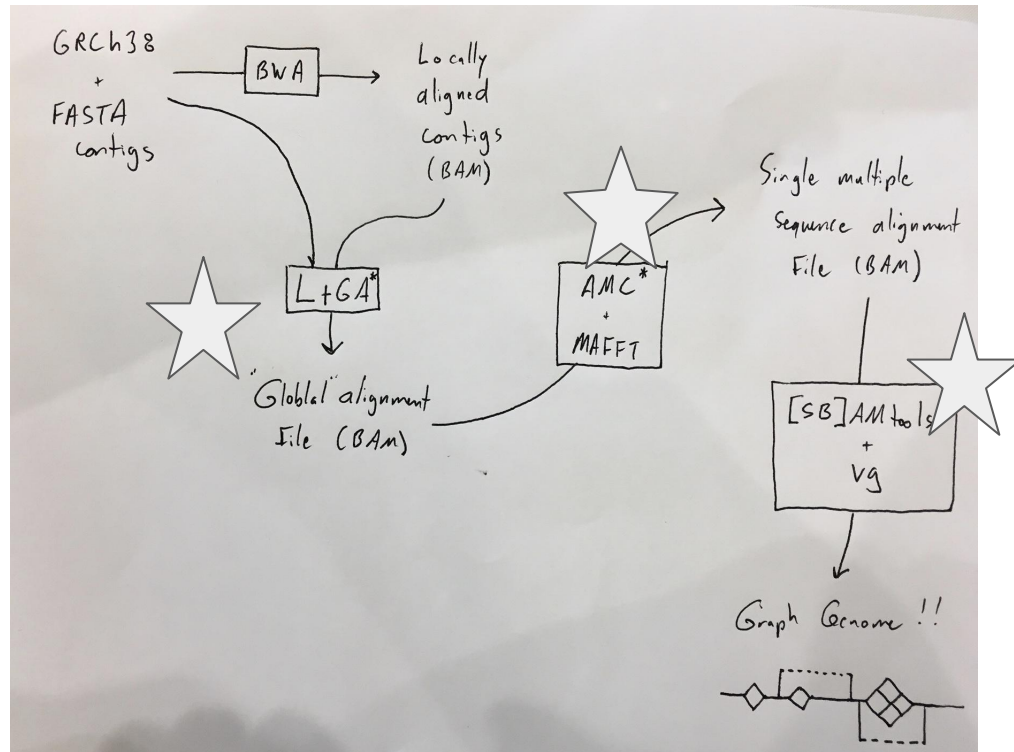# Whole genome graph creation from multiple assemblies

Whole genome graph creation from multiple assemblies

CSHL Graph Genome group

Steps: (1) Align all Contigs to GrCH38

1. Input:
    a. FASTA files (contigs)
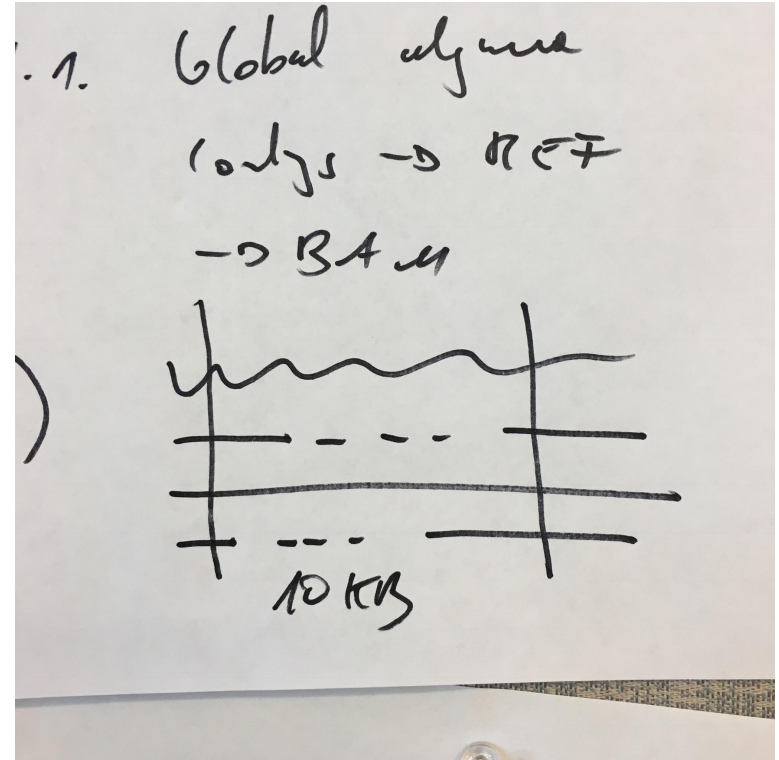   b. Reference FASTA (GRChr38)
2. Output:
    a. BAM
3. Tools:
    a. BWA

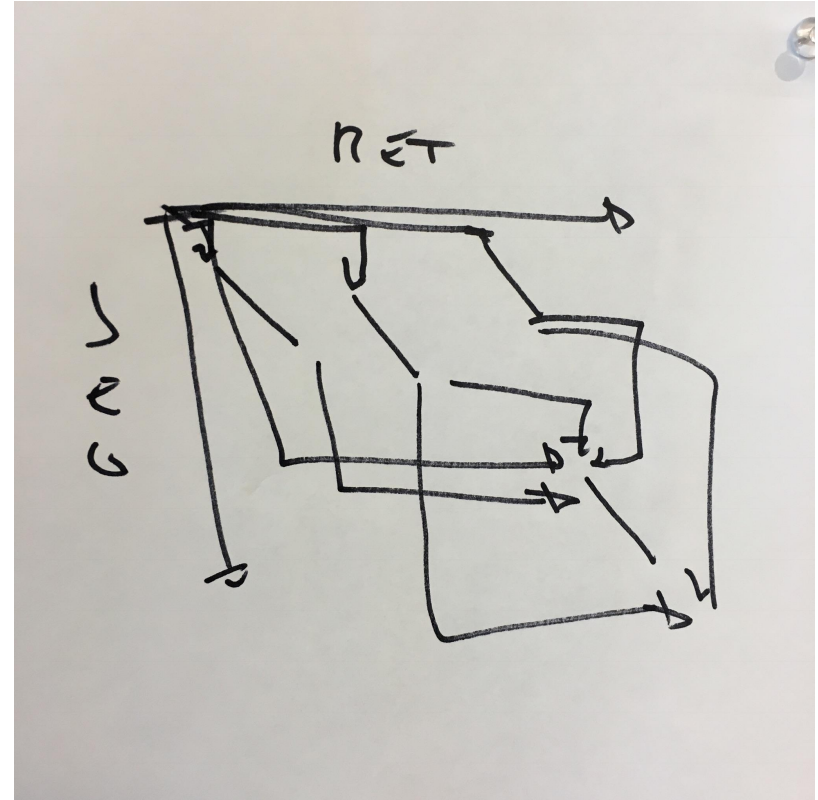# Steps: (1.1) Globally Align Contigs

1. Input:
   a. FASTA files (contigs)
   b. Reference FASTA (GRChr38)
   c. BAM
2. Output: BAM?
   a. each contig has single alignment
3. Tools: **Local to Global Alignment**

Steps: (2) Divide & Conquer MSA

1. Input:
    a. BAM
2. Output:
    a. BAM
3. Tools:
    a. MAFFT (alignment)
    b. **Some C++ code (chopping & reassembling)**

Steps: (2) Divide & Conquer MSA

- Start with large concatenation of "local" alignments in a BAM file.
- Use AMC? code to identify windows, excise each window, do multiple sequence alignments on each window, use AMC to re-assemble MAFFT into a single BAM. [file conversion between FASTA / MAFFT]

Steps: (3) Export

1. Input:
   a. BAM
2. Output:
   a. BAM, VCF, VG
3. Tools:
   a. Vg
   b. Other wrappers