

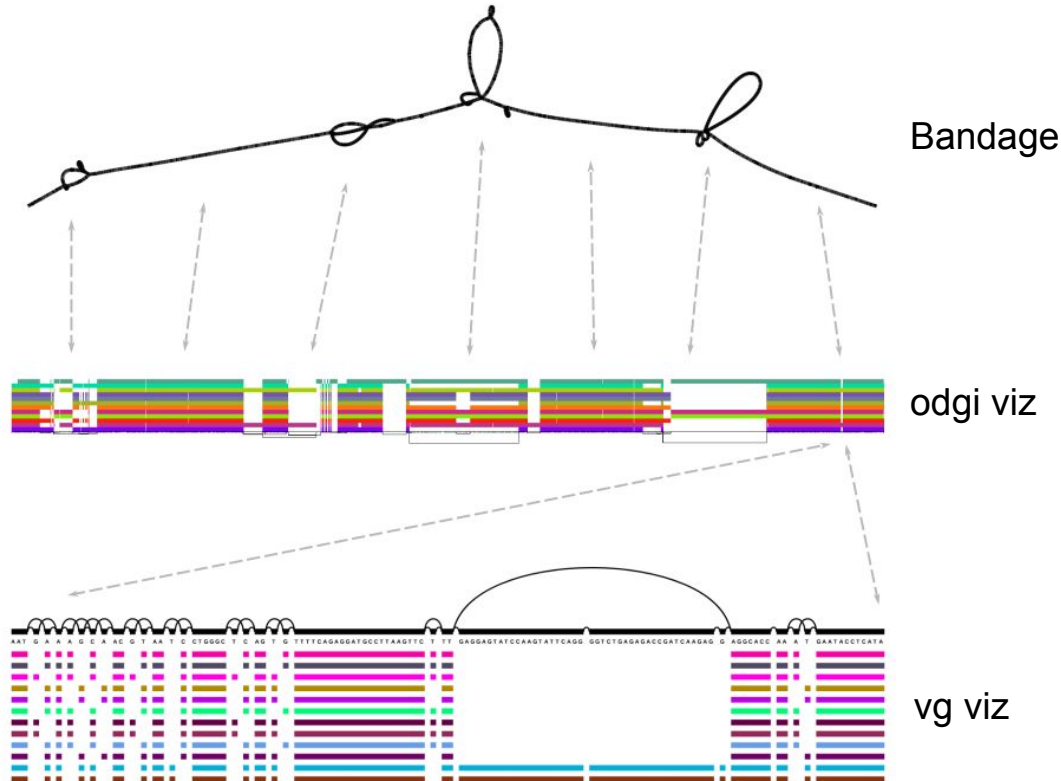
# Computational PANGenomics 2022

## #CPANG22

Instituto Gulbenkian de Ciência, Portugal  
Day 2 - 2022/05/24

Erik Garrison and Andrea Guarracino

# Pangenome graph visualizations



Pangenome graph with 12 ALT sequences of the HLA-DRB1 gene from the GRCh38 reference genome.

Figure from [Eizenga et al., 2020](#).

# Bandage

Bandage - chr6.C4.gfa

File Edit Tools View Select Output Help

**Graph information**

Nodes: 1.748  
Edges: 2.366  
Total length: 51.672

More info

**Graph drawing**

Scope: Entire graph

Style:  Single  Double

Draw graph

**Graph display**

Zoom: 4.0%

Node width: 500,0

Random colours

**Node labels**

Custom  Name  
 Length  Depth  
 BLAST hits  CSV data:

Font  Text outline

**BLAST**

Create/view BLAST search

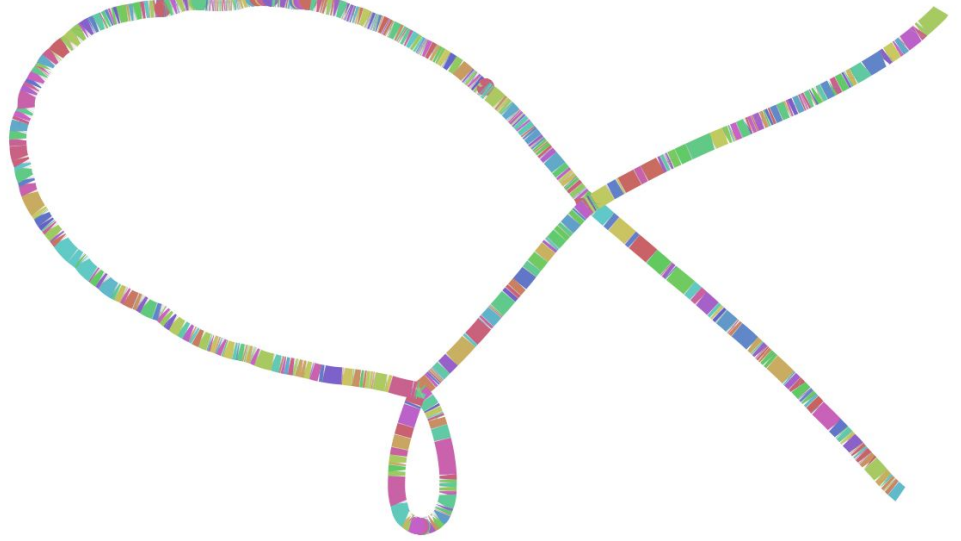
Query: none

**Find nodes**

Node(s):

Match:  Exact  Partial

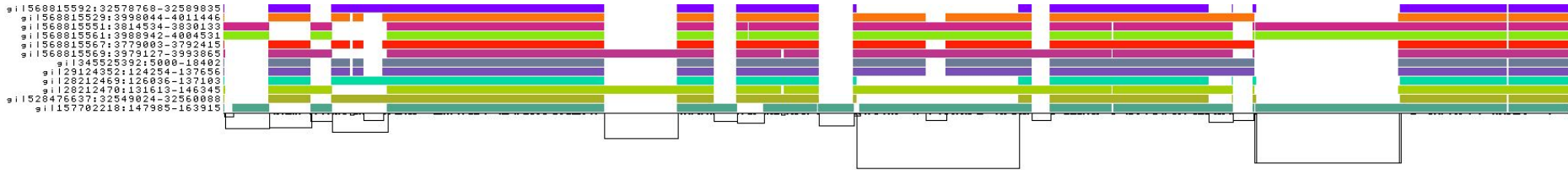
Find node(s)



# Visualizing pangenome graphs in 1D - odgi viz

By visualizing pangenome graphs we can gain insight into the mutual relationship between the embedded sequences and their variation.

Pangenome graph with 12 ALT sequences of the HLA-DRB1 gene from the GRCh38 reference genome.



- The graph nodes are arranged from left to right, forming the pangenome sequence.
- The colored bars represent the paths versus the pangenome sequences in a binary matrix.
- The path names are visualized on the left.
- The black lines under the paths are the links, which represent the graph topology.

# Visualizing pangenome graphs in 1D - odgi viz

Pangenome graph with 12 ALT sequences of the HLA-DRB1 gene from the GRCh38 reference genome.

Colored by path position (light = start, dark = end)

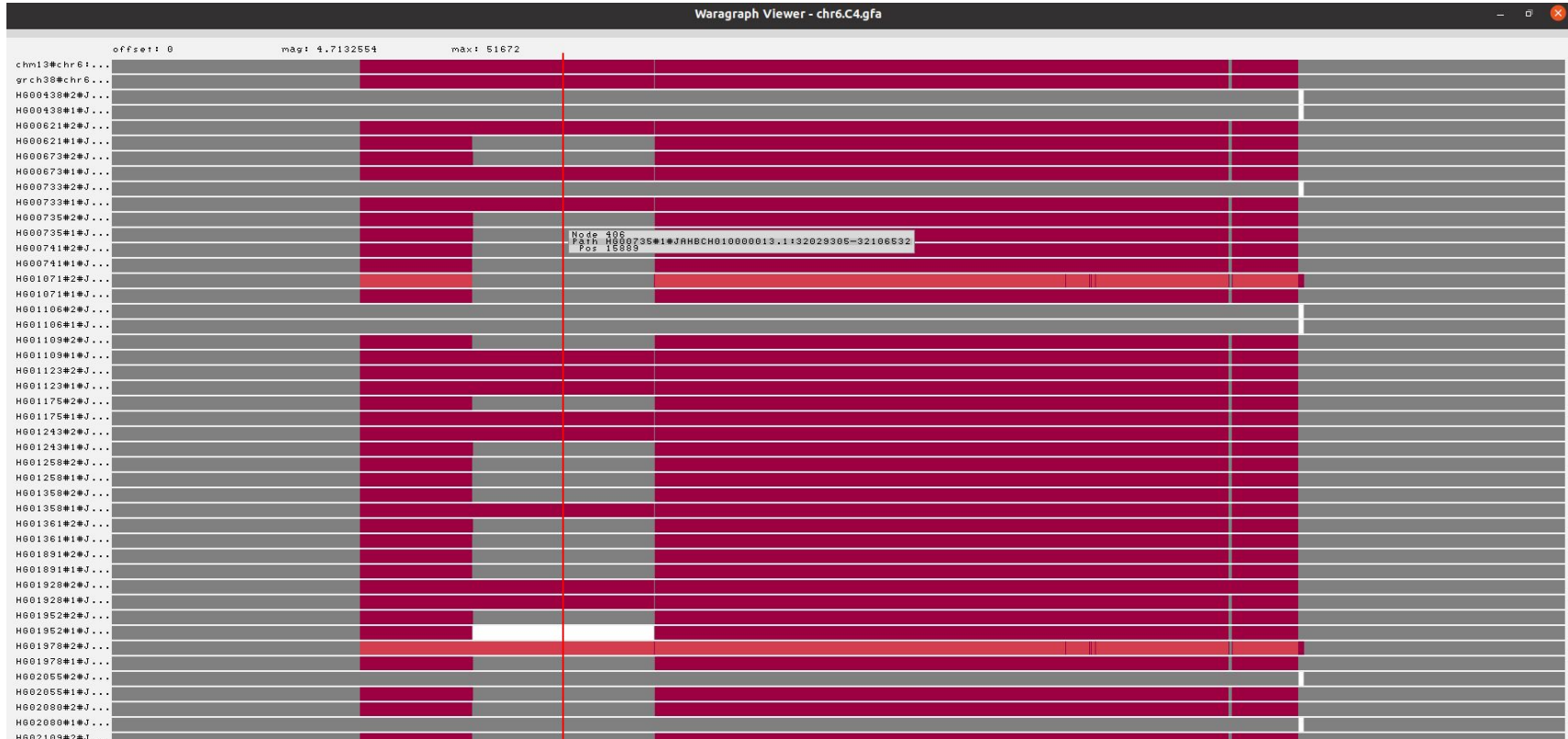


Pangenome graph with 12 ALT sequences of the HLA-DRB1 gene from the GRCh38 reference genome.

Colored by orientation (black = forward, red = reverse)

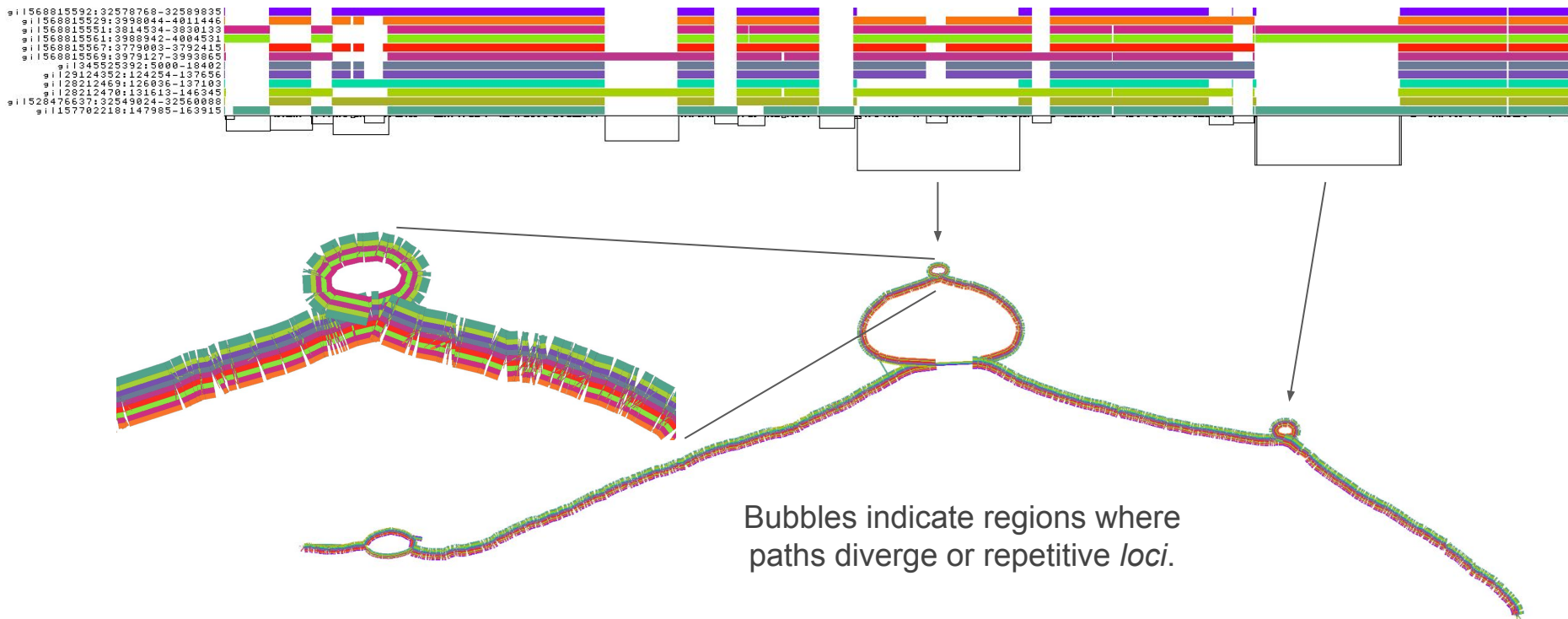


# Visualizing pangenome graphs in 1D - waragraph



# Visualizing pangenome graphs in 2D - odgi draw

Pangenome graph with 12 ALT sequences of the HLA-DRB1 gene from the GRCh38 reference genome.

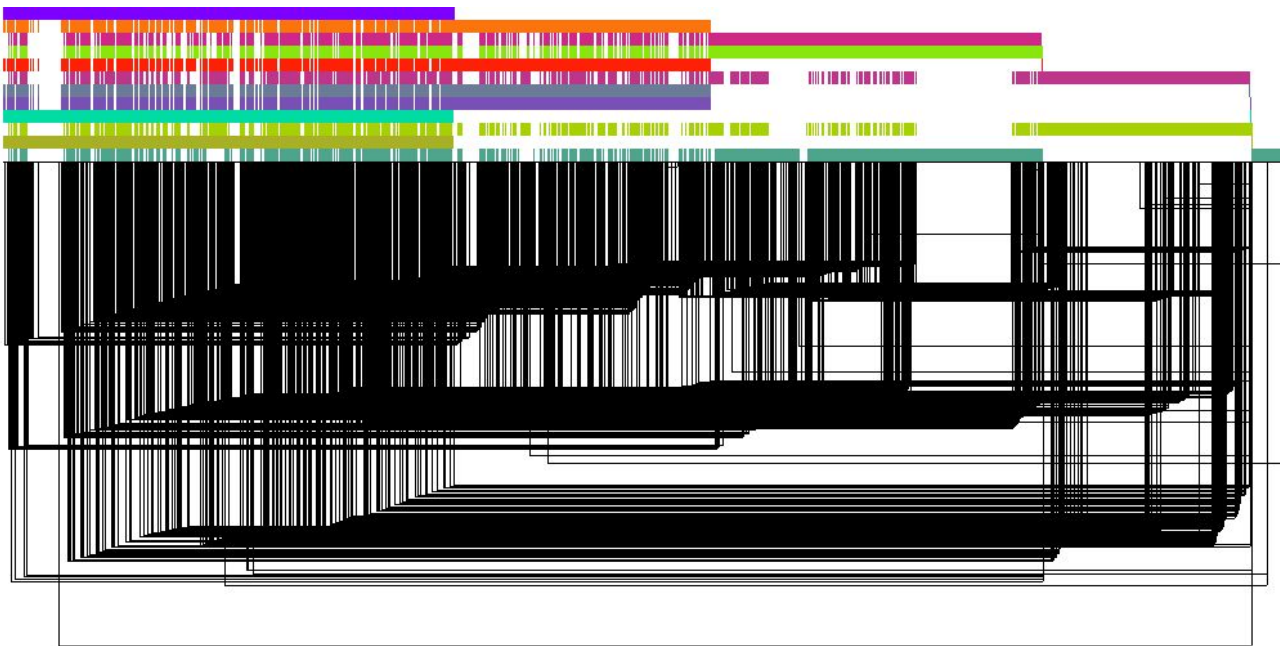


# Finding latent structures in pangenome graphs

Pangenome graphs built from raw sets of alignments may have complex structures which can introduce difficulty in downstream analyses.

Pangenome graph with 12 ALT sequences of the HLA-DRB1 gene from the GRCh38 reference genome.

```
g|1568815592:325787-  
g|1568815529:399804-  
g|1568815551:381453-  
g|1568815561:398894-  
g|1568815567:377900-  
g|1568815569:397912-  
g|1345525392:5000-1-  
g|1291243521:124284-  
g|1282124659:126036-  
g|128212470:131613-  
g|1528476637:325490-  
g|157702218:147985-
```



Raw graph built using [seqwish](#).



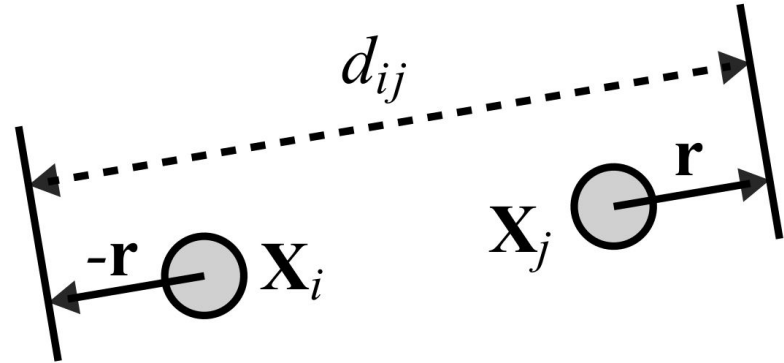
# Graph drawing by stochastic gradient descent

We organize the graph by moving nodes so their position in the layout better-matches their distance through paths in the graph.

**We learn the graph layout.**

So we use a trick that works on our sparse (and large 1M to 1G node) graphs: index the genome paths through them and use that to quickly obtain distances.

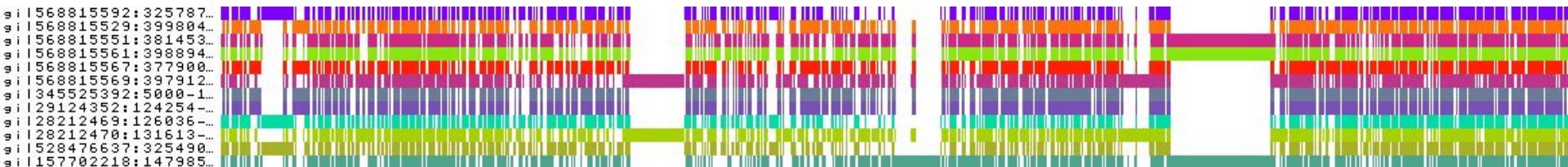
→ **Path-guided SGD**



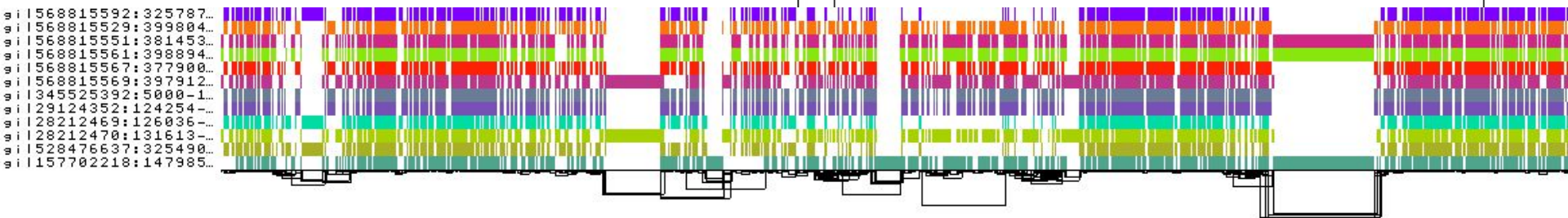
J. X. Zheng, S. Pawar and D. F. M. Goodman,  
"Graph Drawing by Stochastic Gradient Descent,"  
in IEEE Transactions on Visualization and  
Computer Graphics, vol. 25, no. 9, pp. 2738-2748,  
1 Sept. 2019, doi: 10.1109/TVCG.2018.2859997.

# Finding latent structures in pangenome graphs - odgi sort

Pangenome graph with 12 ALT sequences of the HLA-DRB1 gene from the GRCh38 reference genome.

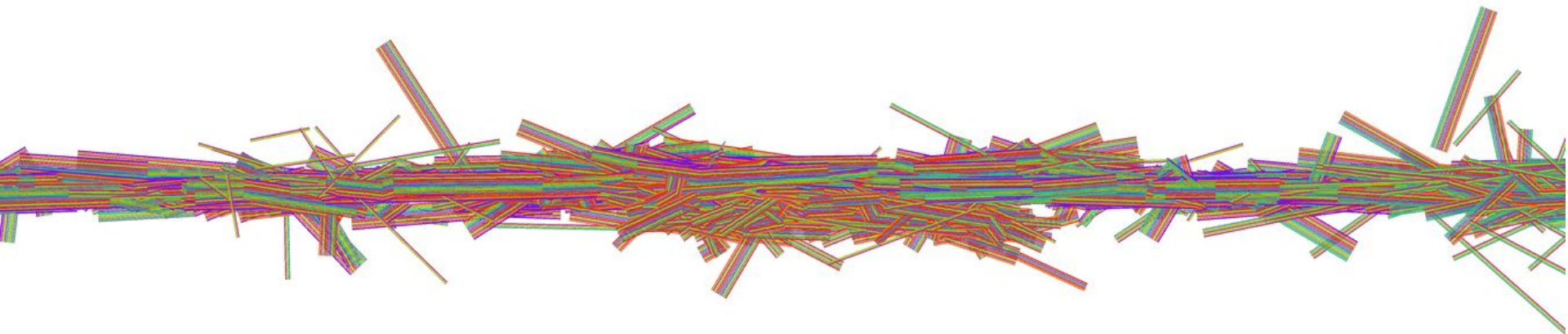


Path-guided stochastic gradient descent algorithm to optimize 1D order to best-match positions in embedded paths.



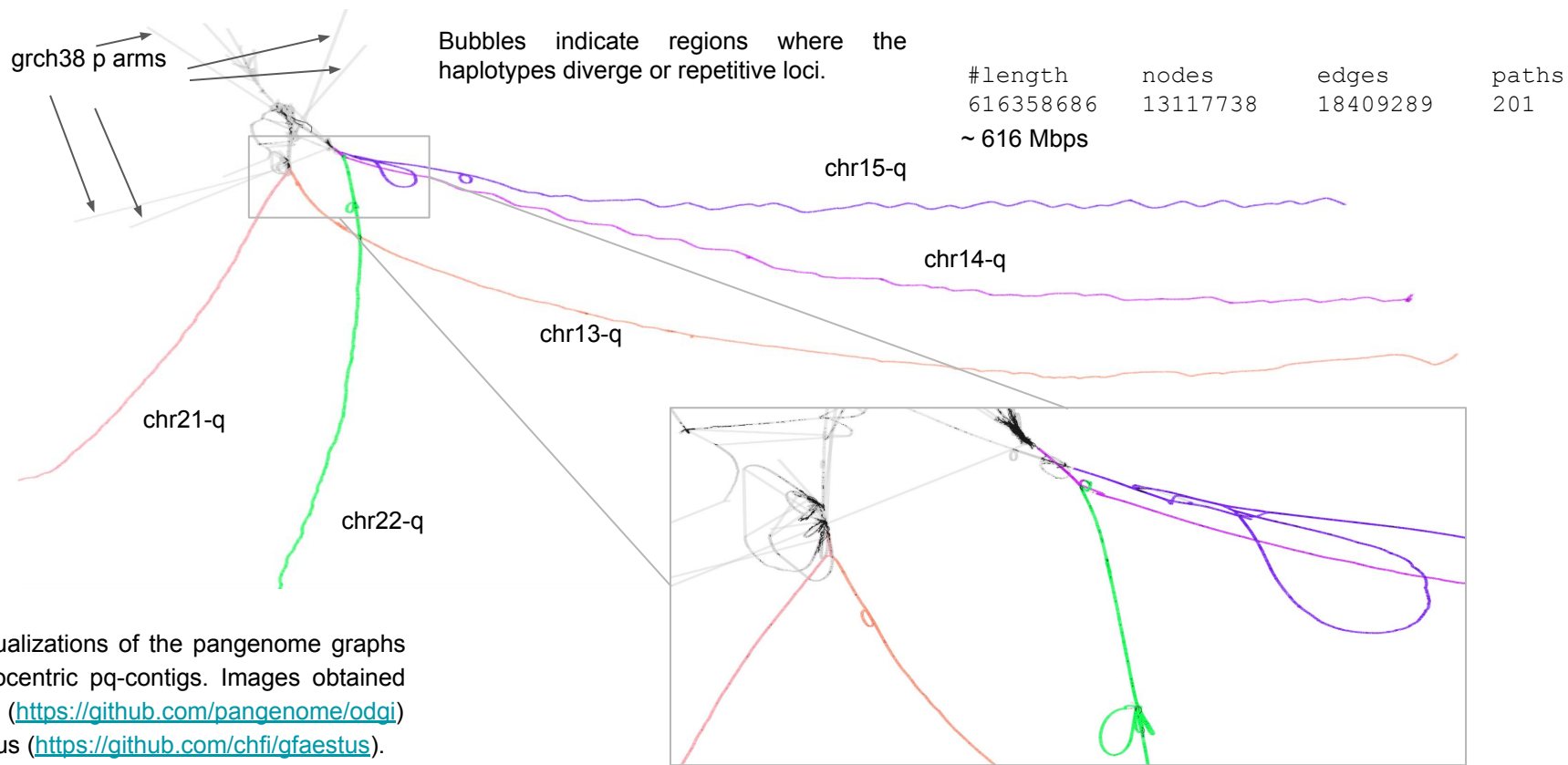
# Finding latent structures in pangenome graphs - odgi layout

Pangenome graph with 12 ALT sequences of the HLA-DRB1 gene from the GRCh38 reference genome.



Path-guided stochastic gradient descent algorithm to optimize 2D layout. Path-labeled rendering with odgi draw.

# Visualizing pangenome graphs in 2D - gfaestus



# Activities

<https://hackmd.io/@AndreaGuarracino/r1dyQA7Dq>