Day 4: Alignments as graphs and the MHC region
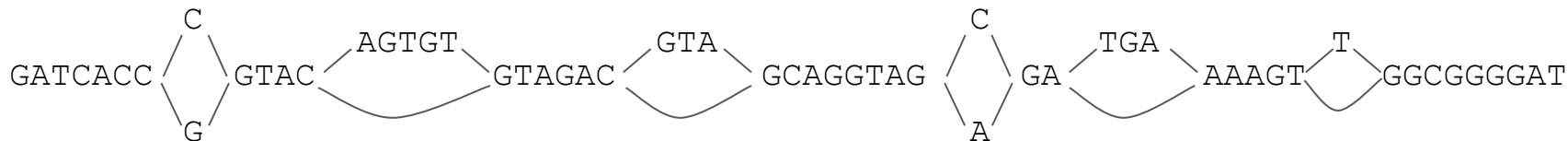
Jordan Eizenga, Tobias Marschall, Erik Garrison
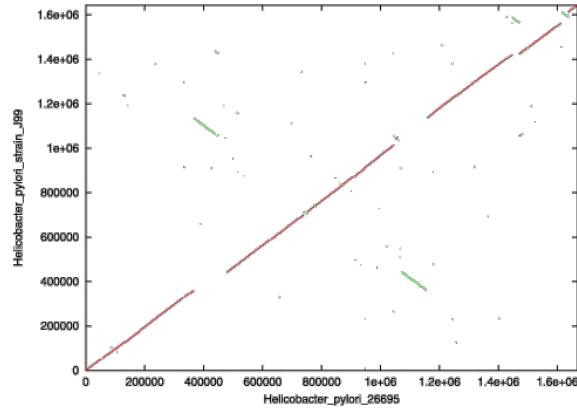
Multiple sequence alignments encode graphs

```
GATCACCCGTACAGTGTGTAGACGTAGCAGGTAGCGATGAAAAGT-GGCGGGGAT
GATCACCCGTACAGTGTGTAGAC---GCAGGTAGCGATGAAAAGT-GGCGGGGAT
GATCACCAGTAC-----GTAGACGTAGCAGGTAGAGATGAAAAGT-GGCGGGGAT
GATCACCAGTAC-----GTTGACGTAGCAGGTAGAGA---AAAGTTGGCGGGGAT
```
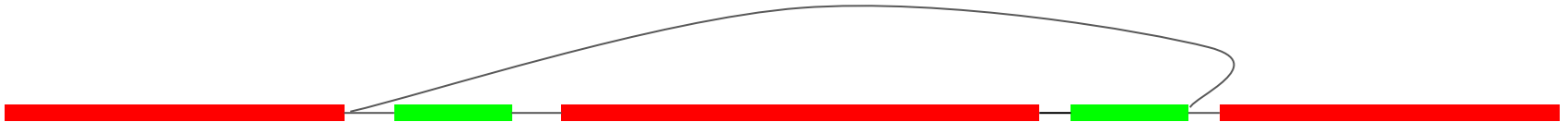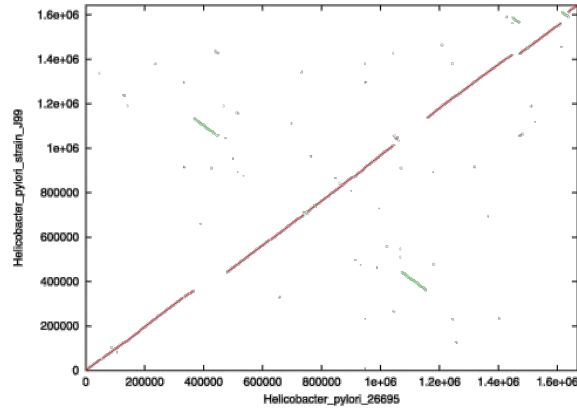
Multiple sequence alignments encode graphs

```
GATCACCCGTACAGTGTGTAGACGTAGCAGGTAGCGATGAAAAGT-GGCGGGGAT
GATCACCCGTACAGTGTGTAGAC---GCAGGTAGCGATGAAAAGT-GGCGGGGAT
GATCACCAGTAC-----GTAGACGTAGCAGGTAGAGATGAAAAGT-GGCGGGGAT
GATCACCAGTAC-----GTTGACGTAGCAGGTAGAGA---AAAGTTGGCGGGGAT
```
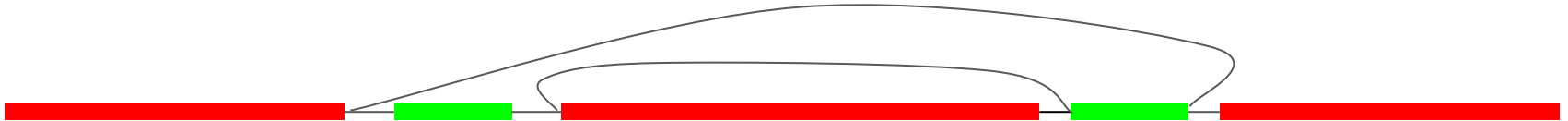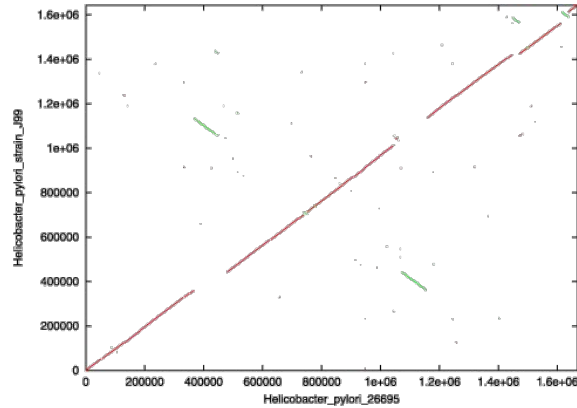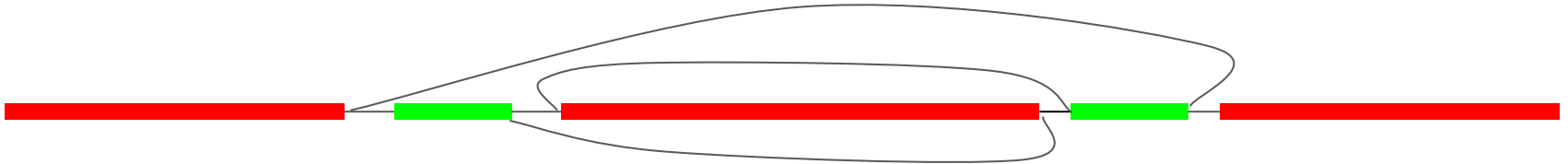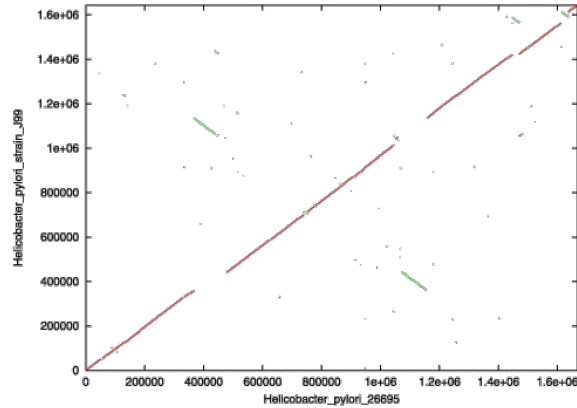
Multiple sequence alignments encode graphs

Multiple sequence alignments encode graphs
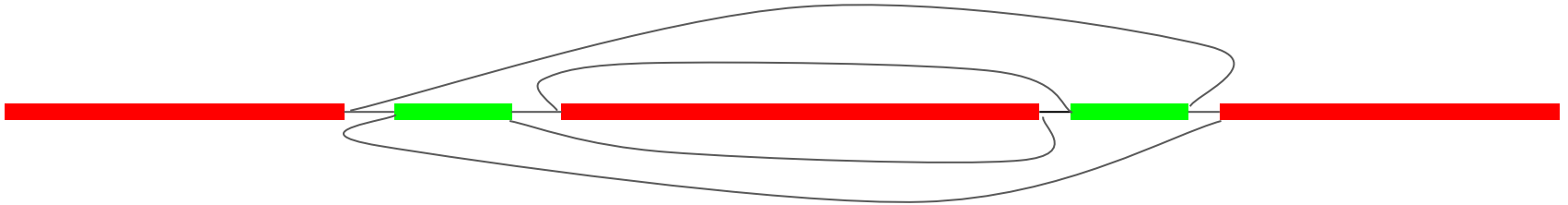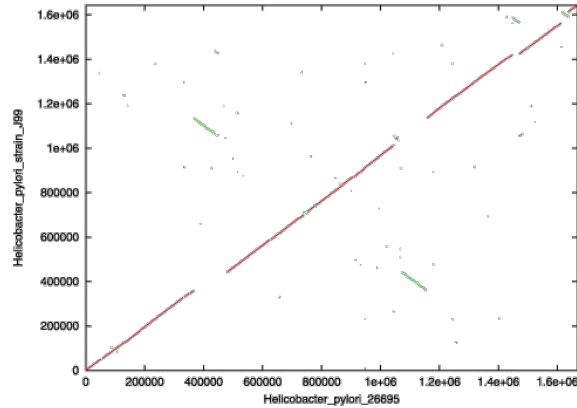
Multiple sequence alignments encode graphs
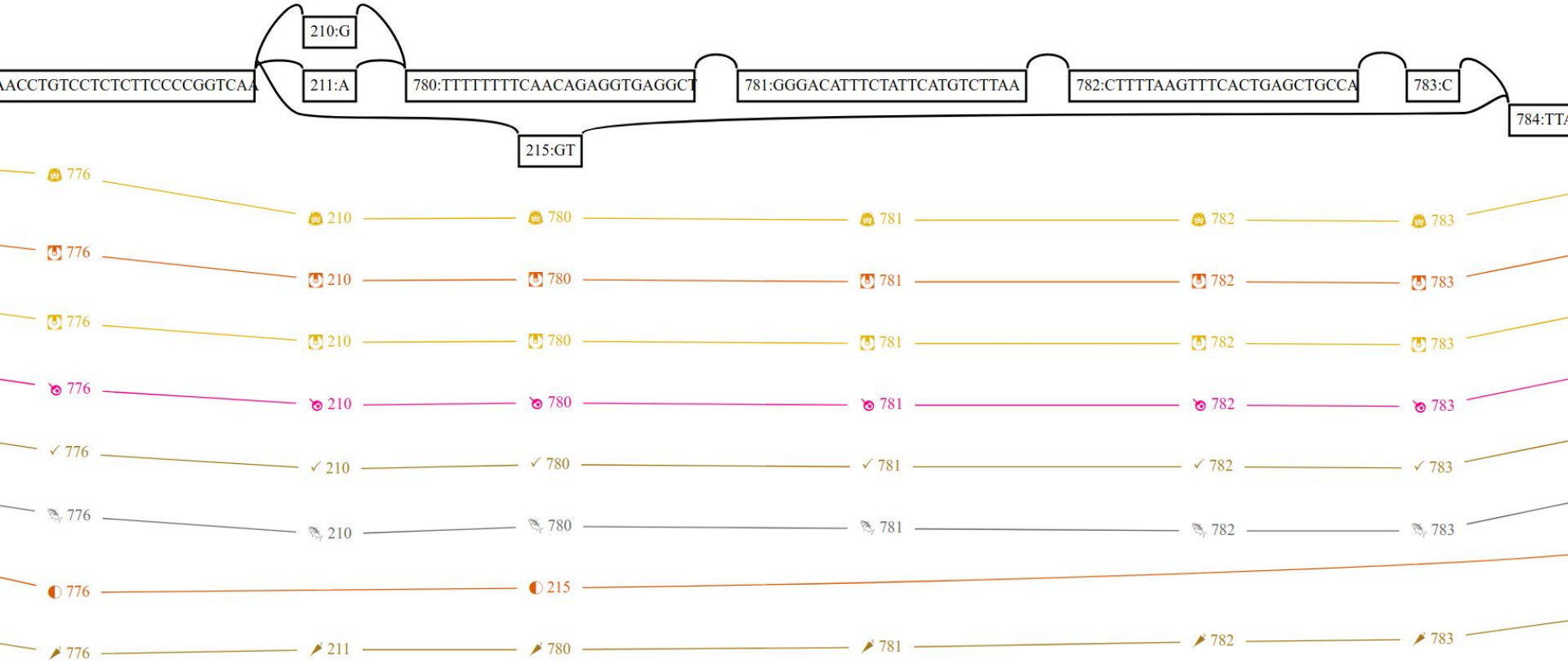
Multiple sequence alignments encode graphs

Multiple sequence alignments encode graphs

Doing it in VG

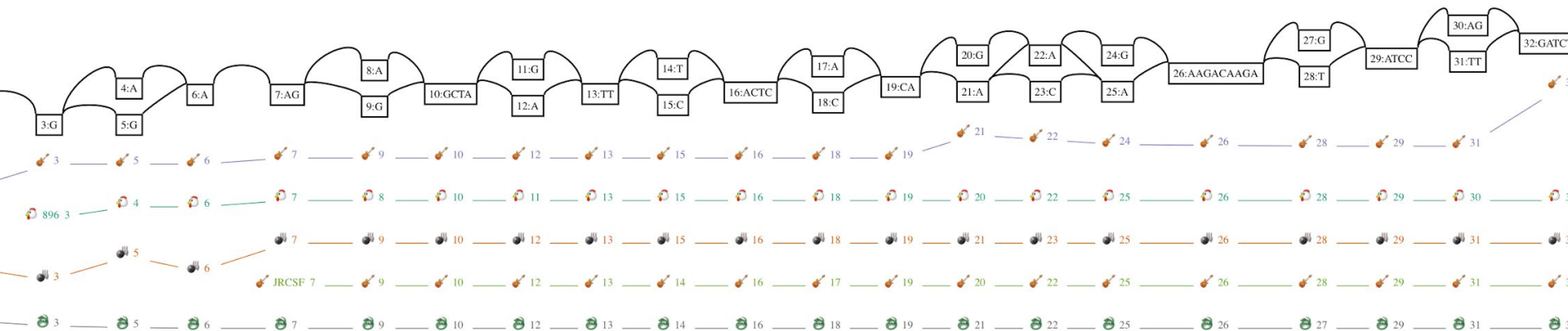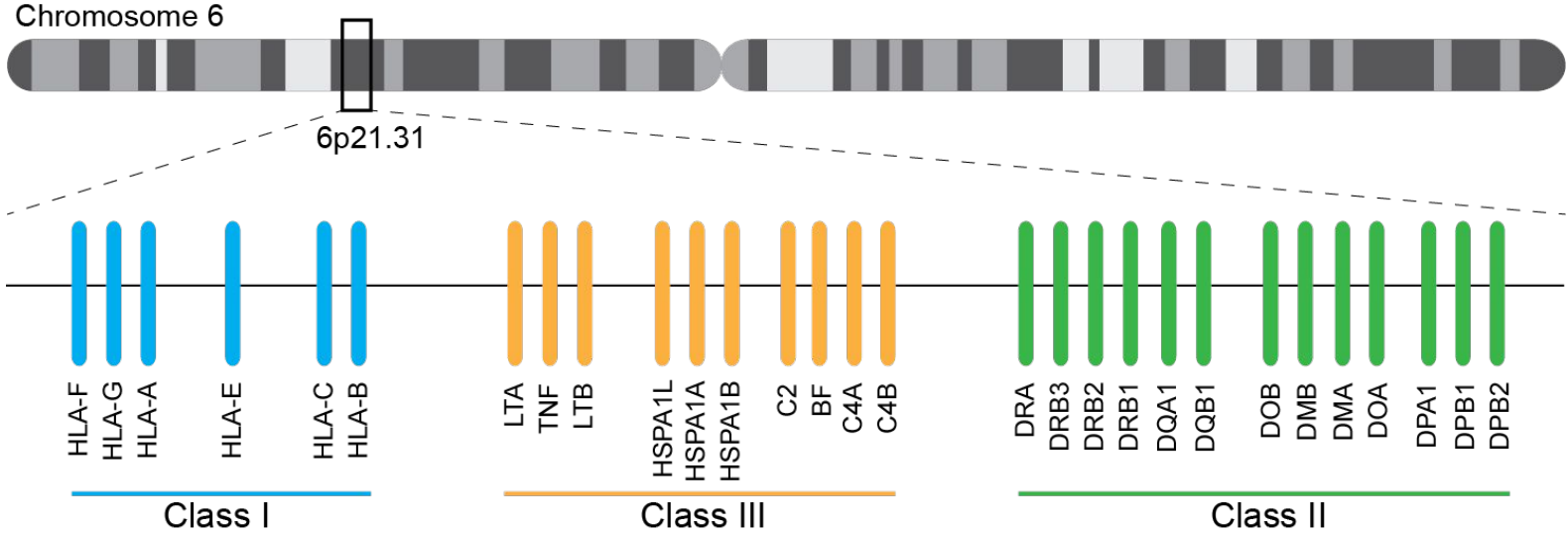# Doing it in VG

Doing it in VG

use Clustal Omega to make a multiple sequence alignment
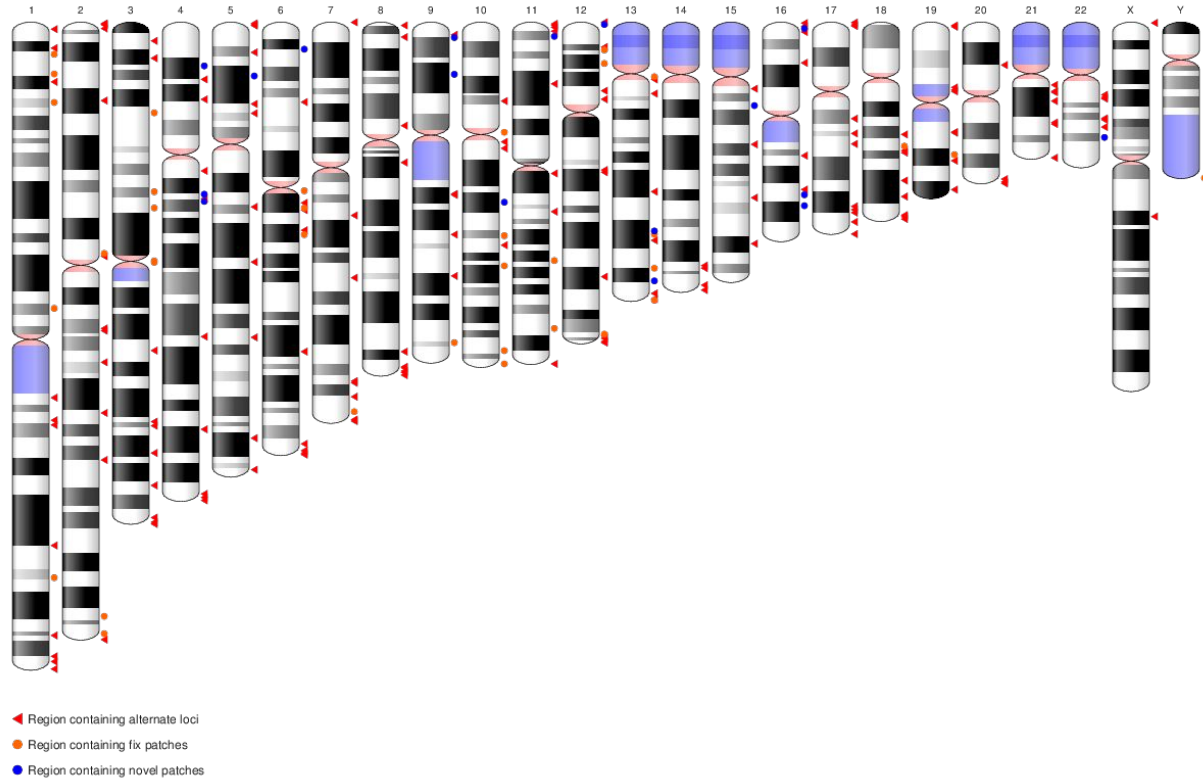clustalo -i sequences.fasta --outfmt clustal > mult_seq_aln.clustal

convert the multiple sequence alignment to a VG file
-M = input file, -F = format, -m = max node size
vg construct -M mult_seq_aln.clustal -F clustal -m 32 > msa.vg

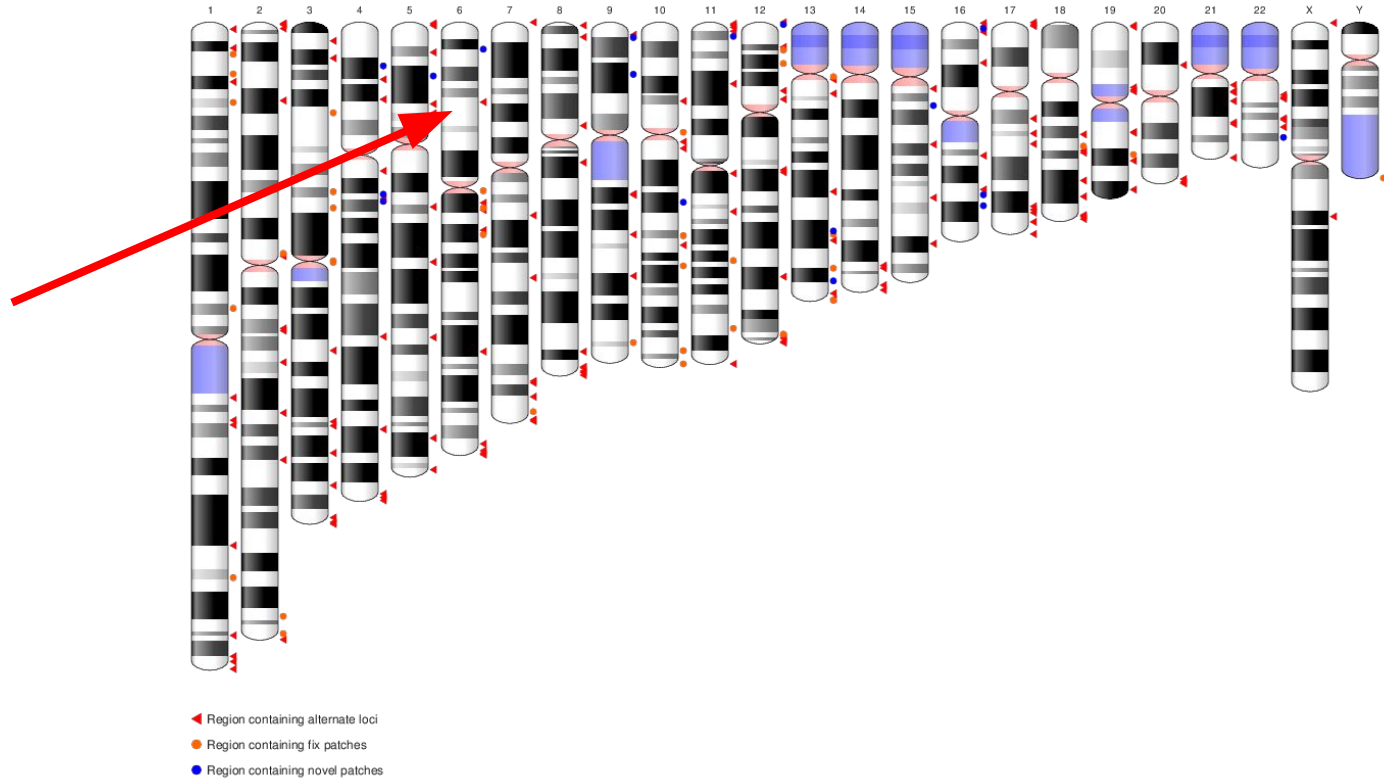# MHC Region

MHC Region

◄ Region containing alternate loci
● Region containing fix patches
● Region containing novel patches

MHC Region

◄ Region containing alternate loci

● Region containing fix patches

● Region containing novel patches

HLA Typing

An important step for matching organ donors and for genetic research

Computationally difficult, and often incomplete

A graph solution?