



Session 3: Beginning manual curation

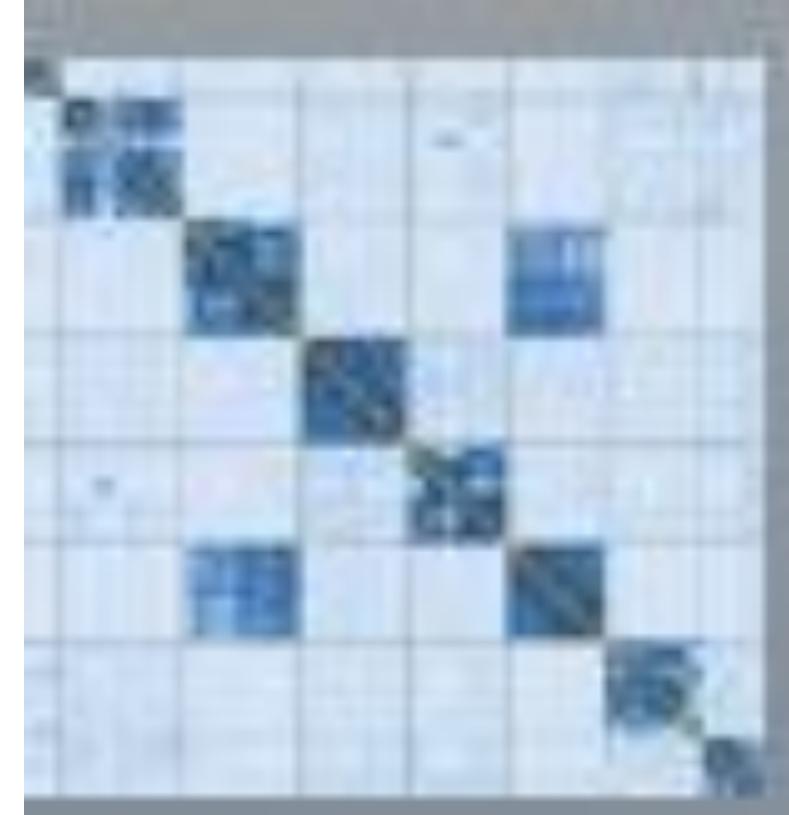
Day 2

Genome Reference Informatics Team (GRIT)
Wellcome Sanger Institute - Tree of Life



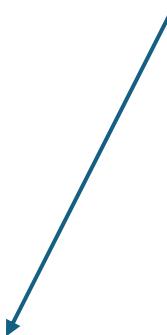
Overview

- **Analysis pipelines**
 - Treeval
-
- **How to use PretextView**





TreeVal as an aid for curation



Jbrowse

TreeVal Browser

- All projects coming into the curation team now go through the TreeVal Pipeline (github.com/sanger-tol/treeval)

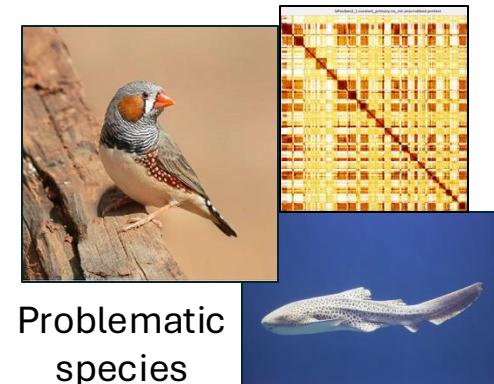
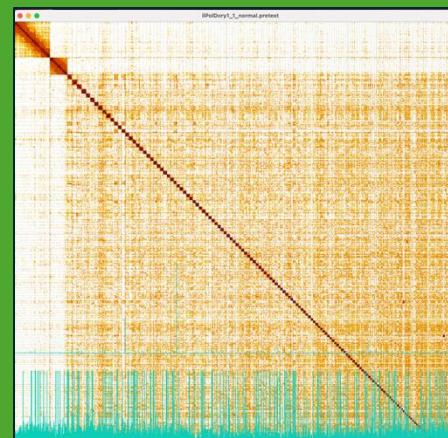
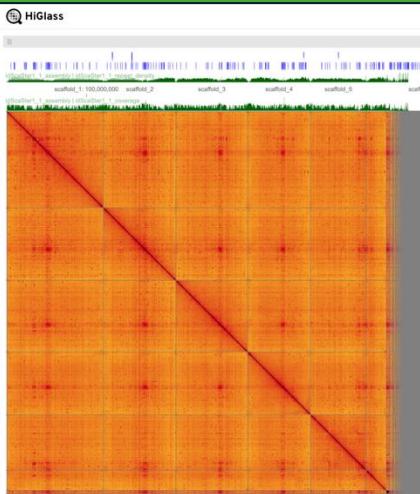
All draft assemblies



Might not be enough for curation

Evidence data for curation

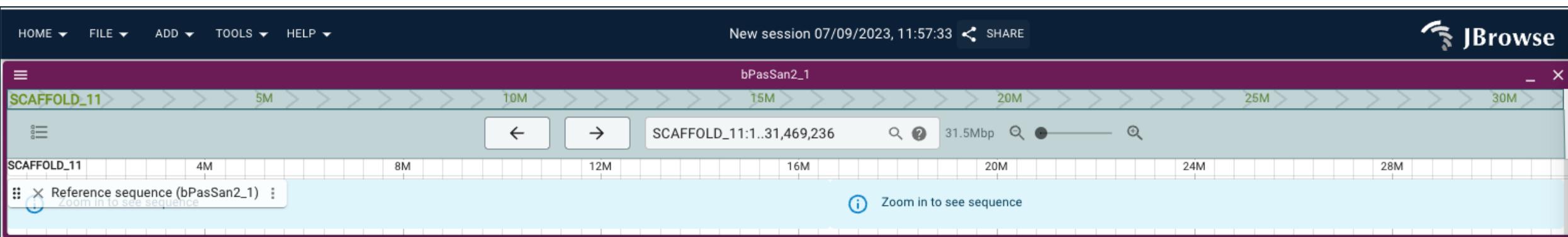
e.g HiC maps, Coverage, telomere, repeat tracks, gene alignments



**TreeVal
Browser**

In depth information is provided

TreeVal Browser



Why Jbrowse?

- Jbrowse provides the perfect platform for visualisation of assemblies:
- Simple to use, web-based, lightweight and interactive
- It's an extensible platform that utilizes multiple data types in a variety of genomic data formats such as FASTA, VCF, BAM, BED/BIGBED, HIC, CRAM, PAF and GFF3
- Highly customisable

How to start using TreeVal (Jbrowse)?

<https://jbrowse.org/jb2/>



JBrowse Genomes NEW! Docs Blog Download Plugins Features Gallery Demos Contact Looking for JBrowse 1? [↗](#) Search...

ⓘ New: JBrowse 2 office hours!
Starting Fall 2023, we are offering 1-on-1 appointments with members of our team via Google Calendar [SCHEDULE APPOINTMENT](#)

JBrowse

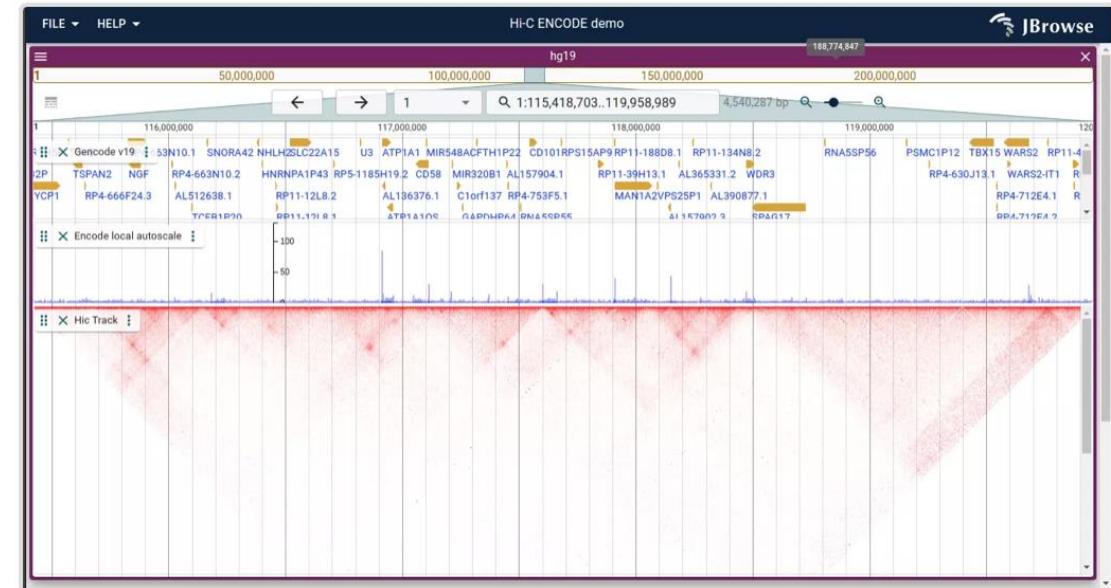
The next-generation genome browser

JBrowse is a genome browser that runs on the web, on your desktop, or embedded in your app.

[DOWNLOAD](#)

[BROWSE DEMO](#)

Check out our [latest release blogpost](#), our [embedded components](#), and our [command line tools](#).



How to start using TreeVal (Jbrowse)?

<https://jbrowse.org/jb2/>



JBrowse Genomes NEW! Docs Blog Download

New: JBrowse 2 office hours!
Starting Fall 2023, we are offering 1-on-1 appointments wit

JBrowse

The next-generation genome

JBrowse is a genome browser that runs on the web, on your

[DOWNLOAD](#)

Check out our [latest release blogpost](#)

Scaffold
Chromosome
level genomes

JBrowse 2 Genomes Genomes Demos About [Looking for the JBrowse 2 homepage?](#) 🔍 🌐 🌐

JBrowse 2 genome hubs

Main genome browsers

Here is a short list of human and mouse genomes, for more species [click here](#)

Name	Scientific name	Organism	Description	JBrowse	UCSC
hg19 (info)	Homo sapiens	Human	Feb. 2009 (GRCh37/hg19)	JBrowse	UCSC
hg38 (info)	Homo sapiens	Human	Dec. 2013 (GRCh38/hg38)	JBrowse	UCSC
hs1 (info)	Homo sapiens	Human	Jan. 2022 (T2T CHM13v2.0/hs1)	JBrowse	UCSC
mm10 (info)	Mus musculus	Mouse	Dec. 2011 (GRCm38/mm10)	JBrowse	UCSC
mm39 (info)	Mus musculus	Mouse	Jun. 2020 (GRCm39/mm39)	JBrowse	UCSC

• [Click here](#) for a more complete listing

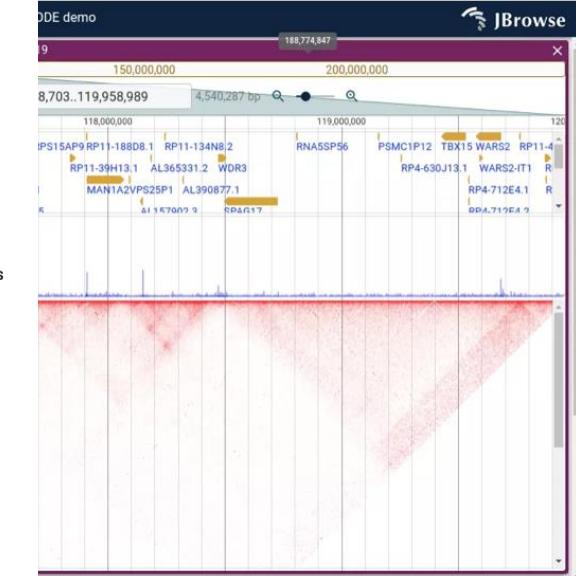
Note that some tracks that are available at UCSC are not yet available here, see [here](#) for more info.

GenArk browsers

We also created JBrowse 2 instances based on the UCSC GenArk browser resources, which is an effort to bulk create genome browsers for numerous plant, animal, and microbial species. See <https://hgdownload.soe.ucsc.edu/hubs/> for more info.

Species categories:

hub gateway	description
primates	NCBI primate genomes
mammals	NCBI mammal genomes
birds	NCBI bird genomes
fish	NCBI fish genomes
vertebrate	NCBI vertebrate genomes
invertebrate	NCBI invertebrate genomes
fungi	NCBI fungi genomes
plants	NCBI plant genomes
viral	NCBI viral genomes
bacteria	NCBI bacteria genomes
archaea	NCBI archaea genomes



How to start using TreeVal?



Click to see available tracks

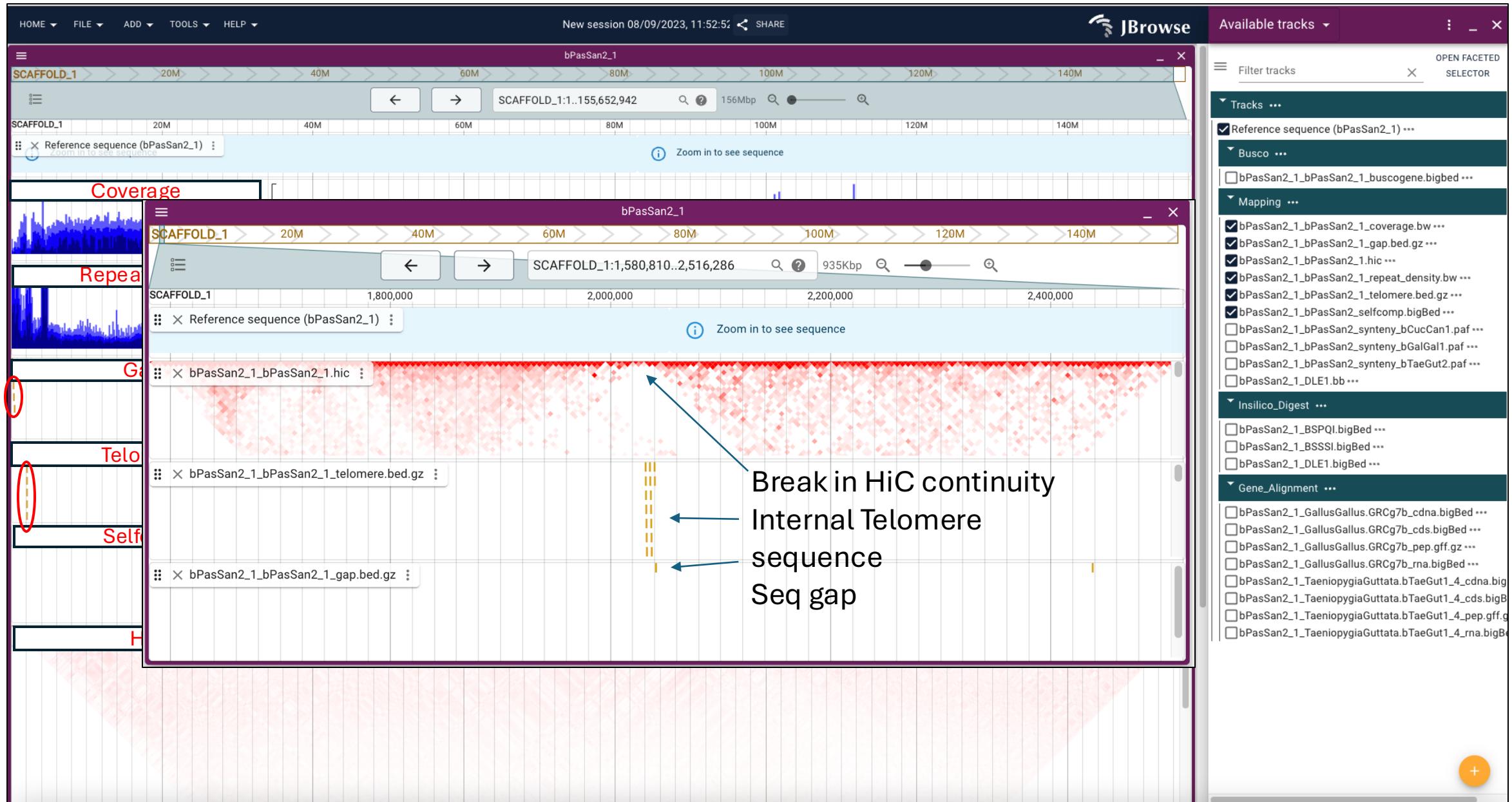
Click or type to select scaffold

Click to zoom in/out

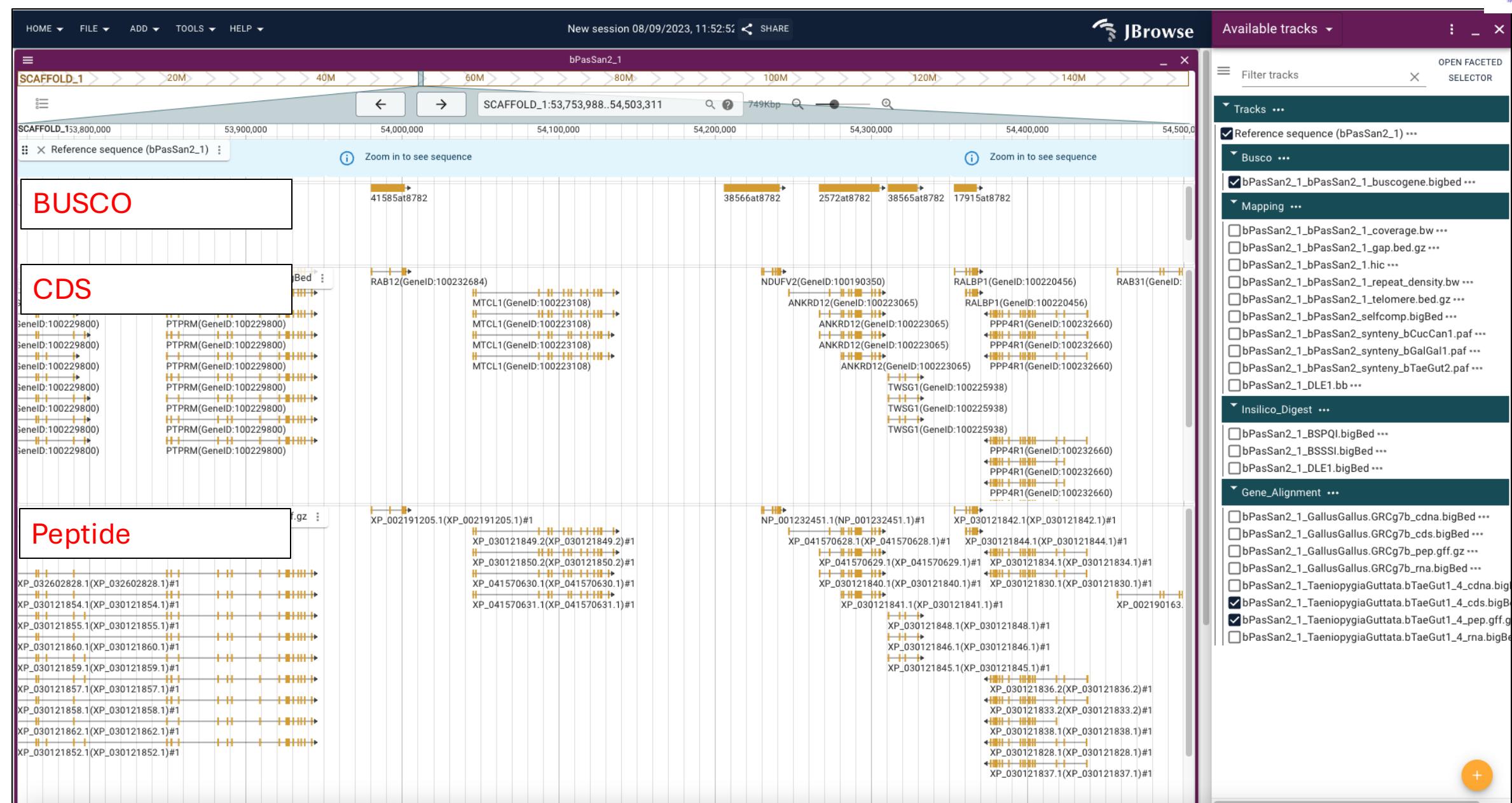
The screenshot shows the JBrowse interface with the following elements highlighted:

- Scaffold Selection:** The scaffold name "HAP1_SCAFFOLD_50M" is highlighted with a red box in the top center.
- Available Tracks Sidebar:** The "Available tracks" sidebar on the right is highlighted with a red box.
- Protein Alignment Track:** The main viewer area shows a track for "bMerAlb1_1_pep_AquilaChysaetos.bAQUChr1_4_pep.gff.gz", which is highlighted with a large red box.
- Bottom Left Annotation:** An annotation "Proteins in the alignment" with a red arrow pointing to the bottom left of the viewer area.

Which data types do we align to our assemblies?



Gene alignment data



How to use PretextView



Download PretextView from: <https://github.com/sanger-tol/PretextView/releases/tag/v1.0.5>

PretextView Public

master 4 Branches 26 Tags Go to file Add file Code

yumisims Merge pull request #53 from sanger-tol/add_box f49d1e3 · last week 388 Commits

.github/workflows v1.0.4 install nsis on Windows 2 months ago

ai_model change the name of the saved ai model, as ':' can not be i... 10 months ago

doc update readme 7 months ago

ico_design icon for mac 9 months ago

include frag cluster, cut with considering gap extesions 6 months ago

python add search boxes last week

src change loading menu key to esc and add a search box un... 3 weeks ago

subprojects frag cluster, cut with considering gap extesions 6 months ago

.gitignore solve the glfw delete segmental fault 6 months ago

.gitmodules frag cluster, cut with considering gap extesions 6 months ago

AGPCorrect 1.0.4 release 2 months ago

CMakeLists.txt fix pixel sort 3 weeks ago

CONTRIBUTORS.md v1.0.0 test version 11 months ago

Info.plist fix pixel sort 3 weeks ago

LICENCE.txt update licence 10 months ago

About OpenGL Powered Pretext Contact Map Viewer

Readme View license Code of conduct Activity Custom properties 50 stars 5 watching 9 forks Report repository

Releases 26

PretextView v1.0.5 Latest last week + 25 releases

Packages No packages published

Contributors 5

- PretextViewAI-1.0.5-Darwin-arm64.dmg
- PretextViewAI-1.0.5-Darwin-x86_64.dmg
- PretextViewAI-1.0.5-Linux-x86_64.deb
- PretextViewAI-1.0.5-Windows-x86_64.exe
- Source code (zip)
- Source code (tar.gz)

How to use PretextView

Press 'ESC' for the Menu



How to use PretextView

- Click ‘Load Map’ and open a map
- Keyboard Shortcuts
- **Press ‘ESC’ to open menu**
- **Press ‘E’ to enter the Edit mode and make edits**
 - Press ‘Q’ to undo an edit
 - Press ‘W’ to redo an edit
 - Click middle button to flip (invert) a scaffold
- **Press ‘S’ to enter the Scaffold mode: Paint scaffolds blocks as chroms**
 - Click middle button to delete painting
 - Press ‘A’ to paint a block of scaffolds at once (flood fill)
 - Press Shift + ‘D’ to delete all the painting in the map



How to use PretextView

- **Press ‘M’ to enter Metadata Tag mode:** add tags such as haplotigs, unloc, HAP1, HAP2, X, Z, Contaminant...
 - Use left/right arrow keys in your keyboard to select the active tag
 - Press Shift + ‘D’ to delete all tags in your map
 - Press ‘I’ to show ID bars: delimit and show orientation of the scaffolds
 - Press ‘L’ to switch the grids on and off
 - Press ‘B’ to show scale bars at the top
 - Press ‘T’ to switch the tool tip on and off: show coordinates on the map and info about active tracks
 - Arrow keys left/right: adjust contrast
 - Arrow keys up/down: change color scheme to the map
- **Press ‘W’ to enter Waypoint edit mode: mark specific regions**



How to use PretextView

- **Press ‘W’ to enter Waypoint edit mode: mark specific regions**
 - Middle click: delete
 - ‘L’: make the line horizontal/vertical

Press ‘X’ for extensions (tracks) and then press:

‘C’: coverage

‘G’: gap

‘R’: repeat content

‘T’: telomere

‘5’: 5p telomere

‘3’: 3p telomere



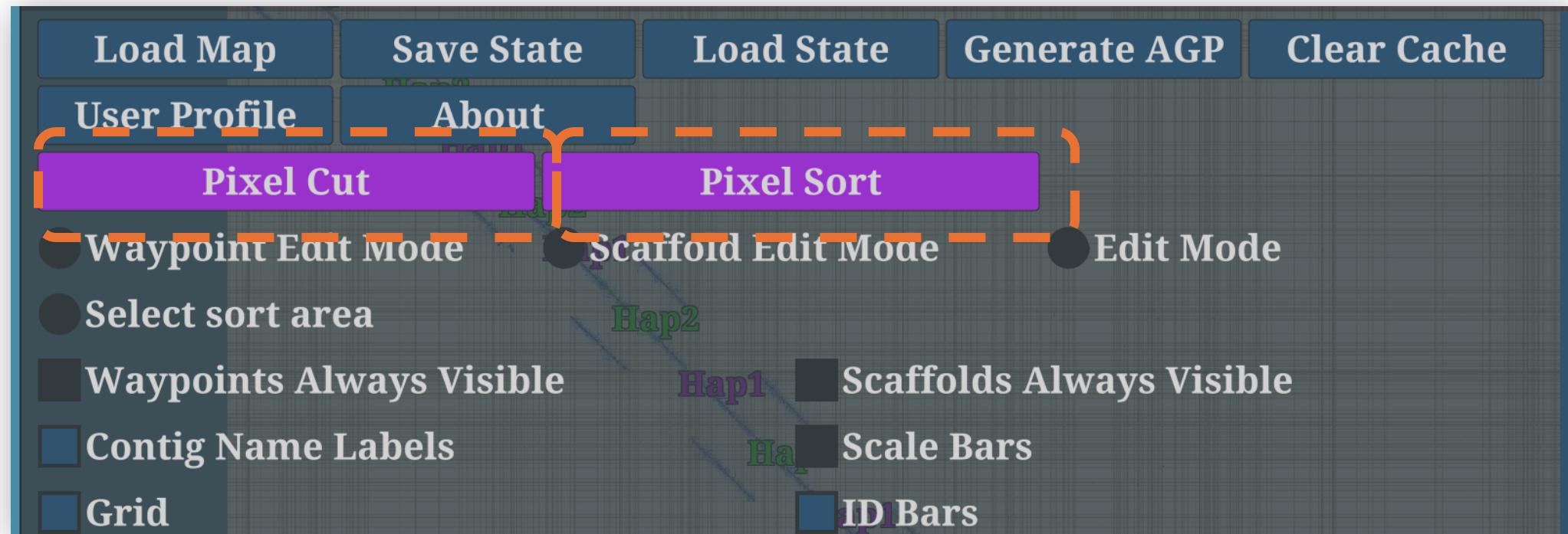
New features

‘C’ and ‘G’ shortcuts work at any time, no need extensions (‘X’) activated

How to use PretextView



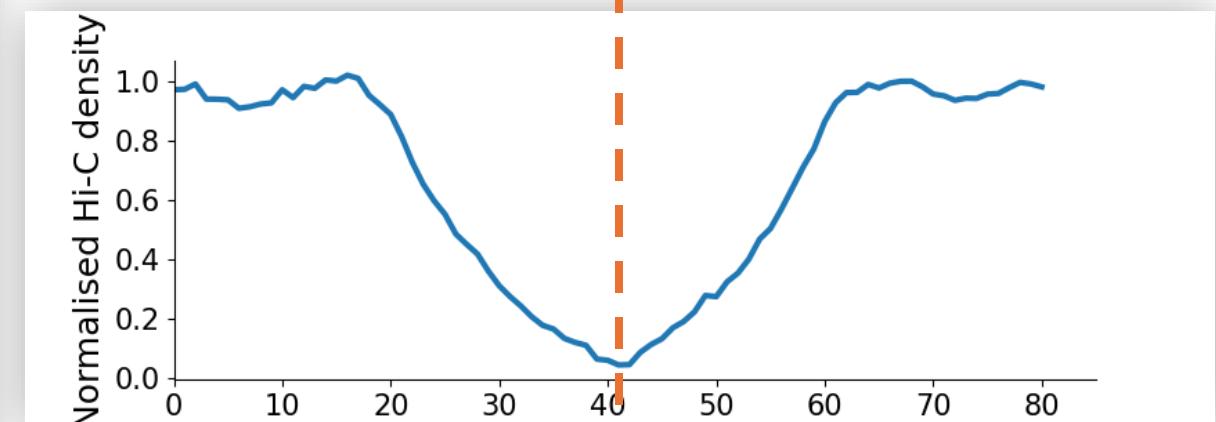
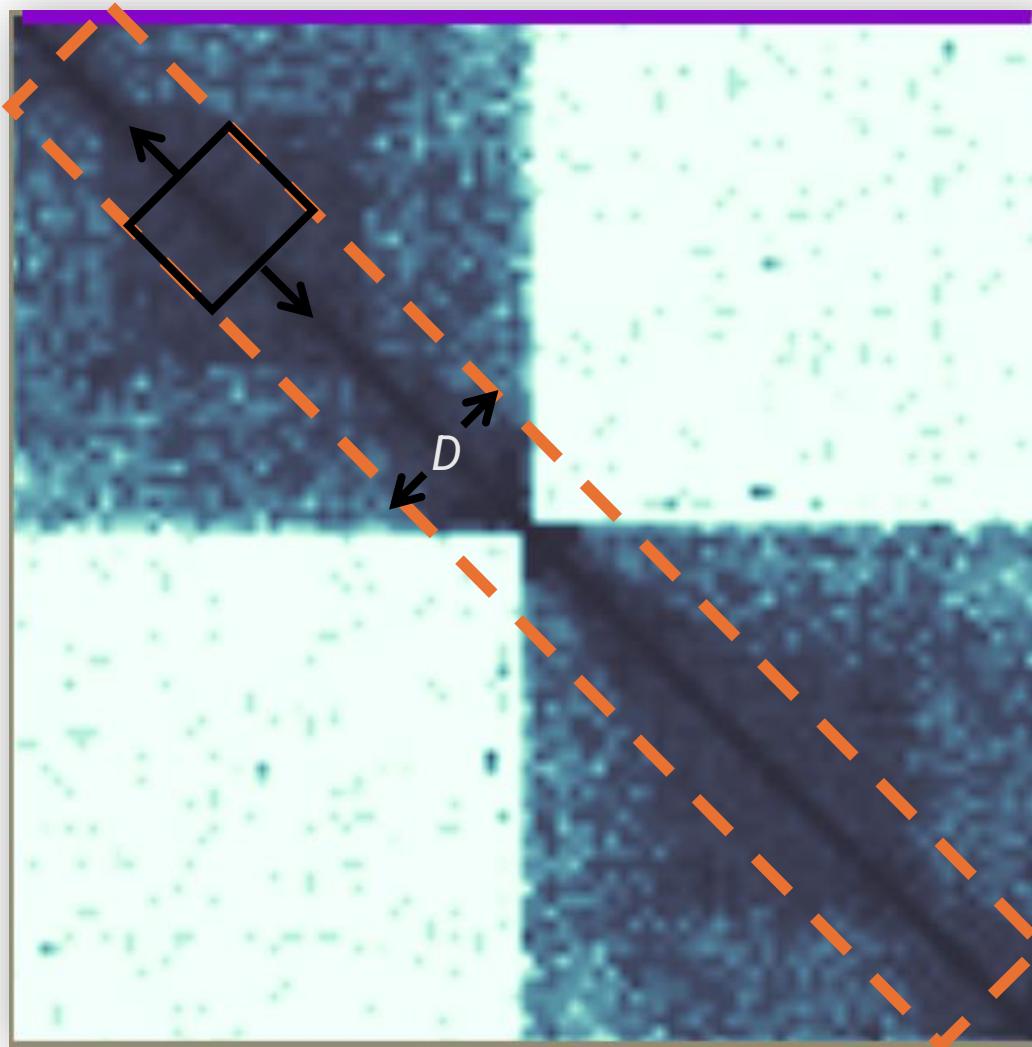
- New AI features for “Automated” curation:
- ‘J’ for jump: scaffold jumps from the shrapnel to the diagonal where the HiC signal is the strongest
- ‘K’ for jumping back to the shrapnel



PretextView

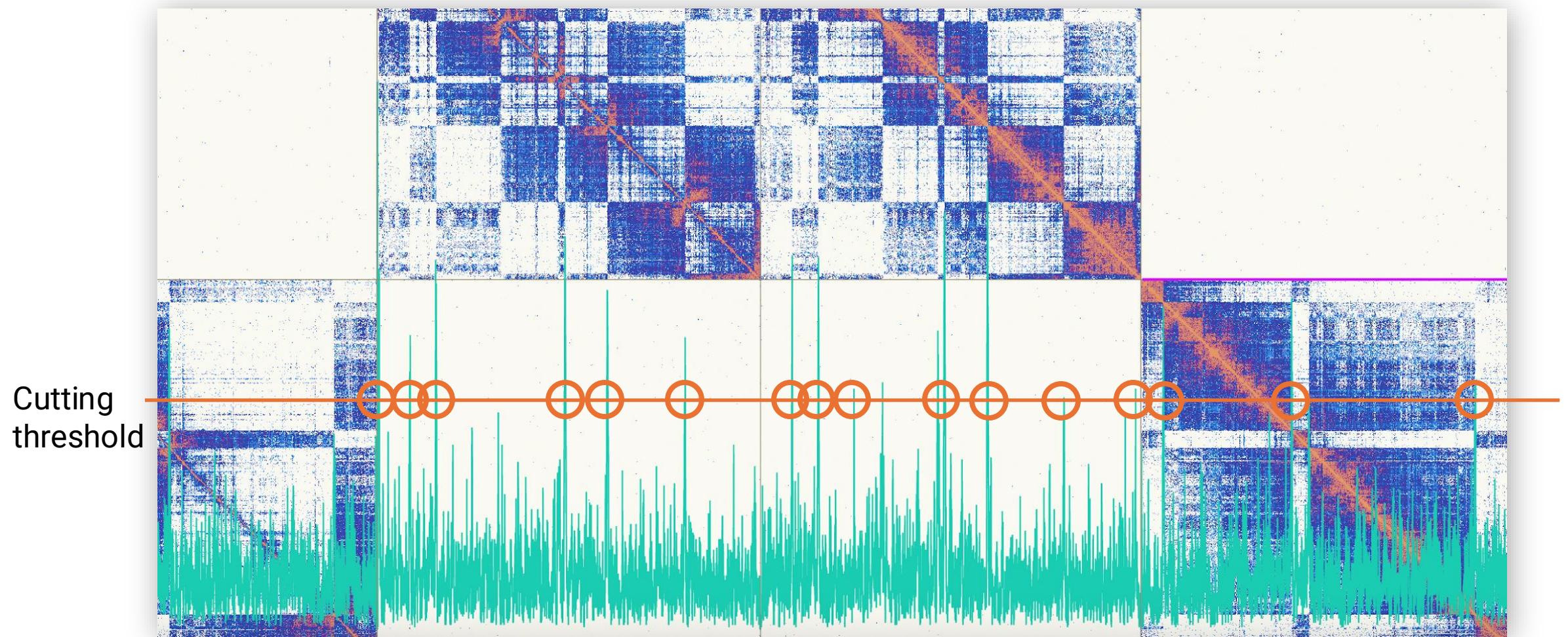


Pixel cut



PretextView

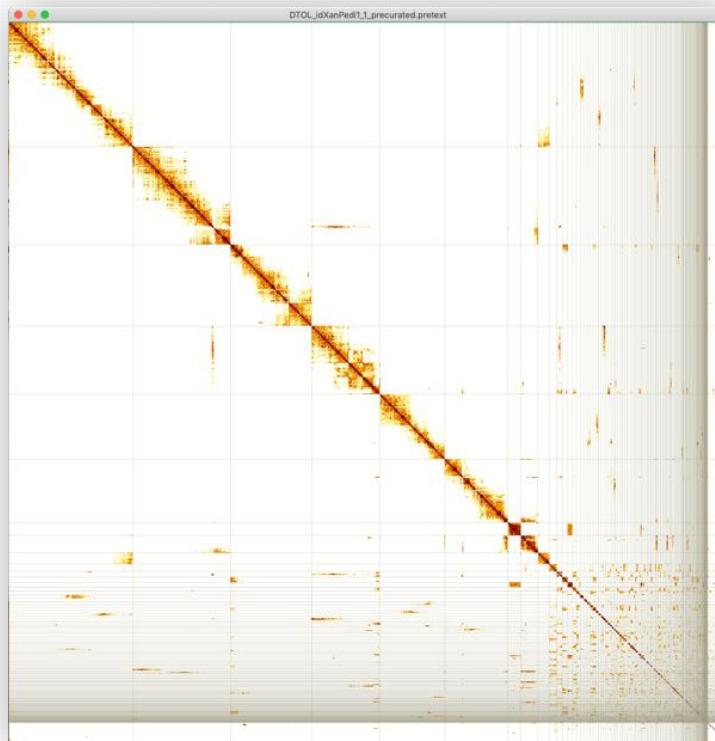
Pixel cut



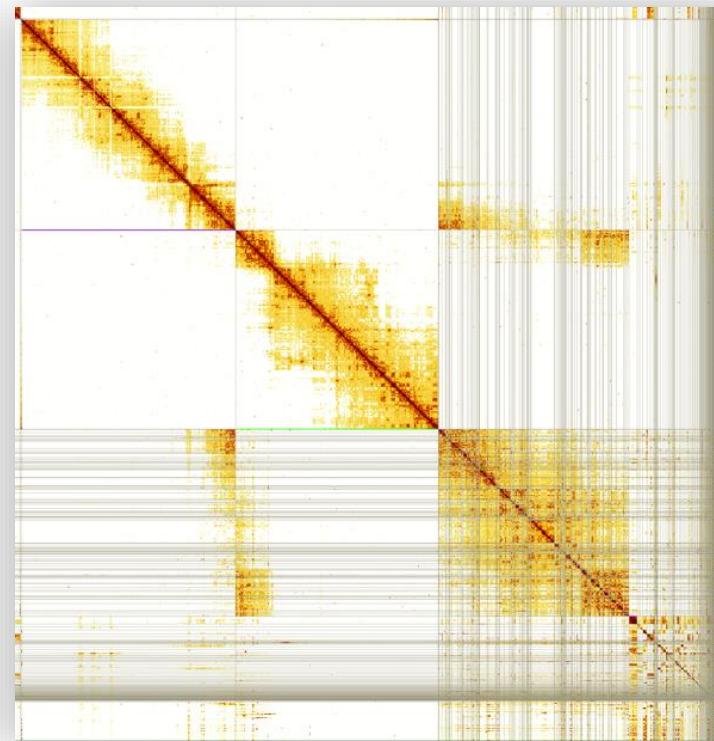
PretextView



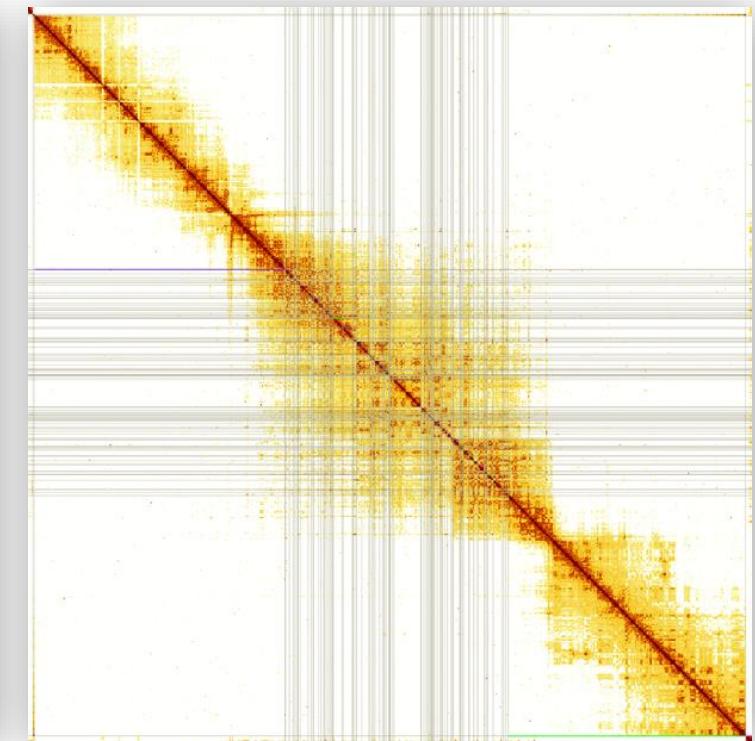
Pixel sort ('F' key shortcut)



Pre-curation



Grouped and ordered



Resolved



Xanthogramma pedissequum (Hoverfly)

PretextView



Three modes for sorting:

Mode	Sensitivity	Computational demand	No. scaffolds
Union find	Low	Low	High
Fuse	Medium	Medium	Medium
Deep Fuse	High	High	Low

Works well for repetitive regions, usually hard to assemble, such as **centromeres and W chromosomes**

Hands-on

<https://github.com/epaule/Physalia-Manual-Genome-Curation/blob/main/Session3.1.md>