```
   ___                       __   __      ______     _      _ __
  / _ | ___ ___ ___ __ _   / /  / /__   /_  __/___(_)__  (_) /___ __
 / __ |(_-<(_-</ -_)  ' \/ _ \/ / -_)   / / / __/ / _ \/ / __/ // /
/_/ |_/___/___/\__/_/_/_/_.__/_/\__/   /_/ /_/ /_/_//_/_/\__/\_, /
                                                            /___/
```

Trinity On TACC For Everyone

Benni Goetz
CCBB Bioinformatics Consultant
Oct 9, 2014

(Joint work with John Hawkins)

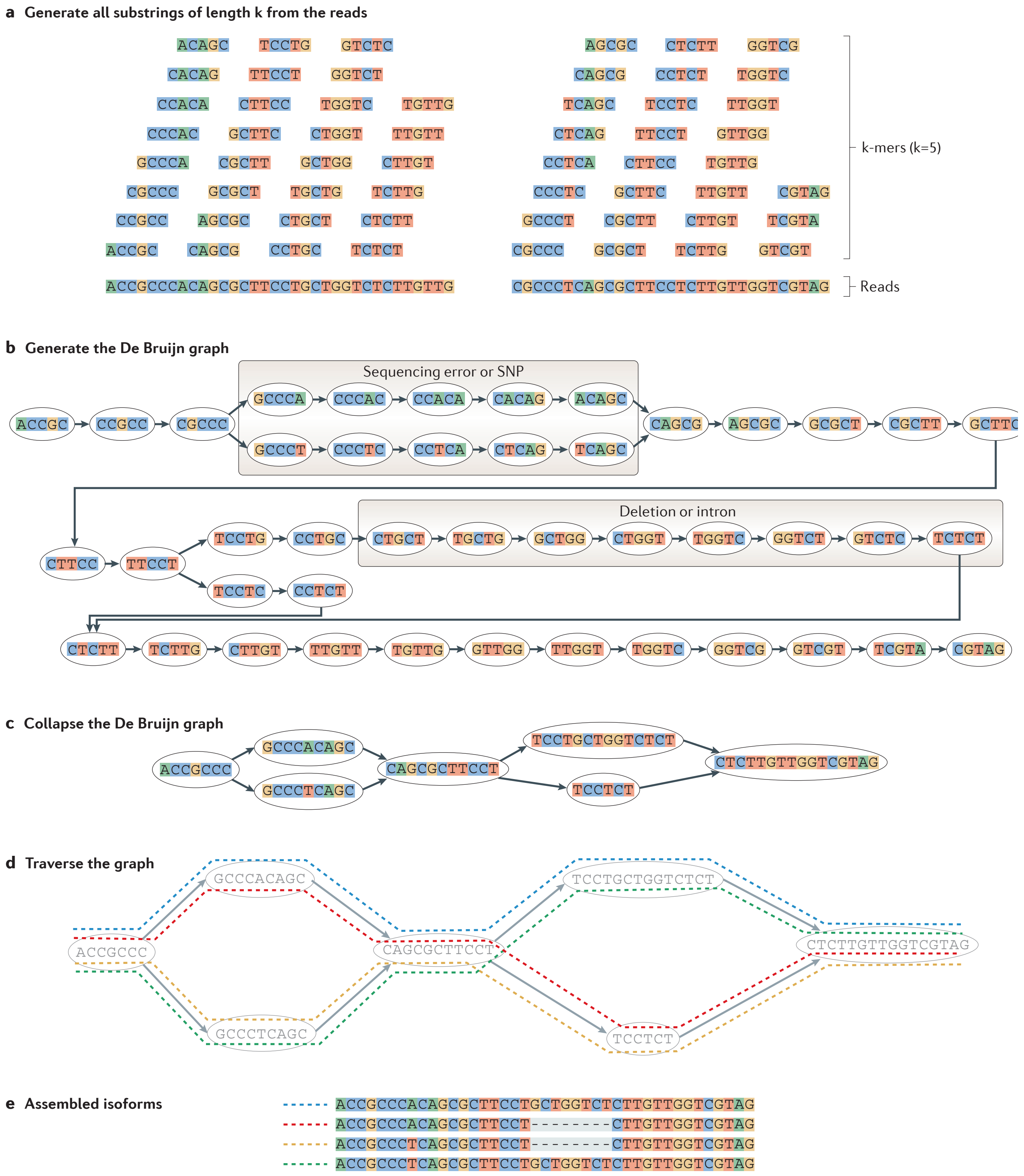
Genomes vs Transcriptomes

- Genomes: mostly uniform coverage across genome; target is usually tens of chromosomes.

- Transcriptomes: widely varying coverage depending on gene expression; splice variants for thousands of transcripts.

- Genome and transcriptome assemblers are based similar ideas, but must use different algorithms.

de Bruijn Graphs: Modern Assembly



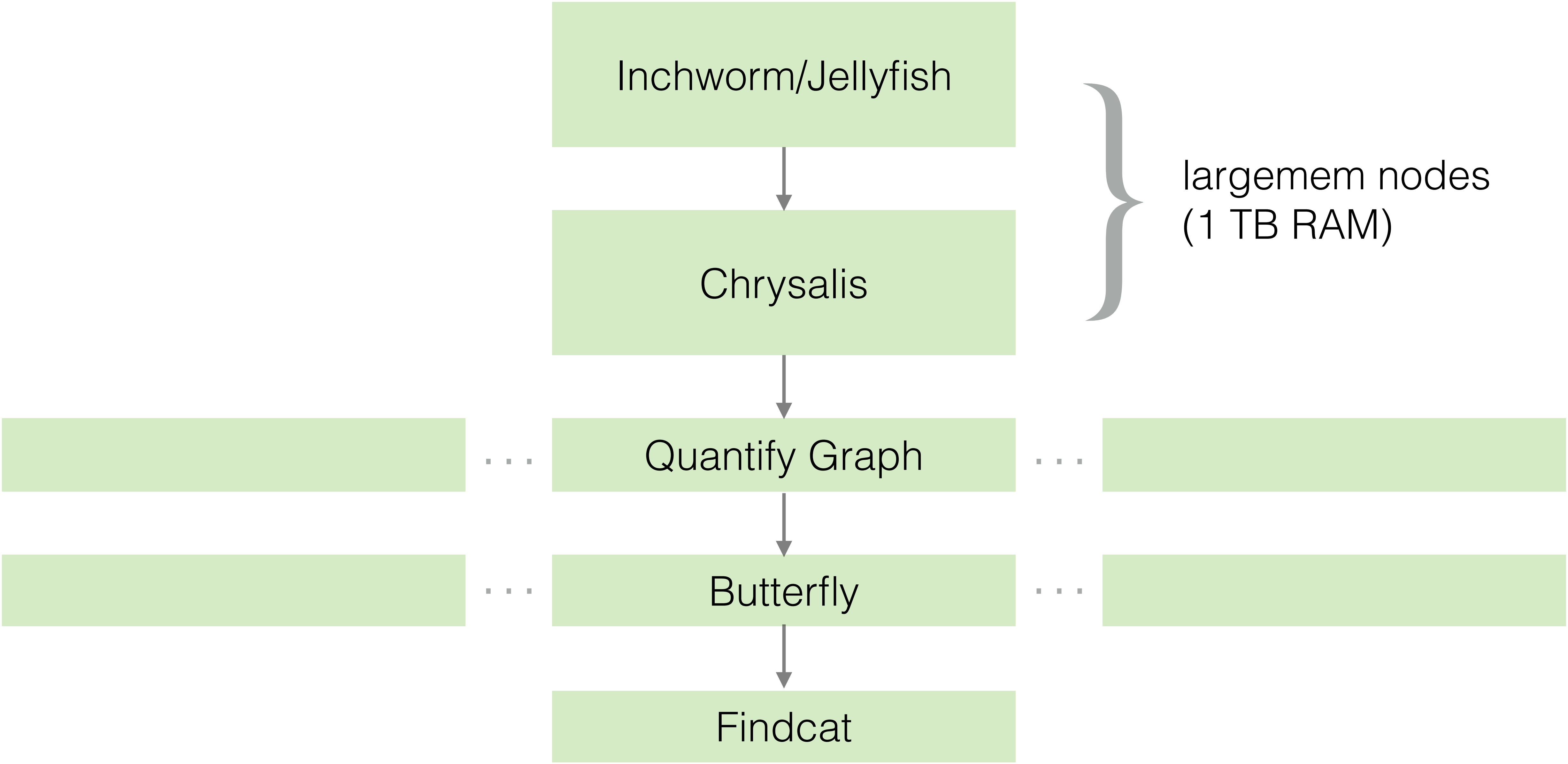
Nature Reviews Genetics **12**, 671-682

Trinity Is A Pipeline

- Inchworm and Chrysalis are responsible for cutting up short reads into k-mers, and generating the de Bruijn graphs (one graph per gene).

- Butterfly determines putative transcripts (and isoforms) from the de Bruijn graphs that Chrysalis generates.

Trinity Parallelized On TACC



Inchworm/Jellyfish

Chrysalis

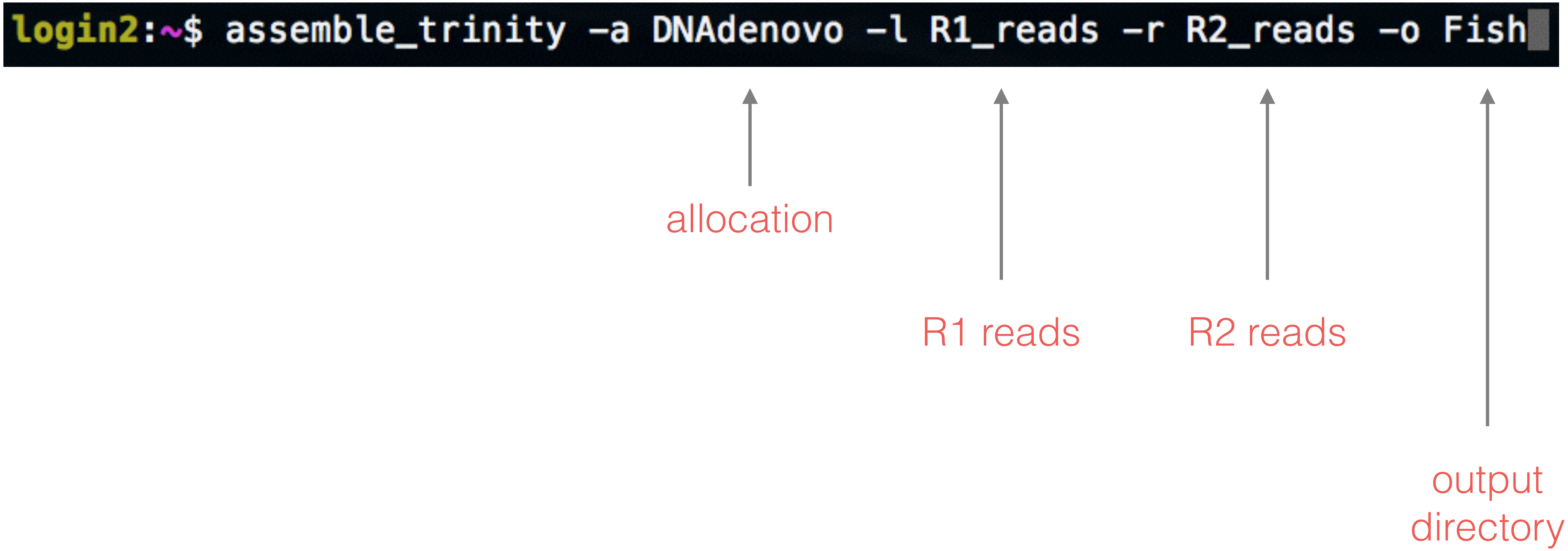
largemem nodes
(1 TB RAM)

Quantify Graph

Butterfly

Findcat

An Actual Command



```
login2:~$ assemble_trinity -a DNAdenovo -l R1_reads -r R2_reads -o Fish
```

allocation

R1 reads

R2 reads

output
directory

Future Development

- Set up `assemble_trinity` to work on Stampede.
- Find a way around the time limits in Lonestar and Stampede. (Some ideas, nothing working yet.)