Assembling complete genomes for oral pathogens and methanogens

Mia Sales, Undergraduate Research Assistant

Jensen Lab

Department of Bioengineering and Carl R. Woese Institute for Genomic Biology

University of Illinois at Urbana-Champaign

Research Summary

My Summer, 2017 research focused on utilizing the ONT MinION sequencer and assembly/polishing programs to determine the complete genomes of *Streptococcus sobrinus* and two species of the genus *Methanosarcina*, none of which have previously been fully sequenced**.**

Abstract

The identification of an organism's complete genome not only provides crucial insight into how genes interact, but it allows for the study of metabolic pathways which can be manipulated and redesigned to meet global needs. In previous years, the scientific community has faced difficulties in genome sequencing primarily due to the lack of long-read technologies. Using the new MinION sequencing device from Oxford Nanopore Technologies along with several bioinformatics tools, I am able to determine the complete genome of *Streptococcus sobrinus* SL1 as well as the complete genomes of two strains of *Methanosarcina thermophila*. With these data, we hope to identify potential drug targets to more efficiently treat dental diseases as well as widen our understanding of metabolic pathways in methane production, progressing the search for alternative energy sources.

Introduction

Streptococcus sobrinus and *Streptococcus mutans* are two particularly dangerous oral pathogens (Loesche 1986). The bacteria consume sugars in the mouth to create acid, which drops the pH and causes teeth to demineralize— bread down—and form cavities. *S. sobrinus* and *S. mutans* thrive in the mouths of children, and are particularly dangerous in individuals who consume large amounts of sugar. They are the two most common bacteria associated with tooth decay because of their ability to survive in and contribute to excessively low pH. When decay permeates the dentin layer of the tooth—just under the approximately 1mm thick enamel—it causes irreversible damage to the tooth.

Currently, the majority of research concerning tooth decay is focused on *S. mutans*, and the complete *S. mutans* genome has been sequenced and extensively manipulated. Due to difficulties working with *S. sobrinus* in the lab, it is far less studied than *S. mutans*, and no complete genome sequence is available. However, *S. sobrinus* is arguably the worse pathogen, as it has been shown to produce higher levels of acid. To understand the differences between these species, our lab investigates the metabolic strategies of the two bacteria and the metabolic pathways found only in *S. sobrinus* (Conrads 2014). During my summer of research, I was able to determine the complete genome of *S. sobrinus,* allowing for the comparison of *S. sobrinus* with the already available genome of *S. mutans* with the ultimate goal of identifying drug targets that can apply to both species.

Difficulties in genome sequencing have previously stemmed from the lack of long-read technologies; standard Illumina sequencing uses “sequencing by synthesis” which allows for sequencing of up to 250 base pairs in one continuous read. Often, this process is inefficient because the inability to sequence long reads results in the output of a very large quantity of short

reads; reassembling a large number of short reads is error-prone and the resulting genome is often hugely incomplete. However, the MinION sequencer developed by Oxford Nanopore Technologies can obtain a much higher sequencing length, dependent only on the length of DNA that can be isolated and inputted to the device.

The MinION is a device used to analyze single molecules such as proteins, RNA, and DNA in real-time. It is composed of a nanopore inserted into an electrically resistant membrane of insulating polymers. A voltage is set across the membrane to create an ionic current flow through the center of the nanopore such that molecules passing through the pore cause a characteristic disruption in the current. The resulting signal can then be analyzed to identify the specific molecule, or in the case of DNA, to identify the sequence of nucleotides passing through the nanopore. The nanopore channels are scaffolded to a microchip sensing array located in a one-time use flowcell.

Results

Using the MinION sequencer, its associated software, and several bioinformatics tools, I was able to assemble the complete genome of *S. sobrinus* SL1. The first step in this process is DNA isolation and library preparation. Specific steps for this process were followed according to the ONT guide. After loading the DNA library into the MinION, the sequencing is allowed to run for a maximum of 48 hours, or until sufficient data is obtained. Following this step, the MinION software, MinKNOW, compartmentalizes the sequence reads into folders designated as "Pass", "Fail", or "Skip". The passed reads are then converted from FASTA to FASTQ format using Poretools. The individual files are then edited such that the first and last 100 base pairs of each read are removed, ensuring that the adaptor sequences ligated to the DNA strands prior to sequencing are no longer present. During this time, any existing barcodes can also be removed

using the same program. Next, the reads are assembled using CANU. Once completed on the *S. sobrinus* data, this process outputted a single contiguous sequence. The rough genome was then polished using a docker program specifically created for Oxford Nanopore data, the ONT Assembly and Polishing Pipeline.

Another group of organisms with incomplete, yet intriguing genomes belong to the genus *Methanosarcina*. These archaea thrive in anaerobic conditions and are responsible for high methane production. At the Field Museum in Chicago during World of Genomics, an outreach event by the Carl Woese Institute for Genomic Biology, three species of *Methanosarcina* were sequenced live using the ONT MinION sequencer as a collaboration with the Metcalf Lab at IGB. The species, *Methanosarcina baltica* and *thermophila***,** were extracted and isolated from the Gotland Deep in the Baltic Sea and from a thermophilic sludge digester, respectively (von Klein, 2017, Catalogue, 2017). Of the three sequencing runs, two were of sufficient quality and coverage to allow for completion of the same genome assembly pipeline through genome polishing. These belonged to the species *M. thermophila*, strains MST-A1 and TM-1.

The final *S. sobrinus* SL1 genome measured 2.23 Mbps and is the first fully assembled *Streptococcus sobrinus* genome. The two *Methanosarcina* genomes measured 3.256 Mbps for MST-A1 and 3.216 Mbps for TM-1. While the two methanogens have been sequenced in the past, these genomes are the first closed-gap assemblies.

Future steps include sequencing three additional strains of *Streptococcus sobrinus.* With these new genomes, I will be able to determine the degree of variability within the *S. sobrinus* species, providing crucial insight into whether the most dangerous qualities found in *S. sobrinus* are specific to one strain, or if they are representative of the species as a whole. I will further my research by comparing the assembled *S. sobrinus* genome with the available *S. mutans* genomes

to identify potential drug targets common to both species, so that we may be able to more efficiently treat dental diseases. The *Methanosarcina* genomes will likely be polished again using the gapped contiguous sequences from previous sequencing runs. Once completed, the two genomes will serve to enhance our understanding of the metabolic pathways responsible for methane production, ultimately providing insight into possible avenues for alternative energy sources.

References

- Loesche, W J. “Role of Streptococcus Mutans in Human Dental Decay.” Microbiological Reviews 50, no. 4 (December 1986): 353–80.
- Nascimento, Marcelle M, José A C Lemos, Jacqueline Abranches, Reginaldo B Gonçalves, and Robert A Burne. “Adaptive Acid Tolerance Response of Streptococcus Sobrinus.” Journal of Bacteriology 186, no. 19 (October 2004): 6383–90. doi:10.1128/JB.186.19.6383-6390.2004.
- Conrads, Georg, Johannes J de Soet, Lifu Song, Karsten Henne, Helena Sztajer, Irene Wagner-Döbler, and An-Ping Zeng. “Comparing the Cariogenic Species Streptococcus Sobrinus and S. Mutans on Whole Genome Level.” Journal of Oral Microbiology 6, no. 0 (2014): 26189. doi:10.3402/jom.v6.26189.
- How it works. https://nanoporetech.com/how-it-works. (2008)
- von Klein D, e. “Methanosarcina baltica, sp. nov., a novel methanogen isolated from the Gotland Deep of the Baltic Sea.” –Pubmed –NCBI. Ncbi.nlm.nih.gov (2017). at http://www.ncbi.nlm.nih.gov/pubmed/12013430
- Catalogue: DSM-1825. *Dsmz.de* (2017). at https://www.dsmz.de/de/kataloge/catalogue/culture/DSM/-1825.html?tx_dsmzresources_pi5%255BreturnPid%255D=304