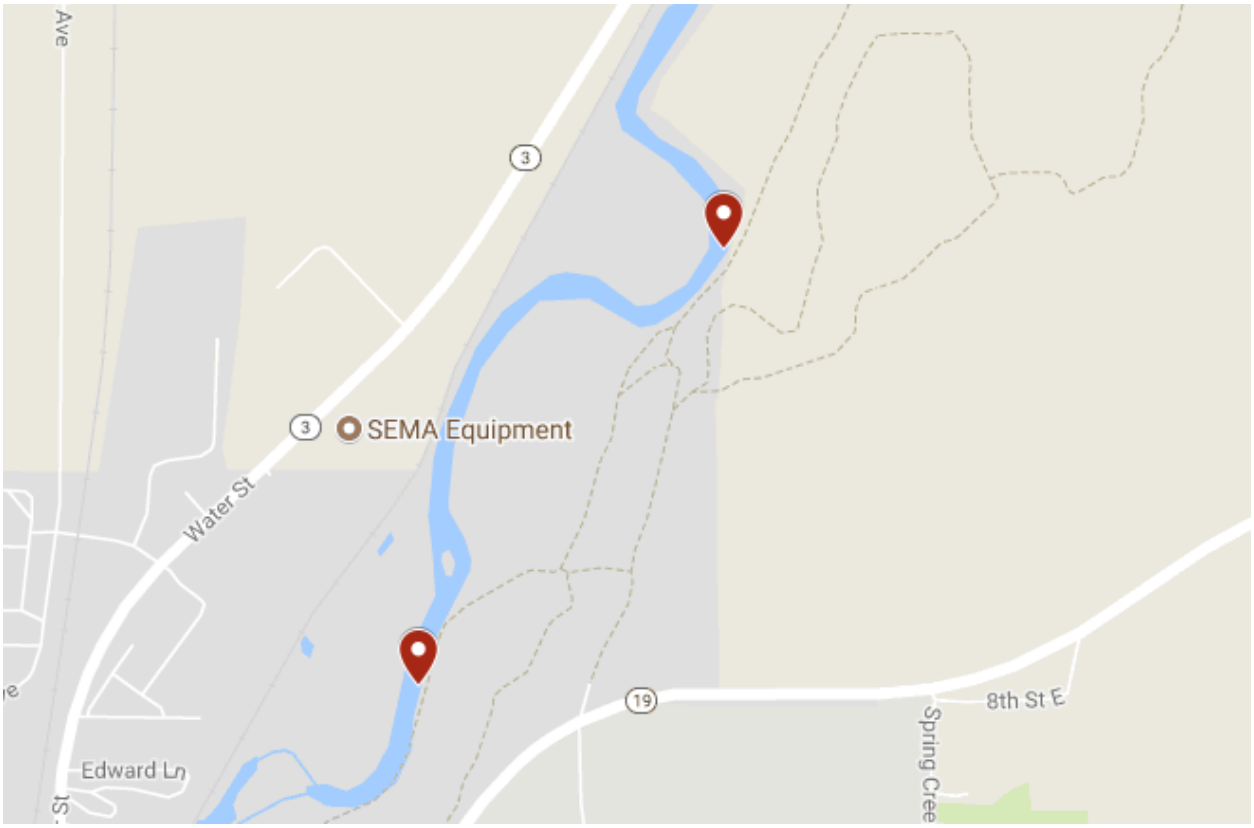
Lab 2: Question Responses
Dustin Michels
26 Sep 2017

1) Sample site and procedure.

In order to monitor microbial ecology in the Canon River—which runs through Northfield, Minnesota—we sampled water from two different locations. The first sample (site 6) was taken upstream from a sewage treatment plant (16:05, September 12, at coordinates 44.46788381, -93.15250309.) The second sample (site 7) was taken downstream of a sewage treatment plant (16:30, September 12, at 44.47493275, -93.14561752). For each sample, we filled a sterile, autoclaved plastic bottle with surface water (~3 inches deep). The sample water was pressed through a syringe into a Sterivex filter, in a quantity of 180-240mL. The filter was capped, and stored at -80˚C for one week at which point we extracted DNA from the Sterivex using the MoBio PowerWater Sterivex Extraction kit according to the manufacturer's directions.



2) Metadata collected and concentration of extract.

In order to properly contextualize any DNA sequencing data, information was recorded about each sampling event. The metadata collected consists of date and time of sample; latitude and longitude of sample location; dissolved oxygen concentration, temperature, pH, and salinity of the water source being sampled; amount of water filtered through the Sterivex, the depth of sample; and general, qualitative descriptions of the site. Unfortunately, our final DNA extract did not have a meaningful concentration, suggesting we performed a step incorrectly.

3) Impact of genome shearing on bioinformatics projects.

It seems that the chemical/mechanical shearing of genomes might be more problematic for a bioinformatics project focused on genome structure than one focused on microbial diversity. This is because it may be possible to identify *which* organism a given piece of DNA belongs to, even if it is a fairly small piece, while the assembly of small DNA fragments into larger chunks necessary to asses genome structure might be challenging, depending on how small the fragments are. In suggesting identification may be possible from small fragments, I am thinking of the small region of DNA that codes for a highly conserved but distinctive RNA molecule, which can often be used to identify genomes (Though not if the DNA fragments are too small.) In suggesting reliable DNA assembly from fragments might be problematic, I am picturing the “bubbles” that can occur when using the De Bruijn graph algorithm, and the possibility of ending up with several viable DNA structures.

4) Appropriate sequencing platform.

In order to evaluate community diversity, it would be useful to retain information about the relative abundance of various genomes, not just identify the types. It seems that quantitative/real-time PCR may be a useful tool, as it would allow us to determine the quantity of a known sequence in a sample. This could be used in conjunction with a sequencing technology that helps us actually determine what the sample is.

5) Effect on analysis of DNA extraction procedures, literature search.

In the first paper I read, the authors examined six different DNA extraction methods (Viltrop et al., 2010). They were interested in testing to what extent saliva sampling could replace blood sampling, motivated by the fact that saliva extraction is less invasive to medical patients. In investigating this question, they quantified differences in the quality of DNA sequences produced by the various methods. The authors used two different kits to extract DNA from saliva-- *Oragene DNA OG-250,* and *PSP SaliGene*-- and three kits were used to extract DNA from blood-- *QIAamp DNA Mini Kit, PAXgene Blood DNA System*, and *Fermentas Genomic DNA Purification Kit*. The sixth method was meant to be reference, and it involved using a phenol–chloroform to extract DNA from fresh blood. Random samples from each extraction method were subjected to amplification via multiplex PCR and genotyping via a "124-plex APEX-2 microarray." Some metrics the authors tracked were:

- **DNA yield** in μg. They found that "The Oragene kit showed remarkable variability in DNA concentration (0.9–36 μg)" while "The QIAamp kit yielded the most reproducible results (3.6–8.6 μg)."
- **DNA purity**, assessed by measuring the $A_{260}/_{280}$ of the samples to determine the amount of organic contamination. They found that "The SaliGene kit samples showed the most variability of A260/280 ratios."
- **PCR efficiency**. They found that "The greatest PCR efficiency (minimal relative PCR inhibition) was obtained with samples from the phenol–chloroform (100%) and Oragene (96%) extraction methods."
- **Ratio of mtDNA to gDNA**. This is investigated because "It has been proposed that high levels of mtDNA decrease amplification of autosomal DNA, especially Y chromosome DNA." They found that "the SaliGene and PAXgene kits gave much higher DNA ratios than the other kits."
- **The cell rate**, defined as percentage of actual genotype calls related to possible number of calls. This metric comes from the genotyping stage, and they found "The highest number of false signals (11/16) occurred with the Oragene extraction method." On the

other hand, “a single false signal occurred with the QIAamp kit. The Fermentas kit did not yield any false signals.”

Overall, the author concluded there were significant differences in DNA quality and purity between the different methods, but not in the efficiency of multiplex PCR or the oligomicroarray signals from the APEX-2 array. As for reproducibility: the paper certainly seems to be focused on saliva and blood sampling. The fact that different kits were used for saliva and blood suggests that those particular kits cannot be widely applied to other kinds of DNA sample types.

A second (slightly more esoteric) article I found was titled “DNA extraction from formalin-fixed laryngeal biopsies: Comparison of techniques” (Torrente et al., 2011). It sought to compare DNA extraction protocols for formalin-fixed paraffin-embedded (FFPE) samples. It seems this entails tissue samples that have been preserved using formaldehyde and paraffin. They explored various techniques for DNA extraction, proteinase K digestion, and organic extraction. They also looked at various “deparaffinization” techniques: alkaline heat, xylene, and no removal. To assess DNA quality, they looked at PCR amplification of the HFE gene, for amplicons of 208 and 390 bp. It seems they were trying to uncover new techniques for DNA extraction, and comparing them against a commercial kit as a baseline. They determined that “Extraction with the commercial kit and proteinase K digestion were more efficient than other techniques.” Once again, the study seems to be fairly specific to particular kinds of sample matter (tissue from biopsies, in this case), and specifically paraffin embedded tissue.

Bibliography:

Torrente, M.C., Ros, C., Misad, C., Ramrez, R., Acua, M., and Cifuentes, L. (2011). DNA extraction from formalin-fixed laryngeal biopsies: Comparison of techniques. Acta Oto-Laryngologica, 2011, Vol.131(3), p.330-333 *131*, 330–333.

Viltrop, T., Krjutškov, K., Palta, P., and Metspalu, A. (2010). Comparison of DNA extraction methods for multiplex polymerase chain reaction. Anal. Biochem. *398*, 260–262.