# Reconstructing each cell's genome within complex microbial communities—dream or reality?

**Scott Clingenpeel , Alicia Clum , Patrick Schwientek , Christian Rinke and Tanja Woyke \***

*Department of Energy Joint Genome Institute, Walnut Creek, CA, USA*

As the vast majority of microorganisms have yet to be cultivated in a laboratory setting, access to their genetic makeup has largely been limited to cultivation-independent methods. These methods, namely metagenomics and more recently single-cell genomics, have become cornerstones for microbial ecology and environmental microbiology. One ultimate goal is the recovery of genome sequences from each cell within an environment to move toward a better understanding of community metabolic potential and to provide substrate for experimental work. As single-cell sequencing has the ability to decipher all sequence information contained in an individual cell, this method holds great promise in tackling such challenge. Methodological limitations and inherent biases however do exist, which will be discussed here based on environmental and benchmark data, to assess how far we are from reaching this goal.

Keywords: multiple displacement amplification, single-cell sequencing, genome completeness, genome quality, microbial ecology, environmental microbiology

## INTRODUCTION

Our current inability to cultivate the bulk of bacteria and archaea in the laboratory combined with a strong drive to analyze microbial communities *in situ*, gave rise to a large array of cultivation-independent methods that have been used for the past 20 years. These have been crucial to study microbial communities by deciphering community structure, organization and function. Commonly used techniques include microarrays, fluorescence *in situ* hybridization (FISH) and PCR-based approaches. With the advent of metagenomics it became possible to examine the genetic content of microbial communities without needing any *a priori* knowledge of the genetic sequences present. Although improved methods for the assembly and binning of metagenome sequences are emerging (Wrighton et al., 2012; Albertsen et al., 2013), linking potential functions to phylogeny still poses a major challenge for metagenomic approaches, especially in complex communities. One way to simplify such complex systems is to focus on the single microbial cell, which is the basic structural and functional unit of living organisms. Single-cell genomics is a method that allows the linkage of function to phylogeny while avoiding the difficulties in cultivating microorganisms.
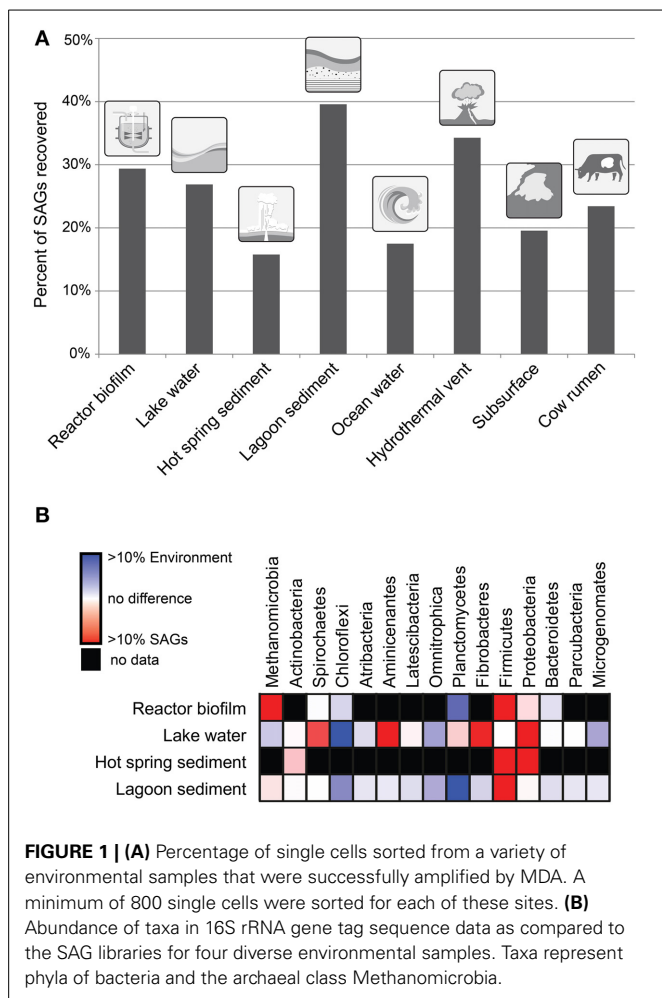
An array of review articles exists describing the state of the art single-cell microbial genomics (de Jager and Siezen, 2011; Lasken, 2012; Stepanauskas, 2012; Yilmaz and Singh, 2012; Blainey, 2013; Lasken and McLean, 2014), outlining key challenges, proposing potential solutions and summarizing the accomplishments that have been achieved with this technology. Using several environmental samples as well as reference organisms for benchmarking, we here place single-cell genomics in perspective and discuss how far we are from reconstructing each individual cell's genome within an environmental sample. We provide some practical implications of using the technology by exposing some of the current biases and limitations while bearing in mind the tremendous window of opportunity.

## LIMITED GENOME ACCESS

The fraction of single-cell genomes that can be recovered from a sample is highly variable (**Figure 1A**) due to technical challenges and biases at multiple steps of the process. The first step involves sample preparation and the isolation of single cells. Each sample may need custom sample preparation methods depending on the nature of the sample. While generalized recommendations do exist (Rinke et al., 2014), research should be done on which methods for dispersing the cells works best with differing sample types. High throughput single-cell isolation is generally performed using fluorescence activated cell sorting (FACS). This typically involves sorting based on the size of the particle (determined by a scatter signal) and fluorescence of a DNA stain such as SYBR Green. In principle every cell from a sample could be sorted, but practical limitations come from difficulties in dispersing the cells (cells attached to particles, aggregated cells), the inability to sufficiently stain all types of cells, and cells that fall outside the sorting window due to odd shapes (e.g., filaments) or unusually large or small size.

The next step of the process involves lysing the cells so that the genome amplification reagents will have access to the cell's DNA. High variability in the composition of microbial cell walls makes it difficult to find a universal method for cell lysis. This is further complicated by the nature of working with single cells. Since there may only be one copy of the genome, any nicks or double-stranded breaks in the DNA prior to genome amplification will likely lead to gaps in the resulting assembly. It therefore is necessary to use lysis methods that are gentle enough to maintain the integrity of the DNA. Moreover, clean-up steps are impractical due to the small sample and reaction size. Thus, reagents added

**FIGURE 1 | (A)** Percentage of single cells sorted from a variety of environmental samples that were successfully amplified by MDA. A minimum of 800 single cells were sorted for each of these sites. **(B)** Abundance of taxa in 16S rRNA gene tag sequence data as compared to the SAG libraries for four diverse environmental samples. Taxa represent phyla of bacteria and the archaeal class Methanomicrobia.

to increase the diversity of cells lysed. One challenge with using enzymes is that they are manufactured by living organisms, which almost inevitably introduces contaminant DNA from the production organism. Thorough quality control and clean-up is thus recommended prior to use for single-cell lysis.

## BIASES IN THE DIVERSITY

The current process of cell sorting involves an anonymous sort meaning that any particle that is within the correct size range and is sufficiently stained is sorted. This results in a random selection of single cells from the sample. Targeted sorting of particular groups of cells is rarely done because it requires a sufficiently bright stain or label. Many stains that work well with microscopy where exposure times can be relatively long are not suitable for FACS since the cell is only in the detection window for less than 100 μs. Thus, only after the genome amplification is completed is there sufficient material to perform screening steps to identify the single amplified genomes (SAGs) obtained; usually by amplification and sequencing of the 16S rRNA gene. Some single-cell MDA products will not yield 16S rRNA gene PCR amplification products. Reasons for this can include the lack of these genomics regions in the amplified DNA due to the amplification bias (see below) or mismatches of the universal 16S rRNA primers, which will prevent amplification of 16S rRNA genes in certain taxa (e.g., Baker et al., 2003, 2010; Youssef et al., 2014).

The 16S rRNA gene identification step could potentially be improved by optimizing the PCR reaction conditions. Multiple primer sets could be used to reduce the bias of any one particular primer set. PCR and sequencing of additional marker genes could be added to account for cases where the MDA amplification bias causes a lack of the 16S rRNA gene. However, these steps would increase the cost and amount of work for each single-cell genome that was amplified and may require separate optimization for each sample. Perhaps a more efficient way to deal with the biases introduced by the 16S rRNA gene identification step is to skip it entirely. The continual decrease in library generation and sequencing costs is making it more feasible to sequence all MDA products without requiring identification by PCR and sequencing of the 16S rRNA gene. Improved methods for targeted sorting of particular groups of interest would also allow one to skip the biases inherent to the 16S rRNA gene identification step since every sorted cell would represent a cell of interest.

All of the difficulties and biases given above in conjunction with 16S rRNA gene PCR biases, lead to a discrepancy in the diversity recovered by single-cell genomics compared to that seen in 16S rRNA gene tag data (**Figure 1B**). Based on four environmental samples analyzed, we found some groups such as the Proteobacteria are consistently overrepresented in the SAG library (see Supplementary Material for methods). These taxa appear to be quite amenable to single-cell genomics by being generally easy to sort and lyse. Other groups such as the Chloroflexi are consistently underrepresented possibly due to their often filamentous morphology, high GC contents, or having unusually tough cell walls. As more data from additional environments is collected in the future, these biases and their underlying cause will likely become clearer.

to the cell for lysis will remain there during the genome amplification step. The lysis reagents therefore have to be compatible with the multiple displacement amplification (MDA) chemistry. Even if a cell is successfully lysed it is possible that the phi29 polymerase will not have access to copy the DNA, possibly due to supercoiling of the DNA or nucleoid-associated proteins or other proteins being bound to the DNA. Research is needed on the magnitude of impact that proteins bound to the DNA such as nucleoid-associated proteins have on blocking the phi29 polymerase. Perhaps an additional step of protease treatment will significantly improve the recovery of genomes from some groups of microbes.

Currently, the most commonly used single-cell lysis method involves an alkaline treatment, which some taxa are not susceptible to, leaving great potential for improvement in this step of the process. Physical methods for lysing cells such as sonication or freeze-thaw cycles are by their nature the most universal ways to lyse cells. However, they are also the methods most likely to shear the DNA. Chemical lysis methods can be less damaging to DNA, but are more sensitive to the composition of the cell wall. Finally, enzymatic methods to degrade the cell wall are the least likely to damage the DNA, but are the most specific to cell wall composition. A cocktail of enzymes could be used
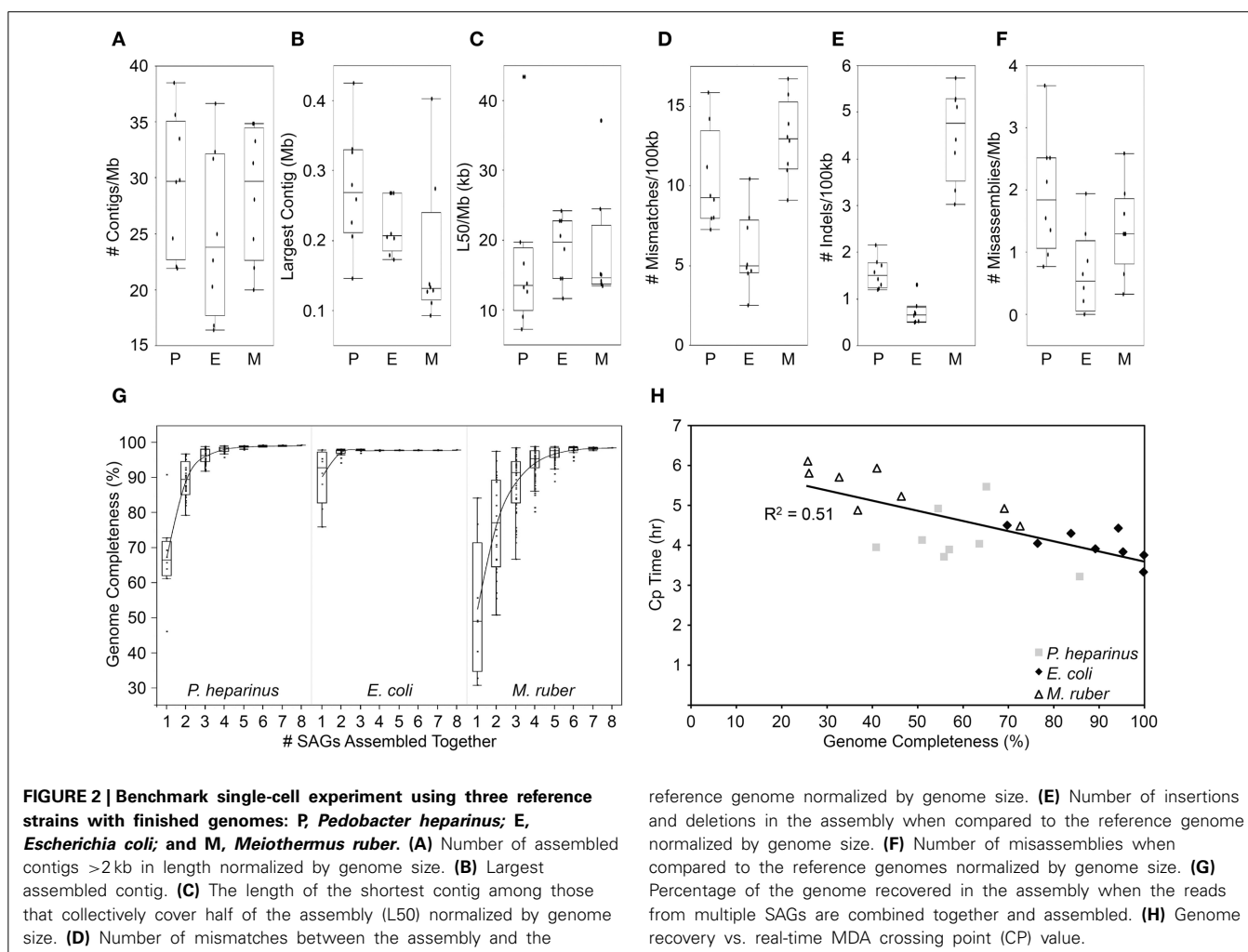
## GENOME RECOVERY

In addition to the limitations on the fraction of cells for which one can generate single-cell genomes, there is a current constraint on quality of the genome that can be recovered by single-cell sequencing. To explore this we produced single-cell genomes from three strains of bacteria with differing genomic G+C contents: *Pedobacter heparinus* DSM 2366, 42% GC; *Escherichia coli* K12-MG1655, 51% GC; and *Meiothermus ruber* DSM 1279, 63% GC (Clingenpeel et al., 2014). These have complete genome sequences available and have the same cell wall structure (Gram negative) to control for major differences in lysis efficiency. For each strain, we sequenced 8 single cells.

The median amount of the genome recovered was 66% for *P. heparinus*, 93% for *E. coli*, and 49% for *M. ruber*. All three strains had comparable numbers and sizes of contigs in their assemblies when normalized to their genome size (**Figures 2A–C**). It is possible that the lower genome recovery in *M. ruber* is due to its higher GC content. Other possible reasons include partial lysis or limited genome access due to DNA-bound proteins. Since *M. ruber* is somewhat more alkaliphilic than the other two strains (optimal growth at pH 8; Nobre et al., 1996) the alkaline treatment used to lyse the cells and denature the contents

may have been less effective than in the other two strains. The MDA reaction itself can provide an indication of the completeness of the genome to aid in selecting which SAGs should be fully shotgun sequenced. If the MDA reactions are monitored in real time (similar to RT-PCR) then the time at which the inflection point of the real-time amplification curves occurs (crossing point; CP value) is correlated with the completeness of the genome obtained (**Figure 2H**).

The amount and quality of a single-cell genome recovered is also partly dependent on how the sequencing reads are assembled. The nature of single-cell sequence is a result of unique complications largely introduced during the whole genome amplification. The MDA process results in uneven coverage across the genome. Parts of the genome can have tens of thousands-fold coverage while other regions have less than ten-fold coverage. Traditional sequence assemblers rely on even coverage across the genome and thus perform poorly on SAG datasets. One way of dealing with this is to normalize the coverage by removing excess sequences from the high coverage regions either in the lab or *in silico* (Rodrigue et al., 2009; Swan et al., 2011; Rinke et al., 2013). More recently, assemblers have been developed specifically to handle the variable sequence coverage including SPAdes, EULER+Velvet-SC



**FIGURE 2 | Benchmark single-cell experiment using three reference strains with finished genomes: P, *Pedobacter heparinus*; E, *Escherichia coli*; and M, *Meiothermus ruber*. (A)** Number of assembled contigs >2 kb in length normalized by genome size. **(B)** Largest assembled contig. **(C)** The length of the shortest contig among those that collectively cover half of the assembly (L50) normalized by genome size. **(D)** Number of mismatches between the assembly and the reference genome normalized by genome size. **(E)** Number of insertions and deletions in the assembly when compared to the reference genome normalized by genome size. **(F)** Number of misassemblies when compared to the reference genomes normalized by genome size. **(G)** Percentage of the genome recovered in the assembly when the reads from multiple SAGs are combined together and assembled. **(H)** Genome recovery vs. real-time MDA crossing point (CP) value.

and IDBA-UD (Chitsaz et al., 2011; Bankevich et al., 2012; Peng et al., 2012). When various assemblers are tested on SAG data from organisms with complete reference genomes SPAdes tends to produce the largest assemblies with the lowest error rates (Nurk et al., 2013). Using SPAdes 3.1.0 to assemble our benchmark SAG datasets, the pattern of errors follows that of genome completeness. *E. coli* had the most complete genomes and the lowest error rates while *M. ruber* had the least complete genomes and the highest error rates (**Figures 2D–G**). When compared to 25 isolate genomes generated at the JGI (data not shown) the *E. coli* error rates are not significantly different from the isolates, but both the *P. heparinus* and *M. ruber* are significantly higher ($P < 0.01$). For those two strains the average number of mismatches per 100 kb is 3–4x higher than the isolates, the average number of indels per 100 kb is 3–9x higher than in the isolates and the average number of misassemblies is 4–5x higher than in the isolates.

The recovery of partial genomes from single cells is due to parts of the genome being significantly over amplified compared to other regions. This is believed to be random process (Marcy et al., 2007; Lasken, 2012; Yilmaz and Singh, 2012; Lasken and McLean, 2014). If the bias is indeed completely random, different single cells from the same strain should recover different portions of the overall genome, thus complementing each other when being co-assembled. We co-assembled all eight SAGs for each strain using SPAdes 3.1.0. The more complete the individual SAGs were, the fewer SAGs were required to be combined to produce a near complete genome. A median completeness of >97% was obtained when 2 *E. coli*, 4 *P. heparinus*, and 5 *M. ruber* SAGs were combined (**Figure 2G**).

While the data described above demonstrate that combined assemblies are feasible with axenic cultures allowing the recovery of more complete genomes, the question of whether they are advantageous for environmental samples remains. Combined assemblies have been performed with single cell genomes from environmental samples (Blainey et al., 2011; Dodsworth et al., 2013; Rinke et al., 2013). The general selection criterion for co-assembly of environmental SAGs has been an average nucleotide identity of ≥95%, which is being used to delineate species (Konstantinidis and Tiedje, 2005; Goris et al., 2007; Richter and Rossello-Mora, 2009). However, one has to bear in mind that the resulting consensus assemblies are composite genomes representing a population and not a single organism, in which single nucleotide polymorphisms (SNPs) may become collapsed in the resulting consensus sequence.

Instead of dealing with the issue of partial genome recovery after the fact by combining multiple sequenced single-cell genomes, it is desirable to try to maximize the amount of genome recovered in the first place by minimizing the amplification bias inherent to MDA. One way to mitigate the amplification bias is to start with multiple copies of the target genome. Two approaches to do this have been successfully demonstrated. The first is to artificially increase the number of copies of the genome in a cell. Such artificial polyploidy involves inhibiting the FtsZ protein which is critical for cell division, while allowing genome replication to continue (Dichosa et al., 2012). This method has the potential to be broadly applicable since FtsZ is found in most bacteria and euryarchaea that have been examined. However, this protein has

not been found in other archaeal groups and it may be absent in uncultivated phyla, which are the taxonomic groups for which single-cell sequencing is most useful. As each FtsZ inhibitor discovered so far has its own range of taxa that it will affect, it is also unclear whether a cocktail of inhibitors can be found that will work with the full diversity of known FtsZ proteins. The growth of cells into microcolonies represents another method to improve the number of copies of the genome (Zengler et al., 2002; Fitzsimons et al., 2013; Dichosa et al., 2014). This process involves isolating individual cells in gel microdroplets (GMDs) and then allowing them to replicate in a co-culture with other cells from the environment while still maintaining a clonal population within each GMD. Even a few cell doublings would result in enough copies of the genome to substantially reduce the amplification bias. The process of co-culturing allows the cells to exchange metabolites and chemical signals with other cells, which has the potential to produce growth in a larger fraction of organisms as compared to traditional cultivation methods.

Other approaches to minimize amplification bias target the amplification reaction itself. Several strategies have been attempted to reduce the amplification bias including shrinking reaction volumes, adding molecular crowding agents, or changing the amplification method itself. Reducing reaction volumes by using a microfluidics system (Marcy et al., 2007) or microwells (Gole et al., 2013) have been reported to reduce the bias. In addition to reducing amplification bias, shrinking the reaction volume limits the risk of reagent-based DNA contamination. Although the use of microfluidics systems allows a significant reduction in reaction volume, such systems are currently not high throughput. Instruments such as the Labcyte Echo liquid handling system permit the accurate transfer of sub-microliter volumes which allows the shrinking of volume of the MDA reaction while maintaining a high throughput (Rinke et al., 2014). The addition of trehalose (Pan et al., 2008) or polyethylene glycol 400 (Ballantyne et al., 2006) to the MDA reaction improves the amount and evenness of the genome recovered. These are thought to work as molecular crowding agents, which increase the initial binding of primers and polymerase throughout the template. This would reduce the stochastic over-amplification of a few regions that started amplifying first in an uncrowded reaction. Although the phi29 polymerase used in MDA has numerous advantages, there is the possibility that other amplification chemistries could prove to be superior. One alternative is multiple annealing and looping-based amplification cycles (MALBAC; Zong et al., 2012). This method reduces the bias from exponential amplification seen in MDA by using a quasilinear preamplification step. While the MALBAC method had favorable results with a human cancer cell line (Zong et al., 2012), it did not perform well compared to MDA when applied to single *E. coli* cells (de Bourcy et al., 2014). While limiting the MDA by reducing the volume of the reaction or the time of the reaction reduces bias, it results in lower amounts of DNA available for sequencing. Improvements in library creation methods such as the Illumina Nextera XT sample preparation kit can create sequencing libraries from sub-nanogram amounts of input DNA, which makes limited MDA a viable approach to reducing amplification bias.

## OUTLOOK

From a genomics perspective, the recovery of complete genome sequences from single cells and from each cell within a complex environment is a desirable goal. Yet current technical difficulties and limitations still prevent this. New innovations will inevitably emerge, which may range from new, improved enzymes and chemistries to amplify a single-cell genome to novel single molecule sequencing technologies that may eliminate the need of sequencing library creation, thus removing the need for whole genome amplification entirely. For the time being, we will have to settle with partial genomes from a fraction of the cells contained within an environmental sample and reconstructing each cell's genome from a complex environment still remains a dream rather than reality. However, these partial single-cell genomes are clearly pushing microbial genomics into exciting, untapped territory, enabling the discovery of unexpected metabolic features, providing insights into population genetics, and improving the phylogeny of microbes (e.g., Swan et al., 2011; Rinke et al., 2013; Kashtan et al., 2014; Roux et al., 2014). We expect important discoveries will continue to be made as single-cell sequencing is applied to more environments.

## AUTHOR CONTRIBUTIONS

Tanja Woyke conceived and managed the project. Scott Clingenpeel and Christian Rinke performed the experiments. Alicia Clum and Patrick Schwientek analyzed the data. Scott Clingenpeel, Alicia Clum, Patrick Schwientek, Christian Rinke, and Tanja Woyke wrote the paper.

## ACKNOWLEDGMENTS

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fmicb.2014.00771/abstract

## REFERENCES

Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K. L., Tyson, G. W., and Nielsen, P. H. (2013). Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat. Biotechnol.* 31, 533–538. doi: 10.1038/nbt.2579

Baker, B. J., Comolli, L. R., Dick, G. J., Hauser, L. J., Hyatt, D., Dill, B. D., et al. (2010). Enigmatic, ultrasmall, uncultivated Archaea. *Proc. Natl. Acad. Sci. U.S.A.* 107, 8806–8811. doi: 10.1073/pnas.0914470107

Baker, G. C., Smith, J. J., and Cowan, D. A. (2003). Review and re-analysis of domain-specific 16S primers. *J. Microbiol. Meth.* 55, 541–555. doi: 10.1016/j.mimet.2003.08.009

Ballantyne, K. N., van Oorschot, R. A. H., Mitchell, R. J., and Koukoulas, I. (2006). Molecular crowding increases the amplification success of multiple displacement amplification and short tandem repeat genotyping. *Anal. Biochem.* 355, 298–303. doi: 10.1016/j.ab.2006.04.039

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021

Blainey, P. C. (2013). The future is now: single-cell genomics of bacteria and archaea. *FEMS Microbiol. Rev.* 37, 407–427. doi: 10.1111/1574-6976.12015

Blainey, P. C., Mosier, A. C., Potanina, A., Francis, C. A., and Quake, S. R. (2011). Genome of a low-salinity ammonia-oxidizing archaeon determined by single-cell and metagenomic analysis. *PLoS ONE* 6:e16626. doi: 10.1371/journal.pone.0016626

Chitsaz, H., Yee-Greenbaum, J. L., Tesler, G., Lombardo, M.-J., Dupont, C. L., Badger, J. H., et al. (2011). Efficient *de novo* assembly of single-cell bacterial genomes from short-read data sets. *Nat. Biotechnol.* 29, 915–921. doi: 10.1038/nbt.1966

Clingenpeel, S., Schwientek, P., Hugenholtz, P., and Woyke, T. (2014). Effects of sample treatments on genome recovery via single-cell genomics. *ISME J.* 8, 2546–2549. doi: 10.1038/ismej.2014.92

de Bourcy, C. F. A., De Vlaminck, I., Kanbar, J. N., Wang, J., Gawad, C., and Quake, S. R. (2014). A quantitative comparison of single-cell whole genome amplification methods. *PLoS ONE* 9:e105585. doi: 10.1371/journal.pone.0105585

de Jager, V., and Siezen, R. J. (2011). Single-cell genomics: unravelling the genomes of unculturable microorganisms. *Microb. Biotech.* 4, 431–437. doi: 10.1111/j.1751-7915.2011.00271.x

Dichosa, A. E. K., Daughton, A. R., Reitenga, K. G., Fitzsimons, M. S., and Han, C. S. (2014). Capturing and cultivating single bacterial cells in gel microdroplets to obtain near-complete genomes. *Nat. Protoc.* 9, 608–621. doi: 10.1038/nprot.2014.034

Dichosa, A. E. K., Fitzsimons, M. S., Lo, C.-C., Weston, L. L., Preteska, L. G., Snook, J. P., et al. (2012). Artificial polyploidy improves bacterial single cell genome recovery. *PLoS ONE* 7:e37387. doi: 10.1371/journal.pone.0037387

Dodsworth, J. A., Blainey, P. C., Murugapiran, S. K., Swingley, W. D., Ross, C. A., Tringe, S. G., et al. (2013). Single-cell and metagenomic analyses indicate a fermentative and saccharolytic lifestyle for members of the OP9 lineage. *Nat. Commun.* 4, 1854. doi: 10.1038/ncomms2884

Fitzsimons, M. S., Novotny, M., Lo, C.-C., Dichosa, A. E. K., Yee-Greenbaum, J. L., Snook, J. P., et al. (2013). Nearly finished genomes produced using gel microdroplet culturing reveal substantial intraspecies genomic diversity within the human microbiome. *Genome Res.* 23, 878–888. doi: 10.1101/gr.142208.112

Gole, J., Gore, A., Richards, A., Chiu, Y.-J., Fung, H.-L., Bushman, D., et al. (2013). Massively parallel polymerase cloning and genome sequencing of single cells using nanoliter microwells. *Nat. Biotechnol.* 31, 1126–1132. doi: 10.1038/nbt.2720

Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., and Tiedje, J. M. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Micr.* 57, 81–91. doi: 10.1099/ijs.0.64483-0

Kashtan, N., Roggensack, S. E., Rodrigue, S., Thompson, J. W., Biller, S. J., Coe, A., et al. (2014). Single-cell genomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus*. *Science* 344, 416–420. doi: 10.1126/science.1248575

Konstantinidis, K. T., and Tiedje, J. M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. U.S.A.* 102, 2567–2572. doi: 10.1073/pnas.0409727102

Lasken, R. S. (2012). Genomic sequencing of uncultured microorganisms from single cells. *Nat. Rev. Microbiol.* 10, 631–640. doi: 10.1038/nrmicro2857

Lasken, R. S., and McLean, J. S. (2014). Recent advances in genomic DNA sequencing of microbial species from single cells. *Nat. Rev. Genet.* 15, 577–584. doi: 10.1038/nrg3785

Marcy, Y., Ishoey, T., Lasken, R. S., Stockwell, T. B., Walenz, B. P., Halpern, A. L., et al. (2007). Nanoliter reactors improve multiple displacement amplification of genomes from single cells. *PLoS Genet.* 3:30155. doi: 10.1371/journal.pgen.0030155

Nobre, M. F., Truper, H. G., and da Costa, M. S. (1996). Transfer of *Thermus ruber* (Loginova et al. 1984), *Thermus silvanus* (Tenreiro et al. 1995), and *Thermus chliarophilus* (Tenreiro et al. 1995) to *Meiothermus* gen. nov. as *Meiothermus ruber* comb. nov., *Meiothermus silvanus* comb. nov., and *Meiothermus chliarophilus* comb. nov., respectively, and emendation of the genus *Thermus*. *Int. J. Syst. Bacteriol.* 46, 604–606.

Nurk, S., Bankevich, A., Antipov, D., Gurevich, A. A., Korobeynikov, A., Lapidus, A., et al. (2013). Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *J. Comput. Biol.* 20, 714–737. doi: 10.1089/cmb.2013.0084

Pan, X., Urban, A. E., Palejev, D., Schulz, V., Grubert, F., Hu, Y., et al. (2008). A procedure for highly specific, sensitive, and unbiased whole-genome amplification. *Proc. Natl. Acad. Sci. U.S.A.* 105, 15499–15504. doi: 10.1073/pnas.0808028105

Peng, Y., Leung, H. C. M., Yiu, S. M., and Chin, F. Y. L. (2012). IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28, 1420–1428. doi: 10.1093/bioinformatics/bts174

Richter, M., and Rossello-Mora, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U.S.A.* 106, 19126–19131. doi: 10.1073/pnas.0906412106

Rinke, C., Lee, J., Nath, N., Goudeau, D., Thompson, B., Poulton, N., et al. (2014). Obtaining genomes from uncultivated environmental microorganisms using FACS-based single-cell genomics. *Nat. Protoc.* 9, 1038–1048. doi: 10.1038/nprot.2014.067

Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N. N., Anderson, I. J., Cheng, J.-F., et al. (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 499, 431–437. doi: 10.1038/nature12352

Rodrigue, S., Malmstrom, R. R., Berlin, A. M., Birren, B. W., Henn, M. R., and Chisholm, S. W. (2009). Whole genome amplification and *de novo* assembly of single bacterial cells. *PLoS ONE* 4:e6864. doi: 10.1371/journal.pone.0006864

Roux, S., Hawley, A. K., Beltran, M. T., Scofield, M., Schwientek, P., Stepanauskas, R., et al. (2014). Ecology and evolution of viruses infecting uncultivated SUP05 bacteria as revealed by single-cell- and meta-genomics. *Elife* 3:e03125. doi: 10.7554/eLife.03125

Stepanauskas, R. (2012). Single cell genomics: an individual look at microbes. *Curr. Opin. Microbiol.* 15, 613–620. doi: 10.1016/j.mib.2012.09.001

Swan, B. K., Martinez-Garcia, M., Preston, C. M., Sczyrba, A., Woyke, T., Lamy, D., et al. (2011). Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. *Science* 333, 1296–1300. doi: 10.1126/science.1203690

Wrighton, K. C., Thomas, B. C., Sharon, I., Miller, C. S., Castelle, C. J., VerBerkmoes, N. C., et al. (2012). Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. *Science* 337, 1661–1665. doi: 10.1126/science.1224041

Yilmaz, S., and Singh, A. K. (2012). Single cell genome sequencing. *Curr. Opin. Biotech.* 23, 437–443. doi: 10.1016/j.copbio.2011.11.018

Youssef, N. H., Rinke, C., Stepanauskas, R., Farag, I., Woyke, T., and Elshahed, M. S. (2014). Insights into the metabolism, lifestyle and putative evolutionary history of the novel archaeal phylum 'Diapherotrites'. *ISME J.* doi: 10.1038/ismej.2014.141. [Epub ahead of print].

Zengler, K., Toledo, G., Rappe, M., Elkins, J., Mathur, E. J., Short, J. M., et al. (2002). Cultivating the uncultured. *Proc. Natl. Acad. Sci. U.S.A.* 99, 15681–15686. doi: 10.1073/pnas.252630999

Zong, C., Lu, S., Chapman, A. R., and Xie, X. S. (2012). Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science* 338, 1622–1626. doi: 10.1126/science.1229164