

**Genome assembly and comparison of *Sulfolobus Islandicus*.** Y. Zhou<sup>1</sup> R. Anderson<sup>2</sup> and R. J. Whitaker<sup>3</sup>,  
<sup>1</sup>yzhou103@illinois.edu <sup>2</sup>rikander@illinois.edu <sup>3</sup>rwhitaker@life.illinois.edu.

**Introduction:** Extremophiles such as the genus *Sulfolobus* thrive in a range of extreme habitats that were once thought to be inhospitable for life. These extreme habitats provide barriers to dispersal, allowing us to better investigate population differentiation and its relationship to ecological conditions. Previous study has shown that *Sulfolobus acidocaldarius* genomes are more conserved than *Sulfolobus islandicus*, despite being closely related and sharing the same geothermal habitats. Our study focus on 42 *S. islandicus* genomes isolated from Nymph Lake, Yellowstone National Park in 2012. We dedicate to shed light on how *S. islandicus* are more subject to inter-species variation.

Since the *S. islandicus* genomes are more variable, and the public genome data of those isolated from Yellowstone National Park are limited, it is likely to introduce bias via mapping the assemblies against a reference to obtain a draft genome. Here we sequenced 42 isolates on an Illumina HiSeq2000 and three isolates from three different hotsprings (Monitor, North/Spunky, and Prosperous Point) using PacBio RSII. The *S. islandicus* genomes were much more difficult to assemble, likely due to the presence of mobile elements. The scaffolds number is around 150-200 after a5 pipeline assembly, which makes it hard to define the orientation as well as the order. We are using a variety of strategies to obtain full draft genome sequences without a reference genome. SPAdes assembler was used to perform a hybrid assembly with illumina reads and PacBio data, while PBJelly2 was used to fill the gaps between scaffolds using PacBio reads as references. So far we managed to reduce the scaffolds size, and identify core blocks of genes using Mauve. Those shared gene blocks are aligned with ClonalFrame for genome comparison purpose. We map variable gene content onto core genome to determine how genes are distributed among strains from different hot spring populations. We find considerable variation in gene content and in core genome variation consistent with *S. islandicus* being a more variable genomic background than *S. acidocaldarius* strains from the same environment.