (3) the DNA inside can be amplified by WGA free of contaminates to produce enough material to support (4) library preparation and (5) high-throughput DNA sequencing (Fig. 1f). WGA is necessary despite the advent of commercial single-molecule sequencing technologies (Helicos, PacBio) and reliable protocols for the PCR amplification of finished sequence libraries. Fundamentally, this is the case because library preparation methods are not sufficiently conservative of the starting material. Given the low efficiency of library creation procedures, each locus in the raw material must be present in high copy number to avoid dropout of that locus in the finished sequence library. In this light, manufacturers currently specify minimum inputs in the nanogram range to prevent the loss of sequence information present in the original sample and to ensure that machine capacity can be utilized while minimizing redundant coverage of the raw input molecules.

In practical terms for single microbial cell sequencing, this means a million-fold amplification of the DNA present at the time of cell selection is required. Such high fold-amplification from subnanogram samples (Dean *et al*., 2002) and individual bacteria (Raghunathan *et al*., 2005) with good representation of the genome were first achieved by the multiple displacement amplification (MDA) WGA chemistry, but produced material with undesirable characteristics such as uneven representation and dislocated sequences. Nonetheless, investigators shortly succeeded in assembling shotgun sequence reads from single WGA-amplified *Escherichia coli* and *Prochlorococcus* (Zhang *et al*., 2006), TM7 (Marcy *et al*., 2007b), and sequencing multiple genes from *E. coli* (Marcy *et al*., 2007a), single marine bacteria (Stepanauskas & Sieracki, 2007), and soil and cultivated archaea (Kvist *et al*., 2007).

Several reviews on single-cell genomics are available that describe popular approaches and catalog the most recent examples that apply these methods (Lasken, 2007, 2012; Binga *et al*., 2008; Ishoey *et al*., 2008; de Jager & Siezen, 2011; Kalisky & Quake, 2011; Kalisky *et al*., 2011; Yilmaz & Singh, 2011; Kamke *et al*., 2012; Stepanauskas, 2012; Lecault *et al*., 2012; Fritzsch *et al*., 2012). The bulk of this review will present emerging approaches to single-cell genome sequencing in depth from a fundamental point of view, highlighting and placing in context the unique features of different methods and potential pitfalls, with the goal of facilitating forward-thinking experimental design for those new to this rapidly developing field.

Configuring experimental approaches for single-cell genomics is now wonderfully complex due to the diversity of experimental techniques and tremendous potential for synergy within integrative approaches considering different data sets or data types. This can be realized in parallel workflows, where comparative analyses evaluate single-cell genomes versus isolate genome sequences and/or composite genomes of differing flavors, or even metagenomic and transcriptomic data sets. More exciting still is the prospect of combined workflows, where, for example, metagenomic reads are incorporated into an assembly of single-cell data (Blainey *et al*., 2011), and single-cell data are used to guide target selection or resolve phase variation in targeted enrichment. A promising approach is the use of single-cell analysis to parameterize and validate the binning of genomes from metagenomic data (Hess *et al*., 2011; J.A. Dodsworth, P.C. Blainey, S.K. Murugapiran, W.D. Swingley, C.A. Ross, del Rio, S.G. Tringe, S.R. Quake and B.P. Hedlund, unpublished data). Conversely, metagenomic data sets can be used to overcome data quality limitations in the assembly of single-cell data sets (Blainey *et al*., 2011; J.A. Dodsworth, P.C. Blainey, S.K. Murugapiran, W.D. Swingley, C.A. Ross, del Rio, S.G. Tringe, S.R. Quake and B.P. Hedlund, unpublished data) or to place single-cell data sets in the broader context of an entire microbial community.

The future of single-cell microbial sequencing is bright, particularly as technical approaches mature and diversify. Single-cell genomic data provide useful insights by themselves and particularly in combination with genomic and metagenomic data sets.

Contamination: the key challenge in single-cell genomics

Contamination suppression is the key challenge in single-cell microbial genomics due to the extremely small quantity of input material available from single cells (roughly 1 fg per megabase of genome size), where representation at a given locus may depend on a single DNA duplex. Contamination arises from three sources: (1) the sample itself, (2) the laboratory environment, and (3) the reagents and instrumentation used for sample preparation (Fig. 2a). A major challenge arising from the ubiquity of microorganisms is the fact that microbial genomic DNA is a common contaminant in all three of these categories and requires greater effort for detection and rejection in single-cell microbial genome-

sequencing projects than would more dissimilar contaminating sequences (e.g. human DNA contaminants).

Single-cell workflows for gene sequencing from single cells take advantage of sequence-specific primers to target (Ruiz Sebastián & O'ryan, 2001; Ottesen *et al*., 2006; Tadmor *et al*., 2011). Such locus-specific approaches are relatively insensitive to fragmented DNA contaminates, but susceptible to back-contamination by the amplicons produced. Segregation of pre- and postamplification work areas and schemes such as uracil incorporation facilitating postreaction product degradation (i.e. by uracil–DNA glycosylase) are measures that can effectively address the back-contamination problem.

The low-input and high-fold genome-wide amplification required for genome sequencing of single cells make the approach exquisitely sensitive to DNA contamination. Commercial suppliers of WGA reagents often require at least 10 ng of template DNA input. This minimum input quantity is not specified due to limited sensitivity of the amplification method, but rather to constrain the fraction of the product mixture that originates from contaminates. For example, the quantity of contaminating DNA in WGA reagents for a 50-(μL reaction is estimated to be of the order of 1 fg (Blainey & Quake, 2011). Given 10 ng input and the yield of such a reaction in the tens of micrograms, the incidence of contaminant sequences in the product mixture will be 10^{-7}, a low level that is acceptable for many applications. However, given a single microbial genome as input, contaminating sequences could make up half the reaction products or the entirety of the products in the case of cell lysis failure.

Of the three sources of contamination, the first two can be effectively addressed by suitably engineering the apparatus used for cell isolation and DNA amplification (Fig. 2a). Specifically, minimizing the real or effective sampling volume suppresses sample-borne contamination. Contaminates from the laboratory environment can be excluded by two classes of engineering controls: integrating the sampling, setup, and reaction steps inside a sealed, disposable microdevice (Marcy *et al*., 2007a, b) that utilizes a minimum sampling volume scheme for selecting cells (Blainey *et al*., 2011), or alternatively, by carrying out these steps using decontaminated equipment and buffers in a very clean environment (Zhang *et al*., 2006; Rodrigue *et al*., 2009; Woyke *et al*., 2009). The third source of contamination can be partially addressed by reducing the volume of the lysis and amplification reactions to the nanoliter scale (Marcy *et al*., 2007a, b). Shrinking these reaction volumes has the effect of concentrating the single genome to be amplified with respect to reagent-borne contaminates in proportion to the volume reduction factor (Fig. 2b and c). Because WGA can produce as much as 1 μg μL^{-1} DNA product (Dean *et al*., 2001), a reaction of only a few nanoliters can support a sequencing run (White *et al*., 2009) or provide a quantity of template sufficient to overwhelm contaminates in a secondary full-scale WGA reaction.

Given the concentration of contaminating fragments present in commercial WGA reagents (varying from 5 to 50 fragments per reaction microliter in the enzyme alone), volume reduction by itself does not necessarily eliminate reagent contamination (Blainey & Quake, 2011). For example, even driving reaction volumes down to the low nanoliter range, a significant fraction of reactions are still expected to carry contaminates from the reagents. Thus, it is necessary to either inactivate contaminates in the reagents or produce reagent sets that are free of contamination. Contaminates in commercial WGA kits have been successfully suppressed by UV exposure with acceptable post-treatment amplification performance (Zhang *et al*., 2006; Woyke *et al*., 2011). Alternatively, reagents for background-free WGA can be produced in batch processes utilizing disposable plasticware produced from virgin materials (Blainey & Quake, 2011). Irrespective of the cleanup approach taken, a key capability for validating reagent lots and cleanup procedures is a rapid

assay for WGA activity and contamination. To be useful, the contamination assay must be both quantitative and extremely sensitive, such that different lots or treatments can be evaluated comparatively. qPCR is insufficiently sensitive as only contaminant molecules with an intact sequence locus matching specific PCR primers can be detected. For example, a PCR assay for the small-subunit ribosomal RNA gene misses thousands of contaminant fragments arising from bacterial genomic DNA for every fragment molecule detected. Alternatively, the digital WGA (dWGA, e.g. digital MDA or dMDA) assay can be used for quantitation down to a few attograms of degraded genomic DNA per microliter (Blainey & Quake, 2011). dWGA is compatible with a variety of off-the-shelf platforms engineered for digital PCR (Baker, 2012).

Cell isolation

Two principal approaches to cell isolation have been applied in single-cell sequencing: random encapsulation and micromanipulation (Fig. 3). Random encapsulation relies on the random partitioning of individual cells at limiting dilution and is typically applied to an entire sample containing a large number of cells. Some random encapsulation workflows include a subsequent cell/ droplet selection step to improve fill factors (e.g. fraction of microwells actually containing a cell) or to target cell types of interest. This approach has been applied in a wide variety of formats for PCR- and WGA-based single-cell genomic analyses. Selected examples focused on microorganisms span manual dilution in standard laboratory ware (Fig. 3a; Zhang *et al*., 2006), arrays of microfluidic chambers (Fig. 3b; Ottesen *et al*., 2006; Love *et al*., 2006; Wang *et al*., 2009), microdroplets in air by flow cytometry (Fig. 3c; Raghunathan *et al*., 2005; Stepanauskas & Sieracki, 2007), and microdroplets in oil (Fig. 3d; Thorsen *et al*., 2001; Agresti *et al*., 2010; Zeng *et al*., 2010). Commercial microfluidic arrays have become popular for PCR-based high-throughput single-cell assays (Kalisky & Quake, 2011; Kalisky *et al*., 2011; Fox *et al*., 2012; Sanchez-Freire *et al*., 2012), but present challenges in sequential addition of reagents for more complicated amplification schemes and for the recovery of products from individual cells necessitated the development of custom preparative microfluidic devices (Marcy *et al*., 2007a, b; Blainey *et al*., 2011; Youssef *et al*., 2011; J.A. Dodsworth, P.C. Blainey, S.K. Murugapiran, W.D. Swingley, C.A. Ross, del Rio, S.G. Tringe, S. R. Quake and B.P. Hedlund, unpublished data; Marshall *et al*., 2012; Pamp *et al*., 2012). Last year, similar microfluidic devices for automated processing of single mammalian cells became available (Fluidigm C1 chips).

Flow cytometry and fluorescence-activated cell sorting (FACS) have become popular platforms for single-cell genomic analysis (Raghunathan *et al*., 2005; Stepanauskas & Sieracki, 2007; Rodrigue *et al*., 2009; Woyke *et al*., 2009; Dupont *et al*., 2012; Hess *et al*., 2011; Swan *et al*., 2011; Yoon *et al*., 2011). Together with automated liquid-handling robots, a FACS-centered workflow for single-cell WGA can be assembled entirely from off-the-shelf instrumentation. Although FACS is fundamentally based on random encapsulation of cells, flow cytometers are able to select and direct cell-containing droplets based on the presence of cells or the presence of cells with characteristics that can be detected optically by the instrument. This capability allows the rejection of droplets containing no cells and droplets containing cells that are not of interest, as well as the addressing of individual droplets into the wells of a microwell plate. A FACS/liquid-handling workflow can process hundreds of single-cell WGA reactions daily, and the use of the microwell plate format makes recovery of reaction products on a cell-by-cell basis straightforward. Besides bacterial cells, FACS has also been applied for the analysis of single cultured virions (Allen *et al*., 2011). This study also applied an interesting alternative approach to the isolation task wherein virions were embedded and amplified within an agarose gel. Limitations of flow cytometry for single-cell genomics include a potential for extrinsic contamination in open-

plate workflow steps, large downstream processing volumes that raise costs and sensitivity to reagent contamination, high shear forces that preclude application to some cell types, and a requirement for large numbers of unaggregated input cells.

Microdroplet, or emulsion technology, provides a capability to rapidly encapsulate thousands of cells in individual picoliter-scale aqueous droplets dispersed in a hydrophobic continuous phase (Thorsen *et al*., 2001). Microdroplets can be formed in bulk by vigorous mixing, or in microfluidic devices, if droplets of uniform size are desired (Thorsen *et al*., 2001). The microdroplets can then be processed *en masse* for genomic analysis (Tewhey *et al*., 2009; Zeng *et al*., 2010). Microfluidics-based emulsion technology combines high speed and high throughput with the advantages of straightforward automation, gentle cell handling, the potential for micrography of the cells analyzed, low potential for contamination, and low reagent consumption. The principal outstanding challenges for single-cell genomic applications of microdroplet technologies are the delivery of cell-processing reagents to the microdroplets (Abate *et al*., 2010), the implementation of desired cell/ droplet selection steps (Link *et al*., 2006), and interfacing the droplets with sequence library preparation procedures, which may require the individual addressing or recovery of microdroplets.

Micromanipulation represents a second class of cell isolation approaches in which individual cells (typically a small minority of the total cells available in the sample) are directly targeted for selection and physically delivered to downstream processing steps. Micromanipulation approaches differ from random encapsulation approaches in that targeted cells are identified first and then isolated. By contrast, cells are encapsulated (isolated) prior to the identification and selection of targeted cells in the random encapsulation approach.

Where cell selection is carried out under continuous, high-resolution microscopy, micromanipulation offers the highest confidence that single cells are in fact selected and delivered to WGA reactions. Because these technologies address particular cells in the sample one at a time (rather than continuously processing a stream of randomly selected cells), they are well suited for samples containing a small number of cells overall. Micropipetting (Fig. 3e; Raghunathan *et al*., 2005; Ishøy *et al*., 2006; Kvist *et al*., 2007; Woyke *et al*., 2010; Grindberg *et al*., 2011) and microfluidic flow (Fig. 3f; Marcy *et al*., 2007a, b) were the first micromanipulation technologies applied for single-cell WGA and microbial genomics, but are relatively slow and low throughput. Microfluidic flow has the advantage of straightforward integration with microfluidic reaction chambers for downstream processing, while micropipetting requires an open platform and transfer of the targeted cell to another vessel for WGA, typically laboratory plasticware. In general, micromanipulation approaches, including optical methods, do not invoke strong shear flows, making them compatible with fragile cell morphologies.

Laser tweezing is an optical method for trapping colloidal particles (such as cells) in a solution with a refractive index that differs from the particles (Fig. 3g; Ashkin *et al*., 1986, 1987). Laser ‘tweezers’ are implemented by tightly focusing a laser beam and allowing cells to be ‘trapped’ at the location of the laser focus. ‘Trapping’ can be understood to occur because the direction of photons is changed by refraction through the particle and scattering from the particle, causing a reactive force that maintains the particle in the center of the laser focus. High laser power is required for optical trapping because the momentum carried by a single photon of light (10^{-29} kg m s^{-1}) is very small compared with the inertia of microbial cells, which have ‘large’ masses (10^{-15} kg). For this reason, on the order of 10^{18} photons s^{-1} are used for cell trapping, which corresponds to a power density of about 10^{11} W m^{-2}. Amazingly, this is more than 1000 times the luminous power density at the surface of the sun (http://nssdc.gsfc.nasa.gov/planetary/factsheet/sunfact.html).

Why are cells not destroyed when optically tweezed, given the high optical power required? In fact, cells can be trapped without ill effect on growth or activity pheno-types when trapped with near-infrared wavelengths of light in the range 750–1100 nm (Neuman *et al*., 1999; Ericsson *et al*., 2000). This is possible because in this wavelength range, only an infinitesimal fraction of the energy of the laser beam is deposited near the cell, with most of the photons passing harmlessly through the cell and the aqueous solution around it. The optical trapping method is suitable for manipulating large and filamentous bacteria (Marshall *et al*., 2012; Pamp *et al*., 2012), as well as very small cells (Blainey *et al*., 2011; Youssef *et al*., 2011; J.A. Dodsworth, P.C. Blainey, S.K. Murugapiran, W.D. Swingley, C.A. Ross, del Rio, S.G. Tringe, S.R. Quake and B.P. Hedlund, unpublished data), and is even scalable to virus particles (Ashkin & Dziedzic, 1987).

Light-driven electrokinetic technologies such as optoelectronic tweezers constitute a second class of optical micromanipulation approaches suitable for application in single-cell genomics, although these have not yet been demonstrated for this purpose (Fig. 3h). Optoelectronic tweezers can be tuned to operate in either dielectrophoretic or electroosmotic modes and depend on the interaction of applied light beams with an electrically charged flow cell (Chiou *et al*., 2003, 2005). These approaches have the advantage of operating at optical power levels several orders of magnitude lower than optical tweezers, which facilitates the implementation of highly parallelized manipulation of cells and complex confinement geometries.

Optical manipulation approaches have two key advantages for single-cell genomics: (1) the best possible sorting volumes and (2) action-at-a-distance. In optical manipulation, the sorting volume can be reduced to equal the cell volume, which allows highly concentrated cell suspensions (up to 10^8 cells μL^{-1}) to be processed directly, sample-borne contamination to be suppressed, and small amplification volumes utilized for the suppression of reagent contamination and realization of low WGA reagent costs. Action-at-a-distance manipulation enables noncontact manipulation under continuous observation inside sealed vessel, preventing extrinsic contamination in a standard laboratory environment.

Cell lysis

Cell lysis or permeabilization is a necessary requirement for genomic analyses as practical genomic analysis workflows require the application of protein, nucleic acid, and/or chemical reagents to the original genomic material to render it in a form amenable to readout.

The susceptibility of different microorganisms' envelopes to rupture by available approaches is tremendously variable. For example, some microorganisms are easily killed by low concentrations of mild surfactants under laboratory growth conditions (Miozzari & Niederberger, 1978), while spores formed by other microorganisms shrug off such mild conditions and are recalcitrant to very harsh disruption treatments (Nicholson *et al*., 2000), reflecting the extreme diversity in the physical make-up of microbial cell envelopes and the environments microorganisms have adapted to endure.

Such variable responses to lysis treatment produce representation bias in bulk experiments (cells that are not lysed drop out) and false-negative WGA results at the single-cell level (no template amplification is observed from sorted cells when lysis fails). Often, several harsh treatments are combined to improve lysis of recalcitrant subsets of cells in bulk samples. These treatments include physical disruption (e.g. sonication, freezing, bead beating, grinding, shearing, high pressure, electrical and thermal disruption), enzymatic degradation of the cell envelope and DNA-binding proteins (e.g. proteinase K, lysozyme), ionic surfactants (e.g. SDS), and extreme pH treatments. A posttreatment purification step is

commonly relied upon to remove added reagents that can interfere with subsequent procedures. While highly effective for processing bulk samples, many of these methods are not suitable for application at the single-cell level.

Single-cell genomics workflows currently exclude purification steps prior to amplification to minimize sample loss. In the extreme, only one copy of each genomic locus is present, and *any* pre-amplification sample loss constitutes locus dropout in the amplified product mixture. The lack of a purification step excludes the application of certain lysis reagents that may interfere with subsequent WGA reactions. Some physical disruption treatments subject cells to strong shear and may result in DNA breakage, which precludes the establishment of linkage across these breaks. This is a problem unique to single-cell analysis. In conventional genome sequencing, each genomic locus is represented by many DNA fragments, so reads spanning every position are present even when many breaks occur in the original sample DNA. In contrast, for single-cell sequencing, a sequence locus may only be represented by a single DNA duplex. If this duplex is broken prior to WGA, no molecules spanning the position of the break will be present in the sequencing libraries, and no reads spanning the break can be generated during sequence runs. In genome reconstruction from single-cell sequence data, contigs adjacent to the location of such a break cannot be joined.

Because some WGA methods underperform on shorter fragments and near the ends of fragments (Panelli *et al*., 2005), the likelihood of poor coverage and dropout is increased at these loci. The effect of DNA-binding proteins on WGA has not been systematically explored, although it is easy to imagine that strongly bound protein complexes inhibit the amplification of template DNA. To prevent this, conditions promoting the cleavage and/or denaturation of template-bound proteins are recommended; proteinase treatments such as proteinase K and trypsin have become standard in WGA protocols for single mammalian cells (Baslan *et al*., 2012; Wang *et al*., 2012a, b). The most popular methods of lysis for single-cell genomics are heat, nonionic surfactants, enzymatic digestion, and alkaline treatment. A complication in devising single-cell workflows is that the lysis methods chosen must be compatible with the sorting and amplification approaches taken. Because of the importance of optimal lysis for single-cell sequencing applications, the sorting technology, lysis procedures, and amplification chemistry must be carefully matched not only to one another, but also the sample type.

Amplification

WGA is an area of active development with a long history. Several approaches have been developed, all based on synthesis by DNA polymerase with various priming strategies that utilize specific, degenerate, and/or hybrid primers (Fig. 4). Single microbial cell WGA has been prosecuted almost exclusively by one method, MDA. This is particularly interesting as a variety of methods have been successfully applied to single-cell WGA of mammalian cells and ongoing development of WGA methods appears to be more active in WGA methods based on degenerate oligonucleotide-primed PCR (DOP-PCR). Over the next few years, it is likely that WGA procedures with improved characteristics and new capabilities will be introduced based on different strategies. Because of the necessity to match amplification chemistry to sorting and lysis procedures and the potential benefit of new single-cell WGA methods for microbiological studies, it is useful to review existing WGA methodologies here in a comprehensive manner.

Two early WGA methods were based on PCR with specific primers. In linker–adapter (also known as ligation-anchored) PCR (LA-PCR), adapter oligonucleotides containing specific PCR priming sites are ligated to sheared template fragments, which are then amplified by PCR (Fig. 4a; Troutt *et al*., 1992; Klein *et al*., 1999). Interspersed repetitive sequence PCR

(IRS-PCR) takes a different approach by targeting previously characterized repeating sequence elements with specific primers (Fig. 4c; Haberhausen, 1987; Ledbetter *et al*., 1990; Lengauer *et al*., 1990; Lichter *et al*., 1990). This approach has been applied to *alu* repeats in human samples, for instance.

Other methods take advantage of degenerate oligonucleotide primers that obviate the need for ligation reactions or prior knowledge of the sequence to be amplified. Primer extension preamplification PCR (PEP-PCR) introduced degenerate primers for whole-genome PCR, applying 15-mer random oligonucleotides as PCR primers under permissive thermocycling conditions, in principle enabling priming at any location in the template sequence (Fig. 4b; Hubert *et al*., 1992; Zhang *et al*., 1992). Degenerate oligonucleotide-primed PCR (DOP-PCR) uses hybrid oligos with degenerate bases at some positions to allow dense priming of the template (Fig. 4d; Telenius *et al*., 1992). Typically, DOP-PCR is run in two stages, with the first PCR stage facilitating primer extension on the template and the second PCR stage favoring amplicon replication. An interesting variant of DOP-PCR, referred to here as ‘displacement DOP-PCR’ (D-DOP-PCR, marketed as PicoPlex by Rubicon Genomics), was developed that allows strand displacement synthesis from hybrid primers during the first stage (in a fashion similar to MDA, described below), followed by the addition of specific primers that amplify the products of the first stage by PCR in the second stage (Fig. 4e; Langmore, 2002). Despite extensive development, the D-DOP-PCR method has only recently been applied to WGA of individual microorganisms (Leung *et al*., 2012).

MDA is the WGA method that has been most commonly applied in single-cell sequencing of microorganisms. MDA works by the extension of 6-mer 3’-protected random primers on the DNA template (Dean *et al*., 2001). In MDA, a polymerase with strong strand displacement activity such as phi29 DNA polymerase or *Bst* DNA polymerase creates and displaces overlapping synthesis products from the template as single-stranded DNA under isothermal conditions (Dean *et al*., 2001; Zhang *et al*., 2001; Aviel-Ronen *et al*., 2006; Fig. 4f). The displaced single-stranded DNA is a substrate for further priming and synthesis (Dean *et al*., 2001; Zhang *et al*., 2001). Phi29 DNA polymerase is typically specified for MDA due to its high accuracy owing to 3–5’ exonuclease-mediated proofreading and exceptionally strong processivity in strand displacement synthesis, which can exceed 10 000 nt (Mellado *et al*., 1980; Blanco & Salas, 1984; Blanco *et al*., 1989; Kim *et al*., 2007; Morin *et al*., 2012). This property of the polymerase evens out amplification on shorter genomic distances to produce high molecular weight products with more uniform amplification across the template than purely PCR-based methods, which typically produce products shorter than 1000 nt and exhibit greater amplification bias (Dean *et al*., 2002). In late 2012, one vendor started marketing a MDA kit that is decontaminated with ultraviolet light treatment and includes a mutant enzyme claimed to improve amplification uniformity and chimera performance (Qiagen REPLI-g Single Cell, read below for detail on chimeric sequences).

Single primer isothermal amplification (SPIA) is an isothermal strand displacement–based method that utilizes partially degenerate primers that contain a specific sequence of RNA nucleotides (Kurn *et al*., 2005). An RNA/ DNA primer, together with RNase H activity, and a strand-displacing DNA polymerase work together to achieve linear amplification under isothermal conditions, where the DNA polymerase extends the primer with DNA bases and displaces earlier product molecules, while RNase H activity degrades the RNA portion of the primer to expose the priming site and allow interaction with another primer molecule for subsequent synthesis. The product molecules are not templates for the RNA/DNA primer in this stage, preventing chain reaction (exponential) synthesis (Fig. 4g).

A new method called multiple annealing and looping–based amplification cycles (MALBAC) was recently demonstrated on individual human cells (Lu *et al*., 2012; Zong *et al*., 2012). The structure of the partially degenerate hybrid primers used in MALBAC is similar to that in D-DOP-PCR, but the sequence of their constant regions is designed to work in concert with thermal cycling during the initial reaction stage to enable quasi-linear amplification of the original template. This is accomplished by allowing products of an initial strand displacement synthesis step to be copied and to form loops by hybridization of complementary sequences on their 3’ and 5’ ends. This looping prevents doubly-tagged products from priming further synthesis or acting efficiently as templates for further synthesis under the conditions for subexponential amplification (Fig. 4h). After several rounds of thermocycled quasi-linear amplification in which priming biases are partially washed out, the thermal program is altered to enable conventional PCR amplification. In the demonstration with human cells, the MALBAC amplification was more uniform than control MDA reactions.

Single-cell WGA applications are far more sensitive to amplification bias and the formation of artificial chimeras (hereafter, simply ‘chimeras’) than multi-cell reactions. With respect to bias, multi-cell bulk WGA reactions suffer from systematic biases such as sequence-dependent priming efficiencies and primer extension rates, but stochastic variations in WGA reaction substeps are evened out, as many copies of each locus are present. On the other hand, single-cell WGA can depend on as few as one or two (double-stranded or single-stranded) copies of each locus as template and is susceptible to the random bias effects as a result. The magnitude of this random bias typically dominates systematic biases driven by sequence content in single-cell WGA applications. Some protocols for PCR-based WGA call for fragmentation of the sample prior to amplification. This practice is not recommended for single-cell WGA applications, as no sequences spanning these original break points will be present in the mixture of products.

Reduced WGA reaction volume and low-shear microfluidic sample handling not only reduce contamination of single-cell WGA products, but have also been associated with improved genomic coverage for the MDA chemistry (Marcy *et al*., 2007a, b), although the relative contributions of lysis quality, reduced DNA shearing during mixing, reduced competition from contaminants, altered reagent stability, and unknown factors to this effect have yet to be systematically investigated.

Chimeric sequences are formed as artifacts in strand displacement-driven and PCR-based DNA amplification reactions when synthesis products prime further synthesis by ‘inappropriately’ hybridizing with the template material or product molecules (Zhang *et al*., 2006; Lasken & Stockwell, 2007). The fraction of molecules in the mixture of products carrying these chimeras can be significant, often exceeding 10% and exceeding 50% in some cases (Wang & Wang, 1996; Zhang *et al*., 2006). Chimeras link template sequences that are not adjacent in the original template, creating artifacts that can be extremely disruptive to downstream analyses. Analogous to the case for amplification bias, the limited copy number of loci in the original template for single-cell WGA reactions aggravates the problem, because when further amplified, chimeras can dominate the product mixture at specific loci. In *de novo* applications, such chimeric sequences are likely to be accepted in reconstructions of the true sequence, corrupting the data set.

Particularly in small-volume reaction configurations, higher WGA product concentration can be advantageous in the context of the whole-sequencing workflow. MDA stands out in this respect as capable of producing single-cell product concentrations up to an order of magnitude higher than PCR-based WGA methods. In MDA, the single-stranded template for priming and synthesis is produced under the priming (annealing) condition, and primers do

not compete directly with amplified product molecules for template hybridization and enzyme, factors which, among others, have been implicated in causing the plateau phase of PCR (Morrison & Gannon, 1994; Kainz, 2000). PCR-based WGA methods are limited by this product inhibition effect and typically produce DNA at concentrations of 50–100 ng μL^{-1}, with the exception of the comparatively inefficient PEP-PCR method, which has been reported to produce relatively low-fold amplification in reactions with small numbers of cells (Zhang *et al*., 1992; Sun *et al*., 1995; Dietmaier *et al*., 1999).

Few systematic comparisons of the various chemistries for WGA have been made, and none that comprehensively address amplification-fold, bias, replication errors, and the incidence of chimerism, or that have been implemented in the low-volume formats that are advantageous for single-cell WGA have been carried out. Side-by-side comparisons of single-cell WGA are necessary to overcome confounding variability in cell preparation, handling, lysis, and approaches to implementing small-volume WGA reactions. Progress in the field is limited for a lack of such comparisons, leaving investigators uncertain about which approach to take, what factors are critical in implementation, and whether observed results constitute typical performance of a given chemistry.

WGA product screening

The per-base cost of sequencing has dropped much more quickly than the cost of library preparation over the last few years. Because of this, and the fact that products of individual single-cell WGA reactions are commonly handled as individual samples for sequencing, the cost and logistics of library preparation presently dominate the investment in single-cell microbial sequencing, given the modest requirement for sequence quantity on a per cell basis. This creates a tremendous incentive for selecting samples of interest and/or high-quality samples prior to library construction.

In any multi-step high-throughput workflow, samples of interest are ideally identified as early in the process as possible, in particular prior to any throughput-limiting or expensive steps. This is the reason why cell isolation approaches that allow selection of cells of interest up front based on information-rich characterization are so highly desirable, particularly in efforts where target cells are rare. Because morphotype underdetermines genotype in many microbial communities, molecular markers such as DNA-based fluorescence *in situ* hybridization (FISH) probes (Langer-Safer *et al*., 1982; N. Qvit-Raz, P.C. Blainey, E.M. Bik, S.R. Quake and D.A. Relman, unpublished data) and stains reporting metabolic capabilities or the presence of specific cell surface molecules are of great utility in this regard. Alternatively, cells of interest can be enriched prior to isolation by virtually any means that preserves the DNA in amplifiable form.

Together with pre-WGA cell selection, or independently, assessment of single-cell WGA reaction quality is very useful in compressing the sample stream prior to library preparation to improve the fraction of libraries yielding useful data. In background-free systems, lysis and reaction success can be easily screened in a binary manner by testing for the presence of DNA products, for example, by the fluorescence of an intercalating dye (Blainey & Quake, 2011). With appropriate control data relevant to the samples of interest, the pace of product appearance may provide a finer measure of reaction quality, for example, by real-time MDA (Zhang *et al*., 2006). PCR-based screening can be implemented utilizing target-specific (binary screen) or universal (sequencing-based screen) primers to identify reactions of interest. Ideally, high-quality WGA reactions on cell types of interest can be strongly enriched to minimize the burden of library creation, sequencing, and data analysis.

However, as single-cell WGA methods move to higher-throughput platforms, sequence barcoding and tagging methods are being developed that allow pooling of samples prior to library construction (Binladen *et al*., 2007; Hoffmann *et al*., 2007; Parameswaran *et al*., 2007; Hiatt *et al*., 2010), and the per-base cost of sequencing continues to drop, it is likely that the marginal value of up-front screening will be diminished, and screening for cells and reactions of interest will progressively shift to computer-based de-multiplexing and evaluation of pooled sequence data.

Sequencing and informatics

The cost, speed, data quantity, and sequence accuracy of DNA sequencing are no longer significant impediments to single-cell microbial sequencing. In addition, the tradeoff between read (or insert) length and the average number of chimeras per read in single-cell sequence data sets mutes the benefit of large insert sizes and sequencing technologies providing multi-kilobase continuous reads. This was evident in the effort to assemble a *Thiovulum* genome from single-cell WGA samples (Marshall *et al*., 2012). The real limitations in *de novo* reconstruction of single-cell genome data are the characteristics of WGA products, not the technology used to read the sequence of library molecules. Thus, while the higher molecular weight WGA products produced by MDA do provide more flexibility in the types of libraries that can be constructed, the incidence of WGA bias, WGA-induced chimerism, WGA error, and strategies to work around and work through these limitations of the data are more impactful considerations. Although associated with greater amplification bias in some cases (Dean *et al*., 2002), PCR-based techniques utilizing hybrid primers provide fragment sizes up to 1000 nt and create opportunities for workflow streamlining through the integration of addressable priming sites.

Many basic aspects of conventional sequence analysis translate to the analysis of single-cell data, for instance the importance of preprocessing data to remove adapters, linkers, barcodes, low-quality bases, and contaminants. However, conventional assembly algorithms break down in *de novo* assembly when applied to single-cell data. Sequence assembly algorithms are generally tuned with the assumption that chimeric sequences are rare and that the read depth along the genome will be Poisson-distributed, or equivalently, that reads are equally likely to come from any position along the genome (Lander & Waterman, 1988). Notwithstanding the minor sequence-content bias of DNA sequencing workflows, this assumption is generally satisfied when sequencing conventional genomic libraries. The assembler can then detect repeats as abrupt variations in average sequence coverage with some confidence and appropriately break contigs to avoid introducing artifacts. When encountering single-cell data sets with dramatic coverage variation, the assumption of Poisson-distributed read depth is violated and conventional assemblers react inappropriately by breaking contigs where coverage changes due to MDA bias. Conversely, the high incidence of chimeric sequences in single-cell data sets violates the assumption that such sequences are rare and can lead conventional assemblers to make erroneous joins.

While there are many applications of single-cell sequencing, fully *de novo* assembly of genomes is an important application class that includes direct reconstruction of microbial genomes without cultivation, generation of reference sequence for conventional metagenomic studies, and validation of genomes binned from metagenomic data sets (Hess *et al*., 2011; J.A. Dodsworth, P.C. Blainey, S.K. Murugapiran, W.D. Swingley, C.A. Ross, del Rio, S.G. Tringe, S.R. Quake and B.P. Hedlund, unpublished data; Wrighton *et al*., 2012). So, how can the bias and artificial chimera defects of single-cell shotgun sequence data be overcome to facilitate *de novo* assembly applications?

A variety of strategies to reduce the impact of WGA bias have been developed. Most simply, as bias presumably increases monotonically through the course of WGA, it seems obvious that lower amplification-fold should reduce the magnitude of bias in the products. Library preparation (Adey *et al*., 2010) and quantification procedures (White *et al*., 2009) have been developed that are more conservative of input material and potentially useful for streamlining the preparation of single cells for sequencing while simultaneously improving data quality by reducing the necessary fold-amplification for single-cell samples. Nonetheless, no significant bias reduction was observed in a paired comparison where the products of 50-nL microfluidic MDA reactions on single *E. coli* cells were split, with a portion used for direct library creation (c. 10^6-fold amplification) and sequencing, and the remaining portion re-amplified in 50-μL MDA reactions (c. 10^9-fold overall amplification), which were subsequently sampled for library creation and sequencing (P.C. Blainey, G. Schiebinger, and S.R. Quake, unpublished data). This indicates that amplification bias is established early in MDA reactions and that amplification-fold must be substantially reduced to realize a meaningful reduction in amplification bias.

Conversely, libraries of nucleic acids can be experimentally normalized to reduce amplification bias (Patanjali *et al*., 1991; Rodrigue *et al*., 2009), although this significantly increases sample preparation costs and is necessarily applied separately for each cell. Alternatively, oversequenced samples can be normalized *in silico* by removing reads that contribute redundant information in high-coverage regions (Swan *et al*., 2011). This approach is attractive given its scalability for large numbers of microbial samples with small genome sizes as costs for sequencing, data storage, and computation continue to drop.

A conceptually different approach takes advantage of the fact that the specific bias profiles and specific chimeric breakpoints arise stochastically and are nearly independent on a cell-to-cell basis (Fig. 5). This independence (and lower fold-amplification) explain why sequence data from many-cell WGA reactions have much lower bias and lower penetrance of chimeric reads at chimeric loci compared with single-cell WGA reactions. To take advantage of this effect using single-cell data, one can combine the reads obtained from several closely related individual cells (Blainey *et al*., 2011; J.A. Dodsworth, P.C. Blainey, S.K. Murugapiran, W.D. Swingley, C.A. Ross, del Rio, S.G. Tringe, S.R. Quake and B.P. Hedlund, unpublished data; Marshall *et al*., 2012; Pamp *et al*., 2012). With respect to bias, high-coverage regions from one cell will overlap with low-coverage regions in other cells, improving joint coverage of the underlying genome with much greater efficiency than would deeper sequencing of a single cell (Fig. 5a).

Recently, approaches to induce higher ploidy in sample cells have also been shown to improve coverage (Dichosa *et al*., 2012). With respect to chimeras, it is unlikely that high-penetrance chimeras would occur at exactly the same location in separate WGA reactions of different cells, because the chimeric reads at a particular locus in one single-cell data set are diluted by accurate reads at that locus in the data sets from other single cells where joint coverage occurs (Fig. 5b). In fact, an active, reference-independent strategy can be devised to deplete chimeric reads from a collection of single-cell data sets in an iterative ‘jackknifing’ procedure whereby reads from one single cell are mapped to a co-assembly of other related cells to identify chimeric reads in the first cell (J.A. Dods-worth, P.C. Blainey, S.K. Murugapiran, W.D. Swingley, C. A. Ross, del Rio, S.G. Tringe, S.R. Quake and B.P. Hedlund, unpublished data; Marshall *et al*., 2012; Pamp *et al*., 2012).

Performance by conventional assemblers improves dramatically when co-assembling four or more closely related single-cell data sets from which many chimeric reads have been filtered and can yield high-coverage assemblies with large contig sizes (Blainey *et al*., 2011; J. A. Dodsworth, P.C. Blainey, S.K. Murugapiran, W.D. Swingley, C.A. Ross, del Rio, S.G.

Tringe, S.R. Quake and B.P. Hedlund, unpublished data; Marshall *et al*., 2012; Pamp *et al*., 2012). In fact, a closed *Sulcia* genome of 243 933 bp was assembled with data obtained from a single highly polyploid cell (Woyke *et al*., 2010). It should be noted that the practice of splitting up template material across several reactions with the intent to reduce amplification bias is not expected to have any effect in the case that loci to be amplified have a copy number of one or are physically linked.

Another promising strategy is the recruitment of database metagenomic reads to single-cell data sets to improve coverage in low-coverage regions of the single-cell data set. This has been successfully applied in cases where DNA for metagenomic sequencing and cells for single-cell WGA were obtained from the same (Blainey *et al*., 2011) or different (J.A. Dodsworth, P.C. Blainey, S. K. Murugapiran, W.D. Swingley, C.A. Ross, del Rio, S.G. Tringe, S.R Quake and B.P. Hedlund, unpublished data) sources, but can be challenging in some cases (Youssef *et al*., 2011).

The informatics community is also active in developing approaches that address the challenges of assembling single-cell data head-on. For example, SmashCell is a shell environment able to wrap complex pre-processing, contamination detection/rejection, assembly, annotation, analysis, and visualization workflows while databasing results on-the-fly (Harrington *et al*., 2010). Focusing on assembly, the Velvet-SC, SPAdes, and IDBA-UD assemblers were all developed in consideration of the uneven coverage encountered in metagenomic and single-cell data sets (Chitsaz *et al*., 2011; Bankevich *et al*., 2012; Peng *et al*., 2012). Velvet-SC introduced the idea of a variable coverage cutoff parameter, while IDBA-UD introduced the concepts of variable relative coverage cutoffs, built-in kmer size iteration, and error correction in high-depth regions. SPAdes extends Velvet-SC to take advantage of paired reads, incorporates k-bimers, and introduces multi-sized assembly graphs to address sequence errors and chimeras.

Although each of these methods has been shown to outperform standard assemblers on hand-picked single-cell data sets, it is difficult to evaluate their real performance potential as unlike bulk data sets, single-cell data sets from WGA samples are so highly variable, even when originating from essentially identical cells. Another confounding factor is the responsiveness of assembly algorithms to species-specific and data set-specific parameter tuning. Naturally, the authors of a given assembler are both the most capable and the most motivated to optimally tune assembly of a given data set. A robust single-cell assembly performance comparison would require testing on many independent single-cell data sets from several organisms with varying genomic structure where standard or algorithmically defined assembly parameter sets are applied. The comparison would be difficult to run reference-blinded, as single-cell data sets from organisms for which secret reference data are available are difficult to come by. Alternatively, the assembly parameters could be fixed prior to the identification of test data sets.

Despite the challenges of assembling single-cell sequence data, sufficiently accurate assemblies can be obtained to allow for the detection of small-scale variability between individual organisms and even differences between cells in small clonal filaments (Pamp *et al*., 2012).

Conclusion

DNA sequencing based on massively parallel clonal DNA amplification is a maturing technology, and per-base sequencing costs no longer dominate the expense of single-cell microbial genomic analyses. Library creation is an area of rapid development, and more efficient preparation procedures have tremendous potential for workflow streamlining and

bias reduction through lower input requirements. Anticipating a growing role of single-cell microbial sequence data in microbial genomics generally, and the desire to apply increasingly demanding analyses, quality control of single-cell WGA products is becoming increasingly important.

Technological advances are improving the reliability and throughput of single-cell WGA and sequencing. The contamination of reagents with bacterial DNA that plagued early efforts in single-cell microbial genomics is now well understood, and effective measures to quantify, control, and eliminate such contamination have been developed. The elimination of background amplification (Blainey & Quake, 2011) not only yields clean data sets, but also allows the facile application of screens for successful WGA reactions. The move to lower-volume WGA reaction platforms further reduces the impact of contaminates and WGA reagent costs, opening the way to massive increases in reaction throughput.

The key drivers of new technology for single-cell genomics will be advances in throughput, integration of isolation and WGA with selection and sequencing procedures, improvements in WGA product quality, expanded sample/cell type compatibility, and contextualization of single-cell genome data by in-line collection of other information about the processed cells (e.g. phenotyping, imaging, RNA, protein, metabolite analyses; Wang & Bodovitz, 2010; Darmanis *et al*., 2012). The ability to obtain sequence data from individual cells from a known biological setting or with a known history (e.g. interaction with other cells prior to the analysis) is expected to be especially informative. Because sample preparation costs already dominate single-cell microbial genomics workflows today, the most impactful advances will scale to large numbers of cells at reasonable cost. Closely related technical advances are also being applied to the sequencing of single human cells in research and biomedical applications (Navin *et al*., 2011; Fan *et al*., 2012; Wang *et al*., 2012a, b; Lu *et al*., 2012; Zong *et al*., 2012).

Single-cell microbial genomics is a rapidly growing field with a significant role to play in both extending and pulling together our understanding of microbial communities in the environment. Single-cell genomics can be applied in different ways, from establishing linkage of key genes, to *de novo* reconstruction of new genomes, to whole-genome population studies targeted at the ultimate biological resolution. The diversity of technological approaches for single-cell genomics that are being developed is advantageous for the field, as no single method for cell isolation or processing is ideal for all the different organisms, sample types, and study designs of interest.

Acknowledgments

Stephen Quake played a key role in the development of several of the microfluidic and optical technologies described in this review as well as the author's introduction to these and single-cell genomic analysis. The NIH supported the author's work in single-cell genomics through R01 HG004863 to David Relman and Stephen Quake, as did the Burroughs Wellcome Fund, through a Career Award at the Scientific Interface to the author. The Stanford Microfluidics Foundry provided substantive support to the author's work in the field, while innumerable collaborators and members of the microbiology community at Stanford supported his introduction to microbial ecology and genomics. In particular, Alfred Spormann and Christopher Francis supported the author's repeated participation in the Hopkins Microbiology Course, while David Relman served as a key long-term collaborator and source of insight into the human microbiome. Geoffrey Schiebinger contributed to work on real-time MDA and analysis of single-cell data from low-input libraries. The author's broader thinking in the single-cell sequencing field was influenced by the work of and discussions with many in the community, particularly Roger Lasken, Stephen Quake, Ramunas Stepanauskas, and Tanja Woyke.

References

Abate AR, Hung T, Mary P, Agresti JJ, Weitz DA. High-throughput injection with microfluidics using picoinjectors. Proc Natl Acad Sci. 2010; 107:19163–19166. [PubMed: 20962271]

Adey A, Morrison HG, Asan XX, et al. Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density *in vitro* transposition. Genome Biol. 2010; 11:R119. [PubMed: 21143862]

Agresti JJ, Antipov E, Abate AR, Ahn K, Rowat AC, Baret JC, Marquez M, Klibanov AM, Griffiths AD, Weitz DA. Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. Proc Natl Acad Sci USA. 2010; 107:4004–4009. [PubMed: 20142500]

Allen LZ, Ishoey T, Novotny MA, McLean JS, Lasken RS, Williamson SJ. Single virus genomics: a new tool for virus discovery. PLoS ONE. 2011; 6:el7722.

Ashkin A, Dziedzic JM. Optical trapping and manipulation of viruses and bacteria. Science. 1987; 235:1517–1520. [PubMed: 3547653]

Ashkin A, Dziedzic JM, Bjorkholm JE, Chu S. Observation of a single-beam gradient force optical trap for dielectric particles. Opt Lett. 1986; 11:288. [PubMed: 19730608]

Ashkin A, Dziedzic JM, Yamane T. Optical trapping and manipulation of single cells using infrared laser beams. Nature. 1987; 330:769–771. [PubMed: 3320757]

Aviel-Ronen S, Zhu CQ, Coe BP, Liu N, Watson SK, Lam WL, Tsao MS. Large fragment Bst DNA polymerase for whole genome amplification of DNA from formalin-fixed paraffin-embedded tissues. BMC Genomics. 2006; 7:312–321. [PubMed: 17156491]

Bai F, Anderson W, Moo-Young M. Ethanol fermentation technologies from sugar and starch feedstocks. Biotechnol Adv. 2008; 26:89–105. [PubMed: 17964107]

Baker M. Digital PCR hits its stride. Nat Methods. 2012; 9:541–544.

Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012; 19:455–477. [PubMed: 22506599]

Baslan T, Kendall J, Rodgers L, et al. Genome-wide copy number analysis of single cells. Nat Protoc. 2012; 7:1024–1041. [PubMed: 22555242]

Bergquist PL, Hardiman EM, Ferrari BC, Winsley T. Applications of flow cytometry in environmental microbiology and biotechnology. Extremophiles. 2009; 13:389–401. [PubMed: 19301090]

Berry AE, Chiocchini C, Selby T, Sosio M, Wellington EM. Isolation of high molecular weight DNA from soil for cloning into BAC vectors. FEMS Microbiol Lett. 2003; 223:15–20. [PubMed: 12798994]

Binga EK, Lasken RS, Neufeld JD. Something from (almost) nothing: the impact of multiple displacement amplification on microbial ecology. LSME J. 2008; 2:233–241.

Binladen J, Gilbert MTP, Bollback JP, Panitz F, Bendixen C, Nielsen R, Willerslev E. The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. PLoS ONE. 2007; 2:el97.

Blainey PC, Quake SR. Digital MDA for enumeration of total nucleic acid contamination. Nucleic Acids Res. 2011; 39:el9.

Blainey PC, Mosier AC, Potanina A, Francis CA, Quake SR. Genome of a low-salinity ammonia-oxidizing archaeon determined by single-cell and metagenomic analysis. PLoS ONE. 2011; 6:el6626.

Blanco L, Salas M. Characterization and purification of a phage phi 29-encoded DNA polymerase required for the initiation of replication. Proc Natl Acad Sci. 1984; 81:5325. [PubMed: 6433348]

Blanco L, Bernad A, Lázaro JM, Martin G, Garmendia C, Salas M. Highly efficient DNA synthesis by the phage phi 29 DNA polymerase. Symmetrical mode of DNA replication. J Biol Chem. 1989; 264:8935–8940. [PubMed: 2498321]

Brussaard L. Biodiversity and ecosystem functioning in soil. Ambio. 1997; 26:563–570.

Bult CJ, White O, Olsen GJ, et al. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii* . Science. 1996; 273:1058–1073. [PubMed: 8688087]

Caro-Quintero A, Konstantinidis KT. Bacterial species may exist, metagenomics reveal. Environ Microbiol. 2012; 14:347–355. [PubMed: 22151572]

Chiou PY, Chang Z, Wu MC. A novel optoelectronic tweezer using light induced dielectrophoresis. IEEE. 2003:8–9.

Chiou PY, Ohta AT, Wu MC. Massively parallel manipulation of single cells and microparticles using optical images. Nature. 2005; 436:370–372. [PubMed: 16034413]

Chitsaz H, Yee-Greenbaum JL, Tesler G, et al. Efficient de novo assembly of single-cell bacterial genomes from short-read data sets. Nat Biotechnol. 2011; 29:915–922. [PubMed: 21926975]

Connon SA, Giovannoni SJ. High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. Appl Environ Microbiol. 2002; 68:3878–3885. [PubMed: 12147485]

Darmanis S, Gallant C, Landegren U. PCR-based multiparametric assays in single cells. Clin Chem. 2012; 58:1618–1619. [PubMed: 23071363]

Dean FB, Nelson JR, Giesler TL, Lasken RS. Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification. Genome Res. 2001; 11:1095–1099. [PubMed: 11381035]

Dean FB, Hosono S, Fang L, et al. Comprehensive human genome amplification using multiple displacement amplification. Proc Natl Acad Sci. 2002; 99:5261. [PubMed: 11959976]

Denef VJ, Banfield JF. *In situ* evolutionary rate measurements show ecological success of recently emerged bacterial hybrids. Science. 2012; 336:462–466. [PubMed: 22539719]

Dichosa AE, Fitzsimons MS, Lo CC, et al. Artificial polyploidy improves bacterial single cell genome recovery. PLoS ONE. 2012; 7:e37387. [PubMed: 22666352]

Dick GJ, Andersson AF, Baker BJ, Simmons SL, Thomas BC, Yelton AP, Banfield JF. Community-wide analysis of microbial genome sequence signatures. Genome Biol. 2009; 10:R85. [PubMed: 19698104]

Dietmaier W, Hartmann A, Wallinger S, Heinmöller E, Kerner T, Endl E, Jauch KW, Hofstädter F, Rüschoff J. Multiple mutation analyses in single tumor cells with improved whole genome amplification. Am J Pathol. 1999; 154:83–95. [PubMed: 9916922]

Dupont CL, Rusch DB, Yooseph S, et al. Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. ISME J. 2012; 78:8555–8563.

Ericsson M, Hanstorp D, Hagberg P, Enger J, Nystrom T. Sorting out bacterial viability with optical tweezers. J Bacteriol. 2000; 182:5551–5555. [PubMed: 10986260]

Fan HC, Gu W, Wang JB, Blumenfeld YJ, El-Sayed YY, Quake SR. Non-invasive prenatal measurement of the fetal genome. Nature. 2012; 487:320–324. [PubMed: 22763444]

Findlay I. Single cell PCR. Methods Mol Med. 1998; 16:233–263. [PubMed: 21390789]

Fleischmann RD, Adams MD, White O, et al. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science. 1995; 269:496–512. [PubMed: 7542800]

Fox BC, Devonshire AS, Baradez MO, Marshall D, Foy CA. Comparison of RT-qPCR methods and platforms for single cell gene expression analysis. Anal Biochem. 2012; 427:178–186. [PubMed: 22617801]

Fritzsch FS, Dusny C, Frick O, Schmid A. Single-cell analysis in biotechnology, systems biology, and biocatalysis. Annu Rev Chem Biomol Eng. 2012; 3:129–155. [PubMed: 22468600]

Gevers D, Cohan FM, Lawrence JG, et al. Re-evaluating prokaryotic species. Nat Rev Microbiol. 2005; 3:733–739. [PubMed: 16138101]

Grindberg RV, Ishoey T, Brinza D, et al. Single cell genome amplification accelerates identification of the apratoxin biosynthetic pathway from a complex microbial assemblage. PLoS ONE. 2011; 6:el8565.

Haberhausen, G. PCR: Overview on Application Formats in Research and Clinical Diagnosis. Vol. Vol. 289. Springer Verlag: Berlin; 1987. p. 327

Hallam SJ, Mincer TJ, Schleper C, Preston CM, Roberts K, Richardson PM, DeLong EF. Pathways of carbon assimilation and ammonia oxidation suggested by environmental genomic analyses of marine Crenarchaeota. PLoS Biol. 2006a; 4:e95. [PubMed: 16533068]

Hallam SJ, Konstantinidis KT, Putnam N, et al. Genomic analysis of the uncultivated marine crenarchaeote *Cenarchaeum symbiosum* . P Natl Acad Sci USA. 2006b; 103:18296–18301.

Harrington ED, Arumugam M, Raes J, Bork P, Relman DA. SmashCell: a software framework for the analysis of single-cell amplified genome sequences. Bioinformatics. 2010; 26:2979–2980. [PubMed: 20966005]

Hess M, Sczyrba A, Egan R, et al. Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. Science. 2011; 331:463. [PubMed: 21273488]

Hiatt JB, Patwardhan RP, Turner EH, Lee C, Shendure J. Parallel, tag-directed assembly of locally derived short sequence reads. Nat Methods. 2010; 7:119–122. [PubMed: 20081835]

Hoffmann C, Minkah N, Leipzig J, Wang G, Arens MQ, Tebas P, Bushman FD. DNA bar coding and pyrosequencing to identify rare HIV drug resistance mutations. Nucleic Acids Res. 2007; 35:e91. [PubMed: 17576693]

Hubert R, Weber J, Schmitt K, Zhang L, Arnheim N. A new source of polymorphic DNA markers for sperm typing: analysis of microsatellite repeats in single cells. Am J Hum Genet. 1992; 51:985. [PubMed: 1415267]

Ishoey T, Woyke T, Stepanauskas R, Novotny M, Lasken RS. Genomic sequencing of single microbial cells from environmental samples. Curr Opin Microbiol. 2008; 11:198–204. [PubMed: 18550420]

Ishøy T, Kvist T, Westermann P, Ahring BK. An improved method for single cell isolation of prokaryotes from meso-, thermo-and hyperthermophilic environments using micromanipulation. Appl Microbiol Biotechnol. 2006; 69:510–514. [PubMed: 16034558]

de Jager V, Siezen RJ. Single-cell genomics: unravelling the genomes of unculturable microorganisms. Microb Biotechnol. 2011; 4:431–437. [PubMed: 21733126]

Kainz P. The PCR plateau phase-towards an understanding of its limitations. Biochim Biophys Acta. 2000; 1494:23–27. [PubMed: 11072065]

Kalisky T, Quake SR. Single-cell genomics. Nat Methods. 2011; 8:311–314. [PubMed: 21451520]

Kalisky T, Blainey P, Quake SR. Genomic analysis at the single-cell level. Annu Rev Genet. 2011; 45:431–445. [PubMed: 21942365]

Kamke J, Bayer K, Woyke T, Hentschel U. Exploring symbioses by single-cell genomics. Biol Bull. 2012; 223:30–43. [PubMed: 22983031]

Kim S, Blainey PC, Schroeder CM, Xie XS. Multiplexed single-molecule assay for enzymatic activity on flow-stretched DNA. Nat Methods. 2007; 4:397–399. [PubMed: 17435763]

Klein CA, Schmidt-Kittler O, Schardt JA, Pantel K, Speicher MR, Riethmuller G. Comparative genomic hybridization, loss of heterozygosity, and DNA sequence analysis of single cells. P Natl Acad Sci USA. 1999; 96:4494–4499.

Koeppel A, Perry EB, Sikorski J, et al. Identifying the fundamental units of bacterial diversity: a paradigm shift to incorporate ecology into bacterial systematics. P Natl Acad Sci USA. 2008; 105:2504–2509.

Koonin EV, Makarova KS, Aravind L. Horizontal gene transfer in prokaryotes: quantification and classification 1. Annu Rev Microbiol. 2001; 55:709–742. [PubMed: 11544372]

Kuppers R, Zhao M, Hansmann M, Rajewsky K. Tracing B cell development in human germinal centres by molecular analysis of single cells picked from histological sections. EMBO J. 1993; 12:4955. [PubMed: 8262038]

Kurn N, Chen PC, Heath JD, Kopf-Sill A, Stephens KM, Wang SL. Novel isothermal, linear nucleic acid amplification systems for highly multiplexed applications. Clin Chem. 2005; 51:1973–1981. [PubMed: 16123149]

Kvist T, Ahring BK, Lasken RS, Westermann P. Specific single-cell isolation and genomic amplification of uncultured microorganisms. Appl Microbiol Biotechnol. 2007; 74:926–935. [PubMed: 17109170]

Lander ES, Waterman MS. Genomic mapping by fingerprinting random clones: a mathematical analysis. Genomics. 1988; 2:231–239. [PubMed: 3294162]

Langer-Safer PR, Levine M, Ward DC. Immunological method for mapping genes on Drosophila polytene chromosomes. Proc Natl Acad Sci. 1982; 79:4381. [PubMed: 6812046]

Langmore JP. Rubicon genomics, Inc. Pharmacogenomics. 2002; 3:557–560. [PubMed: 12164778]

Lasken RS. Single-cell genomic sequencing using multiple displacement amplification. Curr Opin Microbiol. 2007; 10:510–516. [PubMed: 17923430]

Lasken RS. Genomic sequencing of uncultured microorganisms from single cells. Nat Rev Microbiol. 2012; 10:631–640. [PubMed: 22890147]

Lasken R, Stockwell T. Mechanism of chimera formation during the multiple displacement amplification reaction. BMC Biotechnol. 2007; 7:19. [PubMed: 17430586]

Lecault V, White AK, Singhal A, Hansen CL. Microfluidic single cell analysis: from promise to practice. Curr Opin Chem Biol. 2012; 16:381–390. [PubMed: 22525493]

Ledbetter SA, Nelson DL, Warren ST, Ledbetter DH. Rapid isolation of DNA probes within specific chromosome regions by interspersed repetitive sequence polymerase chain reaction. Genomics. 1990; 6:475. [PubMed: 2328990]

Lengauer C, Riethman H, Cremer T. Painting of human chromosomes with probes generated from hybrid cell lines by PCR with Alu and LI primers. Hum Genet. 1990; 86:1–6. [PubMed: 1701413]

Leung K, Zahn H, Leaver T, et al. A programmable droplet-based microfluidic device applied to multiparameter analysis of single microbes and microbial communities. P Natl Acad Sci USA. 2012; 109:7665–7670.

Lichter P, Ledbetter SA, Ledbetter DH, Ward DC. Fluorescence *in situ* hybridization with Alu and L1 polymerase chain reaction probes for rapid characterization of human chromosomes in hybrid cell lines. Proc Natl Acad Sci. 1990; 87:6634. [PubMed: 2395866]

Link DR, Grasland Äê Mongrain E, Duri A, et al. Electric control of droplets in microfluidic devices. Angew Chem Int Ed. 2006; 45:2556–2560.

Liu H, Nolla HA, Campbell L. Prochlorococcus growth rate and contribution to primary production in the equatorial and subtropical North Pacific Ocean. Aquat Microb Ecol. 1997; 12:39–47.

Love JC, Ronan JL, Grotenbreg GM, Van der Veen AG, Ploegh HL. A microengraving method for rapid selection of single cells producing antigen-specific antibodies. Nat Biotechnol. 2006; 24:703–707. [PubMed: 16699501]

Lu S, Zong C, Fan W, et al. Probing meiotic recombination and aneuploidy of single sperm cells by Whole-genome sequencing. Science. 2012; 338:1627–1630. [PubMed: 23258895]

Luo C, Tsementzi D, Kyrpides NC, Konstantinidis KT. Individual genome assembly from complex community short-read metagenomic datasets. ISME J. 2011; 6:898–901. [PubMed: 22030673]

Marcy Y, Ishoey T, Lasken RS, Stockwell TB, Walenz BP, Halpern AL, Beeson KY, Goldberg SMD, Quake SR. Nanoliter reactors improve multiple displacement amplification of genomes from single cells. PLoS Genet. 2007a; 3:1702–1708. [PubMed: 17892324]

Marcy Y, Ouverney C, Bik EM, et al. Dissecting biological “dark matter” with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. P Natl Acad Sci USA. 2007b; 104:11889–11894.

Marshall IPG, Blainey PC, Spormann AM, Quake SR. A single-cell genome for *Thiovulum* sp. Appl Environ Microbiol. 2012; 78:8555–8563. [PubMed: 23023751]

Maryanski JL, Jongeneel CV, Bucher P, Casanova JL, Walker PR. Single-cell PCR analysis of TCR repertoires selected by antigen *in vivo* a high magnitude CD8 response is comprised of very few clones. Immunity. 1996; 4:47–55. [PubMed: 8574851]

Matson, PA.; Vitousek, PM.; Chapin, MC. Principles of Terrestrial Ecosystem Ecology. New York: Springer; 2011.

Mavromatis K, Ivanova N, Barry K, et al. Use of simulated data sets to evaluate the fidelity of metagenomic processing methods. Nat Methods. 2007; 4:495–500. [PubMed: 17468765]

Mellado RP, Penalva MA, Inciarte MR, Salas M. The protein covalently linked to the 5’ termini of the DNA of *Bacillus subtilis* phage phi 29 is involved in the initiation of DNA replication. Virology. 1980; 104:84–96. [PubMed: 6771916]

Miozzari G, Niederberger P. Permeabilization of microorganisms by Triton X-100. Anal Biochem. 1978; 90:220–233. [PubMed: 365019]

Morin JA, Cao FJ, Lázaro JM, et al. Active DNA unwinding dynamics during processive DNA replication. P Natl Acad Sci USA. 2012; 109:8115–8120.

Morrison C, Gannon F. The impact of the PCR plateau phase on quantitative PCR. Biochim Biophys Acta. 1994; 1219:493–498. [PubMed: 7918647]

Namiki T, Hachiya T, Tanaka H, Sakakibara Y. MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. ACM. 2011:116–124.

Navin N, Kendall J, Troge J, et al. Tumour evolution inferred by single-cell sequencing. Nature. 2011; 472:90–94. [PubMed: 21399628]

Neuman KC, Chadd EH, Liou GF, Bergman K, Block SM. Characterization of photodamage to *Escherichia coli* in optical traps. Biophys J. 1999; 77:2856–2863. [PubMed: 10545383]

Nicholson WL, Munakata N, Horneck G, Melosh HJ, Setlow P. Resistance of *Bacillus* endospores to extreme terrestrial and extraterrestrial environments. Microbiol Mol Biol Rev. 2000; 64:548–572. [PubMed: 10974126]

Ottesen EA, Hong JW, Quake SR, Leadbetter JR. Microfluidic digital PCR enables multigene analysis of individual environmental bacteria. Science. 2006; 314:1464–1467. [PubMed: 17138901]

Pamp SJ, Harrington ED, Quake SR, Relman DA, Blainey PC. Single-cell sequencing provides clues about the host interactions of segmented filamentous bacteria (SFB). Genome Res. 2012; 22:1107–1119. [PubMed: 22434425]

Panelli S, Damiani G, Espen L, Sgaramella V. Ligation overcomes terminal underrepresentation in multiple displacement amplification of linear DNA. Biotechniques. 2005; 39:174. [PubMed: 16116788]

Parameswaran P, Jalili R, Tao L, Shokralla S, Gharizadeh B, Ronaghi M, Fire AZ. A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing. Nucleic Acids Res. 2007; 35:el30.

Patanjali SR, Parimoo S, Weissman SM. Construction of a uniform-abundance (normalized) cDNA library. P Natl Acad Sci USA. 1991; 88:1943.

Peng Y, Leung HCM, Yiu S, Chin FYL. Meta-IDBA: a de Novo assembler for metagenomic data. Bioinformatics. 2011; 27:94–il01.

Peng Y, Leung H, Yiu S, Chin FYL. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics. 2012; 28:1420–1428. [PubMed: 22495754]

Podar M, Abulencia CB, Walcher M, et al. Targeted access to the genomes of low-abundance organisms in complex microbial communities. Appl Environ Microbiol. 2007; 73:3205–3214. [PubMed: 17369337]

Prescott, SC.; Dunn, CG. Industrial microbiology. New York: McGraw-Hill; 1949.

Proctor LM. The human microbiome project in 2011 and beyond. Cell Host Microbe. 2011; 10:287–291. [PubMed: 22018227]

Raghunathan A, Ferguson HR Jr, Bornarth CJ, Song W, Driscoll M, Lasken RS. Genomic DNA amplification from a single bacterium. Appl Environ Microbiol. 2005; 71:3342–3347. [PubMed: 15933038]

Relman DA. Microbial genomics and infectious diseases. N Engl J Med. 2011; 365:347–357. [PubMed: 21793746]

Rodrigue S, Malmstrom RR, Berlin AM, Birren BW, Henn MR, Chisholm SW. Whole genome amplification and de novo assembly of single bacterial cells. PLoS ONE. 2009; 4:e6864. [PubMed: 19724646]

Ruiz Sebastián C, O'ryan C. Single-cell sequencing of dinoflagellate (Dinophyceae) nuclear ribosomal genes. Mol Ecol Notes. 2001; 1:329–331.

Sanchez-Freire V, Ebert AD, Kalisky T, Quake SR, Wu JC. Microfluidic single-cell real-time PCR for comparative analysis of gene expression patterns. Nat Protoc. 2012; 7:829–838. [PubMed: 22481529]

Sekar R, Fuchs BM, Amann R, Pernthaler J. Flow sorting of marine bacterioplankton after fluorescence *in situ* hybridization. Appl Environ Microbiol. 2004; 70:6210–6219. [PubMed: 15466568]

Shapiro BJ, Friedman J, Cordero OX, et al. Population genomics of early events in the ecological differentiation of bacteria. Science. 2012; 336:48–51. [PubMed: 22491847]

Smillie CS, Smith MB, Friedman J, Cordero OX, David LA, Alm EJ. Ecology drives a global network of gene exchange connecting the human microbiome. Nature. 2011; 480:241–244. [PubMed: 22037308]

Stein JL, Marsh TL, Wu KY, Shizuya H, DeLong EF. Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. J Bacteriol. 1996; 178:591–599. [PubMed: 8550487]

Stepanauskas R. Single cell genomics: an individual look at microbes. Curr Opin Microbiol. 2012; 15:613–620. [PubMed: 23026140]

Stepanauskas R, Sieracki ME. Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. P Natl Acad Sci USA. 2007; 104:9052–9057.

Sucher NJ, Deitcher DL. PCR and patch-clamp analysis of single neurons. Neuron. 1995; 14:1095–1100. [PubMed: 7541630]

Sun F, Arnheim N, Waterman MS. Whole genome amplification of single cells: mathematical analysis of PEP and tagged PCR. Nucleic Acids Res. 1995; 23:3034–3040. [PubMed: 7659528]

Swan BK, Martinez-Garcia M, Preston CM, et al. Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. Science. 2011; 333:1296–1300. [PubMed: 21885783]

Tadmor AD, Ottesen EA, Leadbetter JR, Phillips R. Probing individual environmental bacteria for viruses by using microfluidic digital PCR. Science. 2011; 333:58. [PubMed: 21719670]

Tanaseichuk, O.; Borneman, J.; Jiang, T. Separating metagenomic short reads into genomes via clustering. In: Przytyckaand, TM.; Sagot, M-F., editors. Algorithms in Bioinformatics. Vol. Vol. 6833. Berlin: Springer; 2011. p. 298-313.

Telenius H, Carter NP, Bebb CE, Ponder BAJ, Tunnacliffe A. Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. Genomics. 1992; 13:718–725. [PubMed: 1639399]

Tewhey R, Warner JB, Nakano M, et al. Microdroplet- based PCR enrichment for large-scale targeted sequencing. Nat Biotechnol. 2009; 27:1025–1031. [PubMed: 19881494]

Thorsen T, Roberts RW, Arnold FH, Quake SR. Dynamic pattern formation in a vesicle-generating microfluidic device. Phys Rev Lett. 2001; 86:4163–4166. [PubMed: 11328121]

Treangen TJ, Koren S, Astrovskaya I, Sommer D, Liu B, Pop M. MetAMOS: a metagenomic assembly and analysis pipeline for AMOS. Genome Biol. 2011; 12:1–27.

Troutt AB, McHeyzer-Williams MG, Pulendran B, Nossal G. Ligation-anchored PCR: a simple amplification technique with single-sided specificity. Proc Natl Acad Sci. 1992; 89:9823. [PubMed: 1409706]

Tyson GW, Chapman J, Hugenholtz P, et al. Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature. 2004; 428:37–43. [PubMed: 14961025]

Venter JC, Remington K, Heidelberg JF, et al. Environmental genome shotgun sequencing of the Sargasso Sea. Science. 2004; 304:66–74. [PubMed: 15001713]

Wallner G, Fuchs B, Spring S, Beisker W, Amann R. Flow sorting of microorganisms for molecular analysis. Appl Environ Microbiol. 1997; 63:4223–4231. [PubMed: 9361408]

Wang D, Bodovitz S. Single cell analysis: the new frontier in ‘omics’. Trends Biotechnol. 2010; 28:281–290. [PubMed: 20434785]

Wang GCY, Wang Y. The frequency of chimeric molecules as a consequence of PCR co-amplification of 16S rRNA genes from different bacterial species. Microbiology. 1996; 142:1107–1114. [PubMed: 8704952]

Wang J, Zhou Y, Qiu H, Huang H, Sun C, Xi J, Huang Y. A chip-to-chip nanoliter microfluidic dispenser. Lab Chip. 2009; 9:1831–1835. [PubMed: 19532955]

Wang JB, Fan HC, Behr B, Quake SR. Genome-wide single-cell analysis of recombination activity and de novo mutation rates in human sperm. Cell. 2012a; 150:402–412. [PubMed: 22817899]

Wang Y, Leung HCM, Yiu S, Chin FYL. MetaCluster 4.0: a novel binning algorithm for NGS reads and huge number of species. J Comput Biol. 2012b; 19:241–249. [PubMed: 22300323]

White RA III, Blainey PC, Fan HC, Quake SR. Digital PCR provides sensitive and absolute calibration for high throughput sequencing. BMC Genomics. 2009; 10:116. [PubMed: 19298667]

Woyke T, Teeling H, Ivanova NN, et al. Symbiosis insights through metagenomic analysis of a microbial consortium. Nature. 2006; 443:950–955. [PubMed: 16980956]

Woyke T, Xie G, Copeland A, et al. Assembling the marine metagenome, one cell at a time. PLoS ONE. 2009; 4:e5299. [PubMed: 19390573]

Woyke T, Tighe D, Mavromatis K, et al. One bacterial cell, one complete genome. PLoS ONE. 2010; 5:e10314. [PubMed: 20428247]

Woyke T, Sczyrba A, Lee J, Rinke C, Tighe D, Clingenpeel S, Malmstrom R, Stepanauskas R, Cheng JF. Decontamination of MDA reagents for single cell whole genome amplification. PLoS ONE. 2011; 6:e26161. [PubMed: 22028825]

Wrighton KC, Thomas BC, Sharon I, et al. Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. Science. 2012; 337:1661–1665. [PubMed: 23019650]

Wu D, Hugenholtz P, Mavromatis K, et al. A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. Nature. 2009; 462:1056–1060. [PubMed: 20033048]

Yilmaz S, Singh AK. Single cell genome sequencing. Curr Opin Biotechnol. 2012; 23:437–443. [PubMed: 22154471]

Yilmaz S, Allgaier M, Hugenholtz P. Multiple displacement amplification compromises quantitative analysis of metagenomes. Nat Methods. 2010; 7:943–944. [PubMed: 21116242]

Yoon HS, Price DC, Stepanauskas R, et al. Single-cell genomics reveals organismal interactions in uncultivated marine protists. Science. 2011; 332:714. [PubMed: 21551060]

Youssef NH, Blainey PC, Quake SR, Elshahed MS. Partial genome assembly for a candidate division OP 11 single cell from an anoxic spring (Zodletone Spring, Oklahoma). Appl Environ Microbiol. 2011; 77:7804–7814. [PubMed: 21908640]

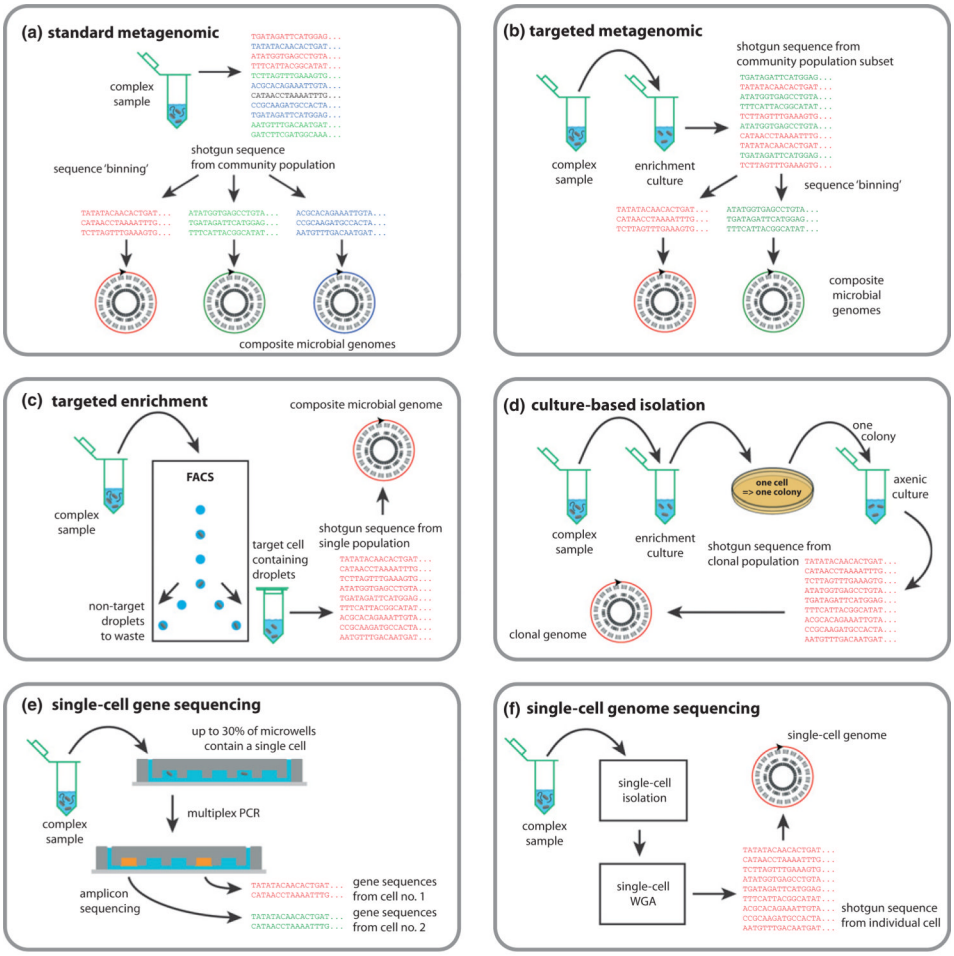
Zeng Y, Novak R, Shuga J, Smith MT, Mathies RA. High-performance single cell genetic analysis using microfluidic emulsion generator arrays. Anal Chem. 2010; 82:3183–3190. [PubMed: 20192178]

Zhang L, Cui X, Schmitt K, Hubert R, Navidi W, Arnheim N. Whole genome amplification from a single cell: implications for genetic analysis. P Natl Acad Sci USA. 1992; 89:5847.

Zhang DY, Brandwein M, Hsuih T, Li HB. Ramification amplification: a novel isothermal DNA amplification method. Mol Diagn. 2001; 6:141–150. [PubMed: 11468700]

Zhang K, Martiny AC, Reppas NB, Barry KW, Malek J, Chisholm SW, Church GM. Sequencing genomes from single cells by polymerase cloning. Nat Biotechnol. 2006; 24:680–686. [PubMed: 16732271]

Zong C, Lu S, Chapman AR, Xie XS. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. Science. 2012; 338:1622–1626. [PubMed: 23258894]

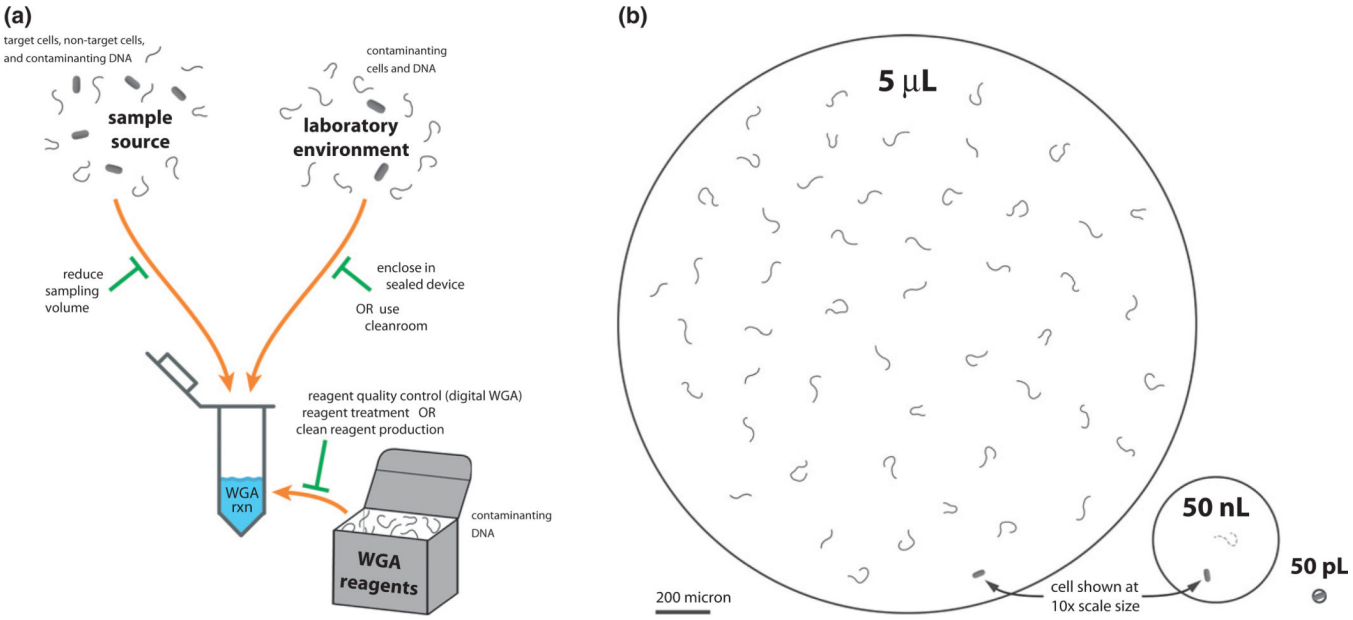


(g) genome sequencing approaches

sample preparation approach	method-specific requirements					
	lab culture	read 'binning'	cell isolation	amplification	enabling technology	data type
standard metagenomic	no	yes	no	no	informatics	composite genome
targeted metagenomic	yes, enrichment	yes	no	no	informatics	composite genome
targeted enrichment	no	no	yes	often WGA	flow cytometry	composite genome
culture-based isolation	yes, axenic	no	often yes	no	traditional method	clonal genome
single-cell (gene sequencing)	no	no	yes	yes, PCR	cell isolation	linked gene sequences
single-cell (genome sequencing)	no	no	yes	yes, WGA	contamination-suppression	single-cell genome

Fig. 1.

Methods for microbial genomics. (a) Standard metagenomics and sequence 'binning' to produce composite microbial genomes. (b) Targeted metagenomics and sequence 'binning' to produce composite microbial genomes. (c) Targeted enrichment of an organism to produce a single composite microbial genome. (d) Culture-based isolation for production of an axenic culture and a clonal microbial genome. (e) Multiplex PCR-based single-cell gene sequencing to obtain the sequence of multiple loci in single cells. (f) Single-cell genome sequencing utilizes cell isolation and single-cell WGA to produce single-cell microbial genomes. (g) Table summarizing characteristics of the genomic methods indicated in parts (a–f).

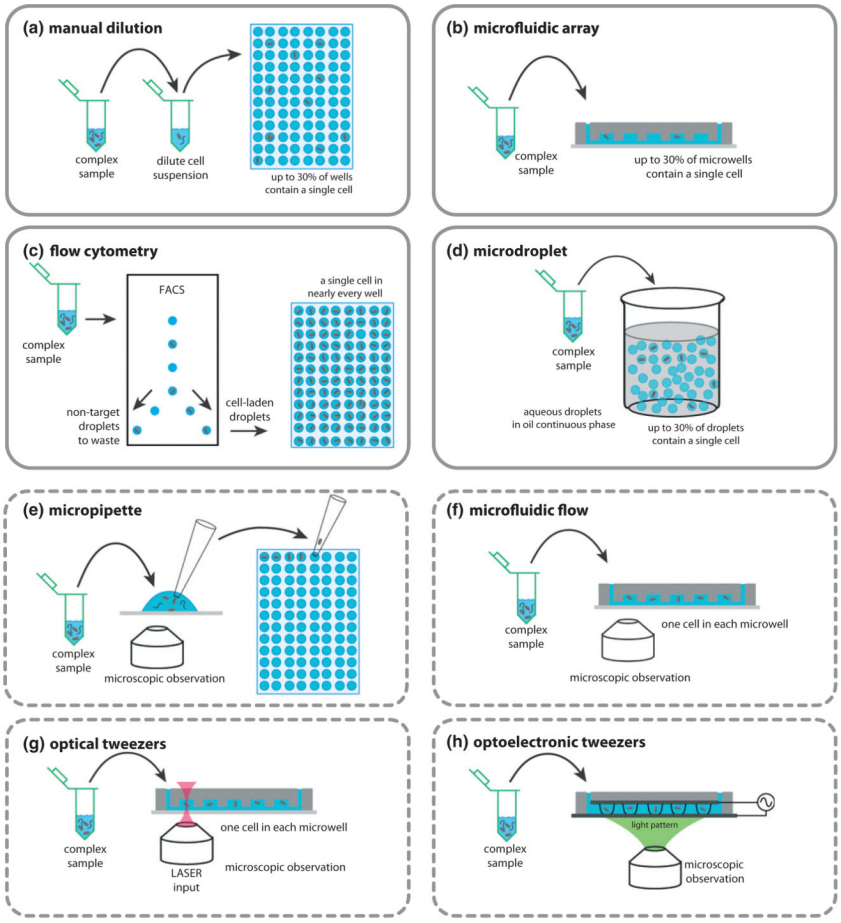


(c)

reaction volume	5 µL	50 nL	50 pL
no. contaminant molecules	~ 50	~ 0.5	~ 0
technology platform	384 well plate	microfluidic chamber	microemulsion droplet
WGA product quantity	up to 5 µg	up to 50 ng	up to 50 pg

Fig. 2.

Sources of contamination and the effect of reaction volume. (a) Three major sources of contamination: sample, laboratory, and reagent. (b) Schematic showing cross section of fluid volumes at the microliter, nanoliter, and picoliter scales. Fixed-concentration contaminant molecules are indicated. (c) Features of WGA reactions at the three volume scales.

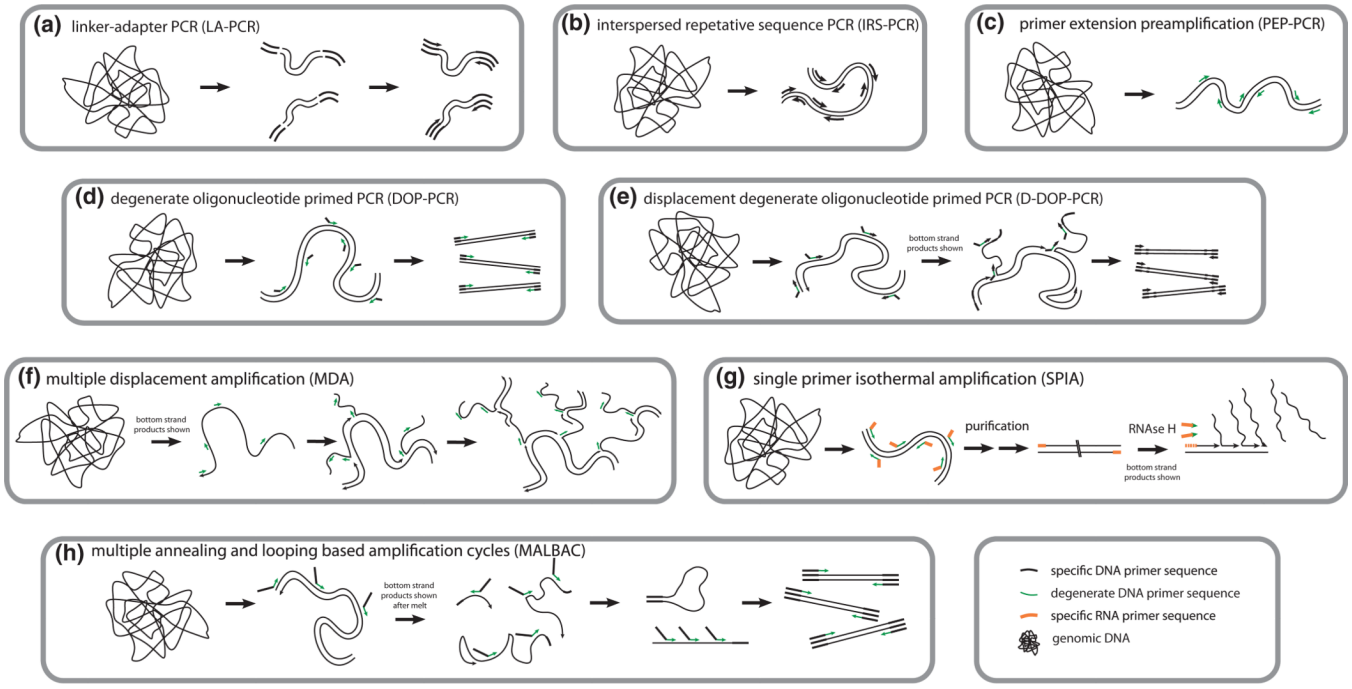


(i) Cell isolation approaches

	implementation	technology platform	commercially available	cell characterization	sampling volume	isolation statistic	cell throughput	potential for parallelization	process volume (reagent cost)	cell/product recovery	notes
random encapsulation	**manual dilution**	standard pipetting	yes	typically none	microliter	poisson	tens	high	microliter (high)	microwell plate	no specialized instrumentation required
	microfluidic array	microfabricated chambers	yes	fluorescence & micrograph possible	nanoliter	poisson	hundreds	low	nanoliter (low)	custom procedure	straightforward, rapid; recovery of products is challenging
	flow cytometry (FACS)	droplets in air	yes	optical scattering, fluorescence intensity	nanoliter	poisson (picked)	hundreds	low	microliter (high)	microwell plate	rapid, cells experience high shear, requires many cells
	microdroplet	droplets in oil, passive microfluidics	emerging	fluorescence & micrograph possible	picoliter	poisson	thousands+	high	picoliter (low)	yet to be shown	highest throughput; rxn setup & recovery methods needed
micromaniupulation	**micropipette**	ultra-precise pipetting	yes	fluorescence & micrograph possible	nanoliter	directed	tens	low	microliter (high)	custom procedure	requires open platform
	microfluidic flow	active microfluidics	yes	fluorescence & micrograph possible	picoliter	directed	tens	low	nanoliter (low)	custom procedure or microwell plate	requires precise flow control
	laser tweezers	laser microscopy	yes	fluorescence & micrograph possible	femptoliter	directed	tens	moderate	nanoliter (low)	custom procedure	microfluidic integration has been demonstrated
	optoelectronic tweezers	opto-electrokinetics	emerging	fluorescence & micrograph possible	femptoliter	directed	hundreds - thousands	high	--	yet to be shown	high-throughput potential; fluidic integration needed

Fig. 3.

Two classes of cell isolation methods, random encapsulation and micromanipulation. Four methods for random encapsulation: (a) manual dilution, (b) microfluidic array, (c) flow cytometry, (d) microdroplet emulsion. Four methods for micromanipulation: (e) micropipetting, (f) microfluidic flow, (g) laser tweezers, (h) optoelectronic tweezers. (i) Table summarizing characteristics of the cell isolation methods indicated in parts (a–h).



(i) Whole-genome amplification (WGA) methods

WGA method	primer ligation	fully degenerate primers	specific primers	partially degenerate primers	RNA/DNA hybrid primers	strand displacement synthesis	PCR	exponential amplification steps	linear amplification steps	amplification of unknown sequence	bias suppression	commercially marketed	commonly applied for single cells	product length (nt)	product concentration (ng/rxn µL)
LA-PCR	yes	no	yes	no	no	no	yes	1	0	yes	poor	no	no	100–1000	50–100
PEP-PCR	no	yes	no	no	no	no	yes	1	0	yes	poor	no	no	100–1000	1
IRS-PCR	no	no	yes	no	no	no	yes	1	0	no	poor	no	no	100–1000	50–100
DOP-PCR	no	no	no	yes	no	no	yes	1 - 2	0	yes	improved	Sigma	yes	100–1000+	50–100
D-DOP-PCR	no	no	yes	yes	no	yes	yes	2	0	yes	improved	Rubicon	yes	100–1000	50–100
MDA	no	yes	no	no	no	yes	no	1	0	yes	improved	GE, Qiagen	yes	> 10,000	1000
SPIA	no	no	no	yes	yes	no	no	0	1	yes	no data	NuGen	no	100–2000	100–150
MALBAC	no	no	yes	yes	no	yes	yes	1	1	yes	improved	no	emerging	500–1500	100

Fig. 4.

Methods for WGA. (a) tinker-adapter PCR. (b) Interspersed repetitive sequence PCR. (c) Primer extension preamplification. (d) Degenerate oligonucleotide-primed PCR. (e) Displacement degenerate oligonucleotide-primed PCR. (f) MDA. (g) Single primer isothermal amplification. (h) MALBAC. (i) Table summarizing characteristics of the WGA methods indicated in parts (a–h).

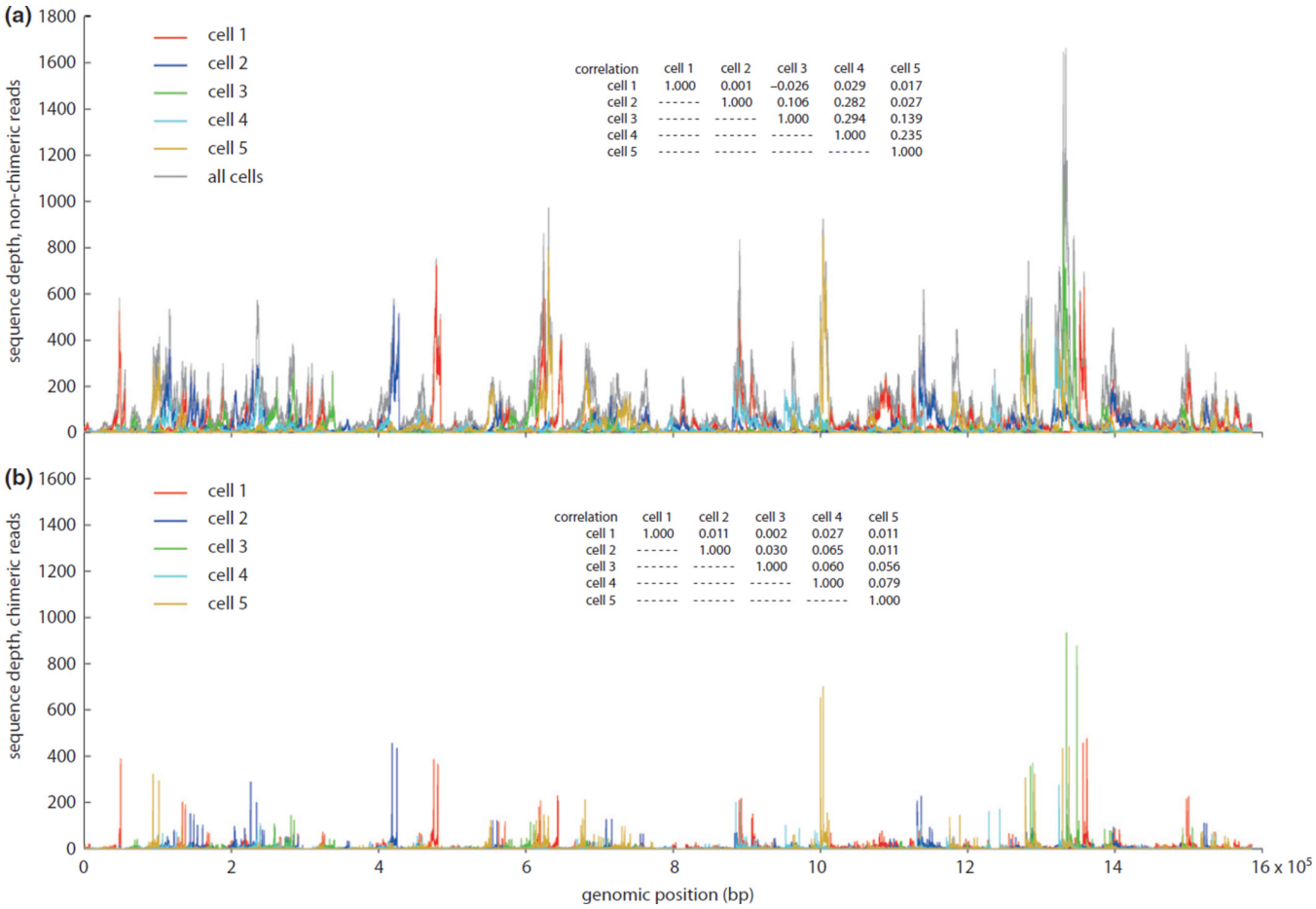


Fig. 5.

Amplification bias and chimerism in five single-cell data sets. (a) Coverage of the segmented filamentous bacteria (SFB) genome by nonchimeric reads based on independent WGA and sequencing of five SFB filaments from mouse gut (Pamp *et al*., 2012). The inset shows Pearson correlation coefficients for the coverage profile of all pairs of cells revealing very weak correlation in the high-quality read bias from cell to cell. (b) Coverage of the SFB genome by MDA-induced chimeric reads from the same five SFB filaments introduced in part (a). The inset shows Pearson correlation coefficients for the coverage profile of all pairs of cells revealing very weak correlation in the distribution of artificial chimeras from cell to cell.