The San Francisco Estuary Microbiome: Circularizing Genomes of Uncultured Bacteria, Archaea, Phage, and Phage Satellites

Lauren M. Lui[1*](lmlui@lbl.gov), Torben N. Nielsen[1*](torben@lbl.gov),
*[1]Lawrence Berkeley National Lab, Berkeley, CA; *contributed equally to this study*

Abstract

To fully understand a microbial system requires knowledge of the genomes of the organisms comprising the system. In the past, this has meant devising a plethora of schemes for sequencing and assembling metagenomes derived from environmental samples followed by attempts to bin the resulting contigs into something that resembles a genome. Generating complete circular genomes was rare due to the short reads typically being used.

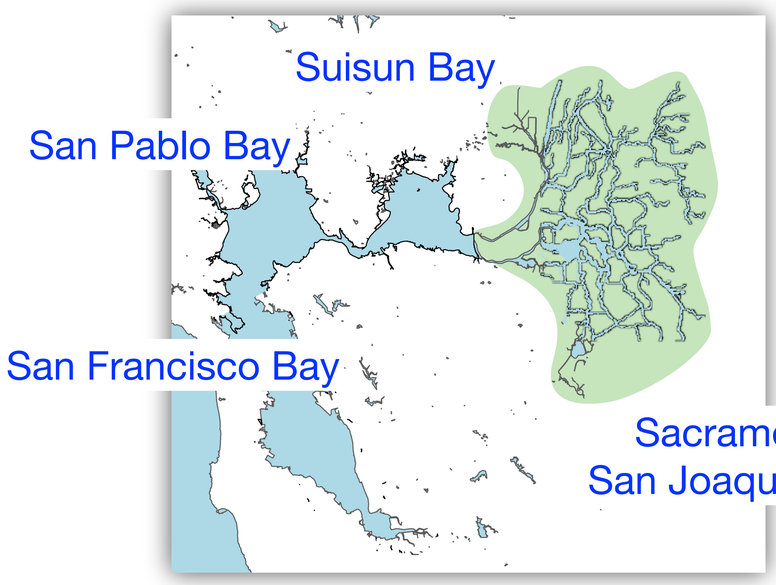
Today we are able to routinely sequence metagenomes using long reads. Oxford Nanopore Technologies provides a platform that is particularly attractive for having a low barrier to entry. We have used long reads to sequence and assemble multiple environmental metagenomes. We commonly extract a number of full length circular bacterial genomes from each metagenome. In some cases, we do not have the required depth of sequencing to complete a genome. However, we can generally use information from the assembly graph to establish reliable bins where the genomes are in just a few contigs.

While our main focus has been on getting complete circular bacterial genomes, long read sequencing and assembly also generates a large number of smaller circular elements. Many of the larger ones can be unambiguously classified as either plasmids or phage, but there are many that cannot.

As part of a Laboratory Directed Research and Development Project at Lawrence Berkeley National Laboratory we are attempting to decompose the entire San Francisco Estuary Microbiome. We have sequenced one sample with approximately 200 Gbases and we have extracted several complete circular bacterial genomes from it. In addition to those, we have found over 5,000 circular elements smaller than 500 Kbases. Using metagenomes sequenced less deeply, we still find large numbers of these elements and they match up across geographically separated samples.

Materials and Methods

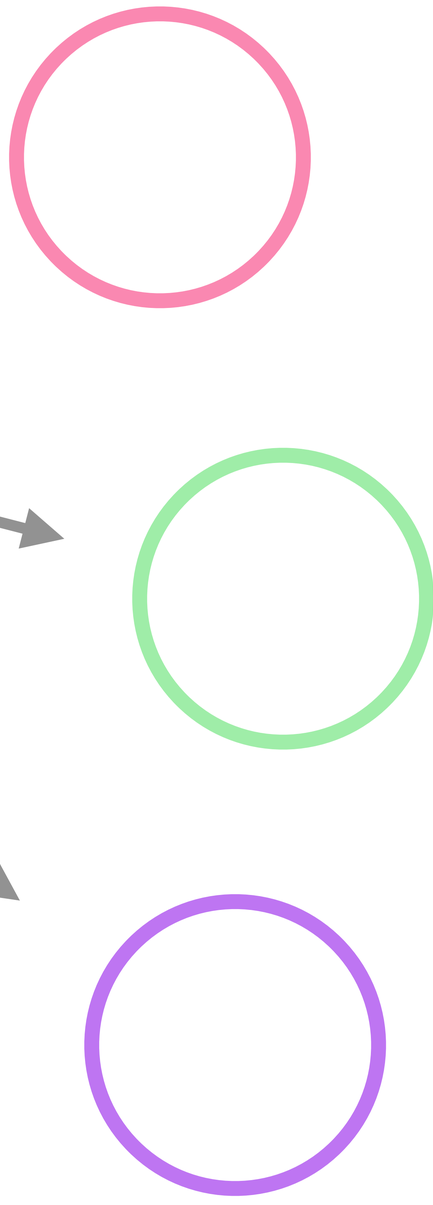
- Filtered water collected by USGS from their regular sampling
- Bead Extraction method to help with obtaining HMW DNA
- Nanopore sequencing on MinION and PromethION



The San Francisco Estuary is a critical nexus of water, commerce, agriculture, and recreation for millions of people in California

Nanopore Sequencing of A San Francisco Estuary Metagenome

Nanopore sequencing helps finish genomes, but creates more questions about strains in the marine environment, and what exists that aren't prokaryotes or eukaryotes.



SAR86

A ubiquitous marine heterotroph with no cultured representatives

Planktomarina temperata

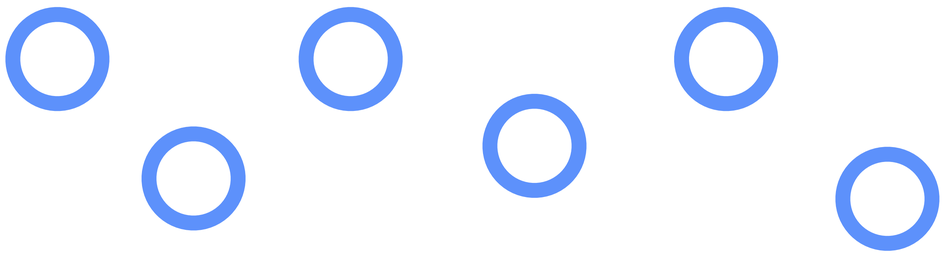
Despite it's pretty name, it's function is similar to a vulture for the marine environment

Actinomarina sp.

Lots of these species in the marine environment, yet very few cultured

Viruses, Plasmids, and Phage Satellites

We found over 5,000 circular DNA elements. Some of them are plasmids, phage, and what we think are phage satellites. Some are only 500bp.



Between the 3 genomes above, they account for ~25% of all prokaryotic abundance in marine environments. We know that SAR11 strains account for another ~25% of everything, but we have yet to complete the genomes.

ACKNOWLEDGEMENTS
We thank the United States Geological Survey (USGS) for sample collection and the Joint Genome Institute for sequencing support.