

## Complete assembly of novel environmental bacterial genomes by MinION™ sequencing

Daniel J Turner<sup>1\*</sup>, Xiaoguang Dai<sup>2</sup>, Simon C Mayes<sup>1</sup>, Sissel Juul<sup>2</sup>

\* Corresponding author: Daniel J Turner, [daniel.turner@nanoporetech.com](mailto:daniel.turner@nanoporetech.com)

Author affiliations:

<sup>1</sup> Oxford Nanopore Technologies Ltd., Oxford, UK

<sup>2</sup> Oxford Nanopore Technologies Inc., New York, USA

**In this study, we adapt a protocol for the growth of previously uncultured environmental bacterial isolates, to make it compatible with whole genome sequencing. We demonstrate that in combination with the MinION sequencing device, complete assemblies can be derived, allowing genomic comparisons to be made. This approach allows rapid, inexpensive and straightforward discovery, and genomic analysis, of previously uncultured prokaryotic genomes, and brings greater ownership of all parts of the sequencing process back to individual researchers.**

There are an estimated  $10^5$  to  $10^6$  microbial species in existence, but only a few thousand such species have been isolated and cultured, because of the difficulty of growing many microbes under laboratory conditions. One reason for this difficulty is thought to be that many microbial species have specific nutritional requirements, without which they cannot grow <sup>1</sup>. This is unfortunate because uncultured microbial species: i) have potential uses in biotechnology and in the development of new antibiotics <sup>2</sup>, ii) contribute significantly to the biodiversity of all ecosystems, and iii) could be pathogenic, and hence warrant further study. Comparison of bacterial genome sequences helps us to understand genome evolution and gene function <sup>3</sup>, and there is consequently a need for a quick, inexpensive and straightforward way to sequence and assemble uncultured prokaryotic genomes.

One way to derive genome sequences from complex samples is by metagenomic assembly, where total DNA is extracted from a sample and sequenced, and the multiple genomes which constitute the sample are assembled together. However, due to the potentially large number of genomes present in such a sample, this approach can make less distinct the whole concept of what constitutes an assembled genome <sup>4</sup>. In contrast to this comprehensive approach, the iChip <sup>5</sup> allows individual, previously unculturable, bacteria to be isolated and grown clonally, by exposing the cells to the environment from which they were taken, in such a way that nutrients can diffuse in, with the bacterial cells immobilised in gel. Following their initial isolation and growth, it has been reported that many previously unculturable bacteria can adapt to being cultured under laboratory conditions <sup>5</sup>.

The iChip device has previously been shown to have utility for the identification of unculturable bacterial species by 16S sequencing <sup>1</sup> and the discovery of new antibiotics <sup>5</sup>. We sought to exploit the device to culture, for whole genome sequencing, bacteria present in a stream close to the Oxford

Nanopore office, Oxford, UK, at the confluence of the Littlemore Brook and the effluent from a sewage treatment facility, which we reasoned would contain a high bacterial count. After two weeks of growth in the iChip, submerged in the stream, we found that we were able to continue growth of the bacteria in low salt culture medium, which had been prepared with sterilised and filtered water from the stream, and also on petri dishes of low salt LB-agar, prepared in the same way. Using this approach we were able to grow, and extract DNA from, individual bacterial strains with the expectation that we would be able to sequence and ultimately assemble these genomes.

Unlike metagenomic assembly, complete assembly of single genomes is relatively straightforward. The criteria for a good quality assembly are well-defined: large contigs with as few misassemblies as possible <sup>4</sup>. Long sequence reads have been shown to facilitate assembly of long contigs, and even complete single contig assemblies, because they span repetitive elements in genomes <sup>6</sup>.

The Oxford Nanopore MinION is a portable, inexpensive single molecule nucleic acid sequencing device, which plugs directly into the USB drive of a standard laptop and requires no additional infrastructure, and which is capable of generating very long reads in real time <sup>7</sup>. Using the Celera assembler <sup>8</sup>, preceded by one or more rounds of error correction, reads from the *Escherichia coli* strain K-12 obtained from the device can be assembled into a single contig assembly <sup>9,10</sup>. We therefore applied this technology to the sequencing of several of the unculturable bacterial strains.

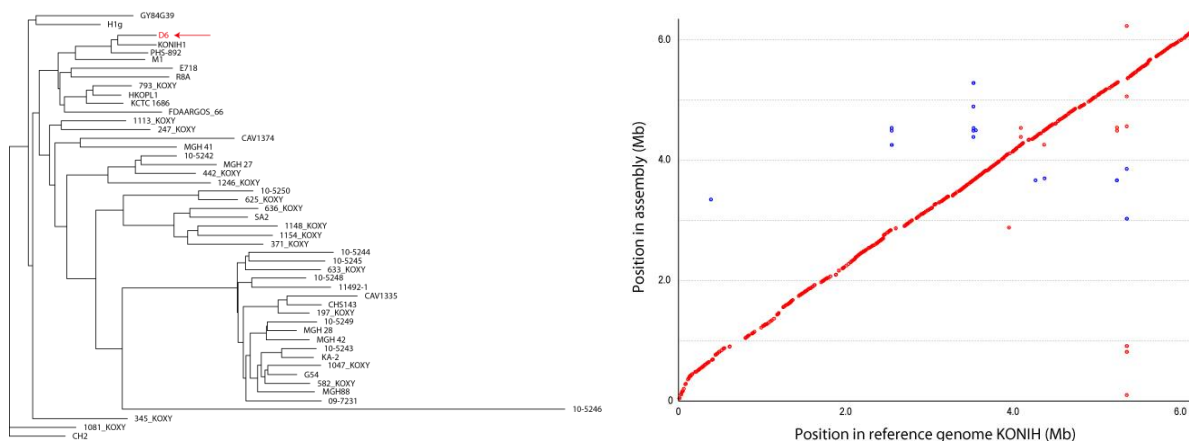
Initially, we attempted to identify the bacterial species present in the different isolates using Kraken <sup>11</sup>, in real time (i.e. as reads were being generated by the MinION, **Table 1**). However, the results were fairly inconclusive, indicating that the majority of these particular bacterial subspecies are not present in the NCBI database. In a similar way to digital PCR <sup>12</sup>, which is quantitative because single template strands are amplified in separate wells or droplets, it is evident from the number of times certain species were detected (**Table 1**.) that those bacterial strains were more abundant in the original sample than others. Here, *Pseudomonas* species were the most abundant.

Well	Family	Genus	Species group
A1, A5, B6, C11, F5	<b>Pseudomonadaceae</b>	<b>Pseudomonas</b>	<i>Pseudomonas fluorescens</i>
A7	<b>Pseudomonadaceae</b>	<b>Pseudomonas</b>	<i>Pseudomonas putida</i>
C5	<b>Enterobacteriaceae</b>	<i>Salmonella</i> <i>Enterobacter</i> <i>Escherichia</i> <i>Klebsiella</i>	<i>Salmonella enterica</i> <i>Enterobacter cloacae</i> complex <i>Escherichia coli</i> <i>Klebsiella pneumoniae</i>
C9	<b>Pseudomonadaceae</b>	<b>Pseudomonas</b>	<i>Pseudomonas aeruginosa</i> <i>Pseudomonas asputida</i>
D1	Alteromonadaceae	<i>Alteromonas</i>	<i>Alteromonas macleodii</i>
D5	<b>Pseudomonadaceae</b>	<b>Pseudomonas</b>	<i>Pseudomonas putida</i> <i>Pseudomonas fluorescens</i>
D6	<b>Enterobacteriaceae</b>	<b>Klebsiella</b>	<i>Klebsiella oxytoca</i>
D11	Flavobacteriaceae	NA	NA
D12, E3	<b>Pseudomonadaceae</b>	<b>Pseudomonas</b>	<i>Pseudomonas fluorescens</i>

**Table 1.** Bacterial isolates identified in the stream water using Kraken

We chose two isolates at random, C9 and D6 (**Table 1.**), which we had identified as strains of *Pseudomonas* and *Klebsiella oxytoca* respectively. We sequenced these genomes using PCR-free libraries with fragment lengths of approximately 8-10 kb, and we generated around 180Mb of data for each, equivalent to around 30x coverage of the *K. oxytoca* genome. After sequencing, we performed error correction using Nanocorrect<sup>9</sup>, assembly using the Celera assembler<sup>8</sup>, and manual finishing. We obtained a circular contig from each strain, covering approximately 4.4 Mb for C9 (N50=4.4Mb) and 6 Mb (N50 = 5.7Mb).

We then performed whole genome BLAST<sup>13</sup> to find out where these strains lie in the taxonomy. By this measure the closest match of D6 was to the strain KONIH1 (**Fig 1a.**). A Mummer plot<sup>3</sup> of the assembled genome against the KONIH1 strain reveals that the two strains are different, but highly similar (**Fig 1b.**). Whole genome BLAST of C9 revealed some similarity to *Pseudomonas alcaligenes* NBRC 14159, though a fully assembled version of this genome is not currently available for further comparison. A mummer plot of C9 against *Pseudomonas pseudoalcaligenes* CECT 5344 indicated significant differences (**Supplementary Figure 1**). We performed BLAST of the 16S gene of C9, and obtained many hits to various *Pseudomonas* species (**Supplementary Table 1**).



**Figure 1.** Genome comparison of one assembled bacterial strain to reference strains. **a)** taxonomy tree obtained by whole genome BLAST, with the assembled genome shown in red, **b)** Mummer plot of the assembled genome against the most closely matching reference genome.

In this study, we have adapted the iChip protocol to allow larger-scale growth of individual bacterial isolates, making it possible to obtain sufficient DNA for whole genome sequencing, and we have demonstrated that, in combination with Oxford Nanopore's MinION, whole genome assemblies can be derived, allowing genome comparisons to be made. This will also enable further study of these bacterial strains, which had not previously been isolated or sequenced. The MinION device itself brings genome sequencing within the reach of small laboratories, and brings greater ownership of all parts of the sequencing process back to individual researchers. The combination of the iChip and the MinION will allow rapid, inexpensive and straightforward discovery and genomic analysis of previously uncultured prokaryotic genomes. It is important to note that the slowest step in this procedure is currently the culturing of the 'unculturable' bacteria: DNA extraction and library preparation can be performed in a morning, and sequence information is generated by the MinION

in real time, starting virtually as soon as the first strand of DNA has translocated through a nanopore. Consequently, removal of the need to culture the bacteria could bring the time-to-result as low as a few hours. As throughput increases, the long read capability of the MinION will make metagenomic approaches the quickest route to assembling and characterising unculturable genomes.

## Method

We made an acrylic iChip<sup>5</sup> by 3D printing (Projet HD), dipped the central panel into 1% low melting point agarose (Invitrogen), allowed the gel to set at room temperature, and removed excess agarose with a scalpel. We then took a sample of water from the stream behind the Oxford Nanopore offices in Oxford, UK., dipped the iChip panel into the water sample, and assembled the device as described<sup>5</sup>. We left the assembled iChip for 14 days in the stream, after which, we cleaned it externally and dismantled it. Using sterile cocktail sticks, we pushed agarose plugs out of iChip wells into separate wells of a deep-well 96-well plate (Bio-Rad) containing low salt LB medium (Sigma) which we had made with water from the stream, filtered (Corning) and sterilised by autoclaving. We sealed the plate with an air-permeable seal (Thermo Scientific) and left it on a plate shaker (Grant Bio) at 300rpm on a lab bench for two days so that the cultures would grow. We prepared Petri dishes containing low salt LB medium, again using filtered and sterilized stream water, but with 15 g/L agar (Sigma). We streaked out the cultures and allowed them to grow inverted on a laboratory bench for 2 days. Following this, we picked single colonies and grew them in 25ml liquid medium as previously described. We extracted DNA using a Qiagen 500Tip, following the manufacturer's protocol, prepared PCR-free sequencing libraries using Oxford Nanopore Technologies' version 6 library preparation kits, following Oxford Nanopore's protocols, and sequenced each genome on separate flow cells following Oxford Nanopore's standard running and analysis protocols.

We error-corrected data from each genome twice using NanoCorrect<sup>9</sup>, and assembled the corrected reads using PBCR<sup>8</sup>. We performed whole genome BLAST using discontinuous megablast<sup>14</sup>, and generated genome alignment plots using Mummer<sup>3</sup>

## Acknowledgements

We are grateful to Ameet Pinto for the stl files used for 3D iChip printing, and to Anthony Jones for printing the iChip.

## References

1. Kaeberlein T, *et al.* Isolating "uncultivable" microorganisms in pure culture in a simulated natural environment. *Science* **296** (5570), 1127-1129 (2002)
2. Ling LL, *et al.* A new antibiotic kills pathogens without detectable resistance. *Nature* **517** (7535), 455-459 (2015)
3. Kurtz S, *et al.* Versatile and open software for comparing large genomes. *Genome Biology* **5** (2), R12 (2004)
4. Scholz M, *et al.* Improved assemblies using a source-agnostic pipeline for MetaGenomic Assembly by Merging (MeGAMerge) of contigs. *Scientific Reports* **4**, 6480 (2014)
5. Nichols D, *et al.* Use of ichip for high-throughput in situ cultivation of "uncultivable" microbial species. *Applied and Environmental Microbiology* **76** (8), 2445-2450 (2010)
6. Koren S, *et al.* Reducing assembly complexity of microbial genomes with single-molecule sequencing. *Genome Biology* **14** (9), R101 (2013)
7. Jain M, *et al.* Improved data analysis for the MinION nanopore sequencer. *Nature Methods* **12** (4), 351-356 (2015)
8. Berlin K, *et al.* Assembling large genomes with single-molecule sequencing and locality-sensitive hashing. *Nature Biotechnology* **33** (6), 623-630 (2015)

9. Loman NJ, *et al.* A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nature Methods* **12** (8), 733-735 (2015)
10. Szalay T & Golovchenko JA. De novo sequencing and variant calling with nanopores using PoreSeq. *Nature biotechnology* (2015)
11. Wood DE & Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology* **15** (3), R46 (2014)
12. Vogelstein B & Kinzler KW. Digital PCR. *Proceedings of the National Academy of Sciences of the United States of America* **96** (16), 9236-9241 (1999)
13. Boratyn GM, *et al.* BLAST: a more efficient report with usability improvements. *Nucleic Acids Research* **41** (Web Server issue), W29-33 (2013)
14. McGinnis S & Madden TL. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Research* **32** (Web Server issue), W20-25 (2004)