

Exploring the full spectrum of genetic variation in Cannabis with Pantograph

Jörg Hagmann, Christian Kubica
Computomics, Tuebingen, Germany

WHAT IS A PANGENOME (GRAPH)?

- Alignments of multiple whole genome assemblies
- Pangenome graphs store the homologous and diverged regions between entire genomes
- A pangenome includes more genes and sequences than a single reference genome

WHY LOOKING AT A CANNABIS PANGENOME?

- Cannabis varieties show **large genetic diversity** and many **gene duplications**, esp. in the synthase genes of the cannabinoid pathway
- A pangenome of Cannabis sheds light on:
 - Complex and large-scale structural variation**
 - Copy number variation**
 - Variation in non-reference sequence**

PANGENOME GRAPH BROWSER PANTOGRAPH

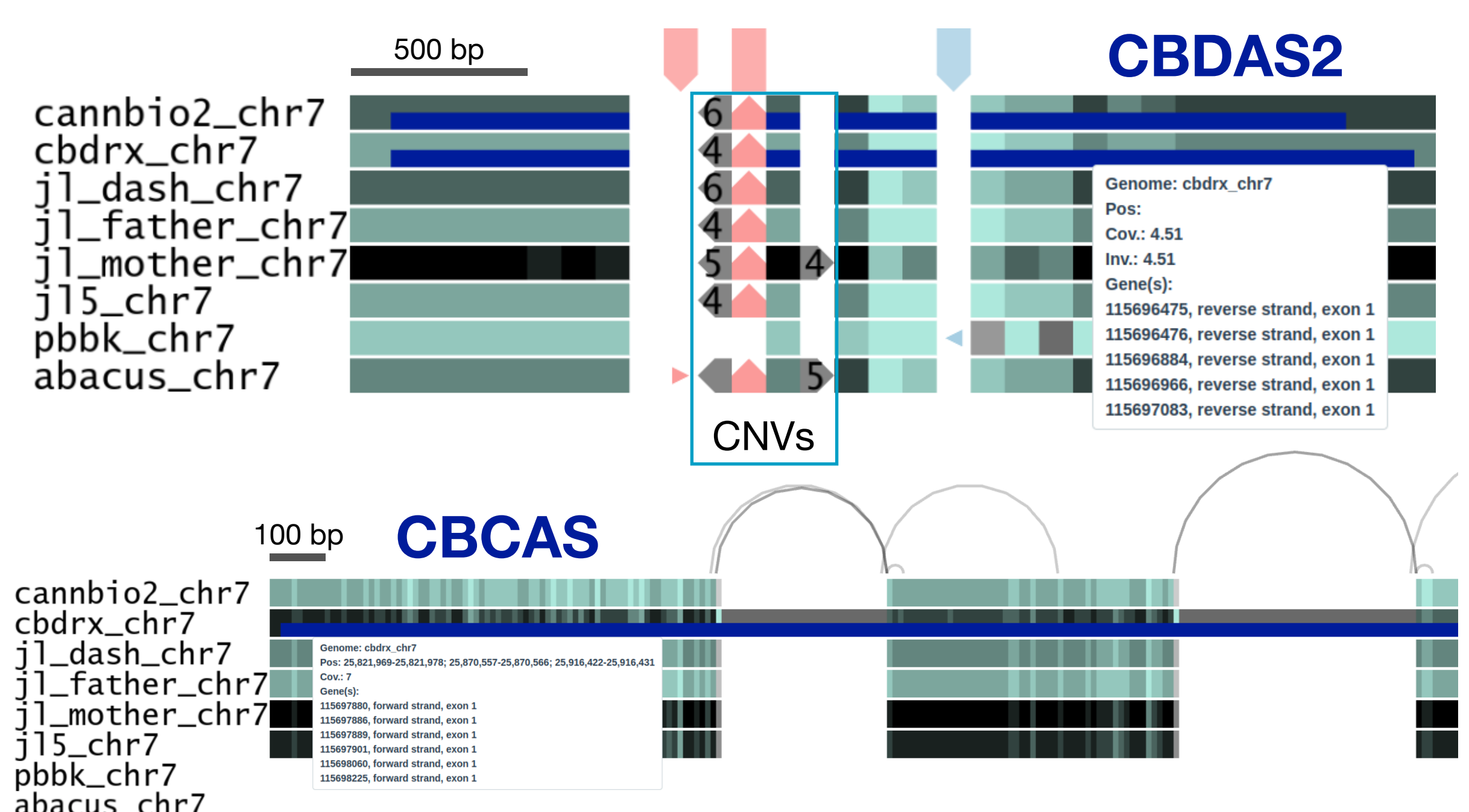
We are developing *pantograph*, an interactive pangenome graph browser with intuitive navigation:

- Visualizes **all-vs-all variety alignments**
- Multiple zoom levels** from single nucleotide level to whole chromosome views
- Inspects QTL/GWAS regions** by showing phenotypic values alongside genetic variation
- Improves marker development** by including structural and non-reference variation **beyond SNPs** potentially leading to accelerated Cannabis breeding

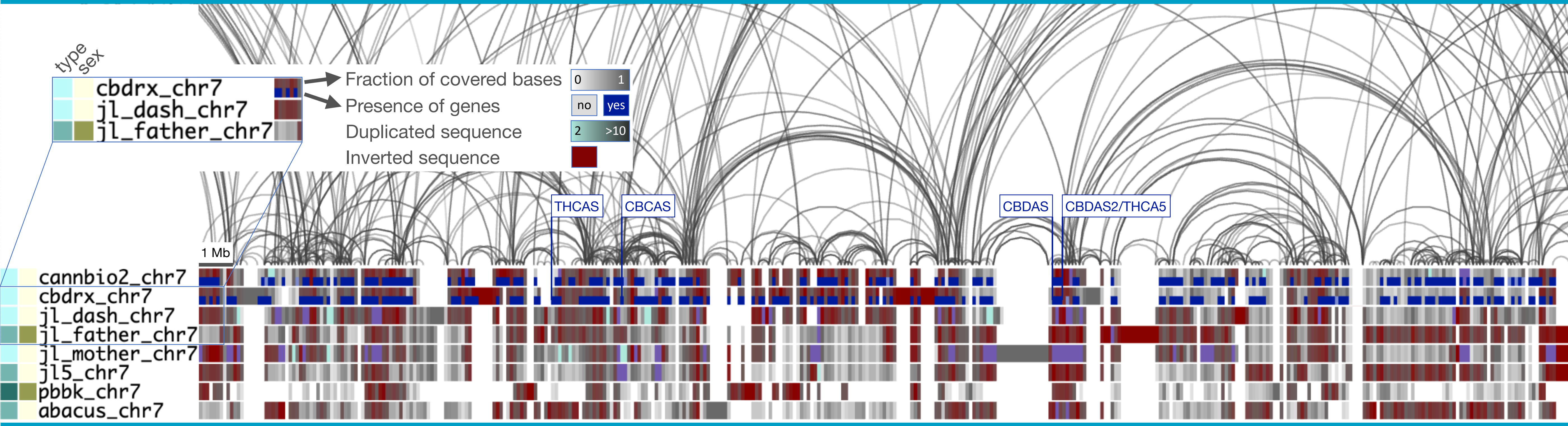
A PUBLIC CANNABIS PANGENOME GRAPH

- We built a pangenome graph of chrom. 7 sequences of 8 publicly available Cannabis assemblies
- The graph was constructed with the pggp pipeline that performs all-vs-all variety alignments (80% pid), free of reference bias (<https://github.com/pangenome/pggb>)

COPY NUMBER VARIATION IN CANNABIS GRAPH



A Pangenome of Cannabis Varieties Reveals large-scale and Copy Number Variation with *pantograph*

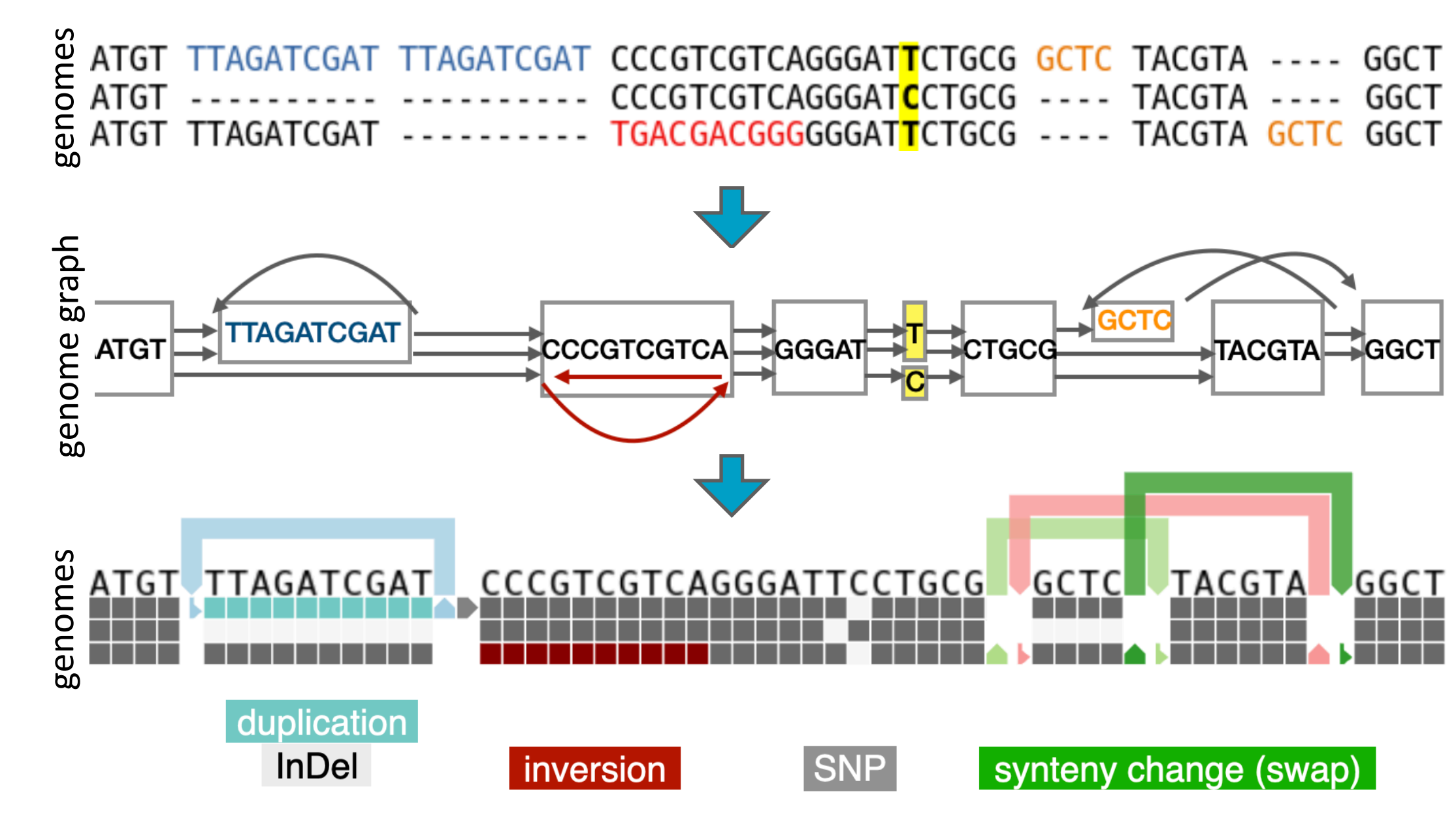


Chromosome-scale view of a pangenome built from eight public Cannabis assemblies

FEATURES OF PANTOGRAPH

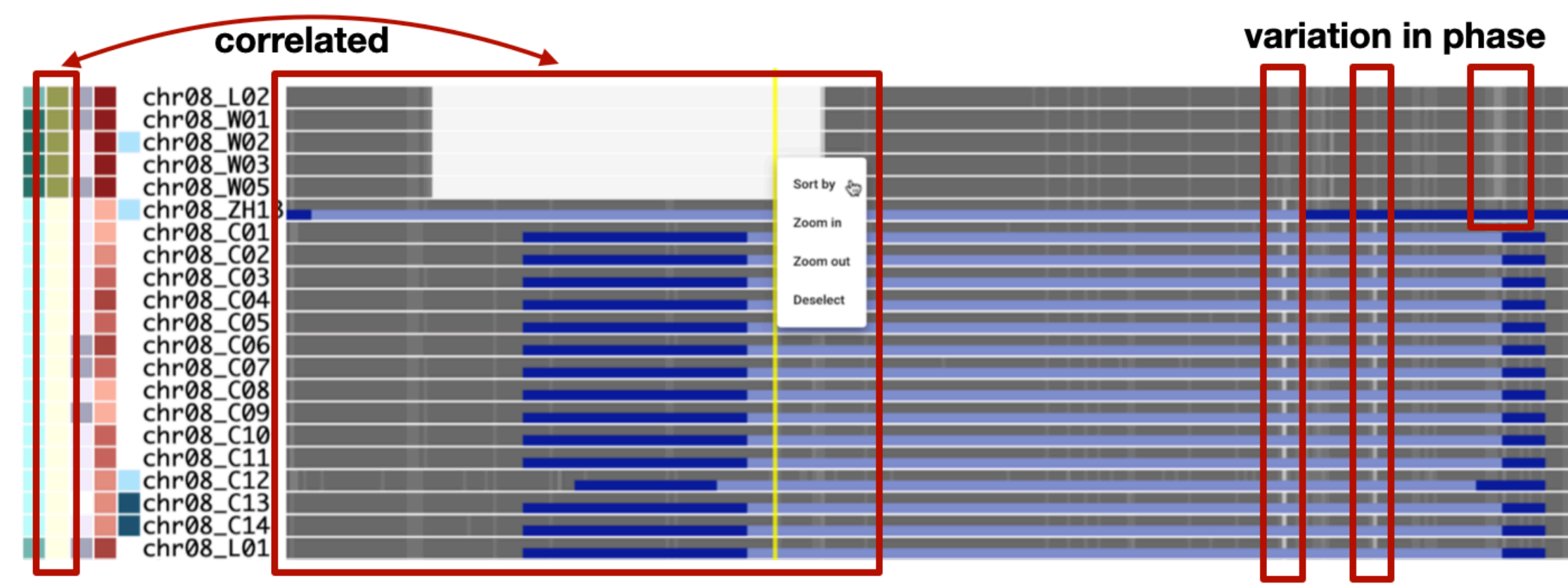
All variant types in a single browser

From an MSA to genome graph to Pantograph:



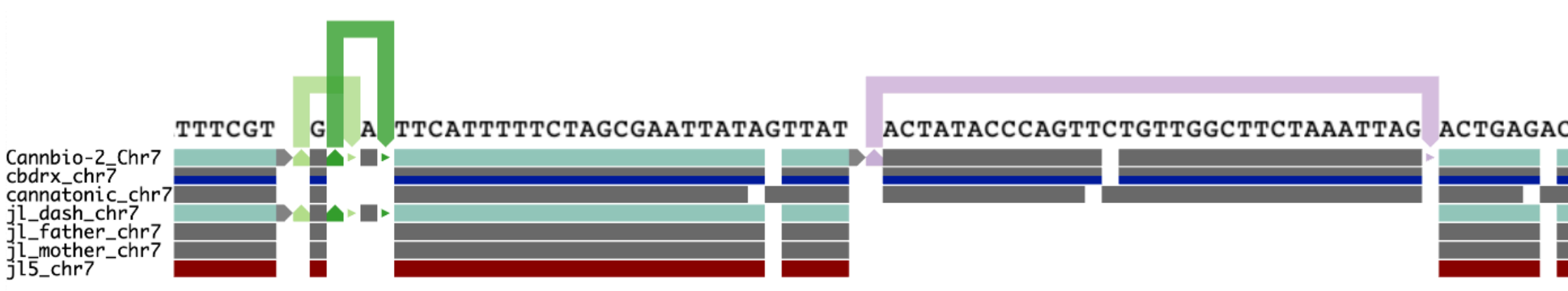
Inspect trait-associated genetic variation

- Sorting by sequence or meta data (phenotypes)
- Coloring by meta data

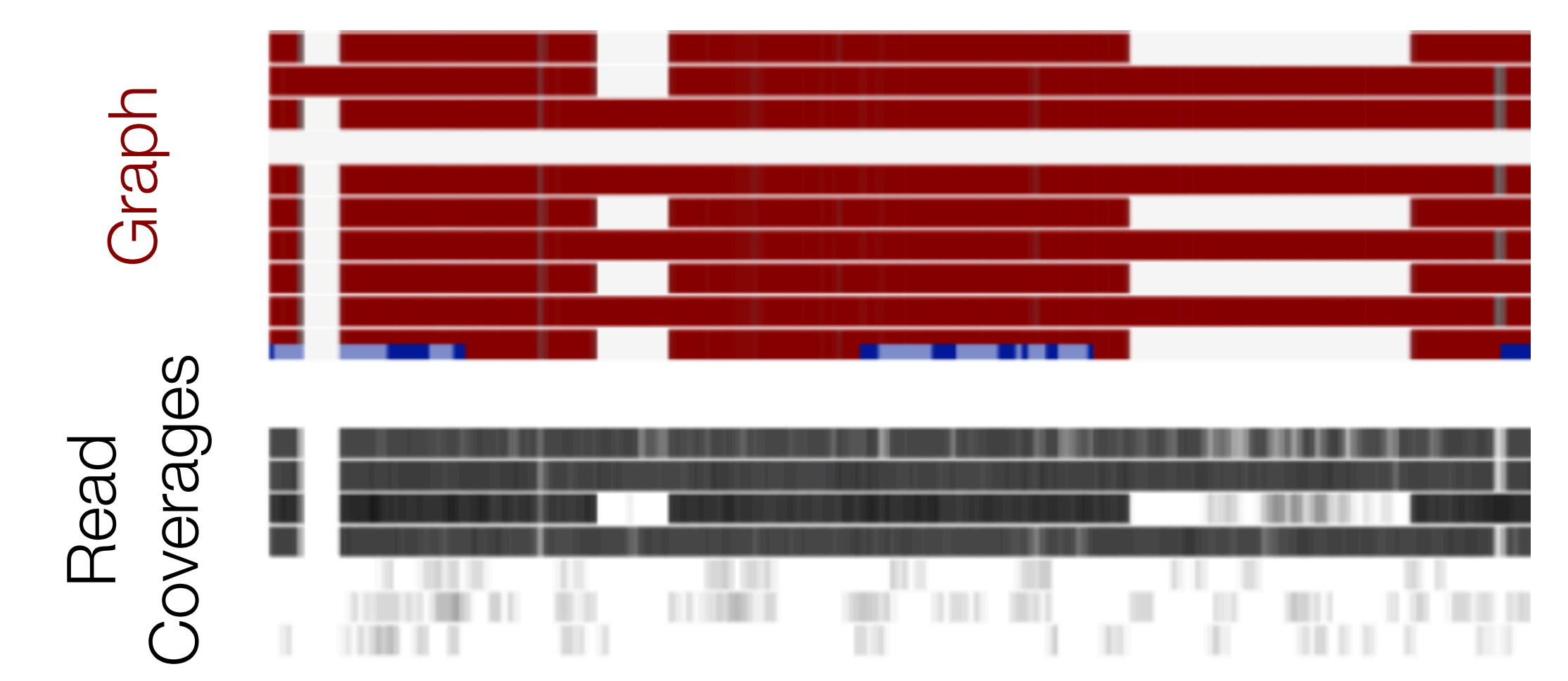


Genetic variants associated with seed coat color in a public soybean pangenome (Liu *et al.* Cell 2020)

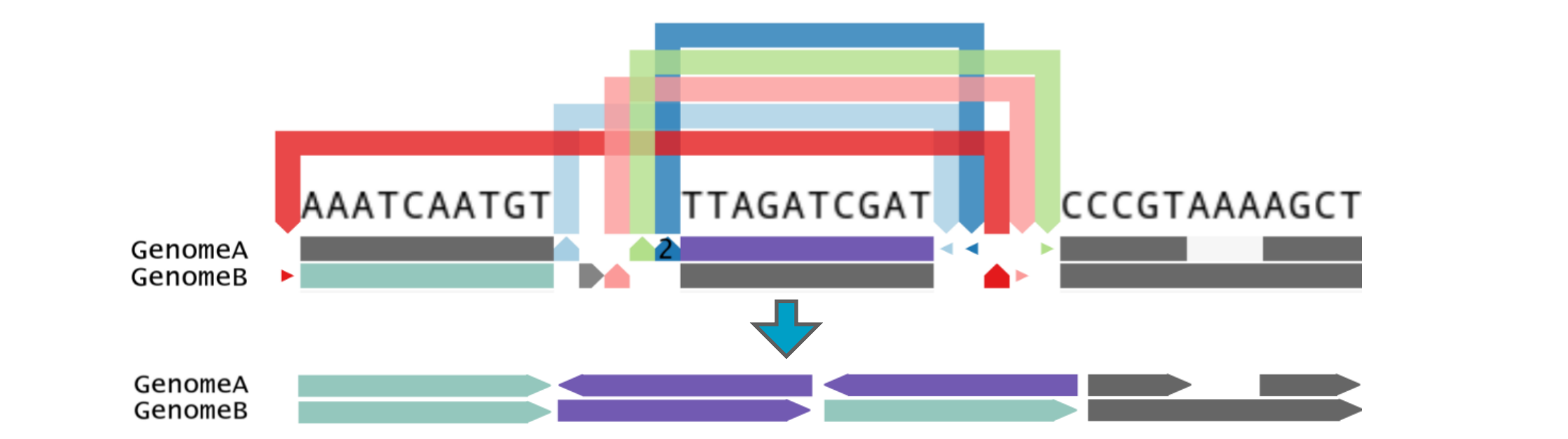
Nucleotide-scale view



Extension: Visualize read mappings/QTL data



Outlook: Visualize repeat synteny



Would you like to see your Cannabis varieties in this way? Talk with us:

joerg.hagmann@computomics.com