Manually editing scaffolds with Juicebox Assembly Tools (JBAT)

What is Juicebox Assembly Tools (JBAT)

- A method to
  - View Hi-C data again a genome assembly
  - Interactively modify scaffolds/contigs

- Start with Hi-C aligned against assembly
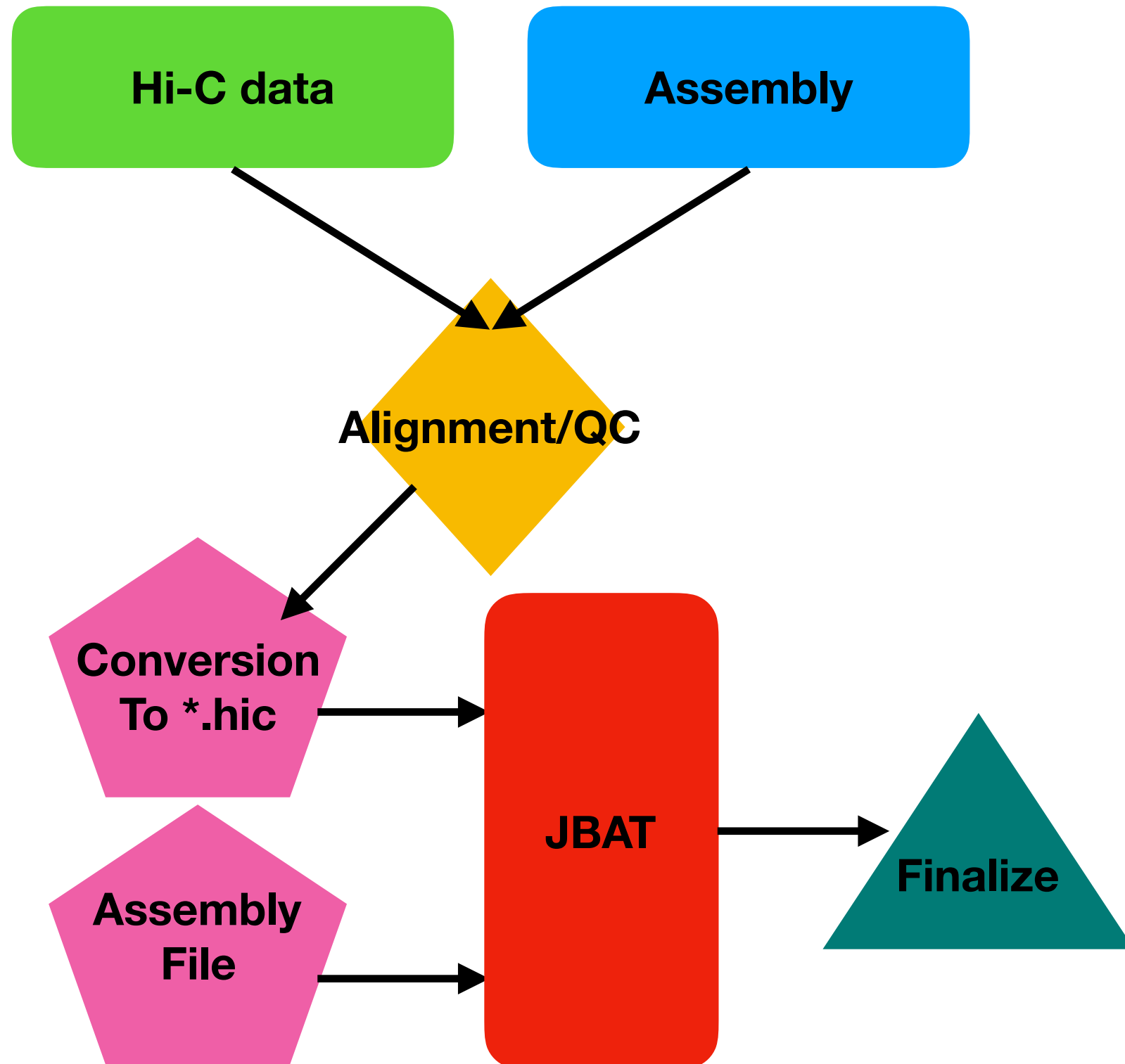 - Juicebox suite offers methods for file conversion
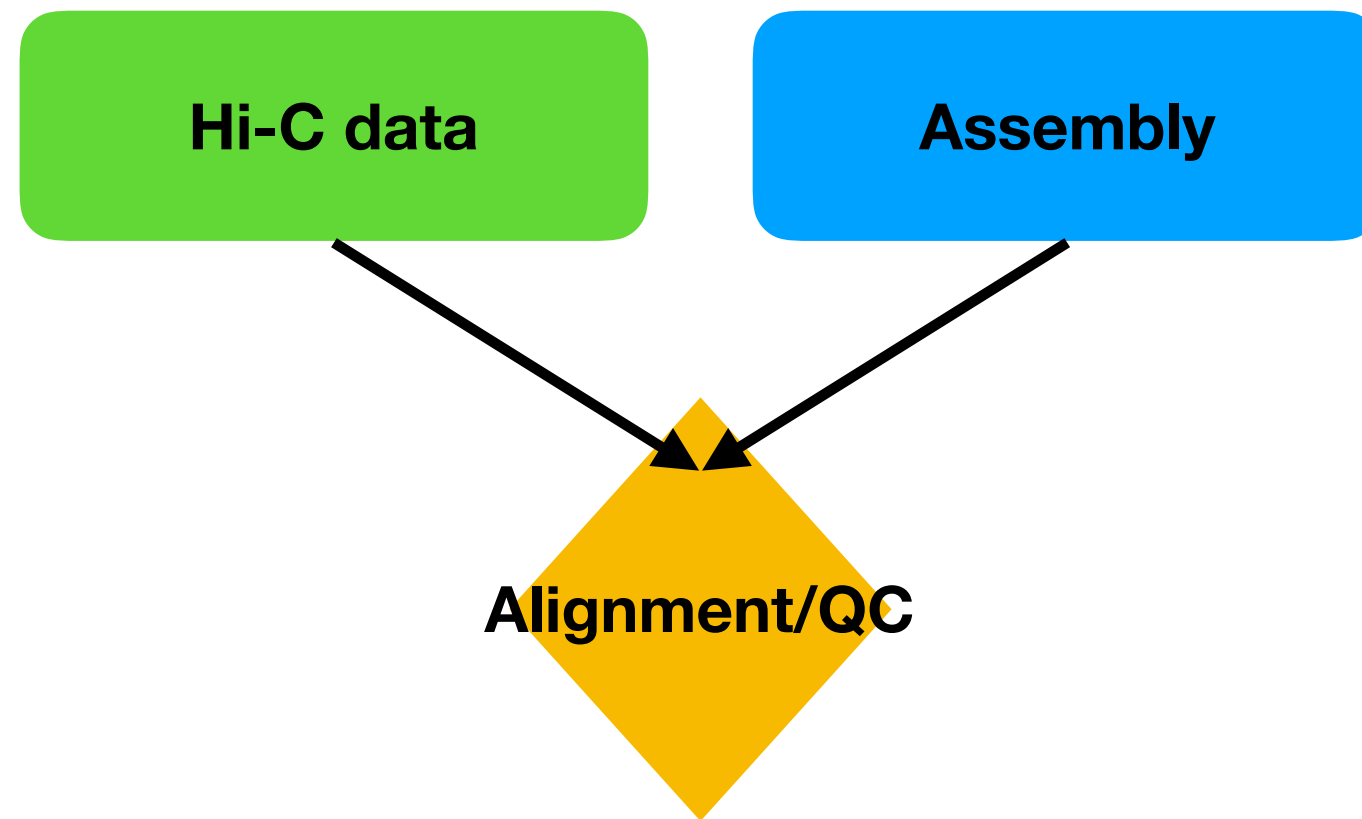
Olga Dudchenko
Postdoc in Aiden lab

DNA Zoo - Aiden lab

Juicebox Assembly Tools Video demo

JBAT Workflow

Hi-C alignment

BWA-mem alignment parameters

Hi-C QC

PHASE GENOMICS

Hi-C library QC report

Assembly statistics

Label	Assembly statistics
BAM file	berberis.bam
Assembly size	1,229,085,329
Contig (CTG) N50	745,698
CTGs	2,785
CTGs > 10KB	2,761

Library statistics

Label	Library statistics	Expected values
Total read pairs (RPs) analyzed	135,545,928	N/A
High quality (HQ)* RPs	35.21%	N/A
RPs >10KB apart	10.03%	1-15%
RPs >10KB apart (CTGs >10KB)	25.35%	1-15%
HQ RPs >10KB apart (CTGs >10KB)	22.87%	1-15%
Intercontig RPs	55.35%	10-60% (contigs) 1-20% (chromosomes)
Intercontig HQ RPs	39.31%	10-60% (contigs) 1-20% (chromosomes)
Same strand RPs	23.02%	2-50%
Same strand HQ RPs	20.47%	2-50%
Split reads	24.20%	1-10% (PG libraries) 30%+ (other libraries)
Zero-distance RPs	0.97%	0-20%
Zero map quality reads	20.85%	0-10%
Duplicate reads**	13.94%	0-10%
Duplicate reads (extrapolated)**	12.10%	0-50%
Unmapped reads	3.50%	0-10%
Subjective Hi-C library judgment	PASS	See Judgment

*High quality (HQ) read pairs have minimum mapping quality >= 20, maximum edit distance <= 5, and are not duplicates.
**If this quantity is zero, see duplicate read section below. If negative, there are too few reads sampled to estimate duplicates.
***Extrapolated to 100,000,000 RPs. If extrapolation fails, it will be -1%.

Description of QC pipeline

Matlock conversion

Conversion To *.hic
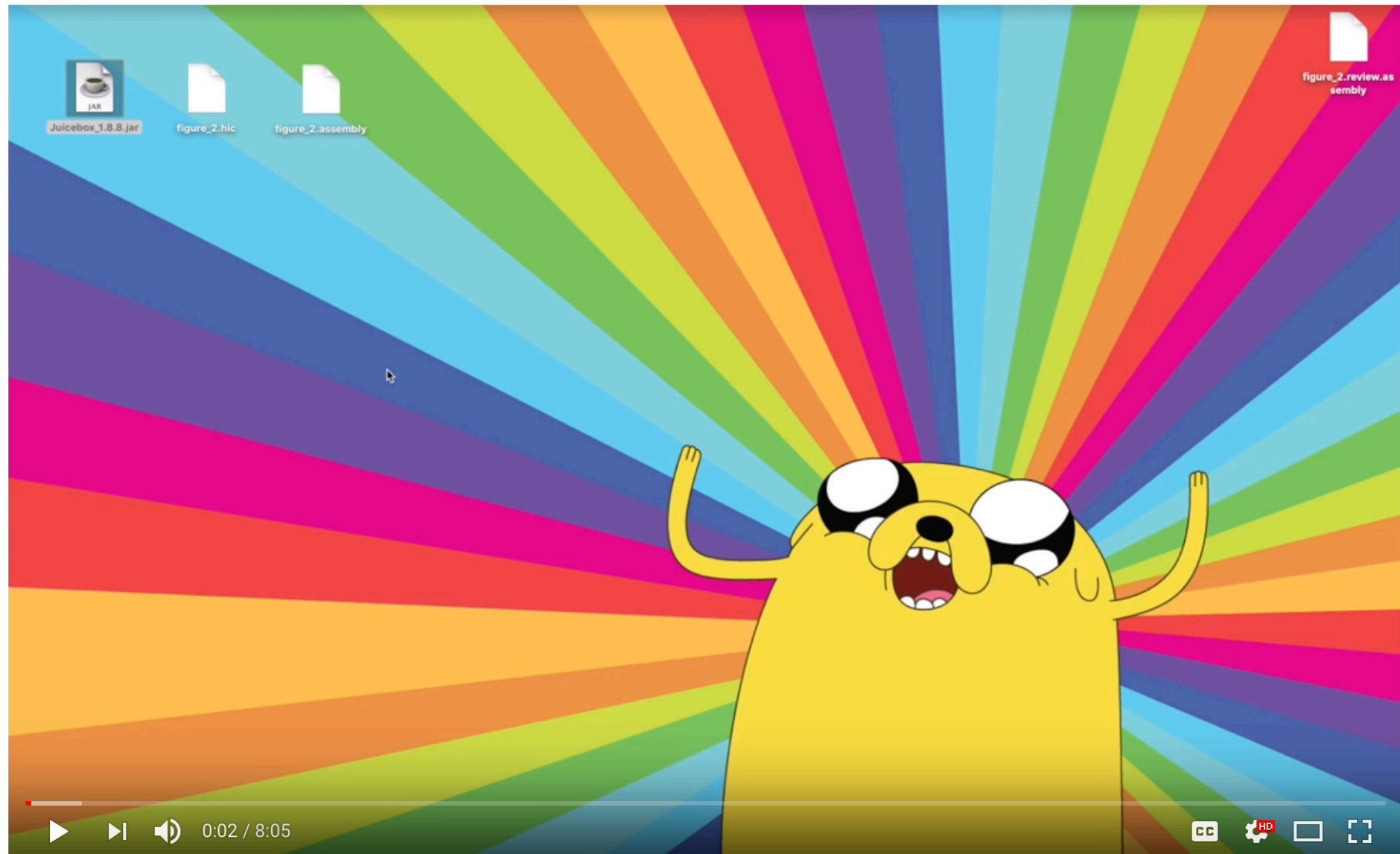
Assembly File

- **Matlock can**
  - **File conversion**
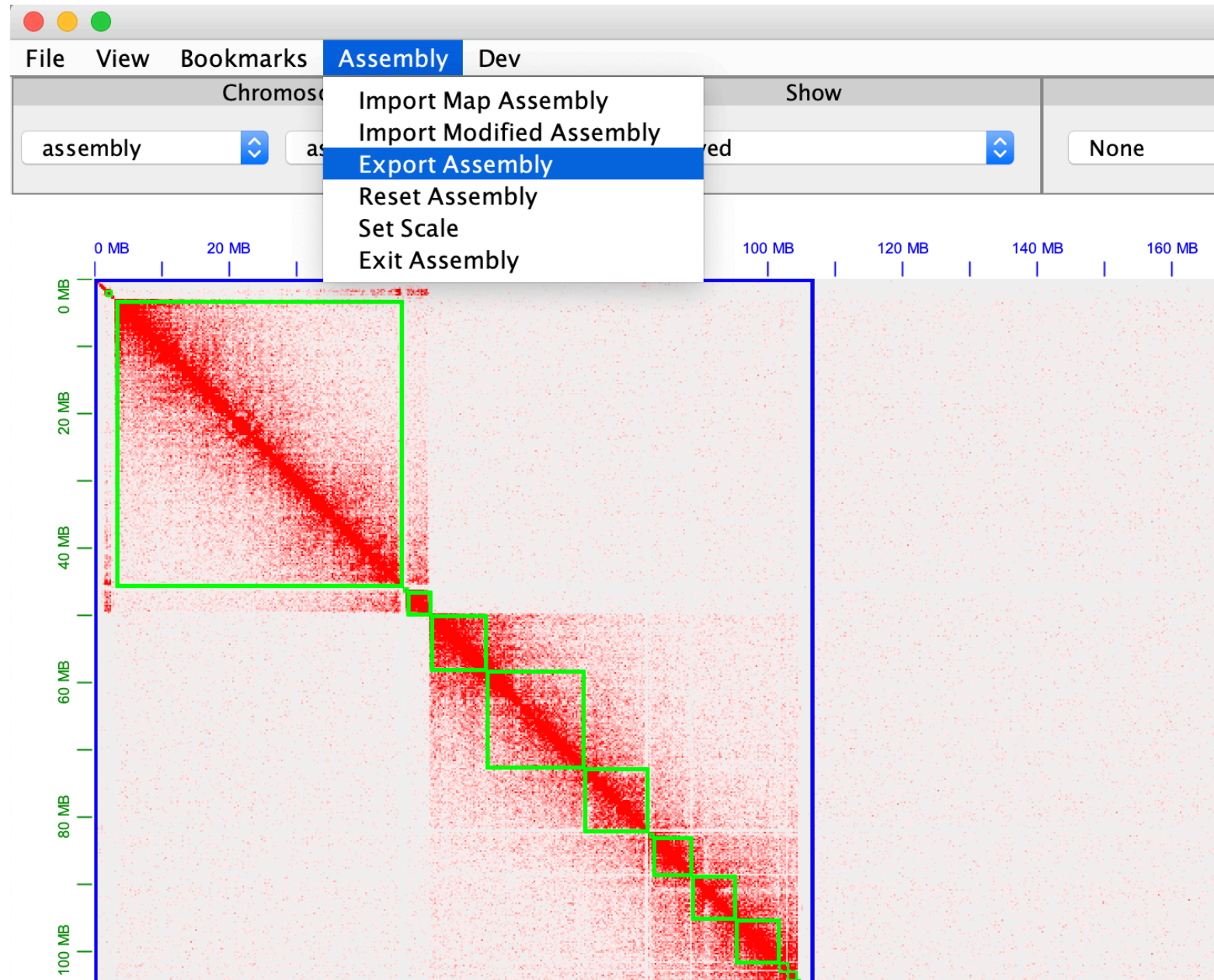 - **Filter Hi-C data**

Link to Matlock

Watch JBAT demo(s)

JBAT

Juicebox Assembly Tools Video demo

Finalizing scaffolds 1

Juicebox assembly scripts (Phase Genomics)

Finalizing scaffolds 2

```
python juicebox_assembly_converter.py -a
PGA_assembly.fasta.review.assembly -f
rmaidis.contig.fa
```

Finalize

Juicebox assembly scripts (Phase Genomics)

Output

```
-rw-r--r--  1 zev   staff    103K Dec 11 13:43 PGA_assembly.fasta.review.agp
-rw-r--r--  1 zev   staff     18K Dec 11 13:42 PGA_assembly.fasta.review.assembly
-rw-r--r--  1 zev   staff     87K Dec 11 13:43 PGA_assembly.fasta.review.bed
-rw-r--r--  1 zev   staff     85B Dec 11 13:43 PGA_assembly.fasta.review.break_report.txt
-rw-r--r--  1 zev   staff    315M Dec 11 13:43 PGA_assembly.fasta[ Preview ].fasta
```
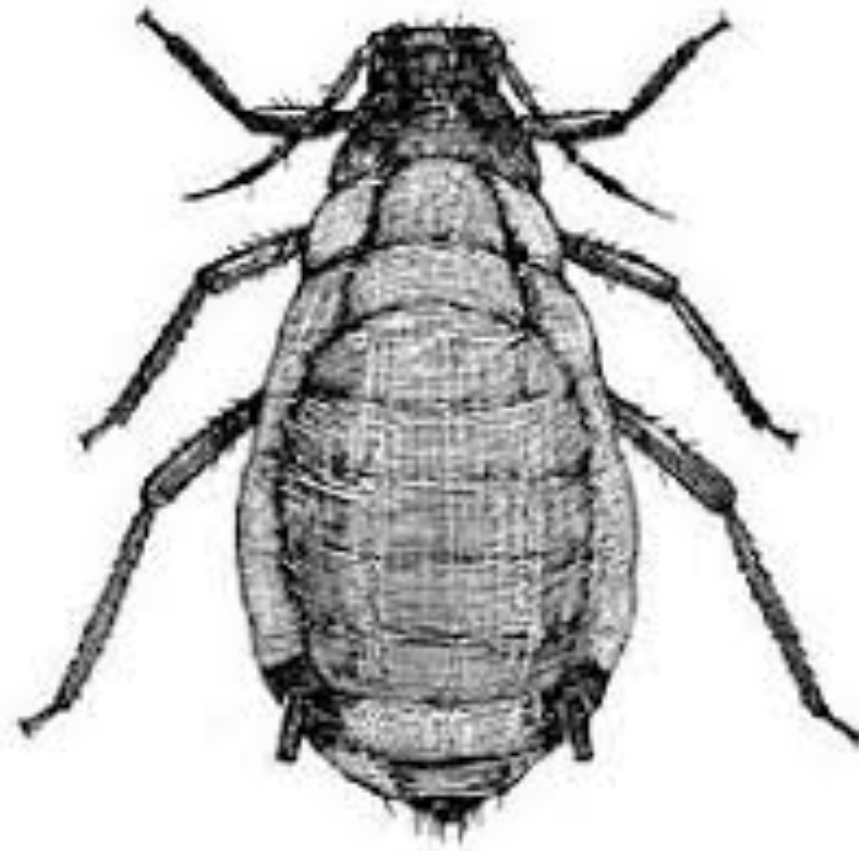
Finalize

Juicebox assembly scripts (Phase Genomics)

Your Turn!

Rhopalosiphum maidis (corn leaf aphid)

Canu assembly: 9Mb N50