ELSEVIER

Single-cell genomic sequencing using Multiple Displacement Amplification

Roger S Lasken

Single microbial cells can now be sequenced using DNA amplified by the Multiple Displacement Amplification (MDA) reaction. The few femtograms of DNA in a bacterium are amplified into micrograms of high molecular weight DNA suitable for DNA library construction and Sanger sequencing. The MDA-generated DNA also performs well when used directly as template for pyrosequencing by the 454 Life Sciences method. While MDA from single cells loses some of the genomic sequence, this approach will greatly accelerate the pace of sequencing from uncultured microbes. The genetically linked sequences from single cells are also a powerful tool to be used in guiding genomic assembly of shotgun sequences of multiple organisms from environmental DNA extracts (metagenomic sequences).

Addresses
J. Craig Venter Institute, 11149 N. Torrey Pines Rd., Suite 200, La Jolla, CA 92037, United States

Corresponding author: Lasken, Roger S (rlasken@jcvi.org)

Current Opinion in Microbiology 2007, **10**:510–516

This review comes from a themed issue on
Genomics
Edited by Claire M. FraseriLiggert and Jean Weissenbach

Available online 17th October 2007

1369-5274/$ – see front matter

DOI 10.1016/j.mib.2007.08.005

Introduction

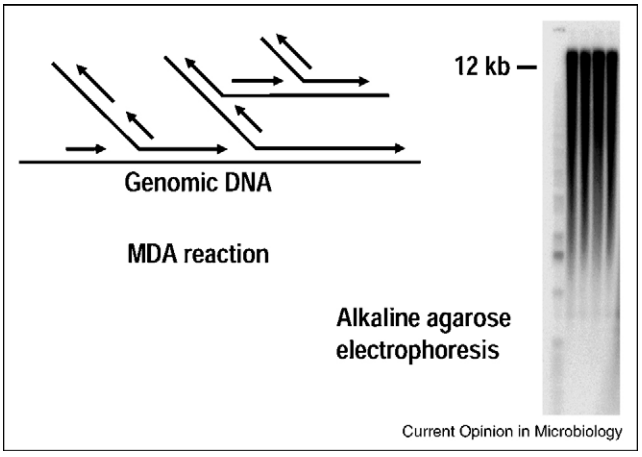
DNA sequencing from individual cells has recently become possible using amplified DNA as template. A method called Multiple Displacement Amplification (MDA) can amplify the few femtograms of DNA in a bacterium up to micrograms to be used in sequencing. MDA is dramatically changing the way that environmental studies can be conducted and is enabling new research strategies for microbial genetics, ecology, and infectious diseases. Previously, newly discovered microbes could only be sequenced if cultured isolates could be grown to sufficient quantities to provide the necessary DNA template. However, less than 1% of microbes have been successfully cultured. Venturing into natural environments will present vast numbers of new uncultured organisms that will be resistant to study using traditional genomic approaches. The advent of single-cell sequencing has opened a new front in this work. DNA template is obtained directly from individual cells without requiring development of culture methods. While amplification bias in the MDA reaction can result in underrepresentation and loss of some sequences, the rewards are great often providing the first genomic sequences obtained from novel species. Furthermore, several studies have already achieved between 50 and 75% completion of genomes from uncultured single cells, and method improvements are promising to improve upon this further. Recent work has also demonstrated the power of single-cell sequencing to reveal the presence of genes predicted from environmental and physiological studies even from partial draft genomes. The genomic data could also assist in the process of eventually developing culture methods, still a paramount goal required to fully understand the biology of organisms. The identification of metabolic pathways can inform the design of culture media and growth conditions. Another advantage of MDA-based single-cell genomics is that it is easily automated and can be combined with high throughput cell sorting of large numbers of microbes. PCR screening of the MDA reactions for informative sequences, such as the 16S rRNA and *recA* genes, can be used to identify the genomes of interest for additional analysis.

DNA amplification from single cells by MDA and genomic sequencing

The MDA reaction [1••,2••,3] uses random primers and the DNA polymerase from bacteriophage phi29 to amplify DNA in a 30 °C isothermal reaction. A 'strand displacement' reaction (Figure 1) makes multiple copies from each template as the polymerase synthesizes new strands while concurrently displacing previously extended strands. Compared with MDA, older whole genome amplification methods, based on PCR with random or degenerate primers, create far greater amplification bias (overrepresentation and underrepresentation of different regions of the template) [2••,4]. They also generate very short amplification products. MDA amplicons average 12 kb in length [1••] (Figure 1) and range up to >100 kb making them suitable to be used in cloning and DNA sequencing. MDA reagents can be obtained from Qiagen (Repli-g kit) and GE Healthcare (GenomiPhi and TempliPhi kits).

Scientists at the Department of Energy had the foresight to offer SBIR grants in 2001 to develop single-cell sequencing, and this was achieved at Molecular Staging Inc. that had developed the MDA method. The first genomic sequencing from single cells was demonstrated using *E. coli* and *Myxococcus xanthus* cells isolated by flow cytometry [5••]. More than a billion-fold amplification by

Figure 1

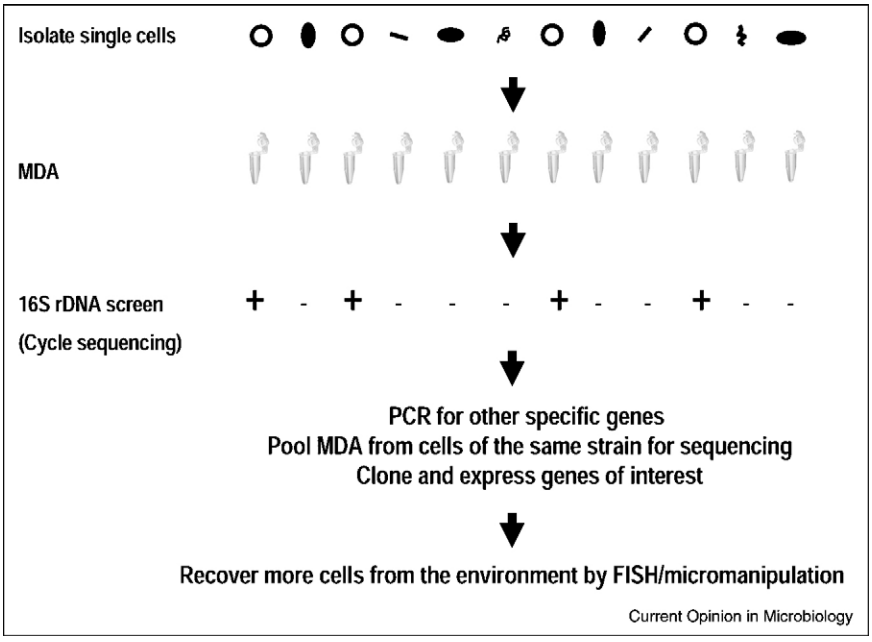


The MDA reaction. DNA template is repeatedly copied by a branching mechanism in which phi29 DNA polymerase extends random hexamer primers while the strong 'strand displacement activity' concurrently displaces previously made copies. Denaturation and resolution of the amplified DNA on an alkaline agarose gel demonstrate that the synthesized strands have an average length of about 12 kb and range up to an estimated >100 kb.

MDA generated micrograms of DNA from the few femtograms in one bacterial cell. DNA sequencing of the amplified DNA was 100% concordant with the known sequences. Quantitative PCR of 10 *E. coli* loci revealed amplification bias in the MDA reaction. Nevertheless, highly accurate DNA sequencing is achieved. Pyrosequencing by the 454 Life Sciences method, using MDA reactions from single *E. coli* cells, correctly mapped to the known *E. coli* genome at >95% for sequence reads and >99% for contigs [6••]. Amplification bias from MDA resulted in genome coverage of only about 70–75% of expected (based on the Lander–Waterman model) for a given sequencing depth. Therefore, greater sequencing depth is required to complete genomes. However, MDA opens a new avenue for sequencing of unculturable organisms that is often not possible by other means.

In the most comprehensive study to date of amplification bias, MDA was carried out with 3000 copies (10 ng) of human DNA as template and evaluated by qPCR for 47 loci, one on the p and q arm of every chromosome [4]. These single copy loci were still represented at between 0.5 and 3 copies per genome after the amplification. This sixfold range of amplification bias is the lowest ever reported for a whole genome amplification method. Amplification from a single bacterial genome copy is a greater challenge with bias ranging over hundreds of folds [5••]. While some sequence may be lost, there are several strategies being tested to complete genomes. Using the MDA-generated DNA as template, PCRs spanning sequence gaps can be used to recover many sequences [5••,7] that may be present but only underrepresented in

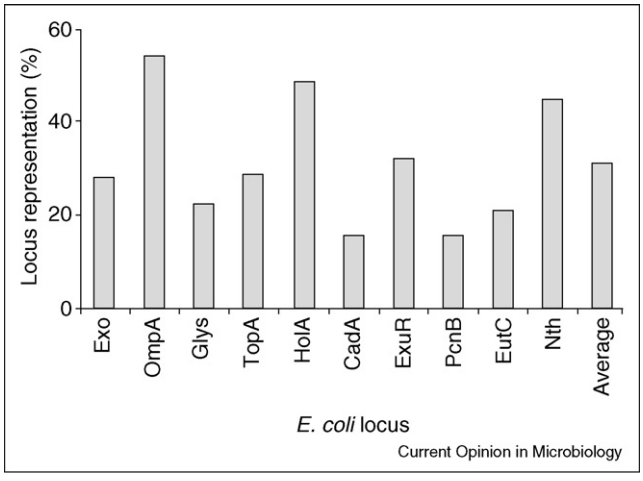
Figure 2



High throughput strategy to screen MDA reactions by PCR. Screening for highly informative sequences, such as the 16S rRNA, *recA*, and *rpoB* genes, can be used to select MDA reactions derived from the same species to be used in completing genomes. A sufficient number of PCR targets must be analyzed to allow confirmation that the MDAs were derived from the same microbial strain. A different use for PCR screening of the MDA reactions is to select for particular genes of interest with PCR primers based on partially known sequences or degenerate primers targeting gene families. Cloning of genes from the MDA reaction will allow *in vitro* expression of the encoded proteins. It will also be possible to use FISH probes based on 16S rRNA gene sequences to obtain additional target cells from the environment.

the MDA. MDAs derived from several different cells of the same strain can also be used to complete genomes. Amplification bias occurs fairly randomly at different sequences from one MDA reaction to the next [5••], and so sequence missing from one MDA will be present in others. To confirm that the MDAs chosen were truly derived from cells of the same strain, PCR screening and cycle sequencing can be carried out on the MDA reactions for the 16S rRNA gene and other markers (Figure 2). Pooling of single-cell MDA reactions generated from the same cell type can provide fairly even representation of all sequences (Figure 3) and is a promising strategy to gain complete coverage in sequencing libraries. Alternatively, genome coverage can be improved if cells can be pooled before the MDA reaction is carried out. It can be possible to identify cells of the same strain (1) by fluorescent *in situ* hybridization (FISH) probes or distinctive auto-fluorescence during cell sorting or micromanipulation, (2) by morphological characteristics observed during micromanipulation, (3) where a single strain predominates such as in partial enrichment cultures or some micro-environments (e.g. where infecting bacteria or symbionts are present in a host), or (4) in bacterial filaments composed of multiple cells, for example, *Beggiatoa* [8]. While pooling single-cell MDA reactions or pooling cells before MDA, it will be important to consider the extreme sequence diversity in natural populations (see reference [9], for example). Even cells having identical sequences for several markers, such as the 16S rRNA gene, may not be identical for all sequences. The resulting genomes would need to consider consensus genomes; however, these may be an important tool with which to make inroads into the formidable problem of genetic heterogeneity in bacteria. An alternative to pooling cells is to carry out MDA on cells from microcolony culturing [10], a method that requires growth, but for only a limited number of cell divisions. This should improve genome coverage by MDA by providing up to several hundred genome copies as template (M Keller, personal communication). Finally, partial sequences obtained from single cells or microcolonies can be combined with other data such as shotgun sequences to complete genomes.

Figure 3



Different loci are represented to a similar level in pooled data from multiple single-cell MDA reactions. Single *E. coli* cells were sorted into 84 microtiter plate wells using a FacsCalibur flow cytometer, lysed and subjected to whole genome amplification by MDA, and assessed by TaqMan quantitative PCR analysis of 10 different loci. For each individual single-cell amplification, bias between the 10 loci ranged over hundreds of folds with some loci dropping out completely [5••]. However, each locus was overrepresented in some MDA reactions and underrepresented in others. Therefore, averaging of the data (bars) over all 84 MDA reactions, for each locus, resulted in improved representation ranging over only a threefold range (18–55% representation). The data suggest that pooling of multiple single-cell MDA reactions will generate DNA libraries with reduced bias between the representations of different loci. The average of all 10 loci of about 30% is represented in the last bar. Reprinted with permission of Applied and Environmental Microbiology.

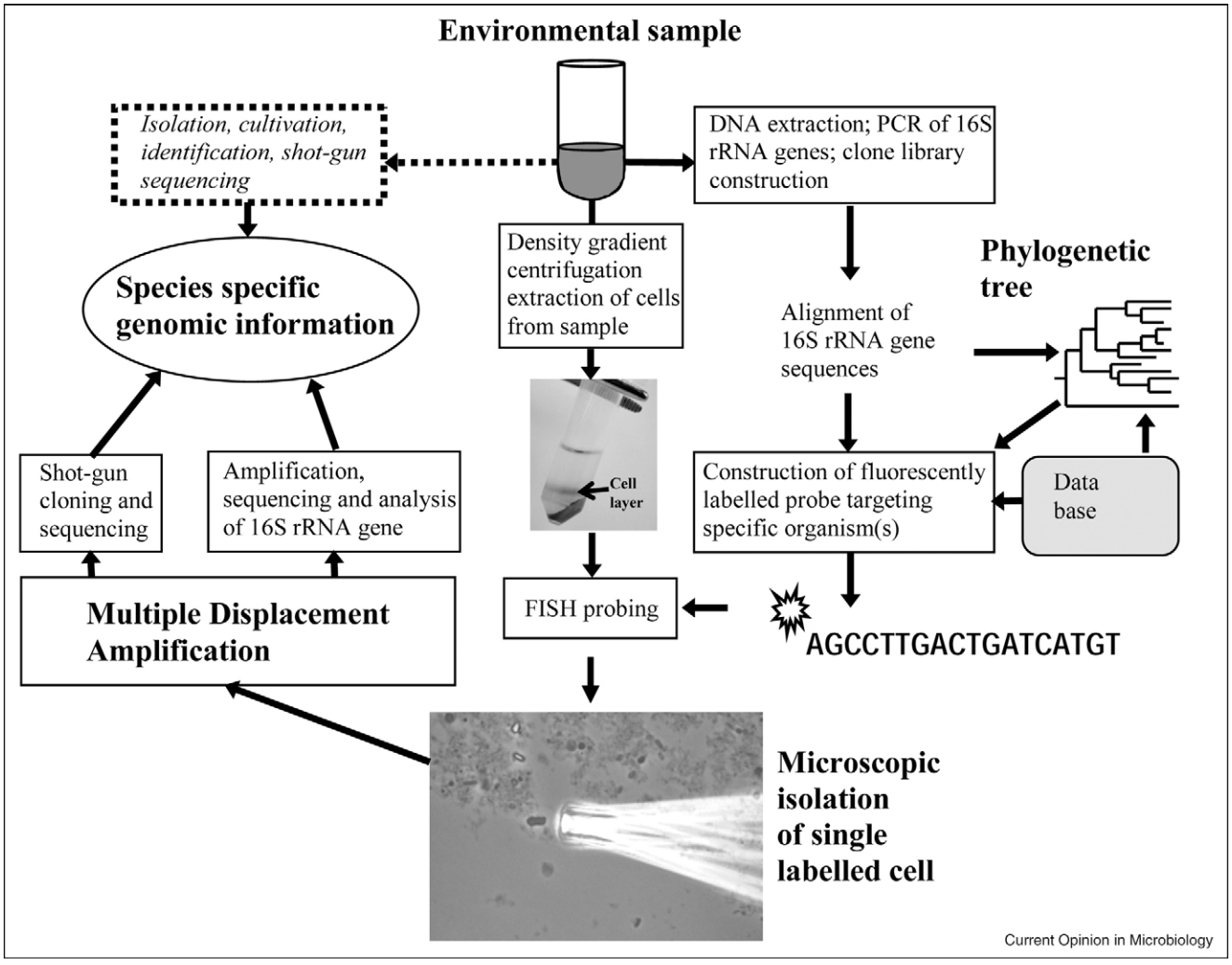
The extremely high fidelity of the proofreading phi29 DNA polymerase [11] gives base calling accuracy indistinguishable from that obtained with unamplified DNA [12]. Mutations accumulate at a rate of only 3 per million bases for MDA [13] compared with an estimated 1 mutation per 900 bases after only 20 cycles of PCR [14]. PCR rapidly accumulates mutations generated by the primer extensions of each cycle with newly synthesized DNA serving in turn as template in subsequent cycles. By contrast, many primers are extended from each template in MDA (Figure 1) such that sequences are not reamplified as many times as in PCR.

Microbial discovery

The first demonstration of sequencing from single uncultured cells obtained from the environment [15••] used soil bacteria isolated by micromanipulation [16]. Micromanipulation provided a high degree of certainty that only one cell was selected and also provided morphological information as the cell was viewed and photographed during isolation. Using the single-cell MDAs as template, PCR of the 16S rRNA gene was carried out to generate gene libraries for sequencing. Two different copies of the 16S rRNA gene were retrieved from a single cell belonging to the bacillus cluster, many of which are known to contain several operons of this gene. This study demonstrates one of the strengths of single-cell sequencing. The conventional method of screening total environmental DNA would not have allowed assignment of several 16S rRNA genes to one species. Assignment of plasmids, commensals, and viruses to the correct organism is also made possible by sequencing from single cells. These DNA sequences do not form contigs with the bacterial chromosome greatly confounding efforts to link them with sequences from total environmental DNA.

Some of the first genomic sequences ever obtained for Crenarchaeota [17••] were obtained using FISH probes specific for both Archaea and Crenarchaeota to select the

Figure 4

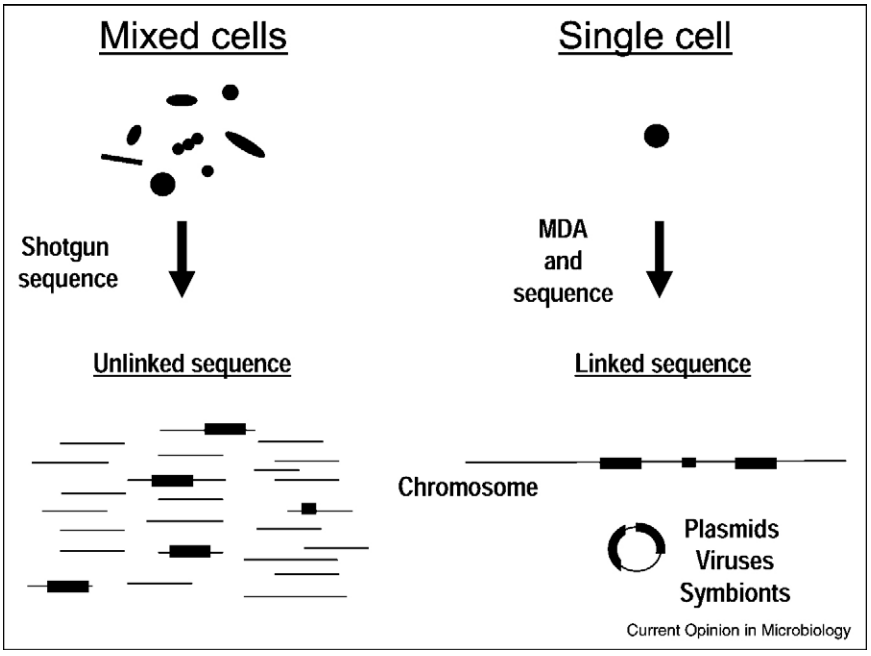


Comprehensive strategy to sequence novel microbes. An environment of interest (soil in this example) is profiled by 16S rRNA gene PCR (right hand pathway). FISH probes based on the sequences present allow isolation of single cells by micromanipulation [16,15••] (or, alternatively, by FACs flow cytometry) to be used in MDA (center pathway). Sequences derived from the MDA are combined with other available sequences from cultured cells and/or metagenomic shotgun databases (left hand pathway) to complete genome assemblies (reprinted from Kvist T, Ahring BK, Lasken RS, Westermann P. *Appl Microbiol Biotechnol* 2006, **74**(March (4)):926–35. Epub 2006 November 16, with kind permission of Springer Science and Business Media).

individual cells. A comprehensive strategy (Figure 4) can be used to (1) prescreen total environmental DNA extracts for 16S rRNA genes by PCR, (2) isolate cells using 16S rRNA FISH probes based on the observed populations, and (3) assemble genomes using the single-cell sequences along with other data such as metagenomic shotgun sequences. Other examples of sequencing from uncultured microbes appear in two recent papers on TM7, a candidate phylum for which no sequenced members had existed. FISH probes were used to isolate candidate cells by flow cytometry [18]. In another study [19] the cells were isolated on a microfluidic chip and MDA carried out in a 60 nl chamber. This method uses the microfluidic channels to achieve cell sorting, and the small volume of the MDA reactions better targets the single copy DNA template [6••].

Another study demonstrates how even partial sequences from single cells can greatly advance biological research. A draft sequence estimated to cover 70% of the genome was obtained for *Beggiatoa* spp. [8], a marine bacterium that has not been successfully cultured. Predicted genes were identified for a number of crucial enzymes for sulfur oxidation, nitrate respiration and oxygen respiration, and CO_2 fixation confirming a putative chemolithoautotrophic physiology. It should also be possible to clone novel genes and express the encoded proteins for study. MDA reactions can be screened with PCR primers for

Figure 5



Use of single-cell sequences in assembly of metagenomic shotgun sequences. Only part of the DNA template (chromosome, plasmid, virus, and symbiont sequence indicated in bold lines) is successfully recovered in MDA and sequencing from single cells (right pathway). However, DNA sequences derived from single cells are genetically linked. These sequences can be used to identify which of the corresponding mixed cell shotgun sequences (bold lines, left pathway) have been derived from the same organism. Sequenced sections are drawn for illustrative purposes only and are not intended to reflect actual sequence gap length or distribution.

known sequences or with degenerate primers to search for novel members of gene families (Figure 2) providing a new avenue to discover proteins of scientific or commercial interest.

Integration of single cell and metagenomic methods

For single-cell studies, all sequences obtained can be attributed to the same species. These genetically linked sequences will be immediately useful to guide the assembly of metagenomic shotgun sequences [20] from environmental samples. Metagenomic studies provide great insight into microbial genomics. Genes for 6 million new proteins were recently sequenced, nearly equaling the total number of all previously known proteins [9]. While the discovery of microbial diversity and novel genes has been greatly accelerated, the resulting metagenomic sequences are contributed by all of the species present and by many cells of each species that have small but important genetic variations between them. It has been exceedingly difficult to correctly assemble the sequences into the thousands of individual species represented. The single-cell sequences provide a solution to this bottleneck. They can be screened against the shotgun sequences to identify those from the same organism (Figure 5).

Recent improvements to MDA methods

Reduction of the MDA reaction volume has recently been found to give greater specificity for a single copy DNA template [21]. Specific amplification was enhanced by reducing amplification of contaminating DNA and nonspecific synthesis such as primer dimers, essentially by eliminating unnecessary reaction volume. We have confirmed that amplification specificity is improved using a 60 nl microfluidic reaction [6••]. Other research has led to a reduction in certain chimeric DNA rearrangements observed when sequencing from MDA reactions in which two segments of the genome are incorrectly joined together. The chimeras can be resolved during genome assembly by obtaining sufficient sequencing depth. However, reducing them before sequencing would simplify the process. Treatment of completed MDA reactions with S1 nuclease to remove single-stranded DNA forms resulted in an 80% reduction of chimeras [7]. It has now been shown that MDA directly causes the rearrangements, and the reaction pathway leading to their formation has been solved [22]. The branched DNA intermediates formed during MDA (Figure 1) can result in some DNA strands being extended on an initial template and then being displaced and extended on a different template resulting in the chimeras. These studies also clarified that the S1 nuclease treatment should reduce chimeric reads even for sequencing methods that do not

use cloning steps (e.g. 454 Life Sciences pyrosequencing) since MDA generates the chimeras rather than subsequent library cloning steps. Insight into the enzymatic pathway of chimera formation also suggests MDA reaction modifications that are currently being tested to reduce the occurrence of chimeras. Identification of the specific types of rearrangements that occur might also lead to improved informatic methods to anticipate and resolve chimeras during the assembly process.

Conclusions

The goal of genomic sequencing from individual bacterial cells has now been achieved with sufficient reliability to enable many exciting new research strategies. MDA is also currently being developed to be used in sequencing from low numbers microbial Eukaryotic cells. Fluorescence-activated flow cytometry was used to collect multiple cells of the same species (A Worden, personal communication). Sequencing from single Eukaryotic cells should be feasible where necessary. High throughput methods for cell sorting and DNA amplification by MDA promise to greatly accelerate the pace of microbial discovery using uncultured microbes. Combining single-cell sequencing with metagenomic data will be a powerful approach with the single-cell sequences guiding assembly of multi-species shotgun sequences into their individual genomes. Discovery of novel microbes in the human biome using single-cell methods is now being proposed by a number of laboratories. For infectious disease research it will be possible to identify the genetic variation of pathogens isolated from clinical specimens without the selective biasing of culture methods. Genetic linkage of virulence factors and other genes is also gained from single cells as apposed to the whole community gene frequencies obtained from metagenomic data. Perhaps the greatest challenge for all environmental genomic methods is the enormous genetic diversity even between members of the same strain. Hereto, single cell sequencing will make an important contribution to revealing the extent and nature of the diversity, the core or consensus genome that is essential to 'species', the role of horizontal gene transfer in evolution, and the relationship of diversity to environmental factors. Finally, progress also continues in the improvement of the MDA reaction enzymology with the goal of achieving even more complete representations of the genome.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. 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.

The first publication on the MDA method.

2. Dean FB, Hosono S, Fang L, Wu X, Faruqi AF, Bray-Ward P, Sun Z,
•• Zong Q, Du Y, Du J *et al.*: **Comprehensive human genome amplification using multiple displacement amplification**. *Proc Natl Acad Sci USA* 2002, **99**:5261-5266.

Use of MDA for whole genome amplification.

3. Lasken RS, Egholm M: **Whole genome amplification: abundant supplies of DNA from precious samples or clinical specimens**. *Trends Biotechnol* 2003, **21**:531-535.

4. Hosono S, Faruqi AF, Dean FB, Du Y, Sun Z, Wu X, Du J, Kingsmore SF, Egholm M, Lasken RS: **Unbiased whole-genome amplification directly from clinical samples**. *Genome Res* 2003, **13**:954-964.

5. Raghunathan A, Ferguson HR, Bornarth CJ, Driscoll M, Lasken RS:
•• **Genomic DNA amplification from a single bacterium**. *Appl Environ Microbiol* 2005, **71**:3342-3347.

The first demonstration of DNA sequencing using MDA from a single bacterial cell.

6. Marcy Y, Ishoey T, Lasken RS, Stockwell T, Walenz B, Halpern A,
•• Beeson K, Goldberg S, Quake S: **Nanoliter reactors improve Multiple Displacement amplification of genomes from single cells**. *PLoS Genetics* 2007, **3(9)**:e155 doi: 10.1371/journal.pgen.0030155.

MDA reactions in 60 nl microfluidic chambers and highly efficient 454 Life Sciences sequencing directly from the MDA reactions.

7. 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.

8. Mußmann M, Hu F, Richter M, de Beer D, Preisler A, Jørgensen B, Glöckner F, Amann R, Koopman W, Janto B *et al.*: **Insights into the genome of large sulphur bacteria revealed by analysis of single filaments**. *PLOS* 2007, **5(9)**:e230 (Epub ahead of print).

9. Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S *et al.*: **The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific**. *PLoS Biol* 2007, **5(March (3))**:e77.

10. Abulencia CB, Wyborski DL, Garcia JA, Podar M, Chen W, Chang SH, Chang HW, Watson D, Brodie EL, Hazen TC *et al.*: **Environmental whole-genome amplification to access microbial populations in contaminated sediments**. *Appl Environ Microbiol* 2006, **72**:3291-3301.

11. Blanco L, Bernad A, Lazaro 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.

12. Paez JG, Lin M, Beroukhim R, Lee JC, Zhao X, Richter DJ, Gabrial S, Herman P, Sasaki H, Altshuler D *et al.*: **Genome coverage and sequence fidelity of Phi29 polymerase-based multiple-strand displacement whole-genome amplification**. *Nucleic Acids Res* 2004, **32**:e71.

13. Nelson JR, Cai YC, Giesler TL, Farchaus JW, Sundaram ST, Ortiz-Rivera M, Hosta LP, Hewitt PL, Mamone JA, Palaniappan C *et al.*: **TempliPhi, phi29 DNA polymerase based rolling circle amplification of templates for DNA sequencing**. *Biotechniques* **(Suppl)**:2002:44-47.

14. Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA: **Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase**. *Science* 1988, **239**:487-491.

15. Lasken RS, Kvist T, Ishøy T, Westermann P, Ahring BK, Boissy R:
•• **Multiple Displacement Amplification from single bacterial cells**. In *Whole Genome Amplification.* Edited by Hughes S, Lasken R. UK: Scion Publishing Ltd.; 2005:119-147.

Complete laboratory protocols and analysis of results for single-cell isolation methods and MDA.

16. Ishoy 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.

17. Kvist T, Ahring BK, Lasken RS, Westermann P: **Specific
•• single-cell isolation and genomic amplification of uncultured microorganisms**. *Appl Microbiol Biotechnol* 2006.

An example of genomic sequencing from an uncultured single cell isolated from an environmental sample.

18. Podar M, Abulencia CB, Walcher M, Hutchison D, Zengler K, Garcia JA, Holland T, Cotton D, Hauser L, Keller M: **Targeted access to the genomes of low abundance organisms in complex microbial communities**. *Appl Environ Microbiol* 2007.

19. Marcy Y, Ouverney C, Bik EM, Lösekann T, Ivanova N, Martin HG, Szeto E, Platt D, Hugenholtz P, Relman DA, Quake SR: **Dissecting biological ''dark matter'' with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth**. *Proc Natl Acad Sci USA* 2007, **104(July (29))**:11889-11894.

20. Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, Nelson W *et al.*: **Environmental genome shotgun sequencing of the Sargasso Sea**. *Science* 2004, **304**:66-74.

21. Hutchison CA III, Smith HO, Pfannkoch C, Venter JC: **Cell-free cloning using phi29 DNA polymerase**. *Proc Natl Acad Sci USA* 2005, **102**:17332-17336.

22. Lasken RS, Stockwell TB: **Mechanism of chimera formation during the Multiple Displacement Amplification reaction**. *BMC Biotechnol* 2007, **7(April (1))**:19 doi: 10.1186/1472-6750-7-19.