

# Comparative Metagenomic Genome Analysis of *Synechococcus* spp. in Microbial Mats Across a Temperature Gradient in Hot Springs From Yellowstone National Park



10-Jun-2022  
Poster #3231

Harihara Subrahmaniam Muralidharan<sup>[1]</sup>, Jacquelyn S Michaelis<sup>[1]</sup>, Justin Tubilla<sup>[1]</sup>, Feiqiao Brian Yu<sup>[2]</sup>, Devaki Bhaya<sup>[3]</sup>, Mihai Pop<sup>[1]</sup>  
[1]University of Maryland, College Park [2]Chan Zuckerberg Biohub, Microbiome Initiative, San Francisco, CA [3]Carnegie Institution for Science, Department of Plant Biology, Stanford, CA

## Abstract

**Background:** Microbial mats from the Mushroom and Octopus hot springs in Yellowstone National Park are well-studied, dynamic systems that contain diverse phototrophs and heterotrophs. The cyanobacterium *Synechococcus* is abundant in these mats along a stable temperature gradient from ~50C to ~70C. They play a key role in fixing carbon and nitrogen. Previous studies have isolated and generated quality reference sequences of two major *Synechococcus* spp.; OS-A and OS-B'. OS-A is more prevalent at higher temperatures and OS-B' is predominant at lower temperatures of the springs.

**Dominant Organisms:** While both species have very high shared genomic content, the extent of diversity within the *Synechococcus* spp under varying conditions remains underexplored, partly because re-assembly of these organisms from metagenomes is challenging due to the high number of rearrangements within, and potentially between, the two species. In this work, we take a systematic approach to explore the genomic diversity of the *Synechococcus* spp. in 34 metagenomic samples from the two hot springs, comparing samples across time and temperature.

**Identifying Variations:** We provide a statistical framework to identify the differential rearrangements between the two species and examine if the rearrangements can be explained by the variation in temperature. Though good quality references to both *Synechococcus* spp. are available, we observe gaps in the reference when mapping back metagenomic reads. We systematically identify the insertion/deletions that are consistent across temperatures and explore potential reasons behind the observed patterns.

## References

Bhaya D, Grossman AR, Steunou AS, Khuri N, Cohan FM, Hamamura N, Melendrez MC, Bateson MM, Ward DM, Heidelberg JF. Population level functional diversity in a microbial community revealed by comparative genomic and metagenomic analyses. ISME J. 2007 Dec;1(8):703-13. doi: 10.1038/ismej.2007.46. Epub 2007 Oct 25. PMID: 18059494.

## Results

### 1. Denovo Assembly of *Synechococcus* Genomes is Fragmented. Many reassembled contigs did not map to the reference genome.

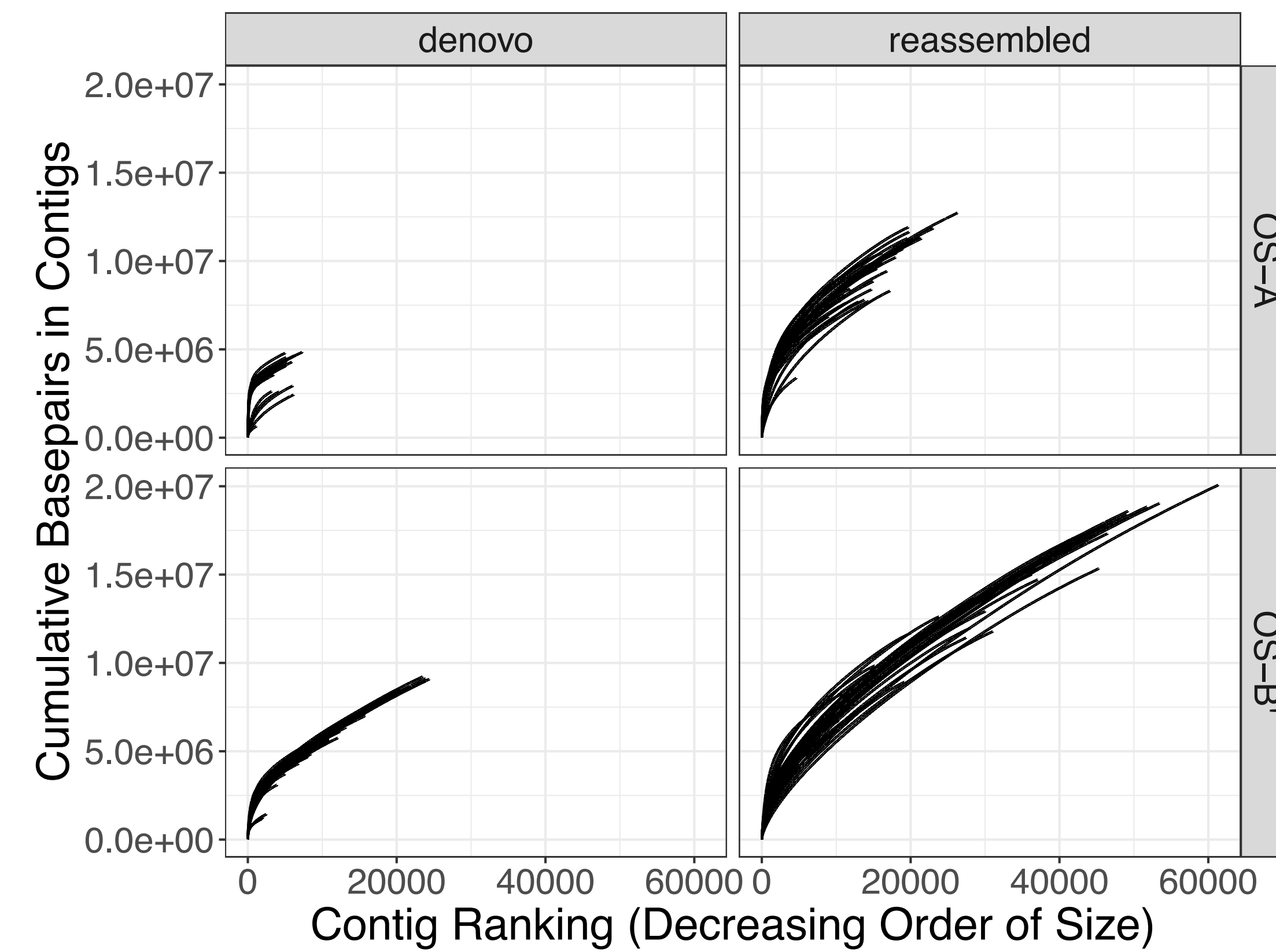


Figure 1. The contigs generated by denovo assembly and reference guided assembly are sorted in decreasing order of their lengths and the y-axis shows the cumulative no. of bases. The plots on the left and right panels shows the cumulative number of bases assembled by Denovo and Reference Guided Reassembly.

### 3. Distribution of Putative Variant Insertions along the *Synechococcus* Genomes.

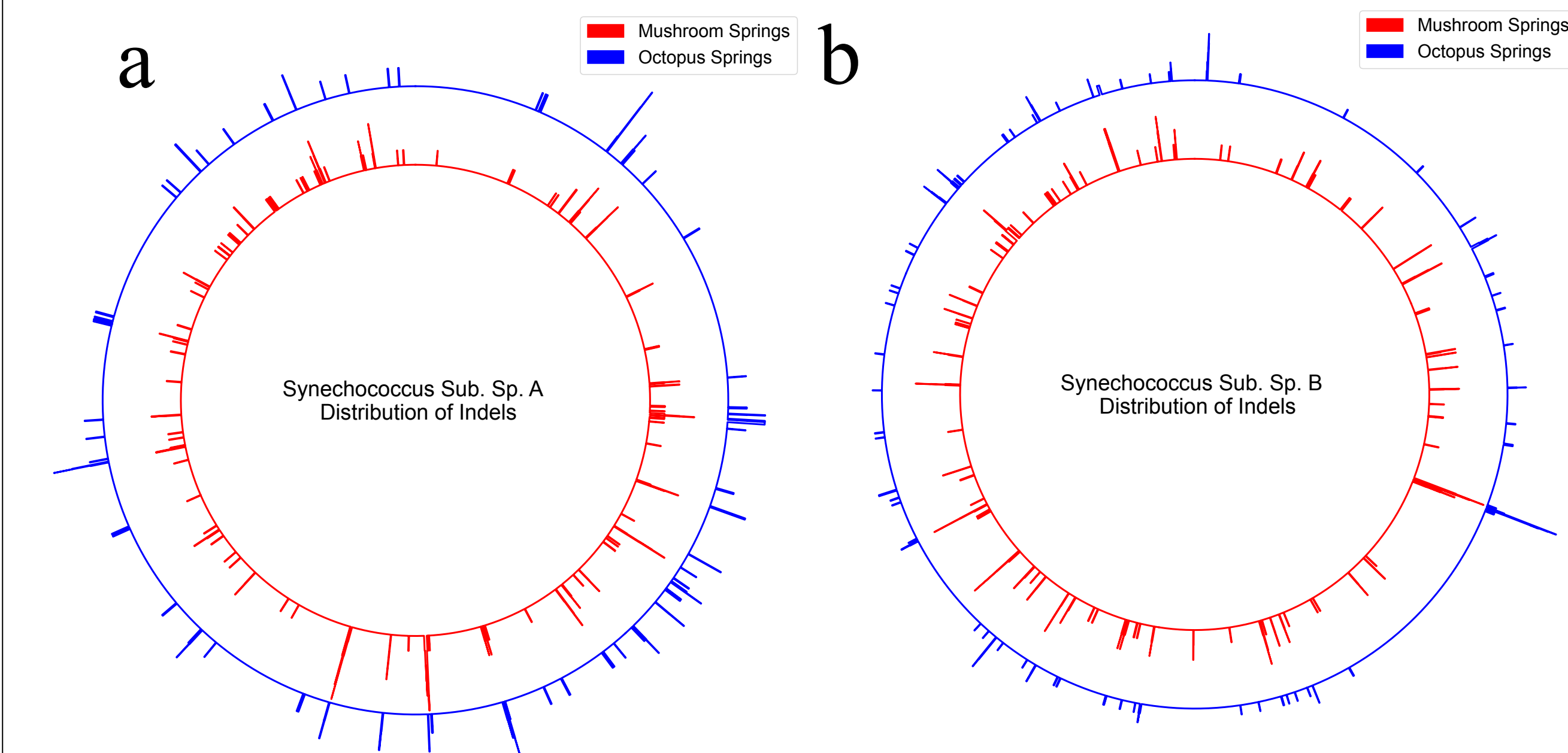


Figure 3. We learn the genomic location of variant insertion contigs through a reference guided scaffolding. We show the landscape of variant insertion along the (a) OS-A and (b) OS-B' genomes, across a temperature gradient for the Mushroom (smaller inner circle) and Octopus Springs (larger outer circle).

### 2. Reads aligning to *Synechococcus* Genomes are Rich in “Orphaned” and “Mismatched” Mates.

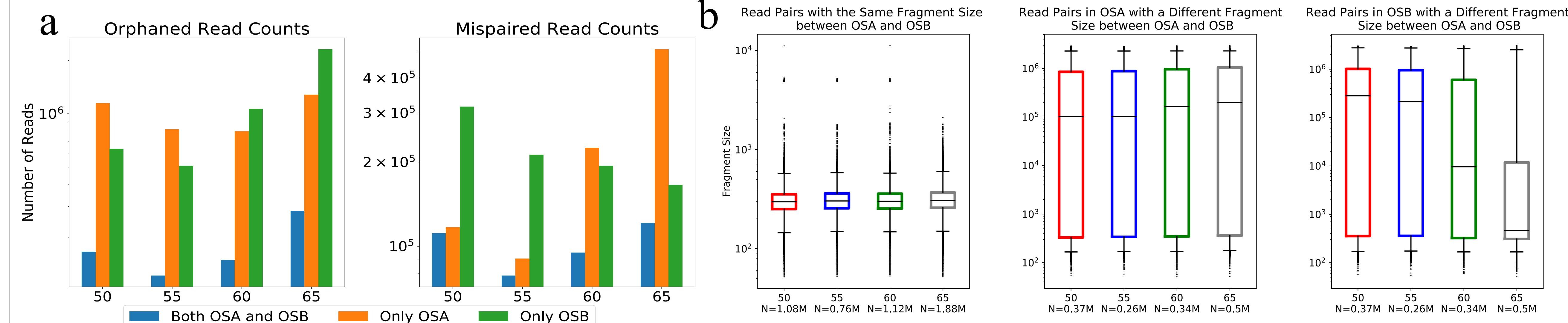


Figure 2(a). Differential orphaned and mismatched read counts for metagenomic reads from Mushroom Springs. Orphaned reads are those where exactly one of the read pairs align with the genomes. Mismatched reads are those whose read pairs are incorrectly oriented in the genome. These are often indicative of indels and rearrangements. (b). Distribution of Fragment Sizes between Read Pairs that map to both OS-A and OS-B'.

## Discussion

- Synechococcus* makes up about 15% of the community and it doesn't assemble well and produces a fragmented assembly. Reassembly is poor even with the help of references [Fig 1]. This could be due to the diversity within the species that cannot be captured by the OS-A and OS-B' reference genomes.
- Even between the reads that map to the reference genomes, many of these mates are orphaned and mismatched pairs [Fig 2a]. We see that the abundance of both mismatched reads and orphaned reads follow the temperature gradient. This could be artifactual of the increased abundances of OSA at higher temperatures and lower abundances of OS-B' at higher temperatures. Alternatively, this could also signal the presence of putative strains of OS-B'. For the reads pairs where both mates align to OS-A and OS-B' we notice that about 30% of the read pairs have different fragment sizes between the genomes [Fig 2b]. These different fragment sizes are indicative of indels or rearrangements.
- We see that some insertions are very specific to the hot springs [Fig 3], indicating the local environment plays a role in enabling the diversity. Gene predictions indicate that these insertions are enriched in phages, transposons and cell-wall repair. We also note that some of these insertions are common to OS-A and OS-B' indicating possible transfer events

Want to know more about microbial community in the hot springs? Stop by Poster #3997 on Sunday (12-Jun-2022).

## Future Work

- We want to establish if the putative variant insertions that we observed also have functional relevance.
- We will repeat a similar analysis on other members of the hot springs community to establish if genome fluidity is a general feature of thermophilic microbes or is specific to *Synechococcus* OS-A and OS-B'.
- We wish to establish if the novel insertion activity we have observed, in these hot spring samples in YNP and elsewhere.

The authors HM and JSM were supported by the NIH (award R01-AI-100947 to Mihai Pop). DB acknowledges support from NSF grant #1921429 and JGI-CSP grant \$503441. The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported under Contract No. DE-AC02-05CH11231.