Trinity

- Read set
  - Extend in k-mer space and break ties
  - Overlap linear sequences by overlaps of k - 1 to build graph components

De Bruijn graph (k = 5)

- Compact graph
- Finding paths
- Extracting sequences

Linear sequences

- a121: len = 5,845
- a122: len = 2,560
- a123: len = 4,443
- a124: len = 48
- a126: len = 66

Transcripts

...CTTCGCA...TOATCGCAT...

...ATTGGCAA...TOATCGGAT...
#!/bin/tcsh
unlimit
Trinity.pl --seqType fq --single psme_allneedle.fq --SS_lib_type R --no_run_butterfly --CPU 1 --min_contig_length 300 --output allneedle_trinity_output.dir
#!/bin/tcsh

SGE_Batch -c 'java -Xmx1000M -jar /raid1/home/bpp/knausb/bin/trinityrnaseq-r20110519/Butterfly/Butterfly.jar -N 238268103 -L 300 -F 300 -C chrysalis/RawComps.0/comp0 --compatible_path_extension --edge-thr=0.26 --stderr -V 10' -m 1G -P 1 -r bfly_1 -p -100

SGE_Batch -c 'java -Xmx1000M -jar /raid1/home/bpp/knausb/bin/trinityrnaseq-r20110519/Butterfly/Butterfly.jar -N 238268103 -L 300 -F 300 -C chrysalis/RawComps.0/comp1 --compatible_path_extension --edge-thr=0.26 --stderr -V 10' -m 1G -P 1 -r bfly_2 -p -100
Visualizing de Bruijn graphs

- GraphViz and *.dot language.
- AMOS *.afg files -> other formats.
Visualizing de Bruijn graphs

- AACAAACCT...CTGATGAAG:207[10][370]
- TTGTCTCAACGGCCTTGAATAA:53[346][24]
- TTCTGGGACAG...TGCAATGGA:111[742][50]
- TAGCACTTG...TGCAATGGCG:97[1137][327]

- CCCGTTCCGC...CCATCGCTCC:93[1307][96]
- 39
- 33

- 23
- 65

- AGCTCTGGAG...TTCTGGCCG:56[792][271]
Visualizing de Bruijn graphs
Visualizing de Bruijn graphs
What about quality?

- Has not been implemented in any of the popular de novo assemblers.
- Coverage as quality.
  - Velvet
  - ABySS
  - Trinity
Other software?

- Oases – extends velvet.
- Trans-ABySS.
- Rannotator.
- RUM – hybrid bowtie and blat.
- Stampy – hybrid bwa.
Introductions

Who are you?
What project(s) are you working on?
What project(s) would you like to work on?
What would you like to get out of these workgroups?