The following full text is a publisher's version.

For additional information about this publication click this link.
http://hdl.handle.net/2066/96349

Please be advised that this information was generated on 2019-09-01 and may be subject to change.
Genomics update

Single-cell genomics: unravelling the genomes of unculturable microorganisms

Victor de Jager1,2,3 and Roland J. Siezen1,2,4,5,6*
1Netherlands Bioinformatics Centre, Nijmegen, The Netherlands.
2Centre for Molecular and Biomolecular Informatics, NCMLS, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands.
3Laboratory of Microbiology, Wageningen University, Wageningen, The Netherlands.
4TI Food and Nutrition, Wageningen, The Netherlands.
5NIZO food research, Ede, The Netherlands.
6Kluyver Centre for Genomics of Industrial Fermentation, Delft, The Netherlands.

Microbial genomics and related transcriptomics methods rely on culturing techniques to obtain enough DNA suitable for high-throughput sequencing without resorting to DNA amplification techniques. A few microgram of DNA is needed for most common next-generation sequencing methods. For transcriptome analysis, sufficient cDNA is needed to measure low abundance mRNA copies in the cell. However, the large majority of microbes on earth resist cultivation, hampering research into their relevant gene pool, ecological niche or industrial relevance. For example, many environmental or gut-related species cannot be grown outside their natural habitat. Even if we isolate the metagenome or the metatranscriptome from these environments, this reveals only a fragmented sequence landscape that is difficult to assign to individual species. Although enrichment techniques or metatranscriptome analysis of previously unculturable species have been shown to assist in directed culturing, e.g. of a Rikenella-like bacterium (Bomar et al., 2011), the unravelling of a complex metagenome into its individual genomes and their organization is impossible using current technologies.

A major challenge is the analysis of bacteria and other organisms living inside a complex matrix, like biofilms. Metagenome or transcriptome analysis of microorganisms has been described for biofilms consisting of a single species by scraping of the biofilm to obtain enough material (Holmes et al., 2006), but for multispecies biofilms this method results in a metagenome or metatranscriptome dataset. The solution to these challenges may be the isolation and genomic analysis of unculturable single cells isolated from such environments. Here we describe in brief the state-of-the-art in single-cell microbial genomics.

Single-cell isolation

Several methods exist to extract and investigate single microbial cells from their environment. Flow cytometry or fluorescence-activated cell sorting (FACS) has been used since the 1970s and its applications in microbiology were recognized early (Fouchet et al., 1993); recent advances are described by Müller and Nebe-von-Caron (2010), Wang and Bodovitz (2010), and Wang et al. (2010). Micromanipulation has been described by Kvist et al. (2007) and more recently by Woyke et al. (2010). Microfluidic device techniques are shown to be effective by combining the separation of cells and subsequently performing biochemical reactions on the device itself, thereby maximizing reaction yield (Marcy et al., 2007a) (Fig. 1).

Single-cell genome sequencing and data analysis

Whereas classical next-generation sequencing to determine an organism’s genome sequence relies on pooling DNA from $10^6$–$10^8$ cells, single-cell genomics relies on whole-genome amplification from a single cell. Most studies rely on Multiple displacement amplification (MDA), a biochemical amplification technique using random primers and Φ29 DNA polymerase (Dean et al., 2001; Raghunathan et al., 2005; Zhang et al., 2006; Marcy et al., 2007a). Other amplification techniques like random primed PCR result in a more over- and under-representation of different regions of the template DNA and generate very short fragments (Dean et al., 2001; Hosono et al., 2003). MDA, however, results in fragments of 12–100 kb rendering them suitable for sequencing. Although the complete microbial genome from a single cell can be amplified to amounts required for current research, the resulting dataset is a mixture of all genomes from a single cell and needs to be processed by computational methods to assign sequences to individual species.
sequencing methods without a priori sequence knowledge, early studies suggested that up to 40% of the genomic sequence was missed (Podar et al., 2007; Marcy et al., 2007b; Woyke et al., 2009) (Table 1). An overview of an MDA set-up using a microfluidic device is shown in Fig. 2, although FACS-based methods are also often reported in literature (Rodrigue et al., 2009; Siegl and Hentschel, 2010). All DNA in the

Table 1. Examples of single-cell genome sequencing.

<table>
<thead>
<tr>
<th>Microorganism</th>
<th>Assembled bases (Mb)</th>
<th>Estimated % genome recovery</th>
<th>Scaffolds</th>
<th>Contigs</th>
<th>GC%</th>
<th>Single cell separation</th>
<th>Isolation source</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>TM7a (new phylum)</td>
<td>2.865</td>
<td>?</td>
<td></td>
<td>1825</td>
<td></td>
<td>34.3 Microfluidics</td>
<td>Human mouth biofilm</td>
<td>Marcy et al. (2007b)</td>
</tr>
<tr>
<td>TM7_GTL1 (new phylum)</td>
<td>0.679</td>
<td>?</td>
<td></td>
<td>132</td>
<td>48.5</td>
<td>FISH/FACS</td>
<td>Soil</td>
<td>Podar et al. (2007)</td>
</tr>
<tr>
<td>Prochlorococcus MED4</td>
<td>95</td>
<td>755</td>
<td></td>
<td></td>
<td></td>
<td>FACS</td>
<td>Sea water; lab culture</td>
<td>Rodrigue et al. (2009)</td>
</tr>
<tr>
<td>Flavobacterium MS024-2A</td>
<td>1.905</td>
<td>91</td>
<td>17</td>
<td>36</td>
<td></td>
<td>Flow cytometer</td>
<td>Coastal water, Maine, USA</td>
<td>Woyke et al. (2009)</td>
</tr>
<tr>
<td>Flavobacterium MS024-3C</td>
<td>1.505</td>
<td>78</td>
<td>21</td>
<td>39</td>
<td></td>
<td>Flow cytometer</td>
<td>Coastal water, Maine, USA</td>
<td>Woyke et al. (2009)</td>
</tr>
<tr>
<td>Cand. Sulcia muelleri DMIN</td>
<td>0.244</td>
<td>100</td>
<td>1</td>
<td>1</td>
<td>22.5</td>
<td>Micromanipulator</td>
<td>Symbiont from insect bacteriome (green sharpshooter)</td>
<td>Woyke et al. (2010)</td>
</tr>
<tr>
<td>Poribacteria</td>
<td>1.885</td>
<td>66</td>
<td>1597</td>
<td></td>
<td>53.4</td>
<td>FACS</td>
<td>Symbiont from marine sponge</td>
<td>Siegl et al. (2011)</td>
</tr>
<tr>
<td>Cand. Nitrosoarchaeum limnia SFB1</td>
<td>1.690</td>
<td>95</td>
<td>26</td>
<td>136</td>
<td>32.4</td>
<td>Microfluidics, laser tweezers</td>
<td>Ammonia-oxidizing enrichment culture; sediment water, San Francisco bay, USA</td>
<td>Blainey et al. (2011)</td>
</tr>
</tbody>
</table>

a. Method validation using strain with known genome sequence.
b. Pooled sequence data from five individual cells; see Table 2.
initial sample will be amplified, which renders the method very prone to DNA contamination. Another disadvantage of the initial method is uneven amplification of the genome, which results in high-coverage sequencing of the amplified genomic regions while remaining sequences may not be sufficiently covered (Zhang et al., 2006). Marcy et al. (2007a) demonstrated that reducing MDA reaction volumes lowers non-specific synthesis from contaminant DNA templates and unfavourable interactions between primers. The work of Rodrigue et al. (2009) demonstrated a biochemical method to normalize the products obtained in MDA reactions. They also discussed the problem of chimera formation linking non-contiguous chromosomal regions in MDA (Dean et al., 2001; Zhang et al., 2006), which may hamper sequence assembly and render mate-pair data less efficient in contig positioning. Several other single-cell techniques are described in recent reviews by Wang and Bodovitz (2010), Kalisky and Quake (2011), and Pan et al. (2011). As data analysis from single-cell amplified genomes is equally challenging, the software framework SmashCell has been developed to automate the main steps in sequence assembly, gene prediction, annotation and visualization (Harrington et al., 2010).

**Single-cell genome sequences of uncultured microorganisms**

Examples of sequencing of single amplified genomes (SAGs) are listed in Table 1. Woyke et al. (2010) describe using a micro-displacement technique to sequence a genome from an uncultured single cell of Candidatus *Sulcia muelleri* DMIN, a symbiont isolated from the bacteriome of the green sharpshooter *Draeculacephala minerva*. This polyploid bacterium has an estimated 200–900 genome copies per cell. Of the 57 Mb of sequence generated, approximately 90% was of contaminant origin, as estimated by mapping to a previously sequenced genome of *Sulcia* and phylogenetic analysis with blastx and MEGAN (Mitra et al., 2009). The remaining reads were assembled into a draft genome, misassemblies due to chimeras were corrected manually, and subsequent application of primer walking, sequencing PCR products and Illumina sequencing resulted in a final finished genome (Fig. 3).

Siegl et al. (2011) used FACS to isolate cells from the candidate phylum *Poribacteria* and subsequently MDA to obtain a SAG. These bacteria are almost exclusively found in marine sponges as symbionts and resist cultivation efforts. The SAG of 1.88 Mb was contained in 1597
contigs, which covered an estimated two-thirds of the total genomic DNA based on the distribution of tRNA genes and their specificities found in the contigs. Nevertheless, a comprehensive overview of poribacterial metabolism could be deduced (Fig. 4). The extensive Sup-type polyketide synthases found in the SAG of *Poribacteria* confirmed the previously proposed assignment of Sup-PKS to this species. With the finding of a second putative PKS system showing high similarity to the lipopolysaccharide type I PKS WcbR from *Nitrosomonas* and *Burkholderia*, as well as RkpA from *Sinorhizobium fredii*, they suggested that *Poribacteria* contain at least two different types of PKS systems and their products may be involved in sponge–microbe interactions. This study showed that single-cell genomics is highly capable of dissecting the genomic information from unculturable bacteria, shedding light on genomic organization, metabolic functions and possibly new insight in the debate on the origin of sponge bioactive compounds.

Ammonia-oxidizing archaea (AOA) are among the most abundant microbes on Earth, and may significantly impact global nitrogen and carbon cycles. Five single cells were isolated from a low-salinity sediment AOA-enrichment culture using a microfluidic device and laser tweezers, and DNA was amplified and sequenced separately from each cell (Blainey *et al.*, 2011) (Tables 1 and 2). Individually, three single-cell datasets gave assemblies of more than 1 Mb at sequencing depths of 10× to 30×, and an estimated 60% genomic coverage each; the low coverage is considered typical due to MDA amplification bias. Surprisingly,
each of the single-cell assemblies represented a different 60% of the target genome, and combining the five datasets led to a single-cell assembly representing >95% of the *Nitrosoarchaeum limnia* genome. Based on nucleotide identity comparisons, this AOA is proposed to represent a new genus of Crenarchaeota. In contrast to other described AOA, this low-salinity archaeum appears to be motile, based on the presence of numerous motility and chemotaxis-associated genes in the genome (Blainey et al., 2011).

**Single-cell transcriptomics, metabolomics and proteomics**

Recent reports on single-cell transcriptomics discuss mainly the analysis of polyadenylated mRNA of eukary-

---

**Table 2.** Assembly statistics for sequencing of three single cells of *Nitrosoarchaeum limnia* SFB1, and consensus genome (reads from metagenome and five single cells).

<table>
<thead>
<tr>
<th>Assembly statistics</th>
<th>Cell 23</th>
<th>Cell 21</th>
<th>Cell 3</th>
<th>Five single cells co-assembly</th>
<th>Consensus single cells and metagenome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw read bases</td>
<td>17 107 411</td>
<td>52 341 561</td>
<td>29 999 202</td>
<td>118 796 782</td>
<td>150 994 537</td>
</tr>
<tr>
<td>Assembly bases</td>
<td>1 094 113</td>
<td>1 039 820</td>
<td>1 041 604</td>
<td>1 690 404</td>
<td>1 769 573</td>
</tr>
<tr>
<td>Scaffolds</td>
<td>68</td>
<td>76</td>
<td>83</td>
<td>26</td>
<td>2</td>
</tr>
<tr>
<td>Unassembled contigs</td>
<td>287</td>
<td>177</td>
<td>265</td>
<td>110</td>
<td>29</td>
</tr>
<tr>
<td>Estimated % genome coverage</td>
<td>62</td>
<td>59</td>
<td>59</td>
<td>95</td>
<td>99</td>
</tr>
</tbody>
</table>

Adapted from Table 1 of Blainey et al. (2011).

© 2011 The Authors
otes. A comprehensive overview of the technologies involved is given by Tang et al. (2011). In short, the single-cell methods exploit reverse transcription using oligo(dT) primers to convert mRNAs with poly(A) tails into cDNAs, followed by uniform amplification and sequencing (RNA-seq). However, currently no single-cell analysis reports are known that exploit protocols for mRNA extraction from bacterial cells, for instance using the MessageAmp II-Bacteria Kit (Ambion) as described by Frias-Lopez et al. (2008). Single-cell metabolome and proteome/peptidome analyses are still in their infancy, as these compounds cannot be amplified and their analysis requires technological breakthroughs in pushing the limits of detection (Rubakhin et al., 2011).

Future

Since the introduction of single-cell genomics (Raghu-nathan et al., 2005), there have been surprisingly few reports of successful reconstruction of whole genomes from single unculturable bacterial cells (Table 1). This undoubtedly reflects the extreme difficulties in the various steps of single-cell isolation, miniaturization, and DNA amplification, avoidance of contamination and data analysis. Nevertheless, the pioneering examples show that it is definitely feasible to sequence genomes of single unculturable cells isolated from complex consortia, and we expect this approach to become more widespread as miniaturization technologies improve.

Recently, it has also been recognized that isogenic microbial populations (pure cultures) contain substantial cell-to-cell differences in physiological parameters such as growth rate, resistance to stress and regulatory circuit output (Ingham et al., 2008; Lidstrom and Konopka, 2010). In this light, adaptation of single-cell genome sequencing using microfluidic approaches towards RNA-seq transcriptome analysis of single cells using next-generation mRNA sequencing should become increasingly important (Siezen et al., 2010).

Acknowledgements

This project was carried out within the research programmes of the Kluver Centre for Genomics of Industrial Fermentation and the Netherlands Bioinformatics Centre, which are part of the Netherlands Genomics Initiative/Netherlands Organization for Scientific Research.

References


