

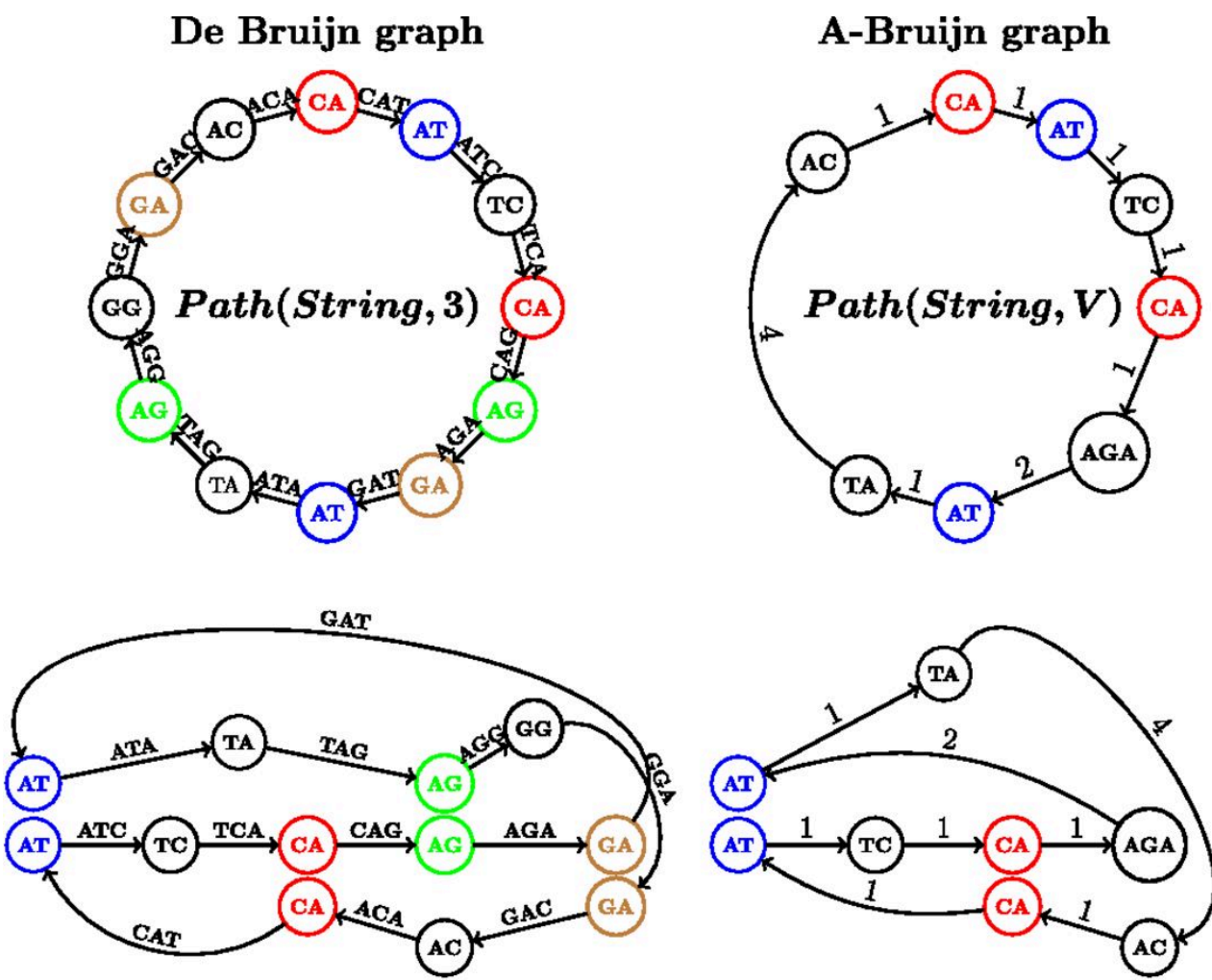
Genome Assembly and Annotation

Jason Stajich
Univ of California, Riverside

Genome Assembly

Assembly approaches

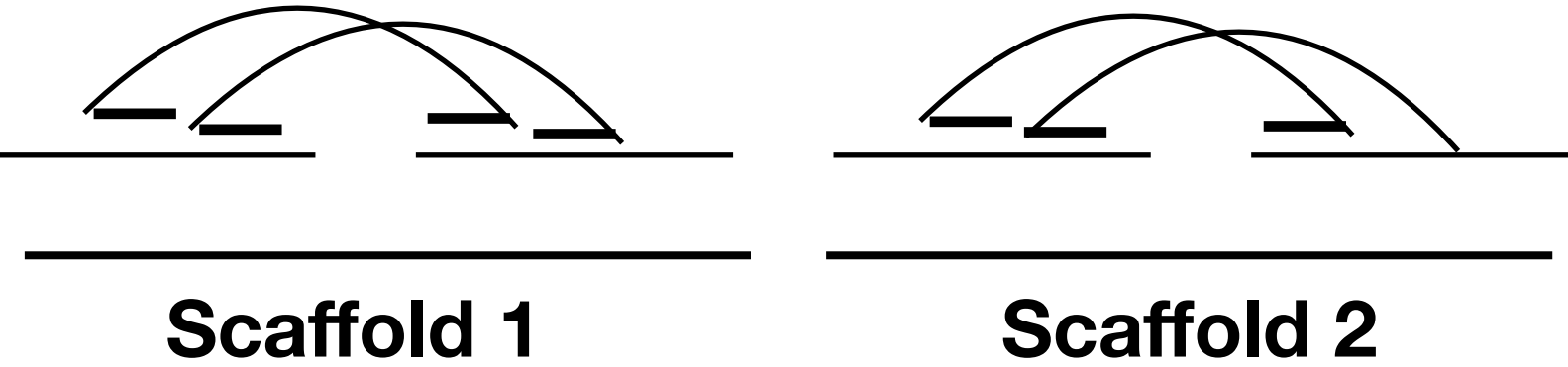
De Bruijn Graphs



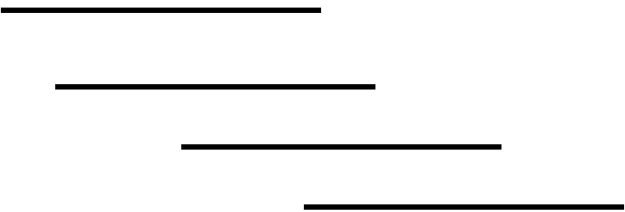
Contigs



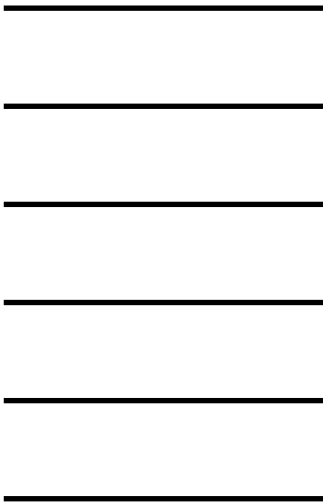
Scaffolds



Overlap Consensus



Raw reads



Annotation I

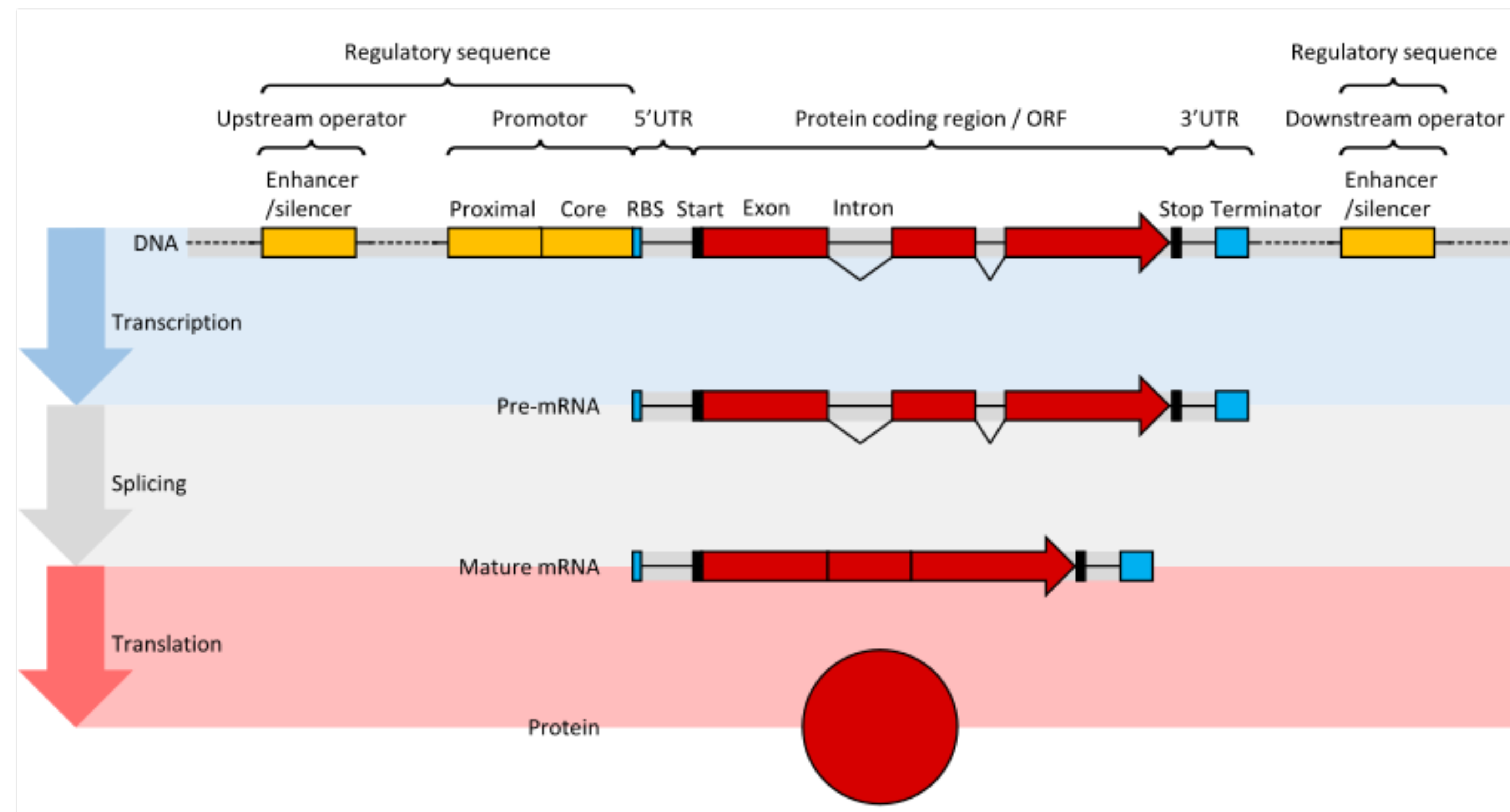
- funannotate mask - Identify repetitive sequences
 - RepeatMasker, tantan
- *funannotate train* - *Train gene predictors*
 - Align mRNA transcripts to genome to identify exon/gene regions exonerate;
 - Align RNA-Seq reads to genome, assemble transcripts with Trinity
 - Refinement with PASA (Program to Assemble Spliced Alignments)
 - Train *ab initio* gene predictor from these spliced models

Annotation II

- funannotate predict
 - Align protein (and mRNA Transcripts) to genome as evidence to support predictions; refine alignment with exonerate to spliced exons
 - Gene prediction (augustus, SNAP, genemark, glimmerhmm)
 - Combine predictions with EVM

Annotation III

- *funannotate update* - refine gene models with mRNA
- Run PASA with models to extend Untranslated regions
- Alternative Splicing



Annotation IV

- funannotate annotate - align to databases to add functional info
 - BLAST against uniprot, swissprot, MEROPs database
 - HMM searches against Pfam, CAZyDB
 - Incorporate InterPro, AntiSMASH (secondary metabolite prediction)
 - Produces annotated genome files ready to upload to NCBI