

Review for:

“The chromosomal genome sequence of the fragile freshwater sponge, *Eunapius fragilis* (Leidy, 1851) and its associated microbial metagenome sequences”, by Sally P. Leys, Ute Hentschel, et al.

Public review by:

Darrin T. Schultz, PhD  
Postdoctoral Scholar  
University of Vienna  
June 6th, 2025

## Statement

Thank you for the invitation and opportunity to review this manuscript. I appreciate the authors' efforts in sequencing and annotating the genome of *Eunapius fragilis*, as well as its associated metagenome. Sponge biodiversity remains underrepresented in chromosome-scale genomic resources, so it's exciting to see this high-quality assembly made available.

## Manuscript Review

This manuscript presents the sequencing, assembly, and annotation of a freshwater sponge, *Eunapius fragilis*, and its symbionts, as part of the Aquatic Symbiosis Genomics Project (McKenna et al. 2021).

### Comments:

- The Background section provides a clear and informative overview of the biology of this species.
- It might be helpful to mention other freshwater sponge or demosponge assemblies published to date, some of which the authors have contributed to, as this would offer newcomers additional context and points of comparison.
- The genome report includes adequate detail for readers interested in technical metrics.
- Given the challenges of generating high-quality contig-level assemblies in demosponges, scaffolding the alternate haplotype to the chromosome scale could strengthen future comparative studies.
- While the report notes 23 scaffolds, it would be useful to also list the number of contigs that comprise these scaffolds.
- The HiGlass instance is a nice inclusion. I attempted to customize the view by adding scaffold names and tracks to the margins but wasn't able to.

- The metagenome report is clear. If possible, a Hi-C heatmap visualizing both host and symbiont scaffolds together would provide helpful insight into their separation, although I recognize this may not be part of the current pipeline.

## Genome Assembly Review

### Technical Comments:

- The dataset includes an impressive 500 million Hi-C read pairs. This level of coverage is an asset for haplotype phasing and symbiont-host contact resolution. That said, it may be more than necessary for scaffolding given the genome size.
- HiFi physical coverage is high (~118x), yet the assembly remains fragmented across ~800 contigs. In future demosponge projects, low-coverage ONT or CLR data could potentially be used with `--u1` mode with *hifiasm* to help improve contiguity.

### Materials and Methods

Hi-C reads used: ERR11641156 (from this project)

Assembly fasta used: <https://www.ebi.ac.uk/ena/browser/api/fasta/links/study?accession=PRJEB70489&result=sequence>

To evaluate the assembly, the Hi-C reads were mapped to the assembly fasta using Chromap v.0.2.7-r494 (Zhang et al. 2021) with quality cutoffs of both 0 and 1. The output was saved to a .pairs files (Open2C et al. 2023). Using a cutoff of 0 allows the visualization of multi-mapping reads and is useful in identifying haplotigs and repetitive sequences. The cutoff of 1 removes these multi-mapping reads. I generated an editable assembly file (Dudchenko et al. 2017) with a script based on the artisanal toolkit (Bredeson, n.d.). I then gzipped these pairs and converted them to the .longp format used by 3d-dna (Dudchenko et al. 2018), then converted that to an editable Hi-C map with the Juicebox Assembly Tools script ‘run-assembly-visualizer.sh’ (Dudchenko et al. 2018). Lastly, I manually inspected the assembly with q values 0 and 1 to look for potential misassemblies.

## Links:

- Script to generate the .assembly file
  - [https://github.com/conchoecia/genome\\_assembly\\_pipelines/blob/master/bin/assembly-from-fasta.py](https://github.com/conchoecia/genome_assembly_pipelines/blob/master/bin/assembly-from-fasta.py)
- Path to snakemake pipeline used to generate the editable .hic heatmap
  - [https://github.com/conchoecia/genome\\_assembly\\_pipelines/blob/master/snakefiles/GAP\\_hic\\_map7\\_genomeReview](https://github.com/conchoecia/genome_assembly_pipelines/blob/master/snakefiles/GAP_hic_map7_genomeReview)
- Path to example config file for the above snakemake pipeline
  - [https://github.com/conchoecia/genome\\_assembly\\_pipelines/blob/master/example\\_configs/config\\_GAP\\_hic\\_map7\\_genomeReview.yaml](https://github.com/conchoecia/genome_assembly_pipelines/blob/master/example_configs/config_GAP_hic_map7_genomeReview.yaml)

## Assembly Comments Foward

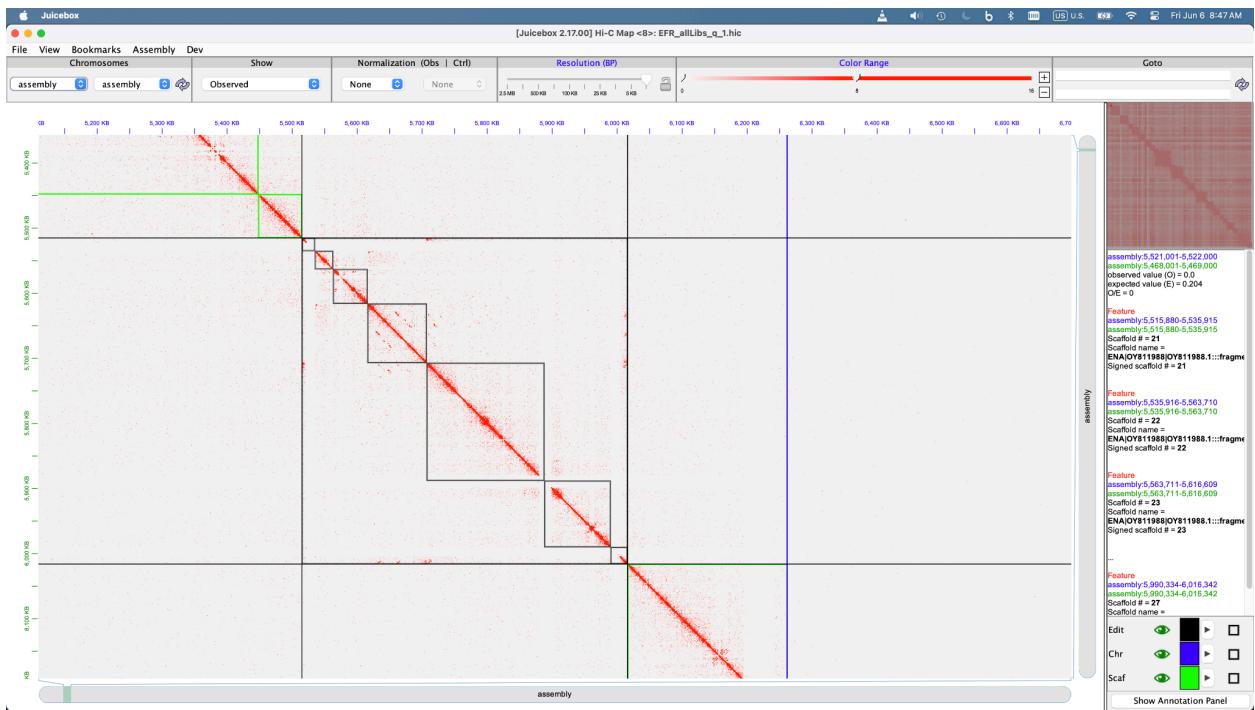
This section includes screenshots from Juicebox. Blue lines represent scaffold boundaries, green lines mark contig boundaries, and black highlights indicate highlighted regions. The fasta header of the highlighted regions are shown on the right-hand side of the screenshots. Screenshots follow the order of the genome FASTA.

## General Assembly Comments

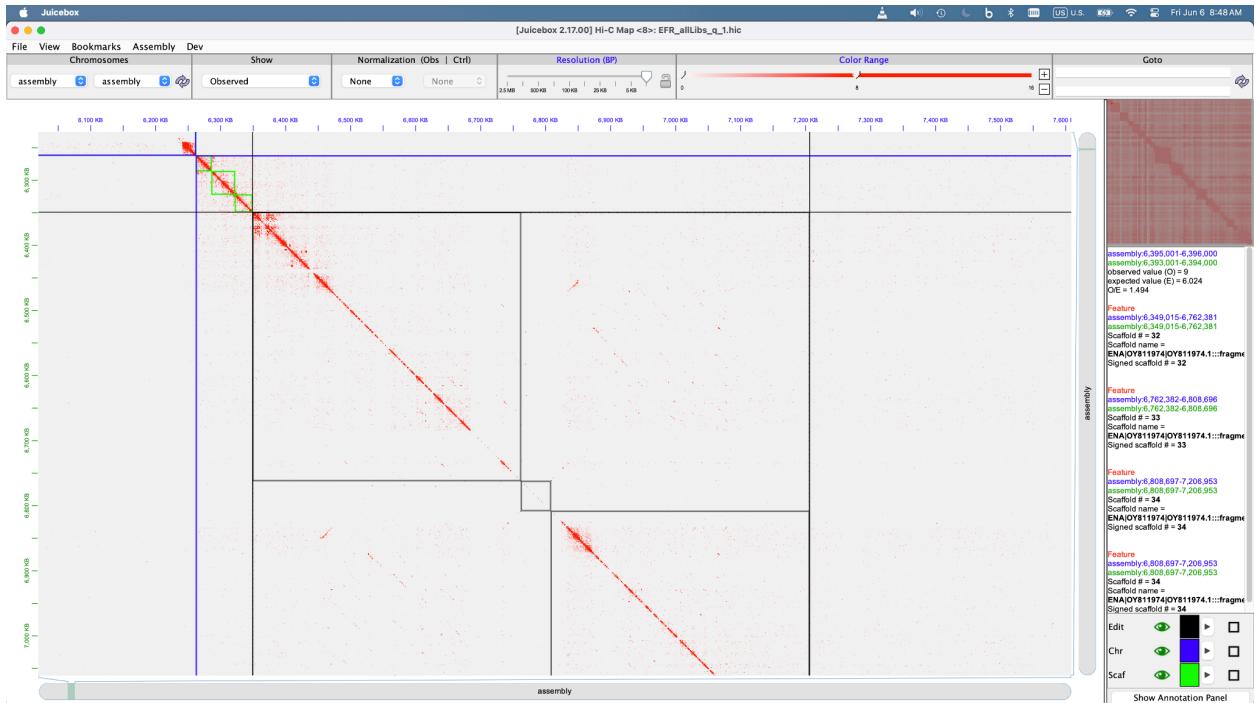
Overall, the assembly is strong. I did note several instances of likely misplacements, overlapping contig ends, and what appear to be redundant haplotigs. These are not exhaustive but represent the clearest examples observed during manual inspection.

## Assembly Screenshots

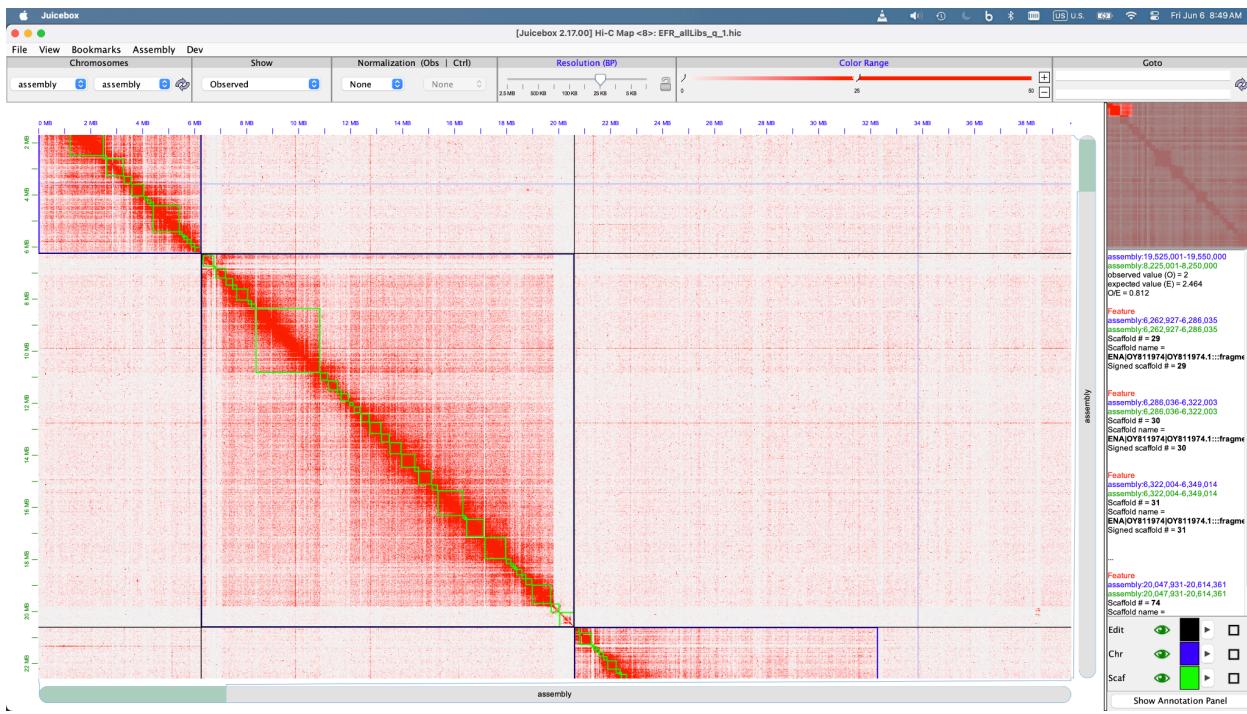
I have attempted to put these screenshots in the order in which they occur in the genome fasta file! This is not an exhaustive list of things that could be fixed, just the examples that were the most clear. For the sake of time I have kept my comments accompanying the screenshots brief.



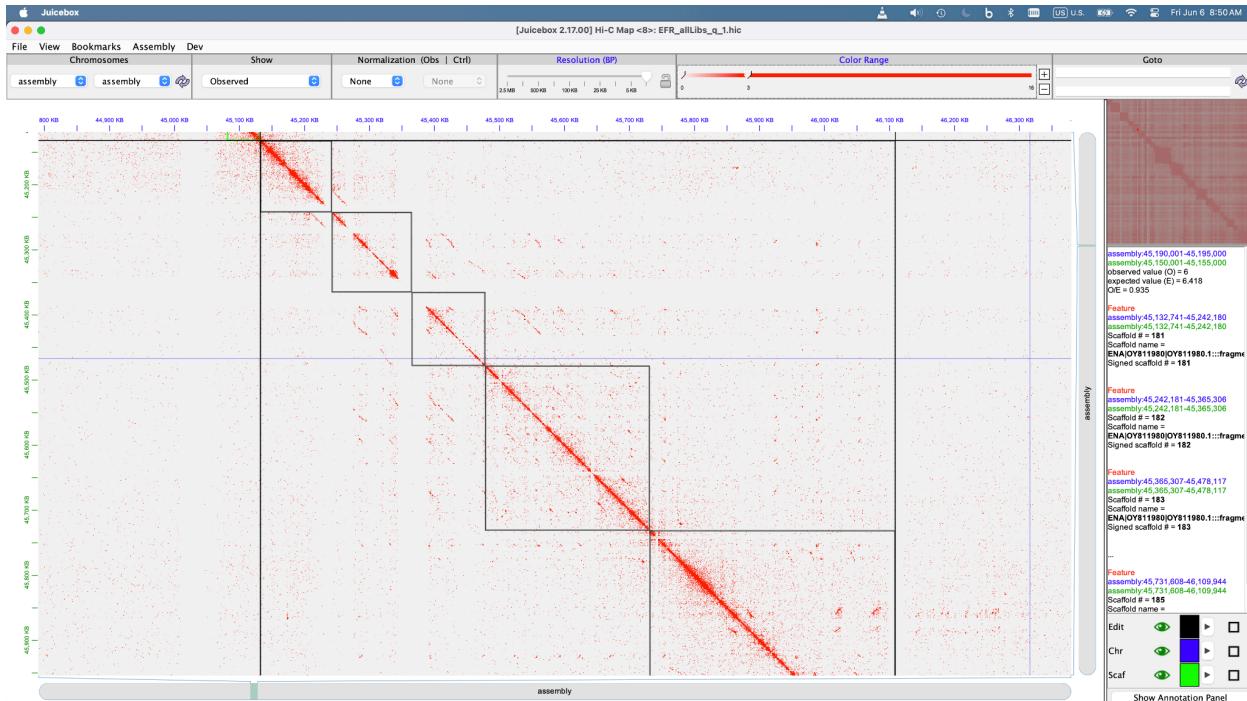
1. The contigs at the beginning and end of this highlighted section appear to be misplaced.



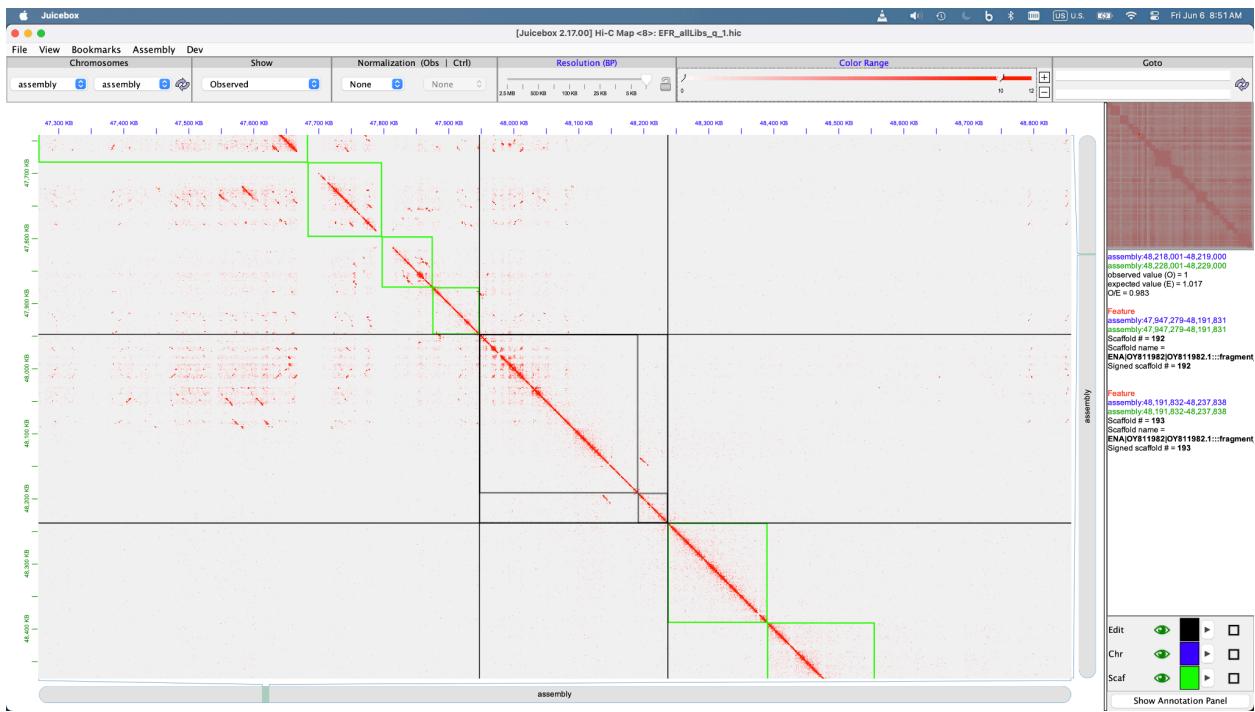
2. The contig in the middle of the highlighted section has little Hi-C support for this position. The 3rd contig has some overlaps with repetitive parts of the first contig. Potential misassembly?



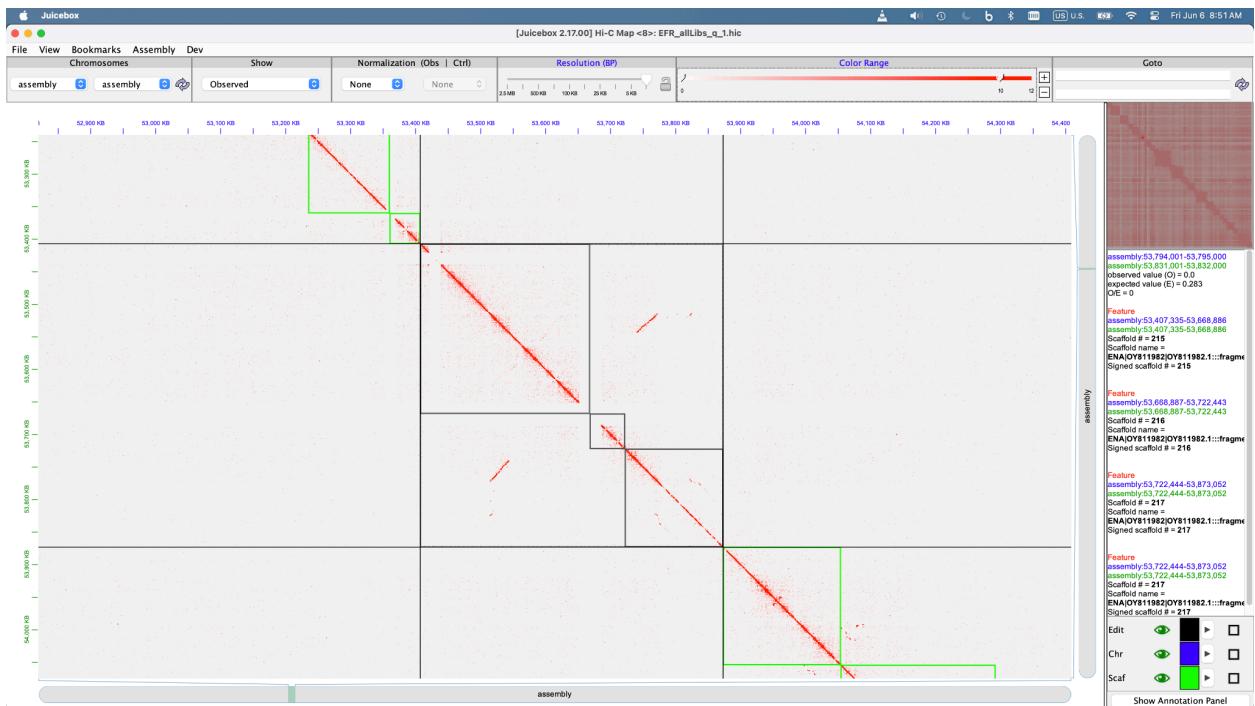
- The last large scaffold on this chromosome has little Hi-C support compared to the rest of the chromosome-scale scaffold, and it looks like the penultimate contig could be the telomeric scaffold. I didn't look at telomeric repeats though, so I could be mistaken.



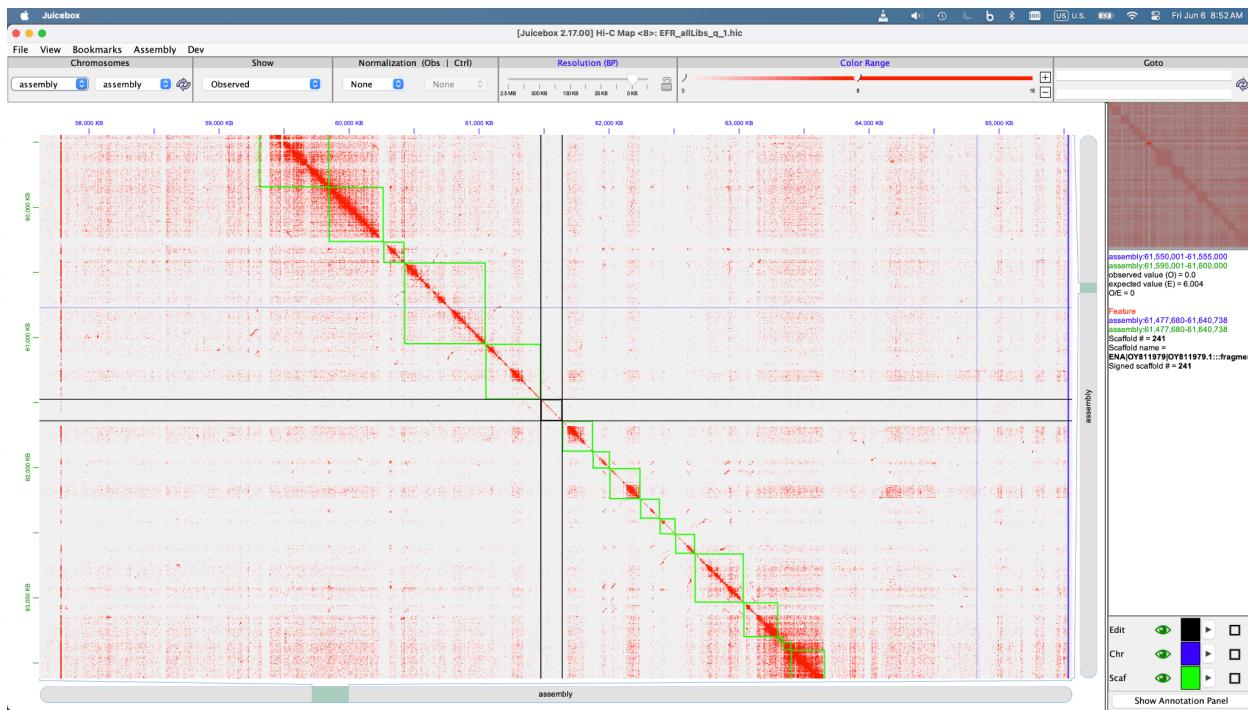
- There are a lot of repetitive sequences in these contigs. Are they all unique? Could the 2nd and 3rd be haplotigs of the 4th?  
The 2nd scaffold looks like a redundant “overlapping end” with the first contig.



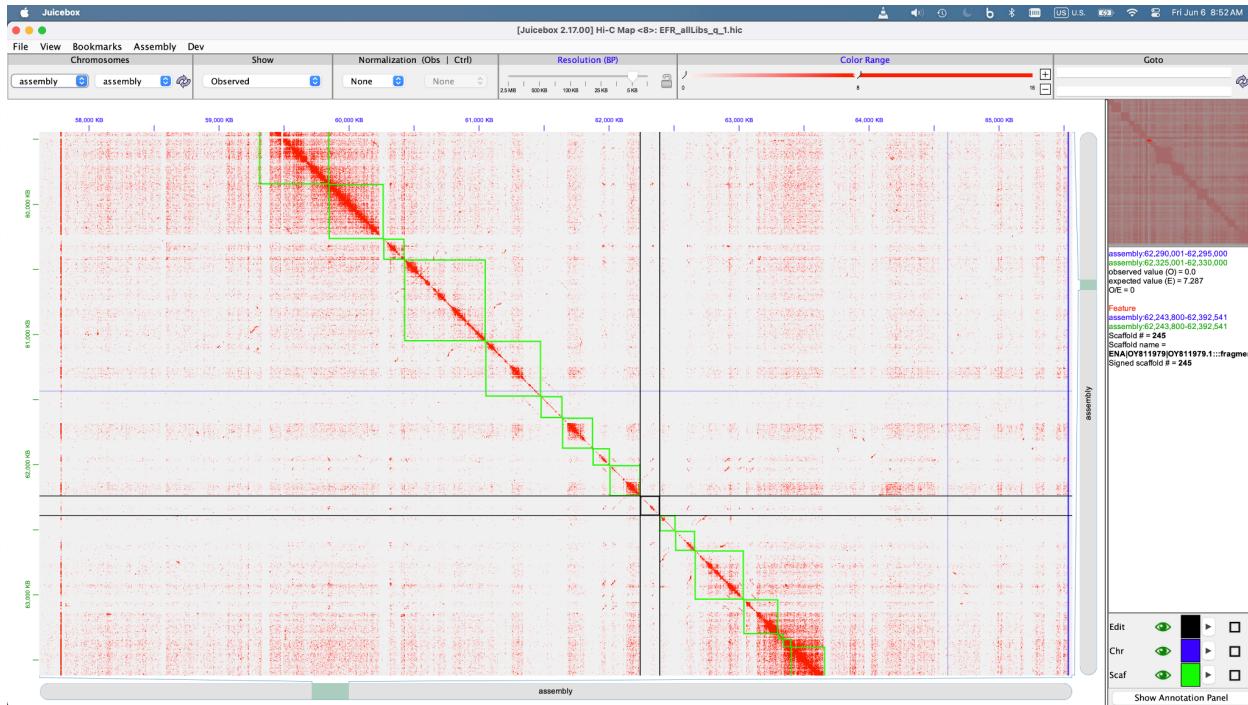
5. This looks like an assembly error, or could be a tandem duplication. The end of the first contig overlaps with the beginning of the second contig.



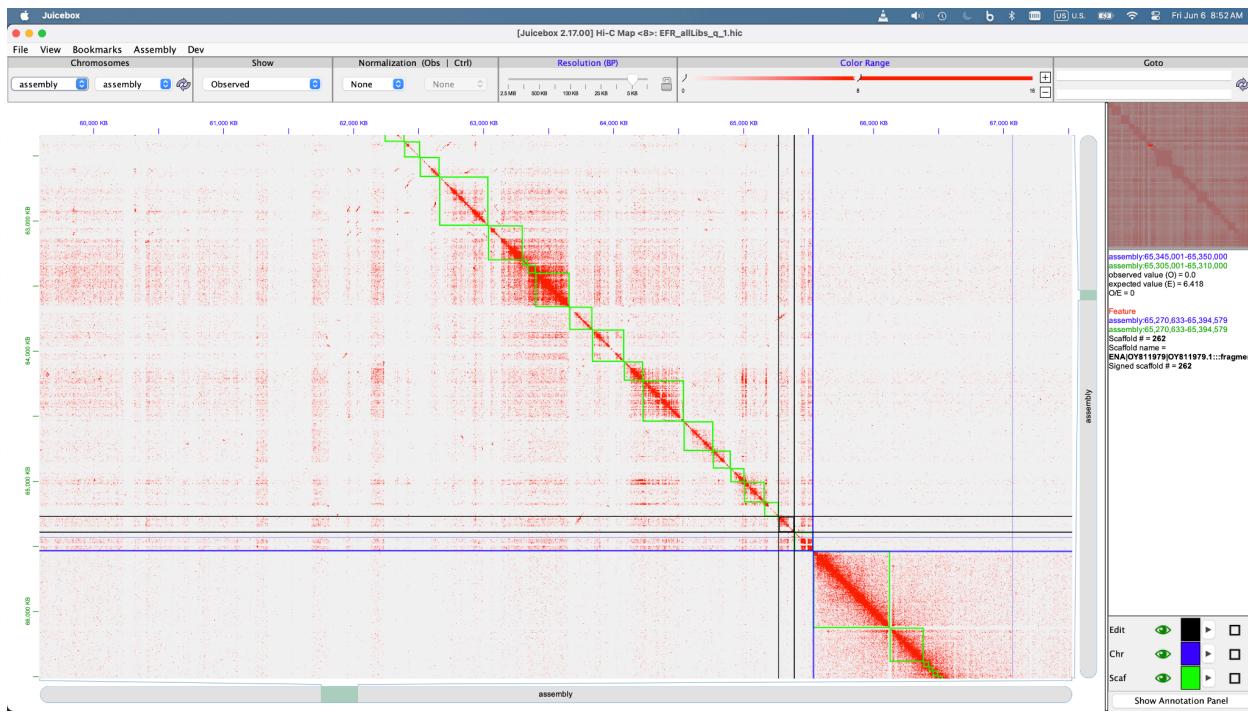
6. Overlapping parts of the 1st and 3rd contigs. Assembly error or duplication?



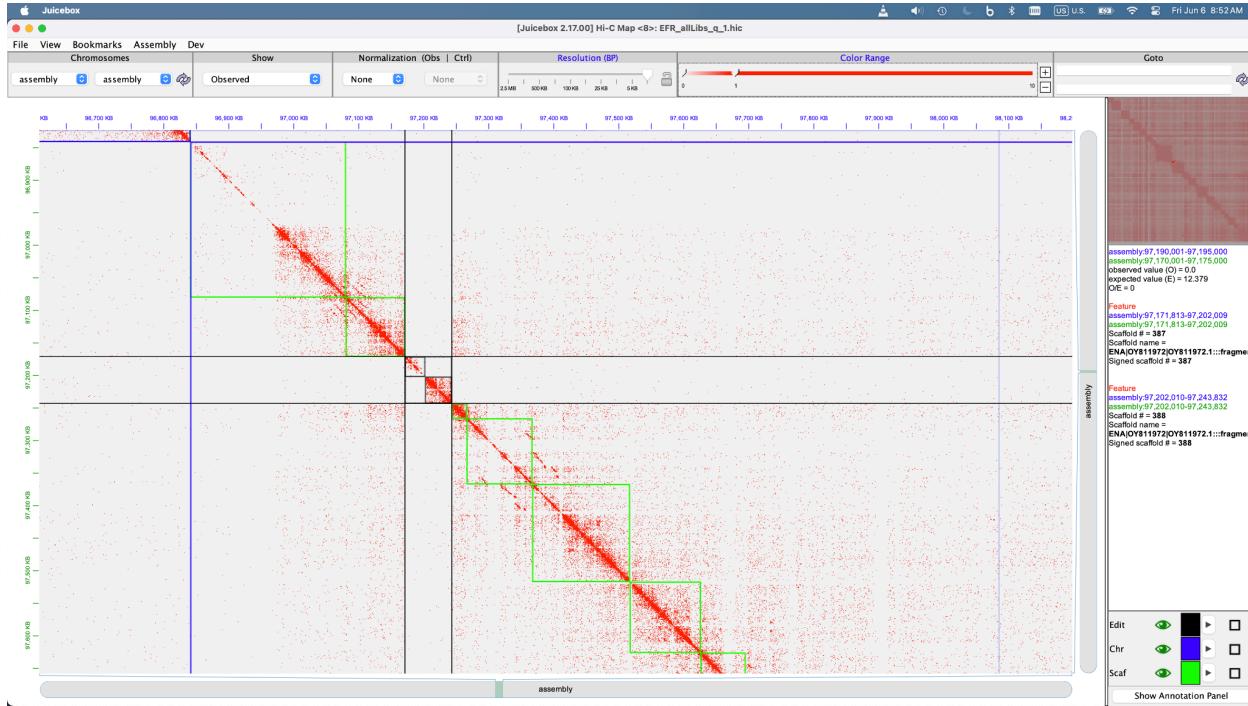
7. This looks like a redundant haplotig of the preceding contig.



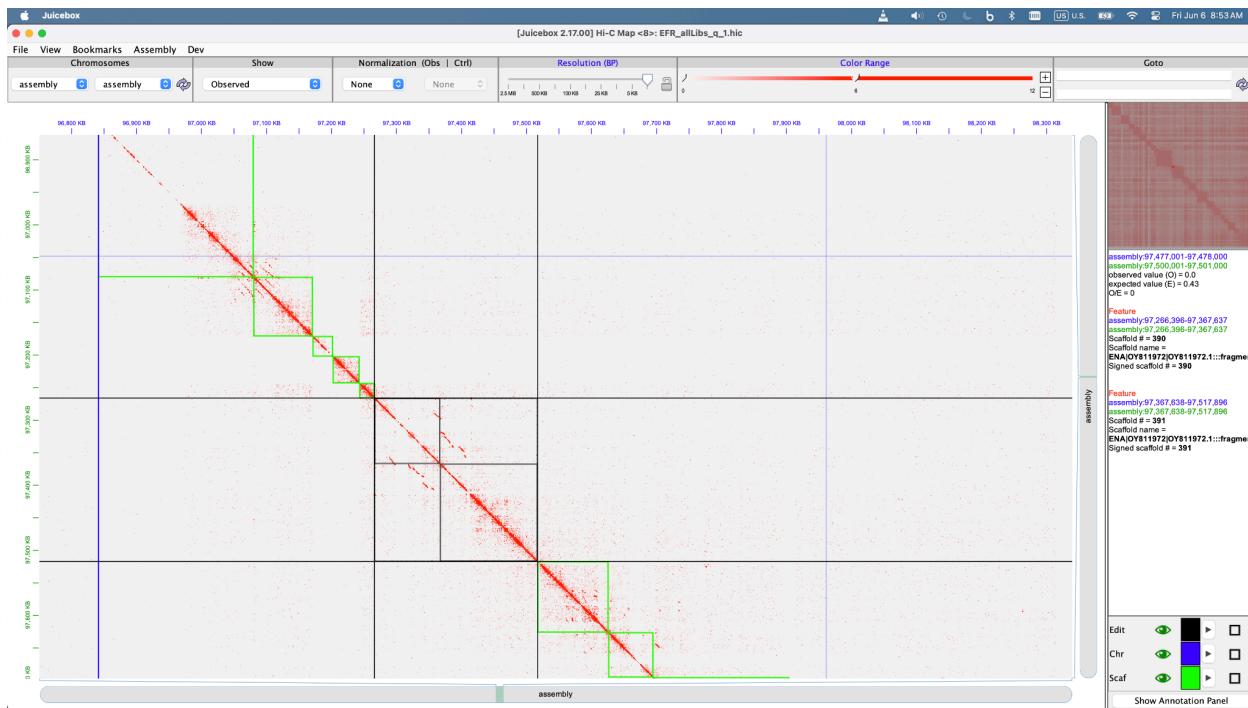
8. Looks like a redundant haplotig of the preceding contig.



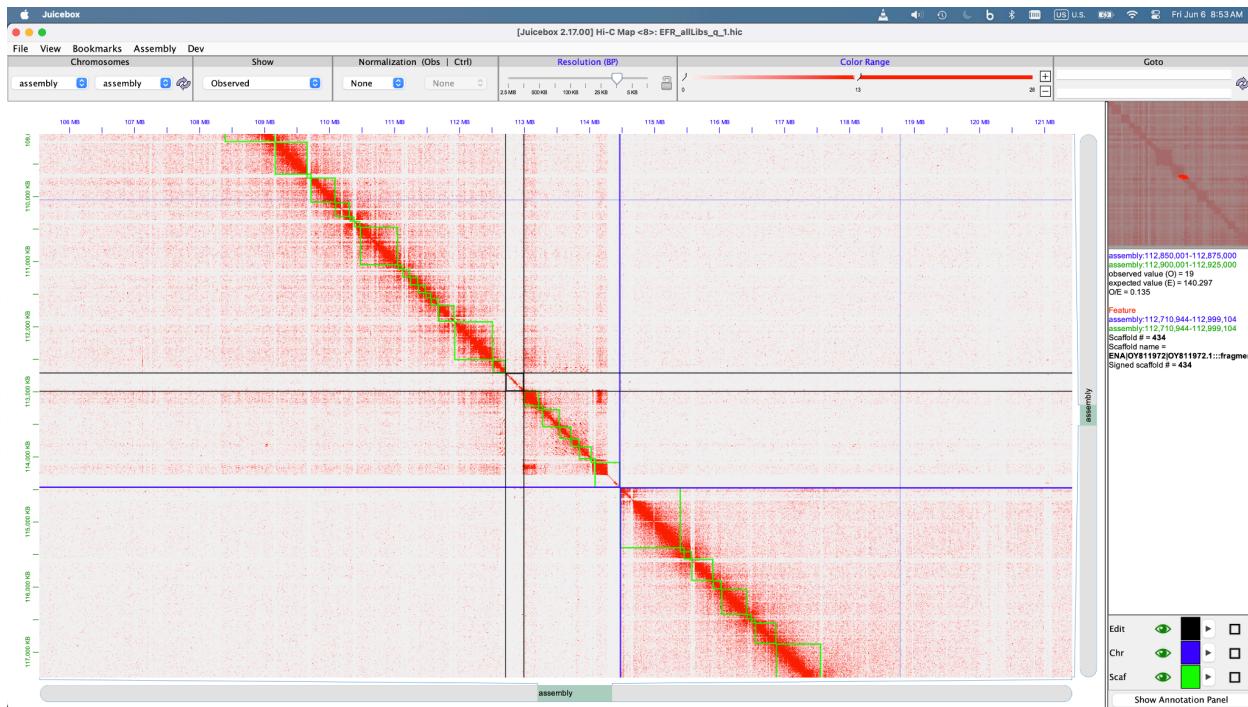
### 9. Overlap with another contig. Misplaced/ partial haplotig?



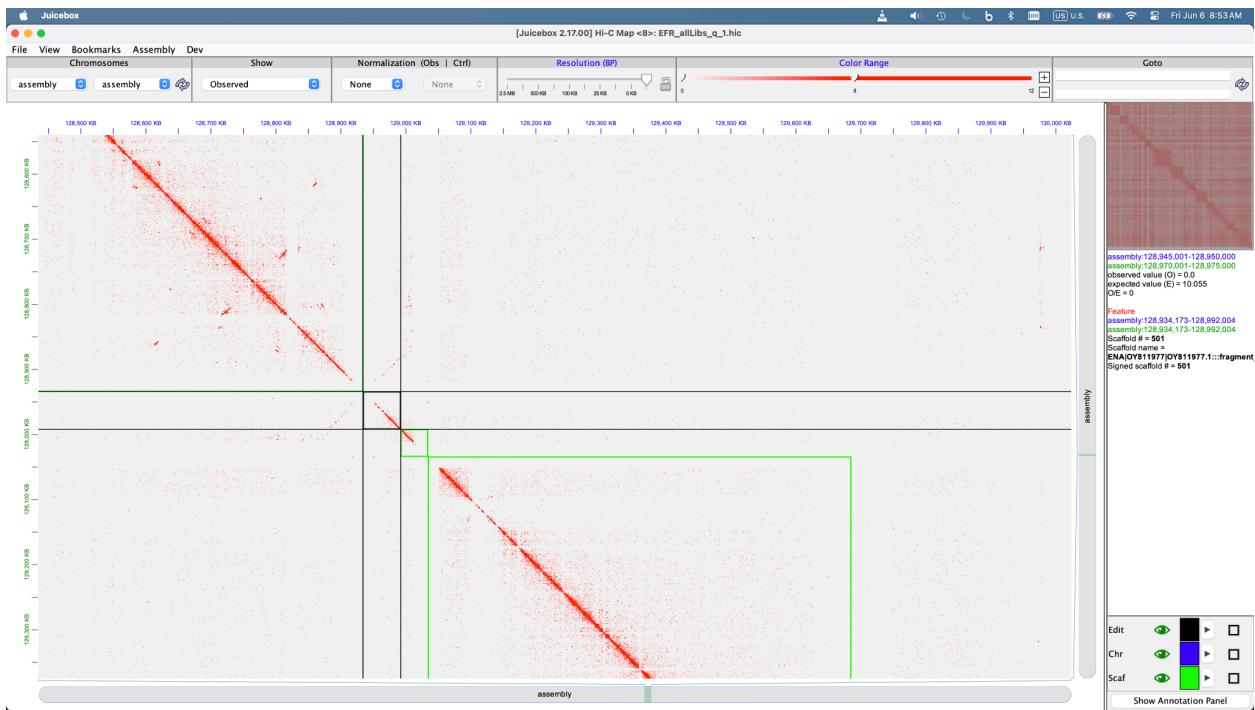
### 10. Little Hi-C evidence for these two contigs to be here. Could these by from a symbiont, or do they go elsewhere in the genome?



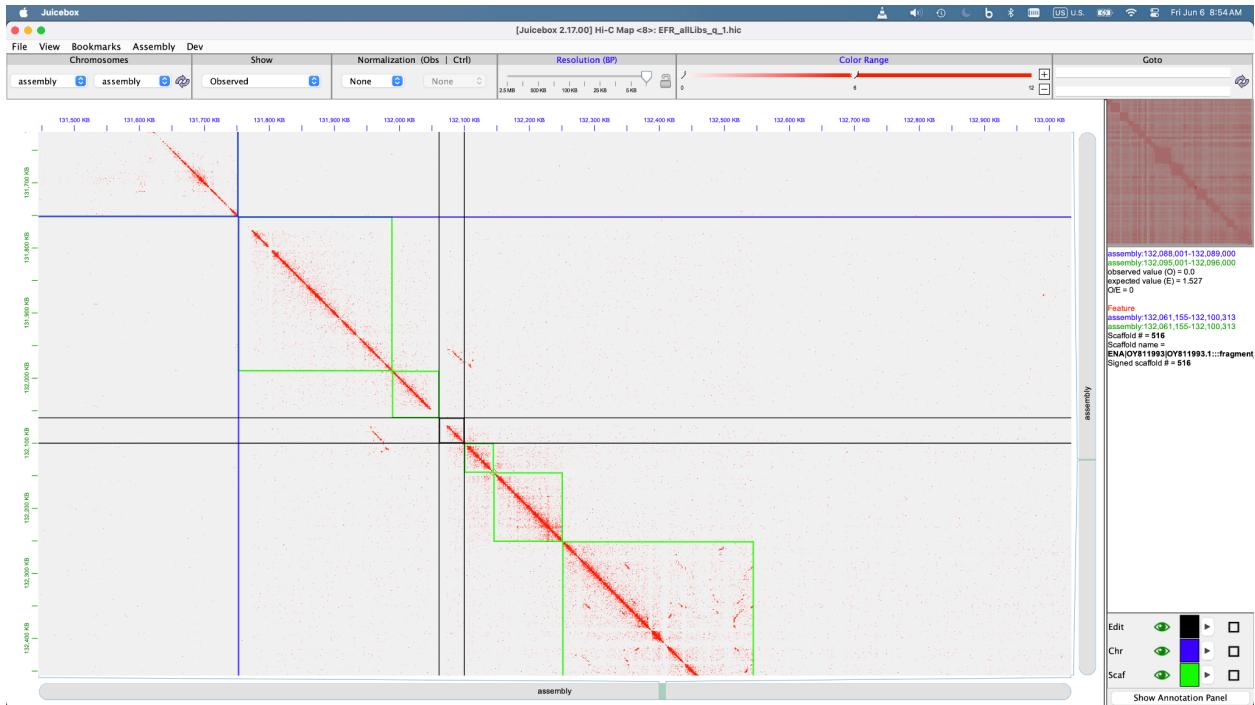
11. Looks like assembly errors or tandem duplications.



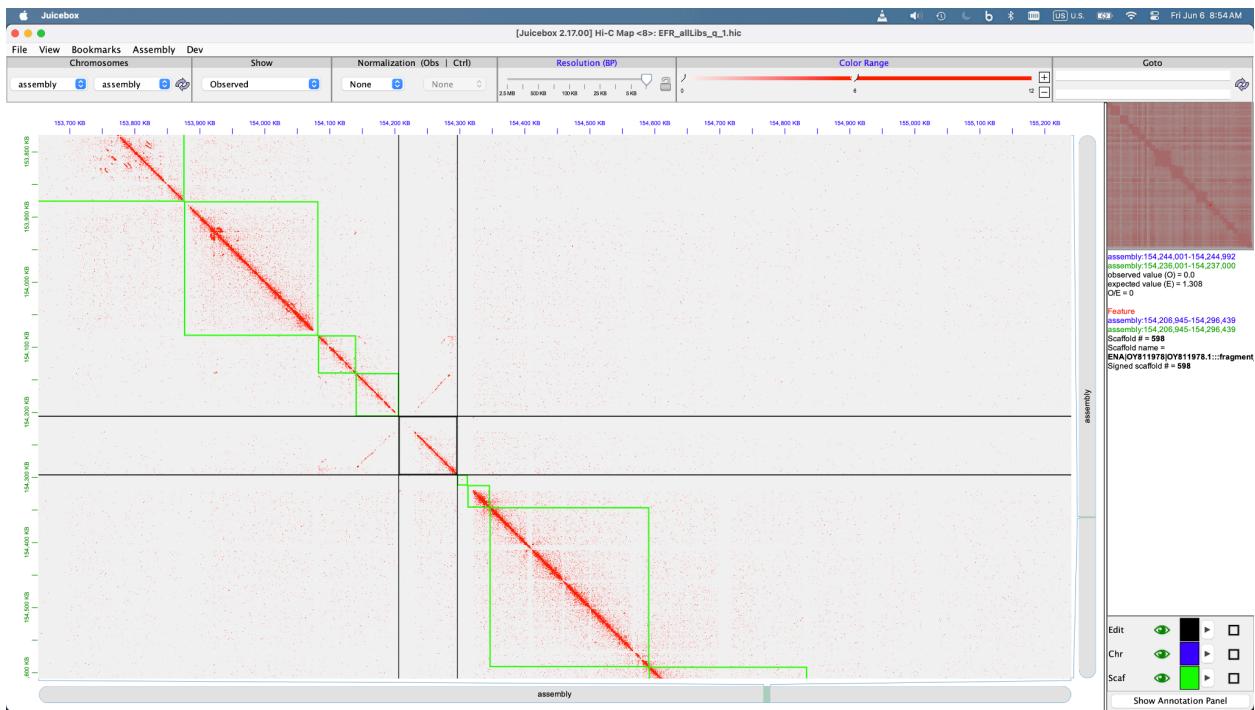
12. Little Hi-C evidence that this contig goes here - interrupts the signal between the flanking contigs.



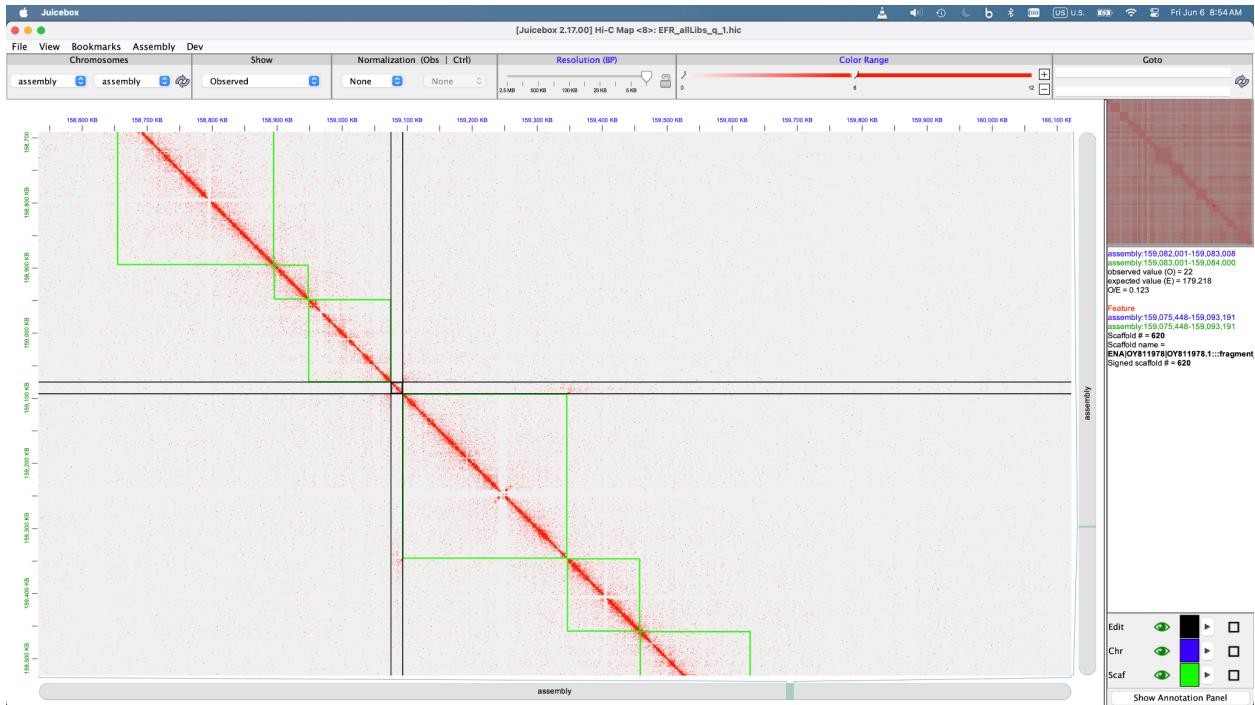
13. Looks like a redundant haplotig



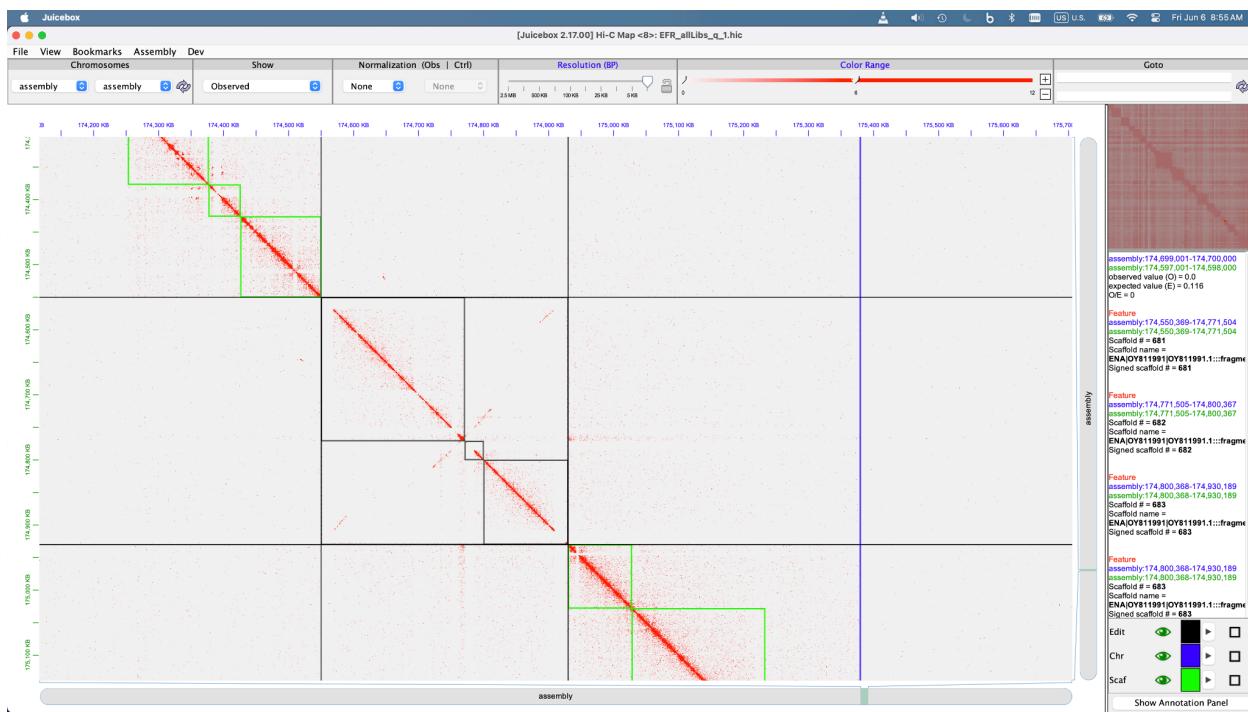
14. Looks like a redundant haplotig



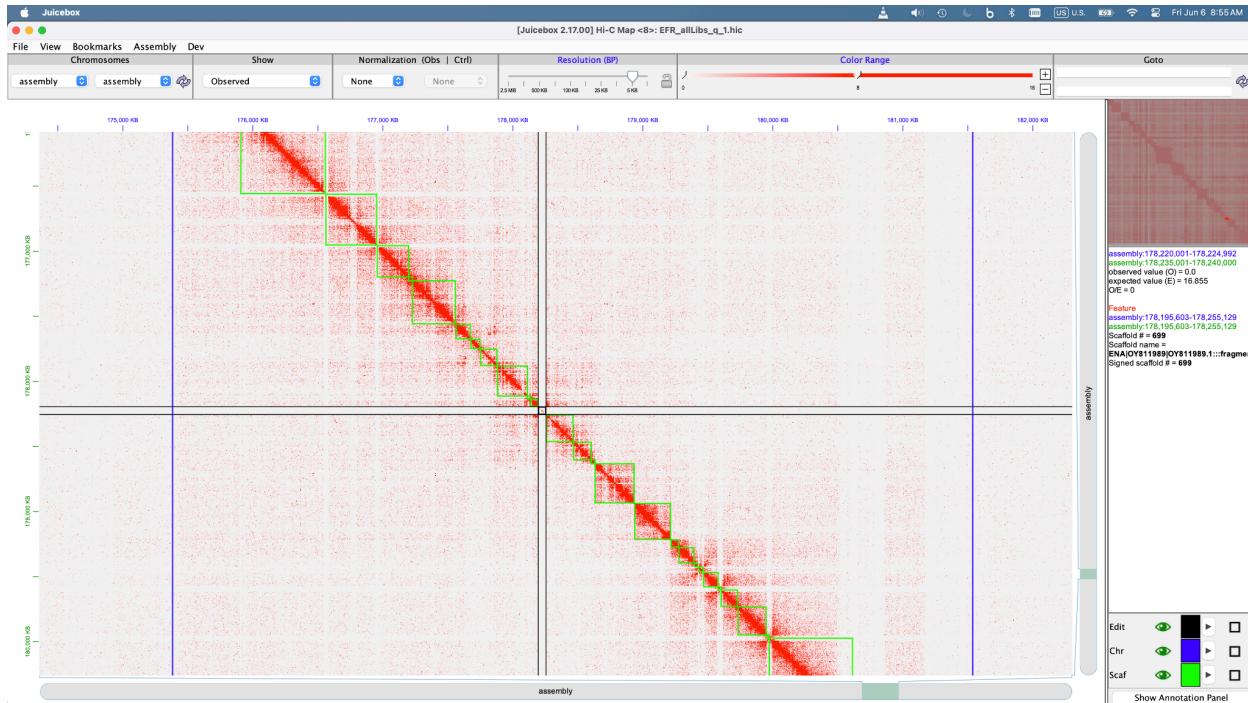
15. Looks like a redundant haplotig



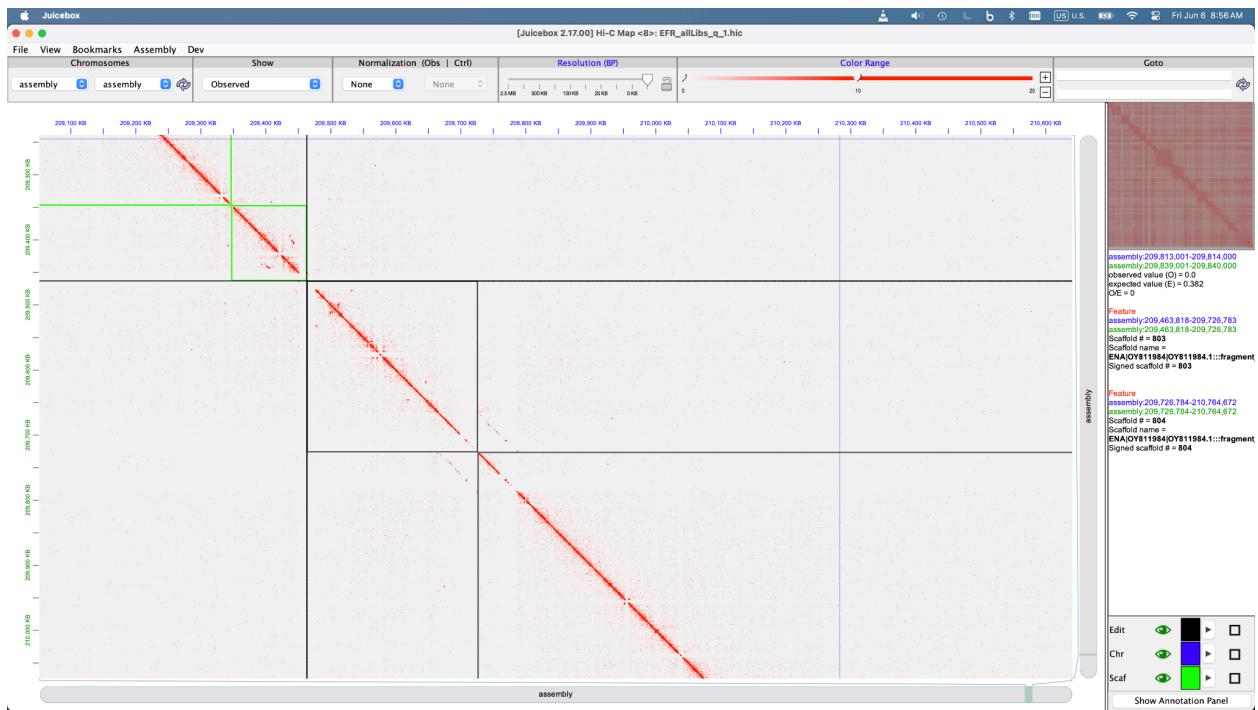
16. This contig should go after the next contig. Looks misplaced, and there is more Hi-C signal after the following contig.



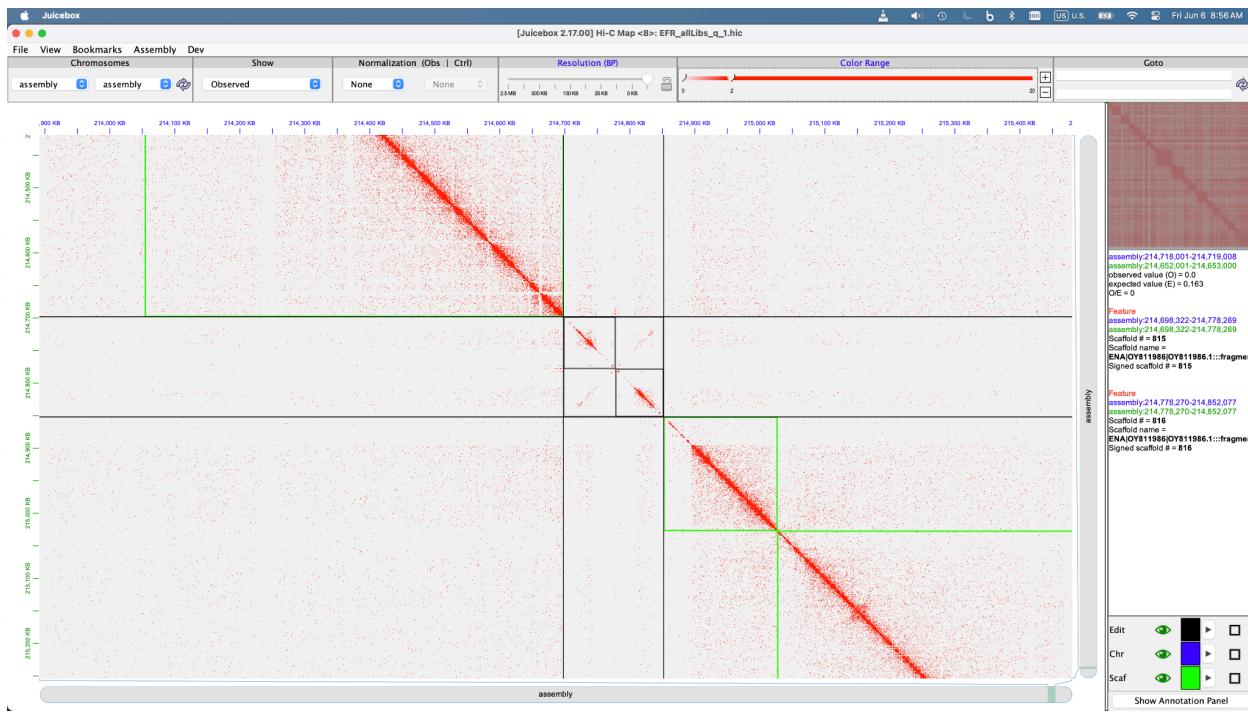
17. Looks like the middle contig could be a potential haplotig



18. Little Hi-C evidence for this contig to go here



19. Overlapping contig ends - looks like the beginning of the second contig should be trimmed.



20. These look like redundant sequences. I think one could be safely deleted. They are repeats, followed by nearly identical sequence.

## References

- Bredeson, Jessen V. n.d. "Artisanal - Assembly Review Tools, Including Sequence Analysis'N Annotation Liftover." Accessed February 26, 2023.  
<https://bitbucket.org/bredeson/artisanal/src/master/>.
- Dudchenko, Olga, Sanjit S. Batra, Arina D. Omer, Sarah K. Nyquist, Marie Hoeger, Neva C. Durand, Muhammad S. Shamim, et al. 2017. "De Novo Assembly of the *Aedes Aegypti* Genome Using Hi-C Yields Chromosome-Length Scaffolds." *Science* 356 (6333): 92–95.
- Dudchenko, Olga, Muhammad S. Shamim, Sanjit S. Batra, Neva C. Durand, Nathaniel T. Musial, Ragib Mostofa, Melanie Pham, et al. 2018. "The Juicebox Assembly Tools Module Facilitates de Novo Assembly of Mammalian Genomes with Chromosome-Length Scaffolds for under \$1000." *bioRxiv*. <https://doi.org/10.1101/254797>.
- McKenna, Victoria, John M. Archibald, Roxanne Beinart, Michael N. Dawson, Ute Hentschel, Patrick J. Keeling, Jose V. Lopez, et al. 2021. "The Aquatic Symbiosis Genomics Project: Probing the Evolution of Symbiosis across the Tree of Life." *Wellcome Open Research* 6 (October):254.
- Open2C, Nezar Abdennur, Geoffrey Fudenberg, Ilya M. Flyamer, Aleksandra A. Galitsyna, Anton Goloborodko, Maxim Imakaev, and Sergey V. Venev. 2023. "Pairtools: From Sequencing Data to Chromosome Contacts." *bioRxiv : The Preprint Server for Biology*, February. <https://doi.org/10.1101/2023.02.13.528389>.
- Zhang, Haowen, Li Song, Xiaotao Wang, Haoyu Cheng, Chenfei Wang, Clifford A. Meyer, Tao Liu, et al. 2021. "Fast Alignment and Preprocessing of Chromatin Profiles with Chromap." *Nature Communications* 12 (1): 1–6.