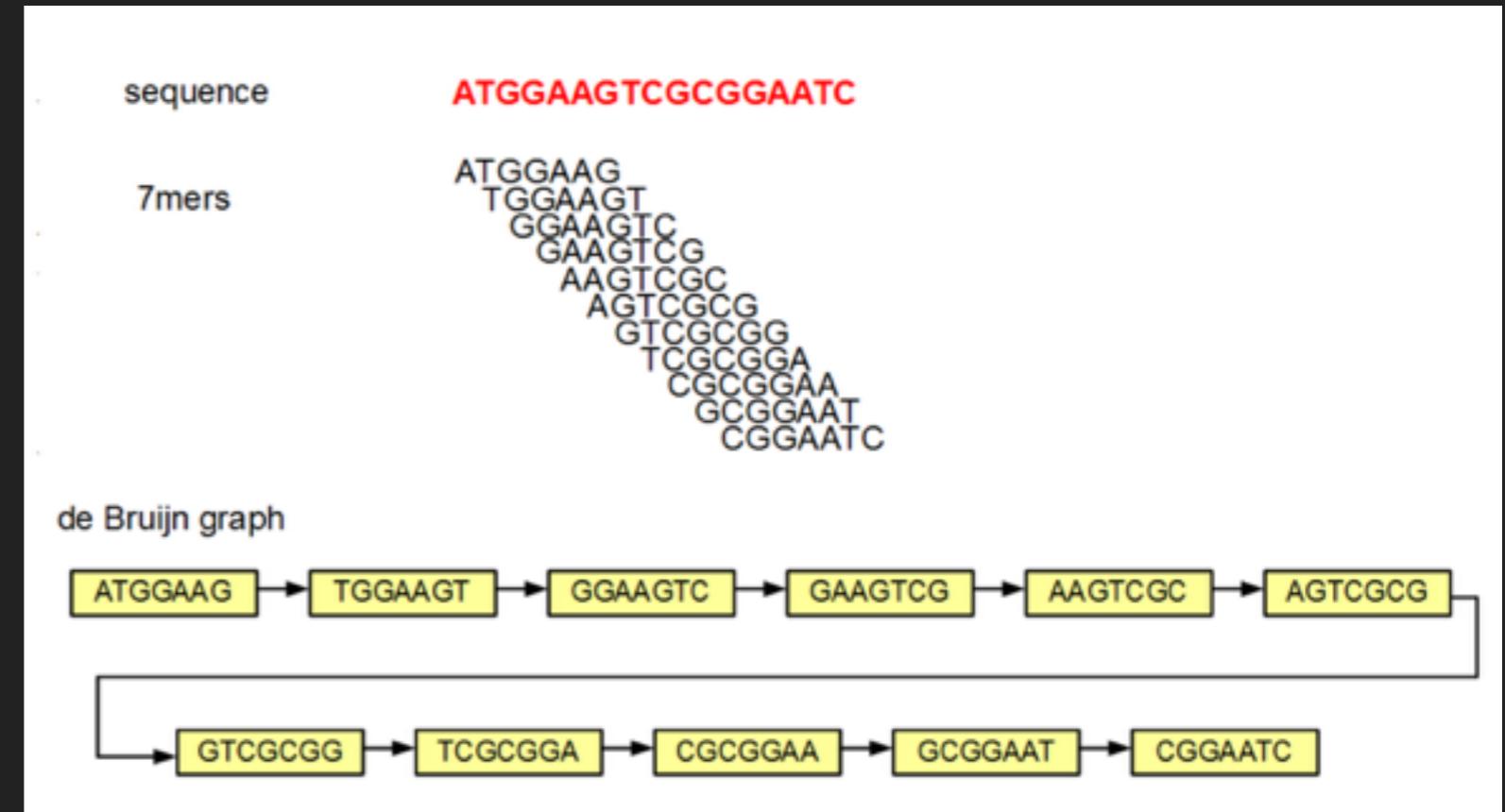

DE NOVO TRANSCRIPTOME ASSEMBLY WITH TRINITY

DE NOVO TRANSCRIPTOME ASSEMBLY

- ▶ Trinity is a de bruijn graph assembler
- ▶ Trinity is written for use on short Illumina reads
- ▶ Short reads are broken into even shorter kmers
- ▶ kmers are constructed into a graph



TRINITY

- ▶ Paper: “Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data”

Nature Biotechnology doi:
10.1038/nbt.1883

- ▶ Trinity takes place in three phases:

- ▶ Inchworm
- ▶ Chrysalis
- ▶ Butterfly



HOW IS TRANSCRIPTOME ASSEMBLY DIFFERENT FROM GENOME ASSEMBLY? (E.G. WHY USE A PURPOSE-BUILT TRANSCRIPTOME ASSEMBLER?)

Genome Assembly

- Uniform coverage
- Single contig per locus
- Double-stranded

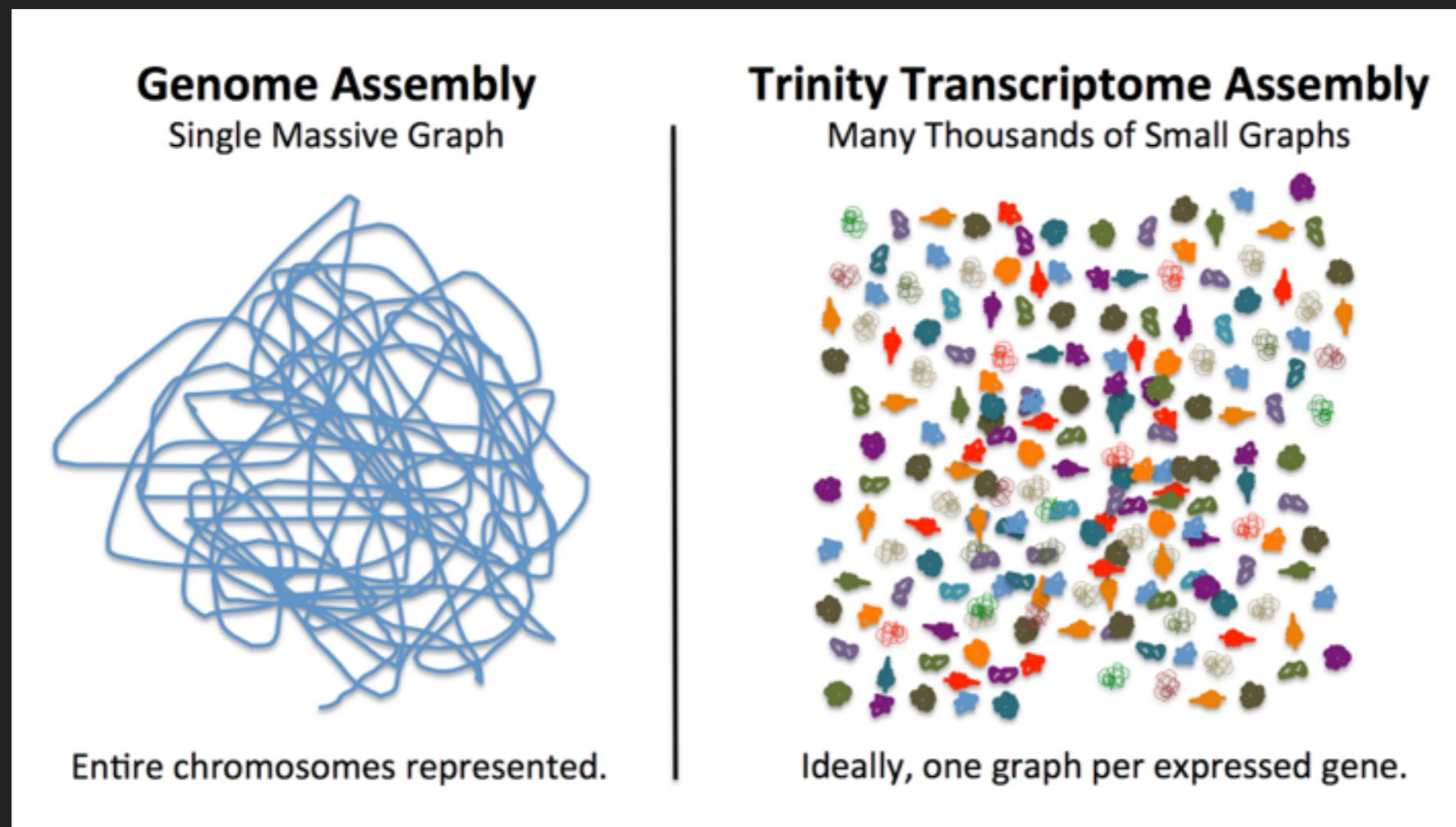
Transcriptome Assembly

- Exponentially distributed coverage levels
- Multiple contigs per locus (alt splicing)
- Strand-specific



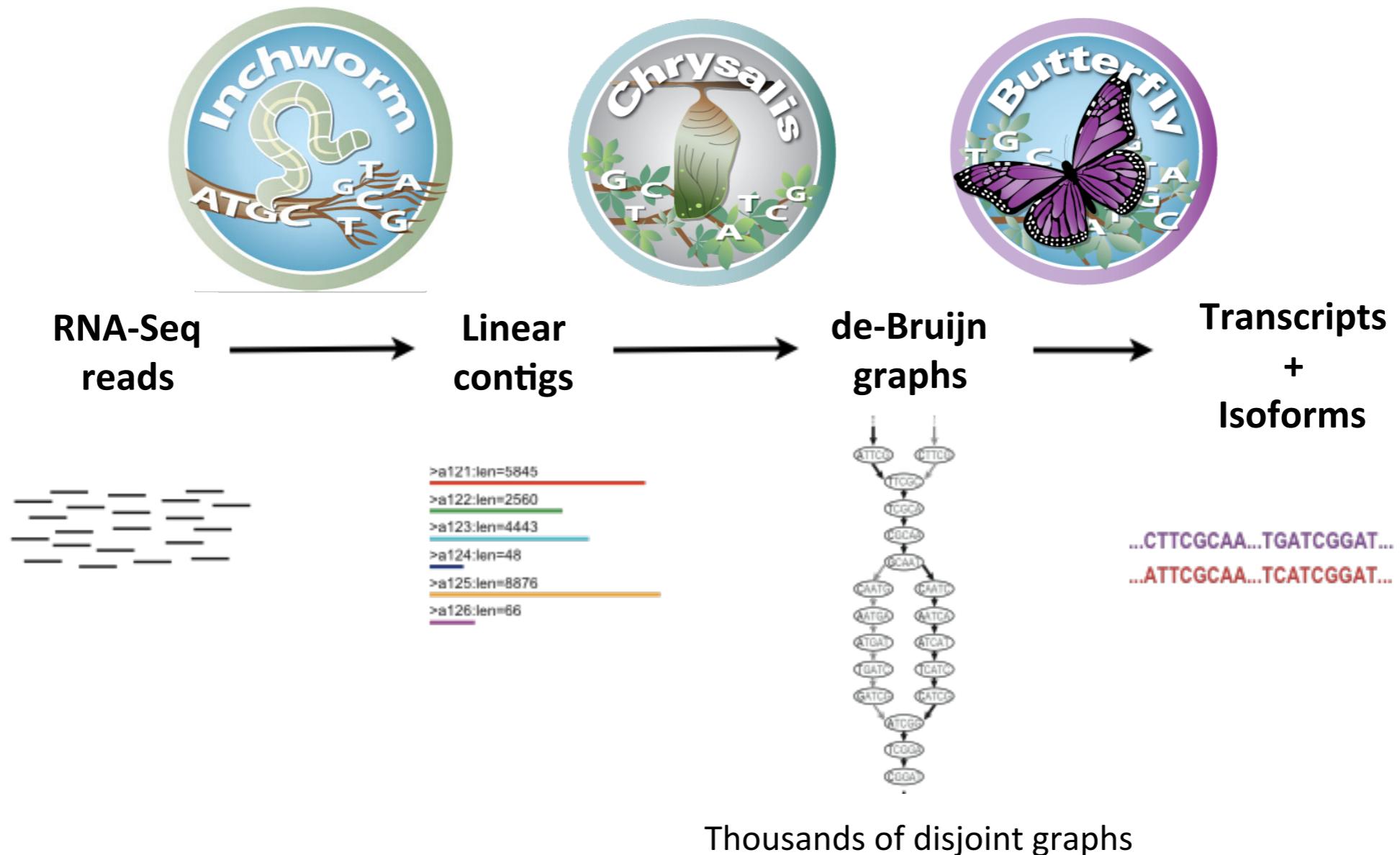
Trinity slides from: https://github.com/trinityrnaseq/RNASeq_Trinity_Tuxedo_Workshop/blob/master/docs/rnaseq_workshop_slides.pdf

HOW IS TRANSCRIPTOME ASSEMBLY DIFFERENT FROM GENOME ASSEMBLY? (E.G. WHY USE A PURPOSE-BUILT TRANSCRIPTOME ASSEMBLER?)



DE NOVO TRANSCRIPTOME ASSEMBLY WITH TRINITY

Trinity – How it works:



Trinity slides from: https://github.com/trinityrnaseq/RNASeq_Trinity_Tuxedo_Workshop/blob/master/docs/rnaseq_workshop_slides.pdf

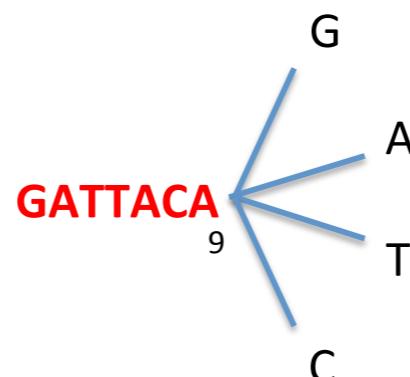


Inchworm Algorithm

Decompose all reads into overlapping Kmers (25-mers)

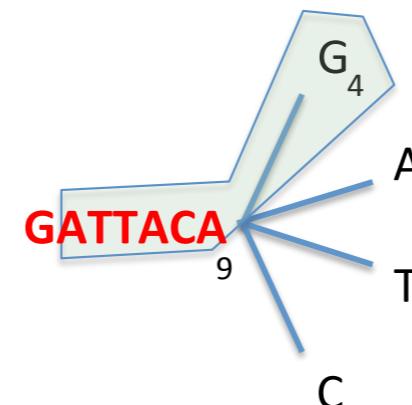
Identify seed kmer as most abundant Kmer, ignoring low-complexity kmers.

Extend kmer at 3' end, guided by coverage.





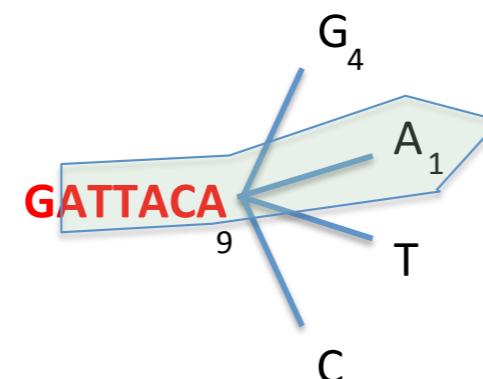
Inchworm Algorithm



Trinity slides from: https://github.com/trinityrnaseq/RNASeq_Trinity_Tuxedo_Workshop/blob/master/docs/rnaseq_workshop_slides.pdf



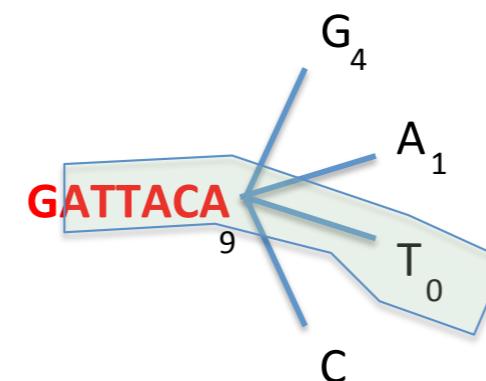
Inchworm Algorithm



Trinity slides from: https://github.com/trinityrnaseq/RNASeq_Trinity_Tuxedo_Workshop/blob/master/docs/rnaseq_workshop_slides.pdf



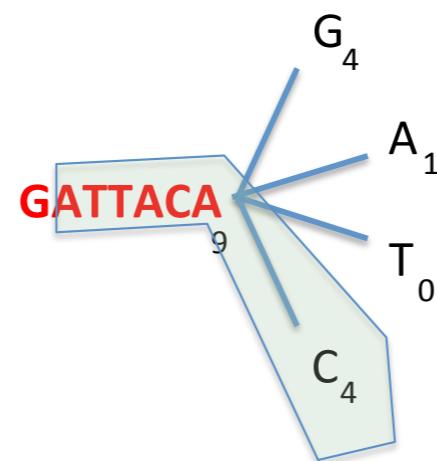
Inchworm Algorithm



Trinity slides from: https://github.com/trinityrnaseq/RNASeq_Trinity_Tuxedo_Workshop/blob/master/docs/rnaseq_workshop_slides.pdf



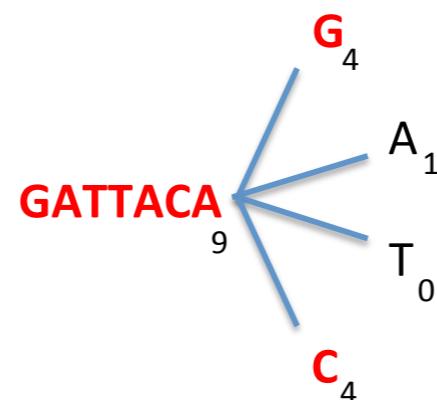
Inchworm Algorithm



Trinity slides from: https://github.com/trinityrnaseq/RNASeq_Trinity_Tuxedo_Workshop/blob/master/docs/rnaseq_workshop_slides.pdf



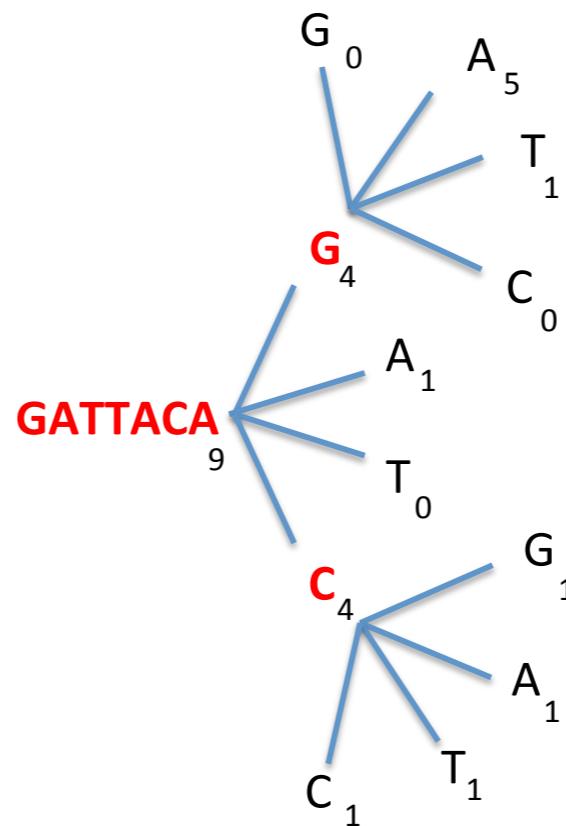
Inchworm Algorithm



Trinity slides from: https://github.com/trinityrnaseq/RNASeq_Trinity_Tuxedo_Workshop/blob/master/docs/rnaseq_workshop_slides.pdf



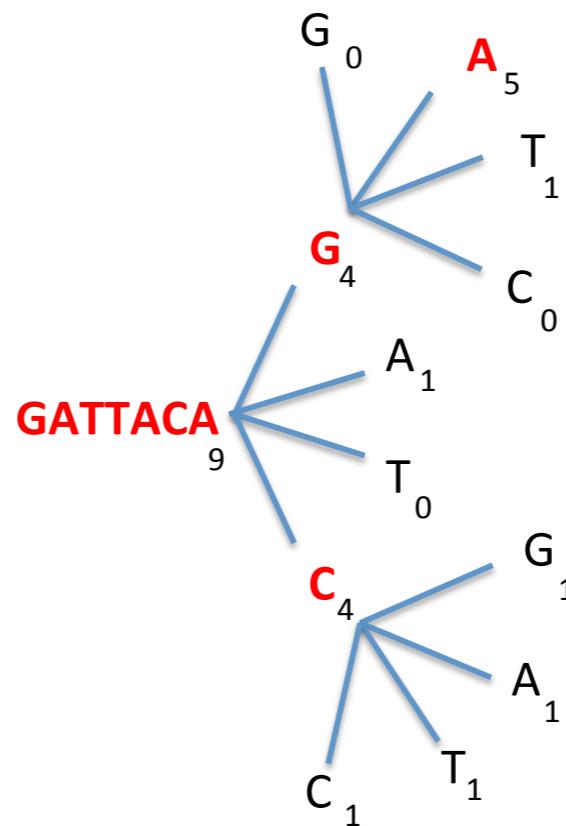
Inchworm Algorithm



Trinity slides from: https://github.com/trinityrnaseq/RNASeq_Trinity_Tuxedo_Workshop/blob/master/docs/rnaseq_workshop_slides.pdf



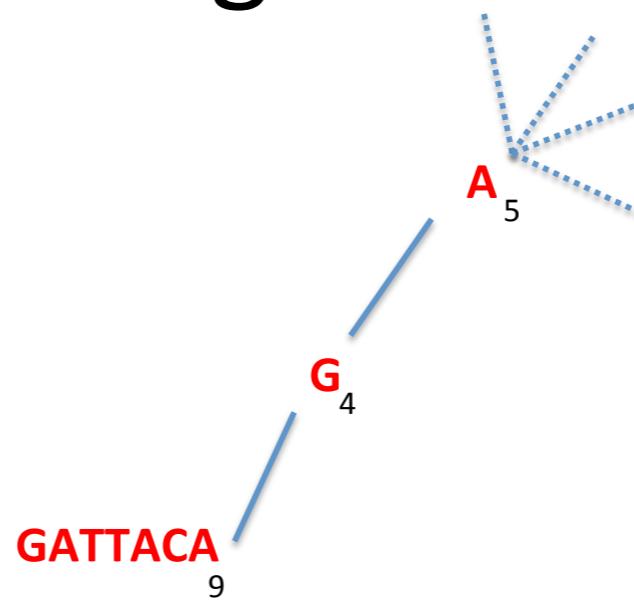
Inchworm Algorithm



Trinity slides from: https://github.com/trinityrnaseq/RNASeq_Trinity_Tuxedo_Workshop/blob/master/docs/rnaseq_workshop_slides.pdf



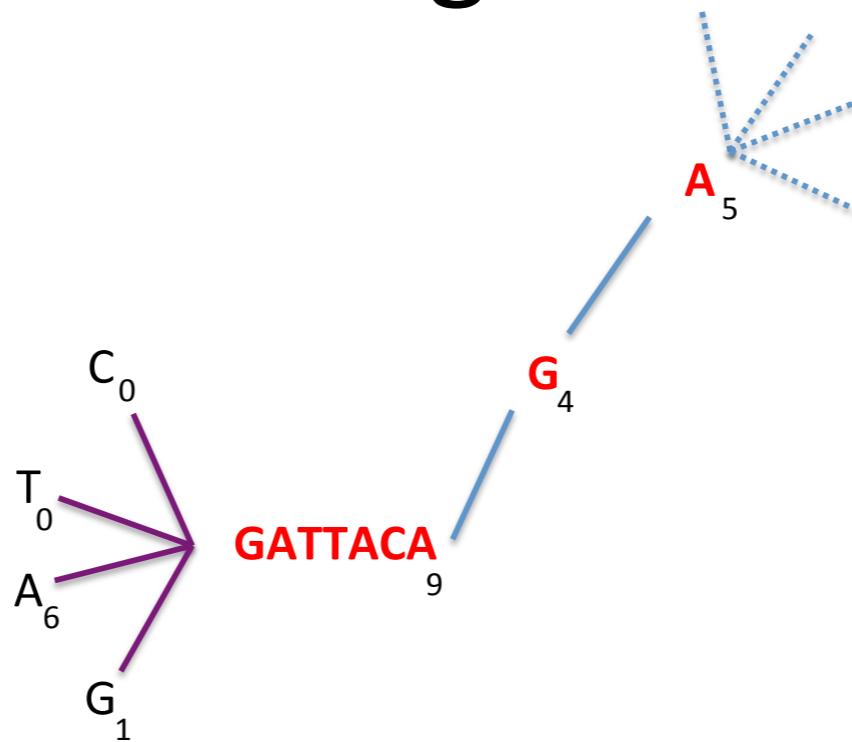
Inchworm Algorithm



Trinity slides from: https://github.com/trinityrnaseq/RNASeq_Trinity_Tuxedo_Workshop/blob/master/docs/rnaseq_workshop_slides.pdf



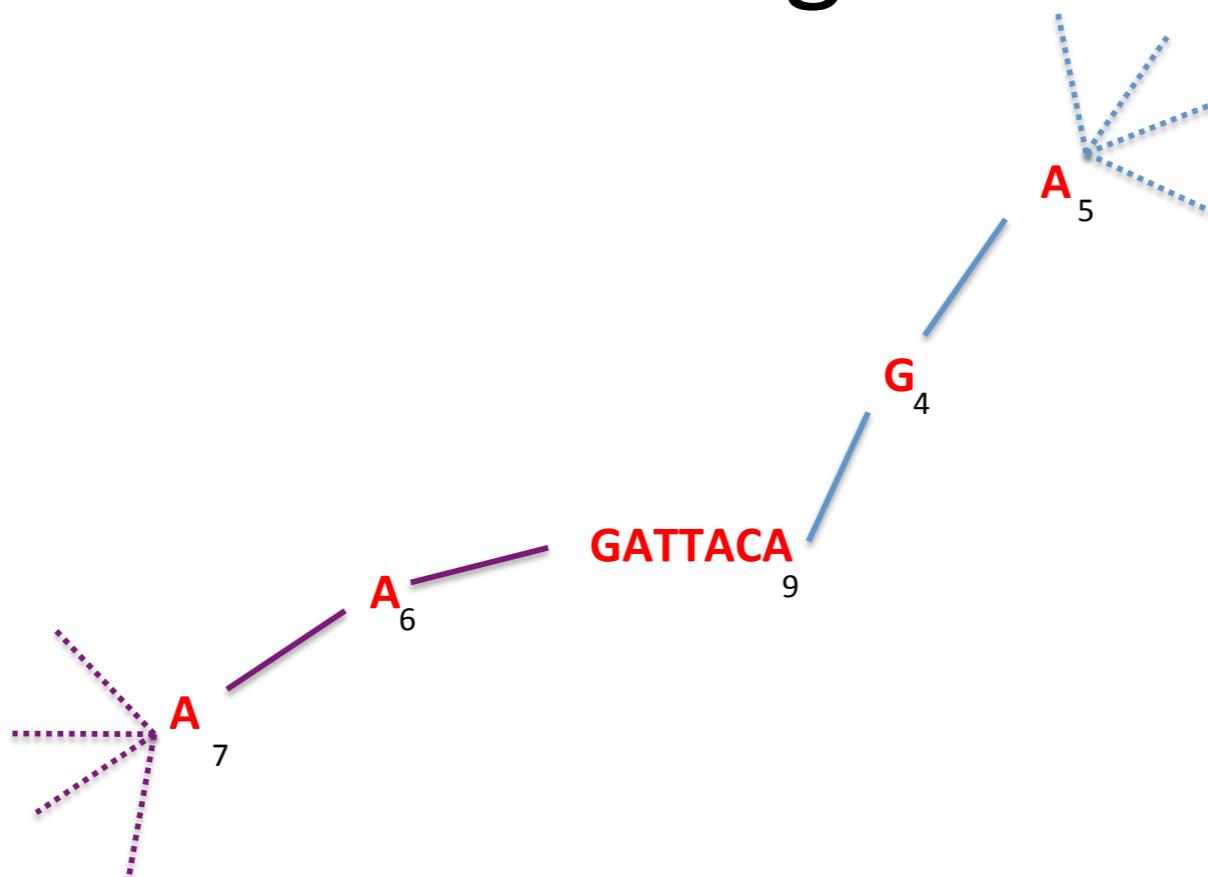
Inchworm Algorithm



Trinity slides from: https://github.com/trinityrnaseq/RNASeq_Trinity_Tuxedo_Workshop/blob/master/docs/rnaseq_workshop_slides.pdf



Inchworm Algorithm



Report contig:**AAGATTACAGA**....

Remove assembled kmers from catalog, then repeat the entire process.

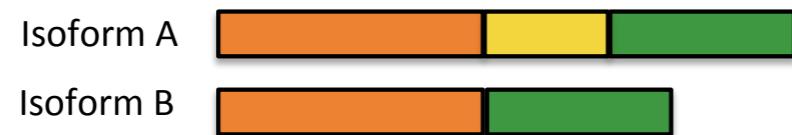
Trinity slides from: https://github.com/trinityrnaseq/RNASeq_Trinity_Tuxedo_Workshop/blob/master/docs/rnaseq_workshop_slides.pdf

DE NOVO TRANSCRIPTOME ASSEMBLY WITH TRINITY



Inchworm Contigs from Alt-Spliced Transcripts

Expressed isoforms



Trinity slides from: https://github.com/trinityrnaseq/RNASeq_Trinity_Tuxedo_Workshop/blob/master/docs/rnaseq_workshop_slides.pdf

DE NOVO TRANSCRIPTOME ASSEMBLY WITH TRINITY



Inchworm Contigs from Alt-Spliced Transcripts

Expressed isoforms

Isoform A

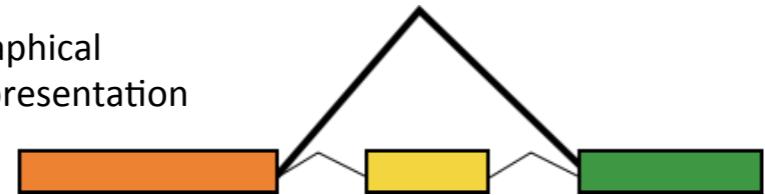
Isoform B

Expression

(low)

(high)

Graphical representation

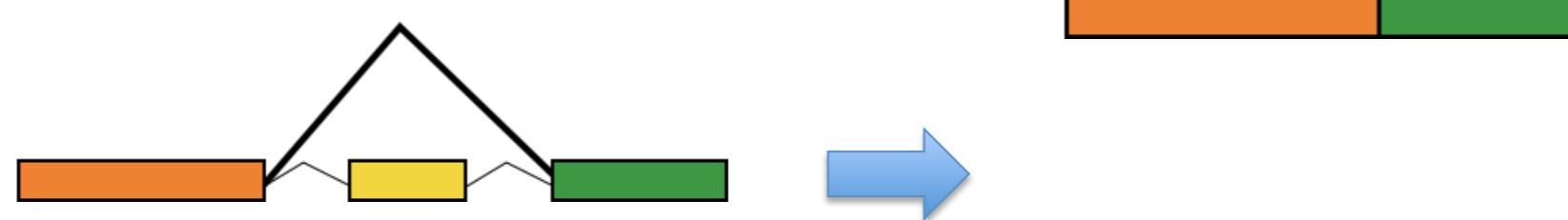


Trinity slides from: https://github.com/trinityrnaseq/RNASeq_Trinity_Tuxedo_Workshop/blob/master/docs/rnaseq_workshop_slides.pdf

DE NOVO TRANSCRIPTOME ASSEMBLY WITH TRINITY



Inchworm Contigs from Alt-Spliced Transcripts

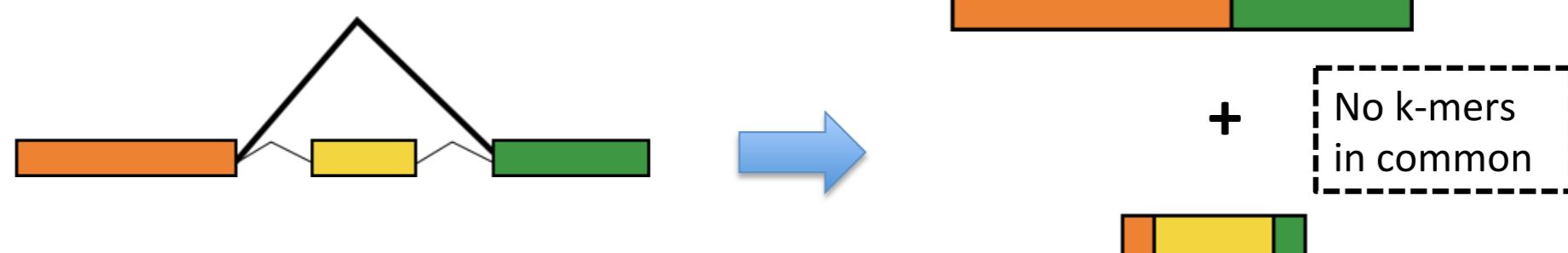


Trinity slides from: https://github.com/trinityrnaseq/RNASeq_Trinity_Tuxedo_Workshop/blob/master/docs/rnaseq_workshop_slides.pdf

DE NOVO TRANSCRIPTOME ASSEMBLY WITH TRINITY



Inchworm Contigs from Alt-Spliced Transcripts

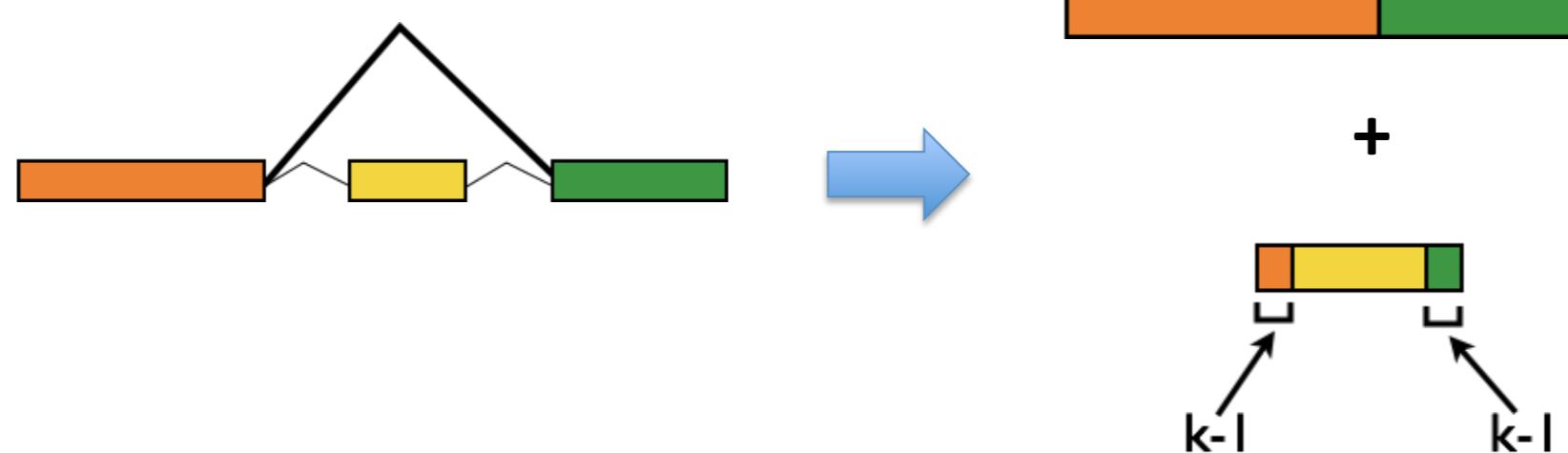


Trinity slides from: https://github.com/trinityrnaseq/RNASeq_Trinity_Tuxedo_Workshop/blob/master/docs/rnaseq_workshop_slides.pdf

DE NOVO TRANSCRIPTOME ASSEMBLY WITH TRINITY

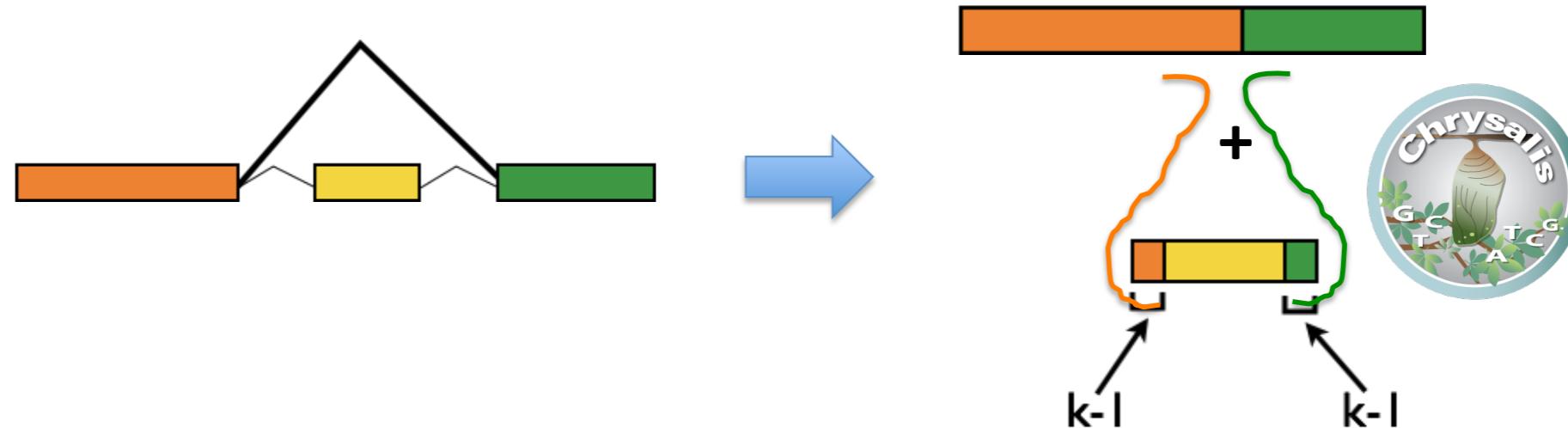


Inchworm Contigs from Alt-Spliced Transcripts



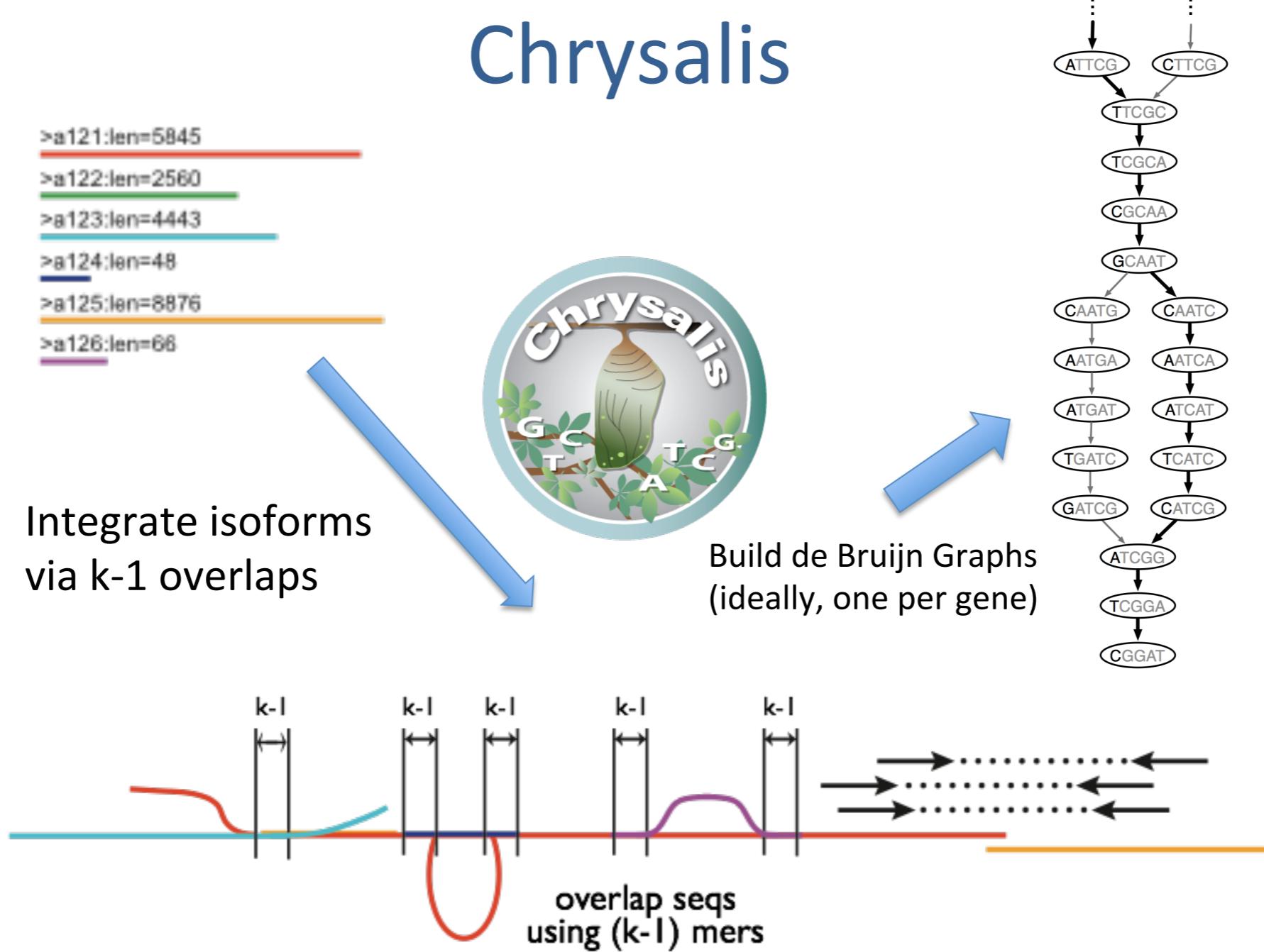
Trinity slides from: https://github.com/trinityrnaseq/RNASeq_Trinity_Tuxedo_Workshop/blob/master/docs/rnaseq_workshop_slides.pdf

Chrysalis Re-groups Related Inchworm Contigs



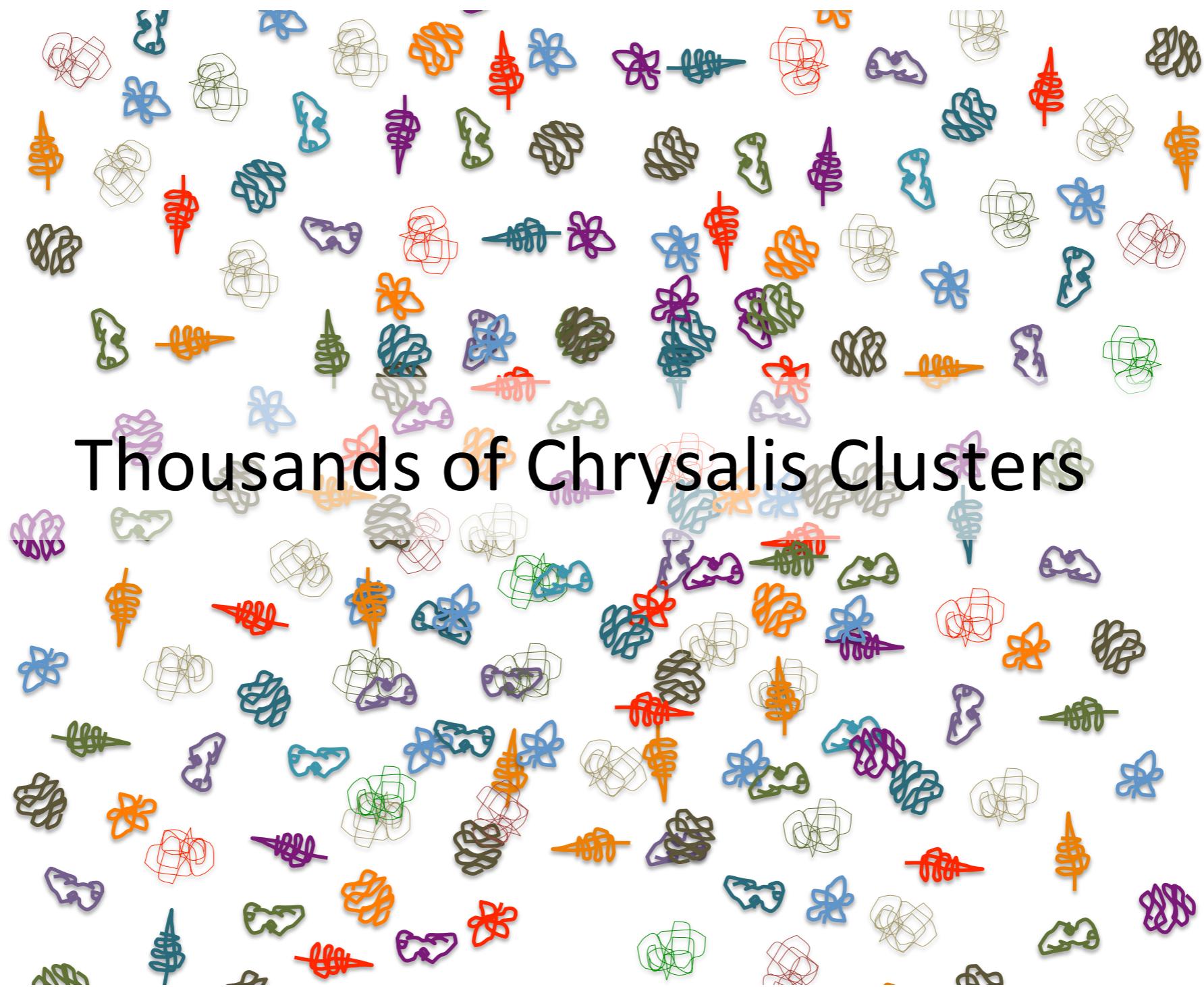
Chrysalis uses $(k-1)$ overlaps and read support to link related Inchworm contigs

DE NOVO TRANSCRIPTOME ASSEMBLY WITH TRINITY



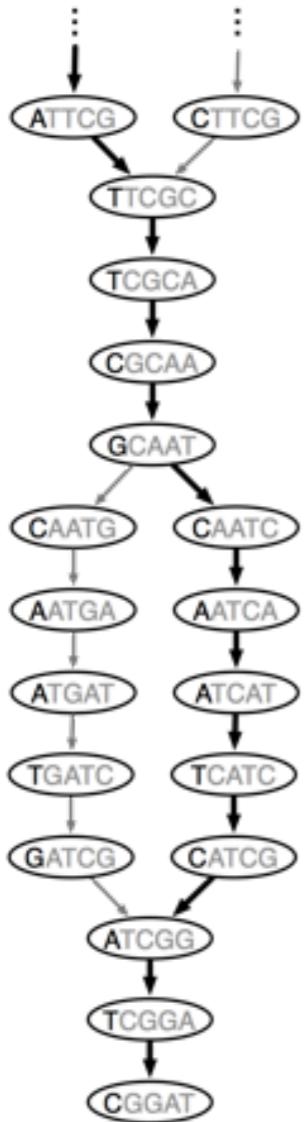
Trinity slides from: https://github.com/trinityrnaseq/RNASeq_Trinity_Tuxedo_Workshop/blob/master/docs/rnaseq_workshop_slides.pdf

DE NOVO TRANSCRIPTOME ASSEMBLY WITH TRINITY



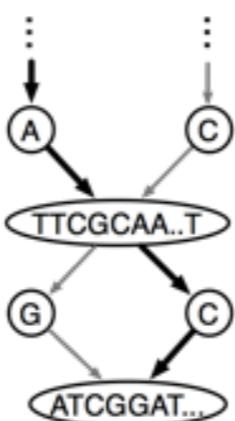
Trinity slides from: https://github.com/trinityrnaseq/RNASeq_Trinity_Tuxedo_Workshop/blob/master/docs/rnaseq_workshop_slides.pdf

DE NOVO TRANSCRIPTOME ASSEMBLY WITH TRINITY



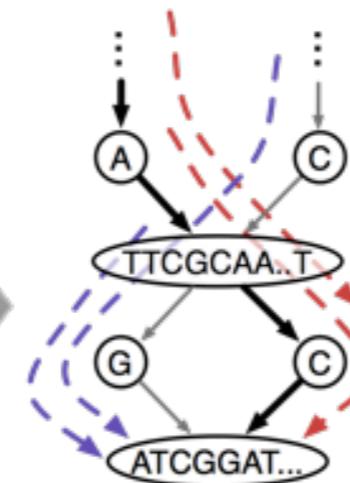
de Bruijn
graph

compacting



compact
graph

finding paths



compact
graph with
reads

extracting
sequences
sequences
(isoforms and paralogs)

..CTTCGCAA..TGATCGGAT...
..ATTCGCAA..TCATCGGAT...



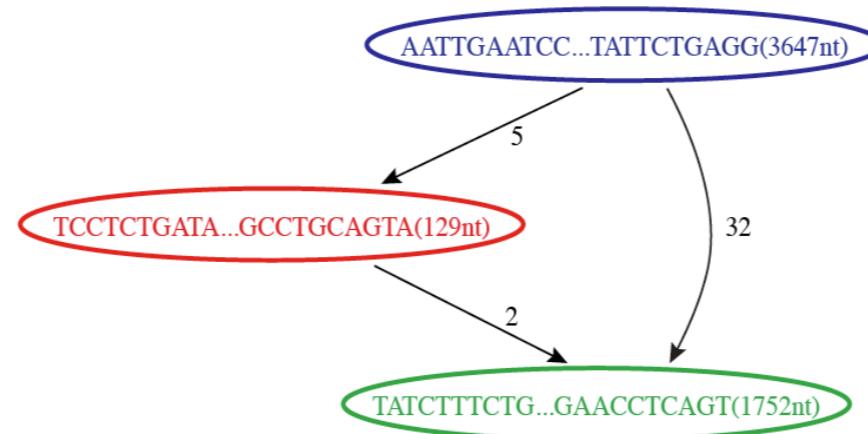
Butterfly

Trinity slides from: https://github.com/trinityrnaseq/RNASeq_Trinity_Tuxedo_Workshop/blob/master/docs/rnaseq_workshop_slides.pdf

DE NOVO TRANSCRIPTOME ASSEMBLY WITH TRINITY

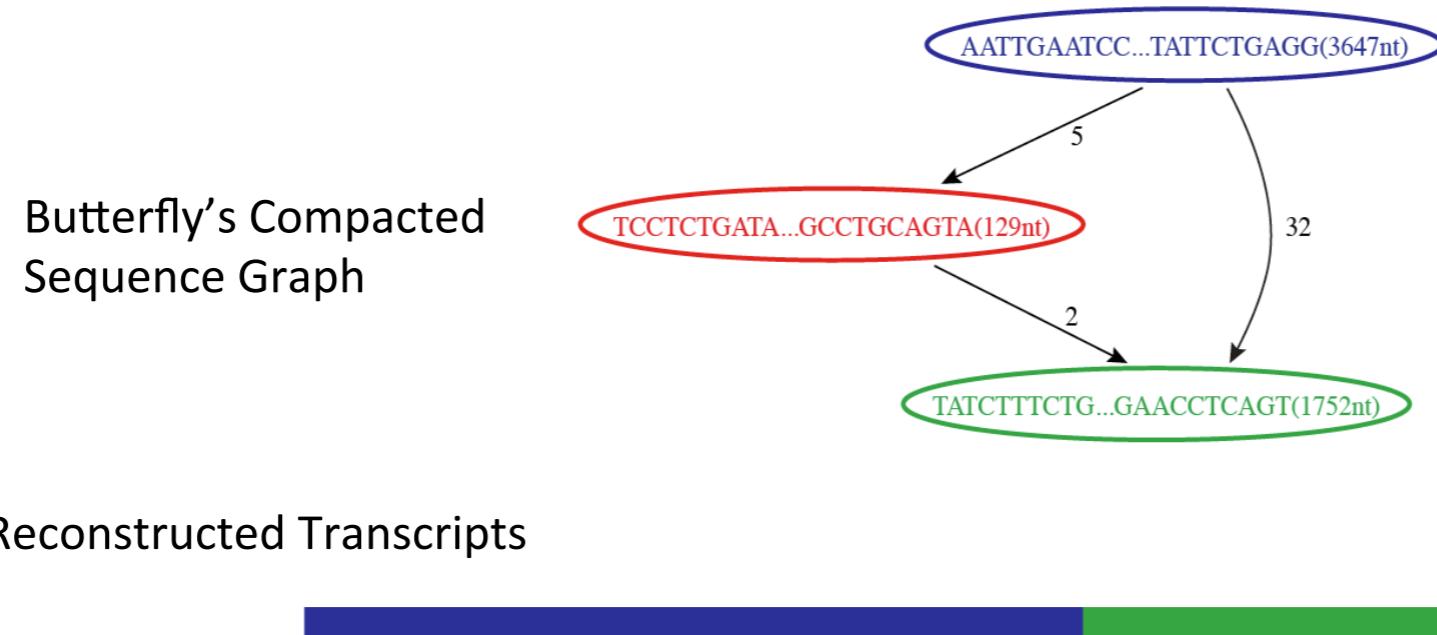
Butterfly Example 1: Reconstruction of Alternatively Spliced Transcripts

Butterfly's Compacted
Sequence Graph



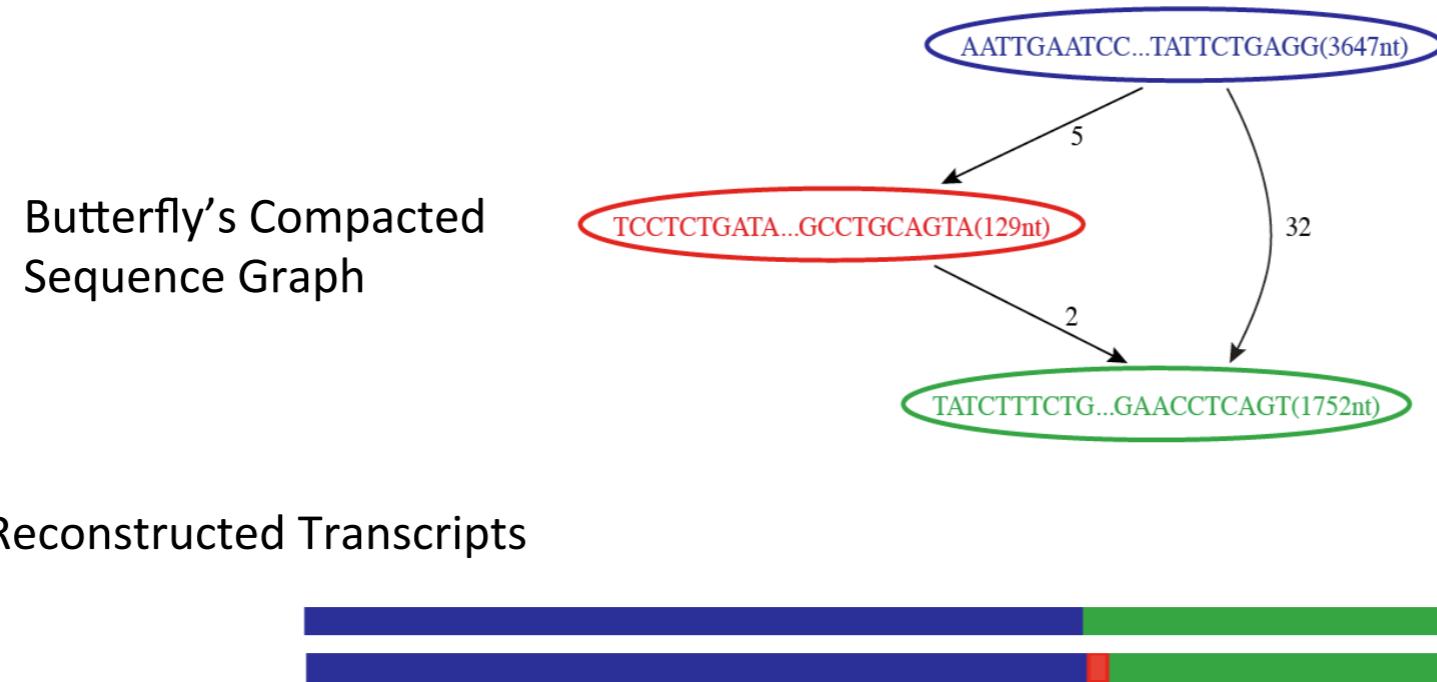
Trinity slides from: https://github.com/trinityrnaseq/RNASeq_Trinity_Tuxedo_Workshop/blob/master/docs/rnaseq_workshop_slides.pdf

Reconstruction of Alternatively Spliced Transcripts



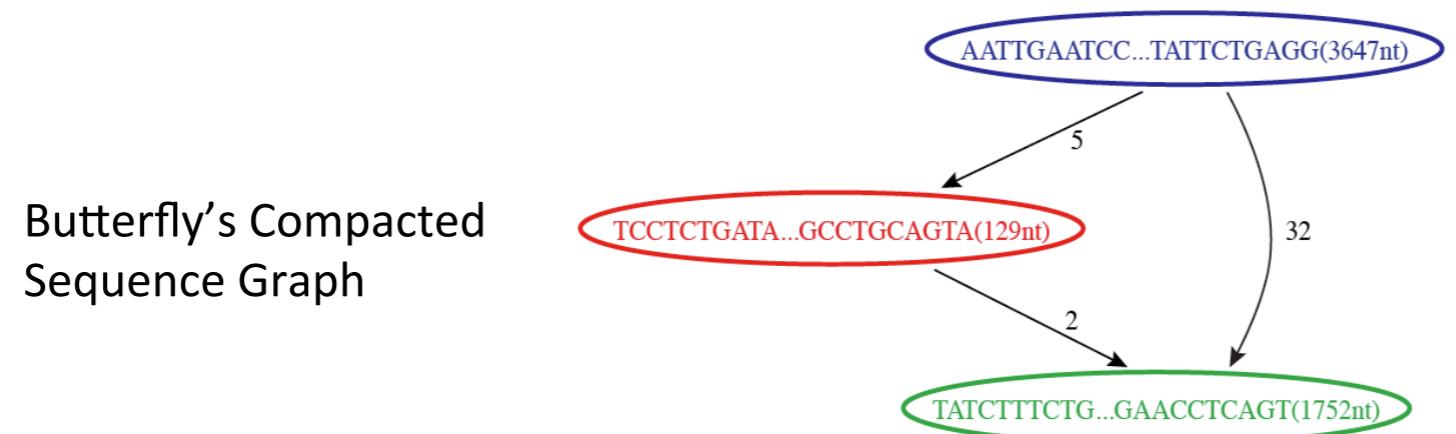
Trinity slides from: https://github.com/trinityrnaseq/RNASeq_Trinity_Tuxedo_Workshop/blob/master/docs/rnaseq_workshop_slides.pdf

Reconstruction of Alternatively Spliced Transcripts



Trinity slides from: https://github.com/trinityrnaseq/RNASeq_Trinity_Tuxedo_Workshop/blob/master/docs/rnaseq_workshop_slides.pdf

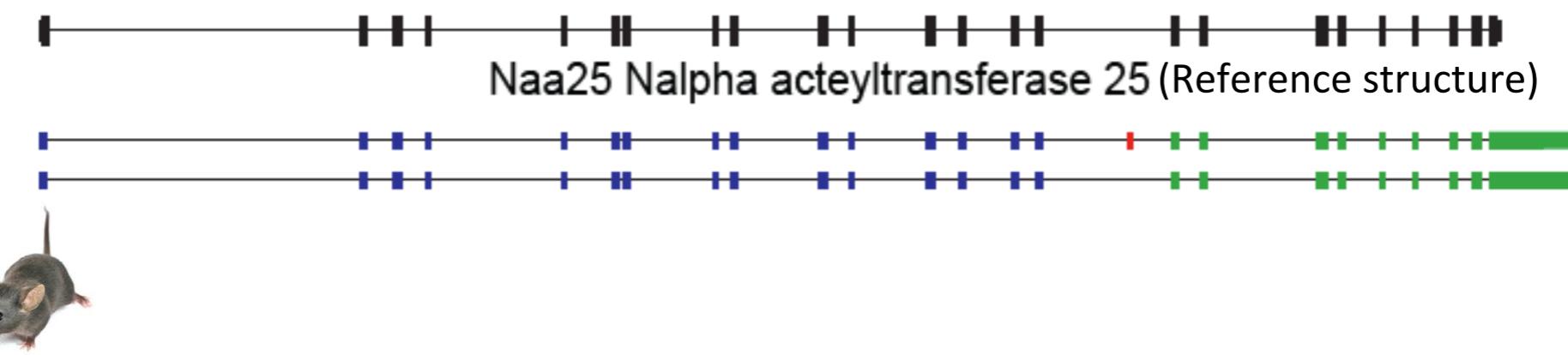
Reconstruction of Alternatively Spliced Transcripts



Reconstructed Transcripts

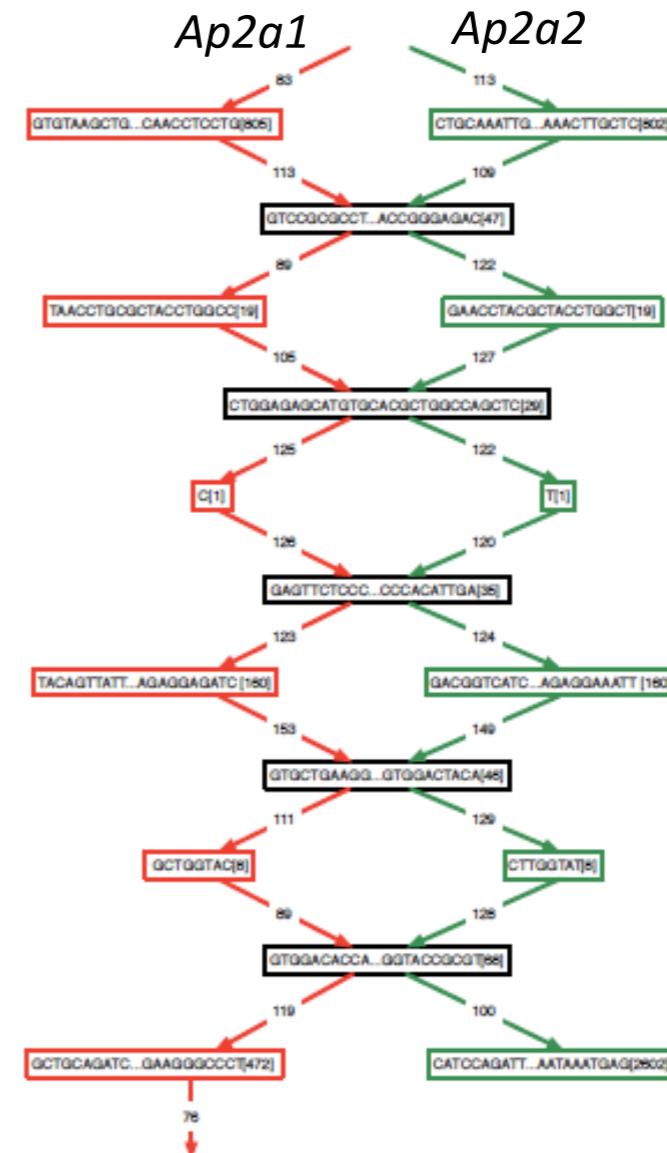


Aligned to Mouse Genome



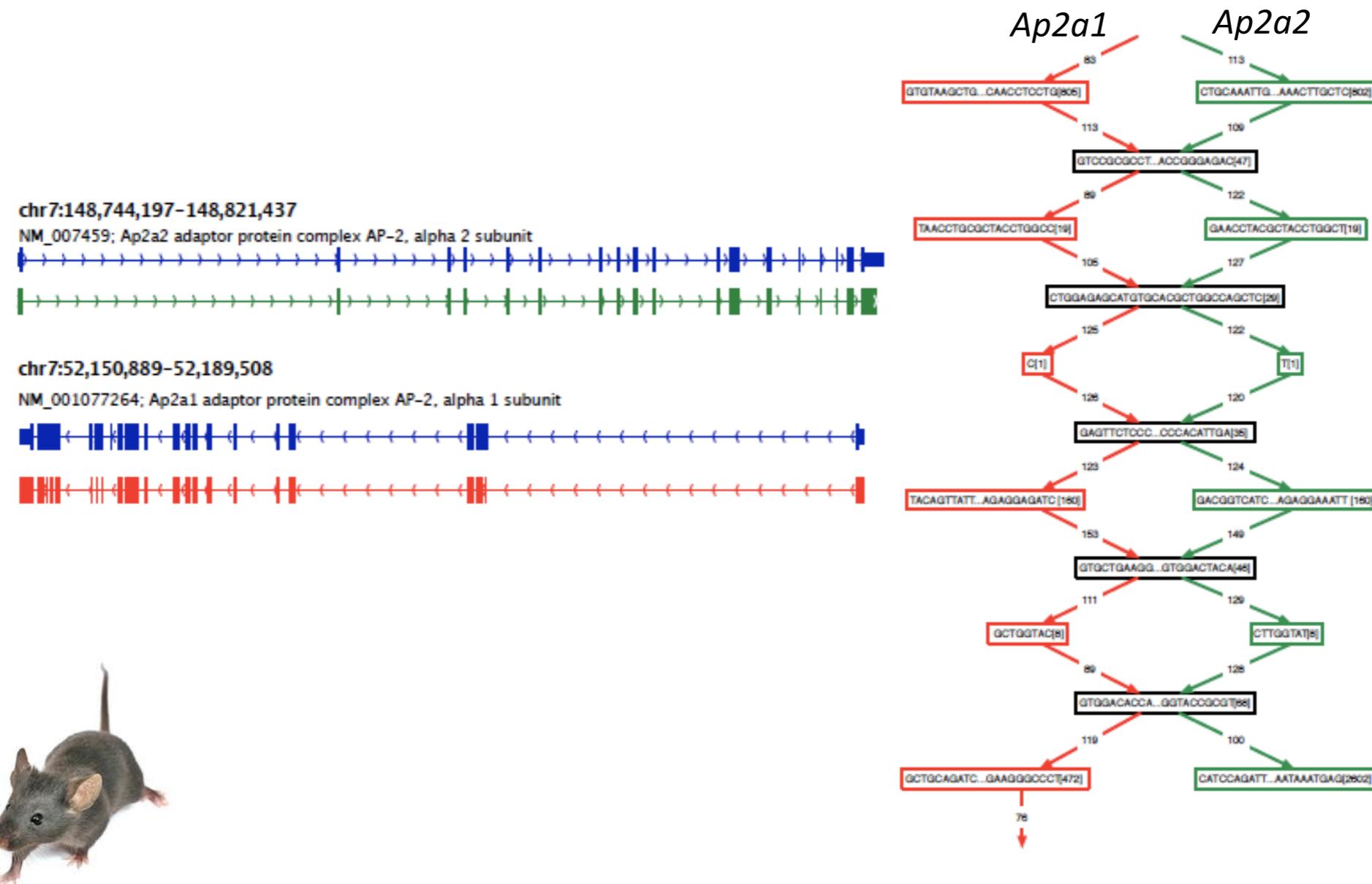
Trinity slides from: https://github.com/trinityrnaseq/RNASeq_Trinity_Tuxedo_Workshop/blob/master/docs/rnaseq_workshop_slides.pdf

Butterfly Example 2: Teasing Apart Transcripts of Paralogous Genes



Trinity slides from: https://github.com/trinityrnaseq/RNASeq_Trinity_Tuxedo_Workshop/blob/master/docs/rnaseq_workshop_slides.pdf

Teasing Apart Transcripts of Paralogous Genes



Trinity slides from: https://github.com/trinityrnaseq/RNASeq_Trinity_Tuxedo_Workshop/blob/master/docs/rnaseq_workshop_slides.pdf

DE NOVO TRANSCRIPTOME ASSEMBLY WITH TRINITY

Trinity output: A multi-fasta file

Trinity slides from: https://github.com/trinityrnaseq/RNASeq_Trinity_Tuxedo_Workshop/blob/master/docs/rnaseq_workshop_slides.pdf

HANDS-ON

- ▶ Go to [https://github.com/
SmithsonianWorkshops/RNA-seq](https://github.com/SmithsonianWorkshops/RNA-seq)
- ▶ Click on "3a_De novo assembly
with Trinity.md"

