Genome Assembly

Lecture 22
Oct 24, 2016

Announcements

Final Project: The final project will consist of an oral presentation and written report (e.g., the methods section) related to an assembly project. Projects must incorporate an implementation of the computational techniques we've learned about. The final project will be worth 100 points (75 written/25 oral). Oral presentations will occur during the last 3 days of class. Written reports will be due on the last day of class. More details will be provided later in the semester.

Announcements

http://oyster-river-protocol.readthedocs.io/en/v2/

https://www.ebi.ac.uk/ena/

Probably Transcriptome
Illumina, PacBio, Nanopore
Error Correct
Trim
Assembly
Filter
Annotate

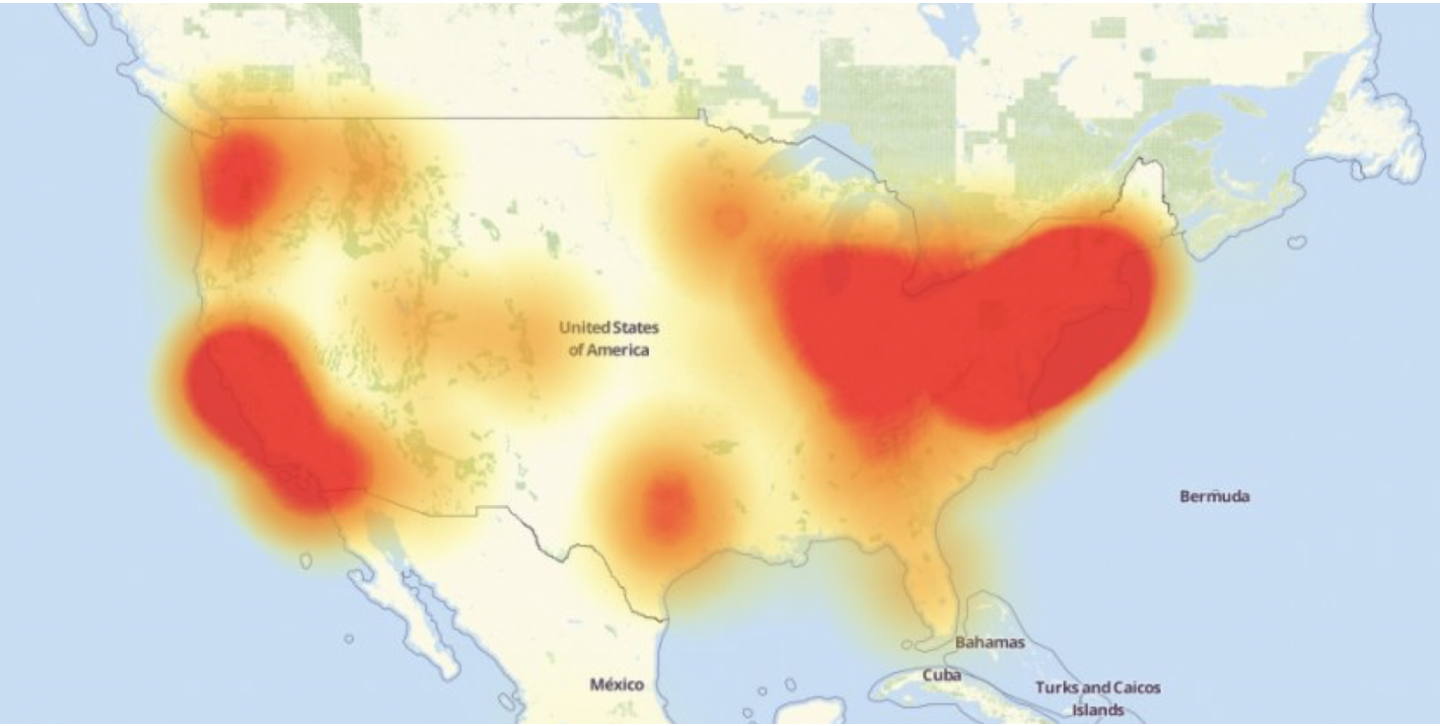
Announcements

Due: 12/9
Presentations 12/5, 7, 9

Read Dataset Approval: 11/11?

Last few weeks of lab dedicated to project work.

A **DDoS attack** uses a variety of techniques to send countless junk requests to a website. This boosts traffic to the website so much that it gets overwhelmed, making it nearly impossible for anyone to load the page.



Assembly – De Bruijn

Hamiltonian Path Problem

Eulerian Path Problem

Assembly – De Bruijn

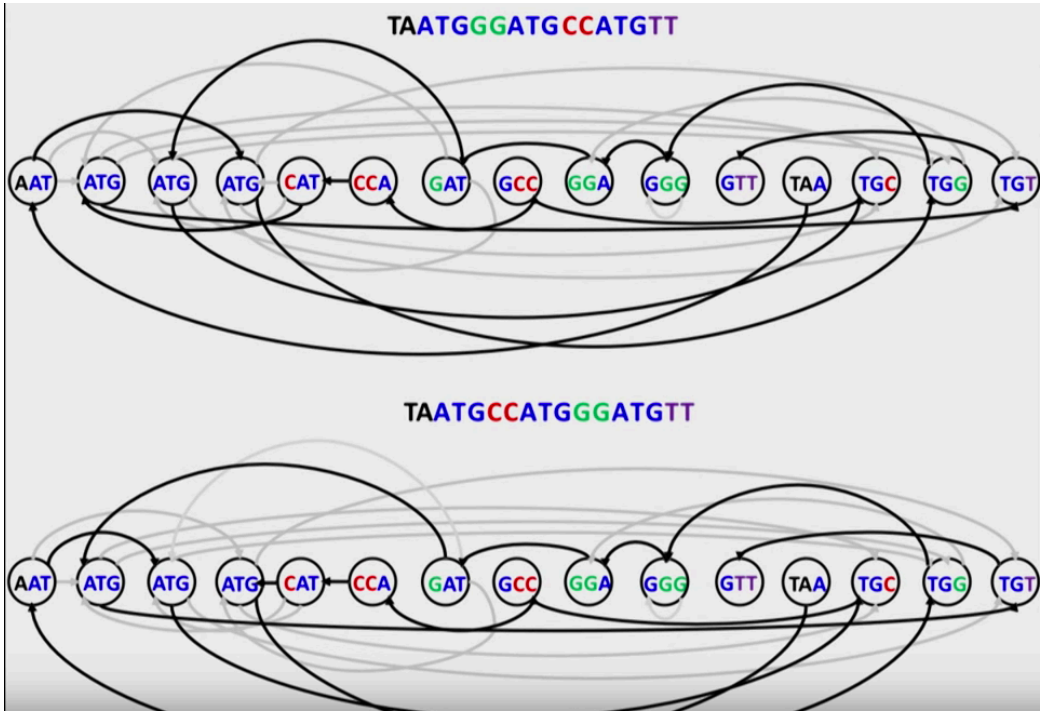
Hamiltonian Path Problem

$Composition_3$(TAATGCCATGGGATGTT) =

Can we construct this genome path without knowing the genome **TAATGCCATGGGATGTT**, only from its composition?

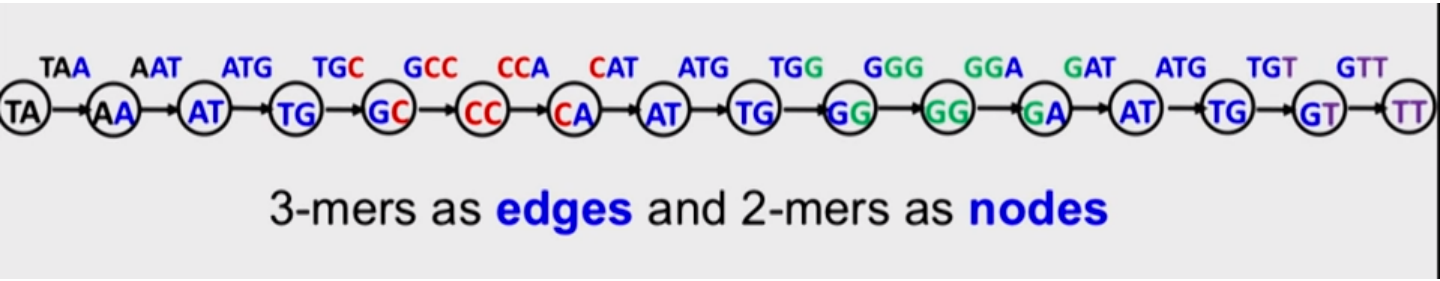
Assembly – De Bruijn

Hamiltonian Path Problem



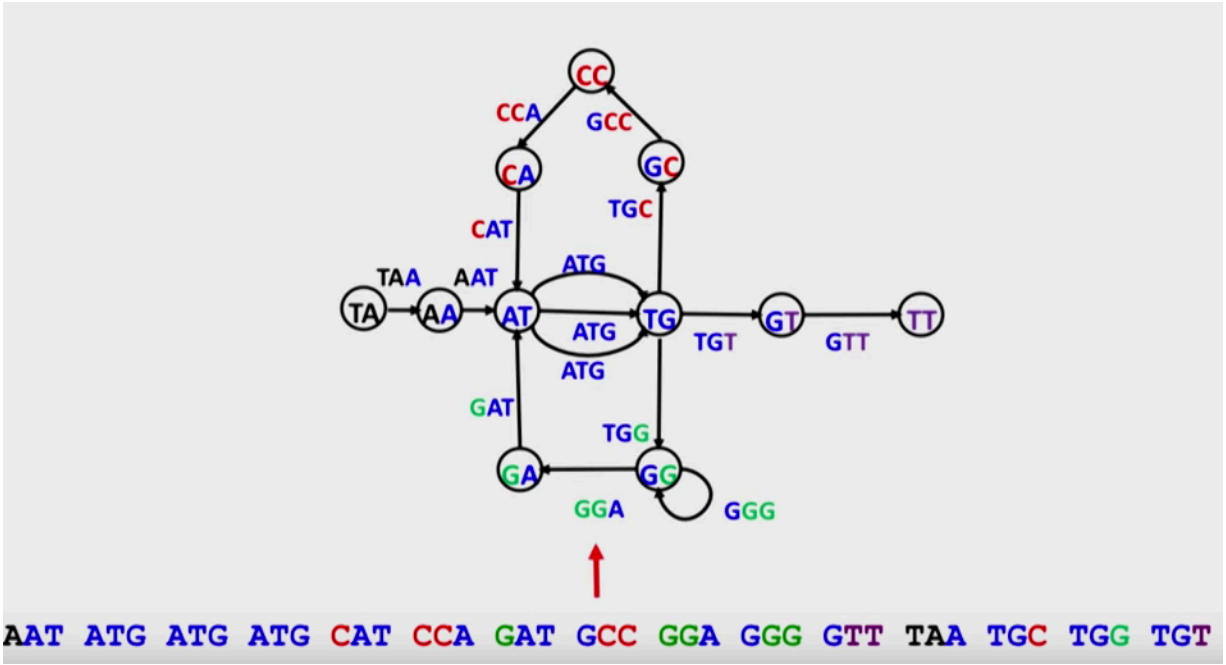
Assembly – De Bruijn

Eulerian Path Problem



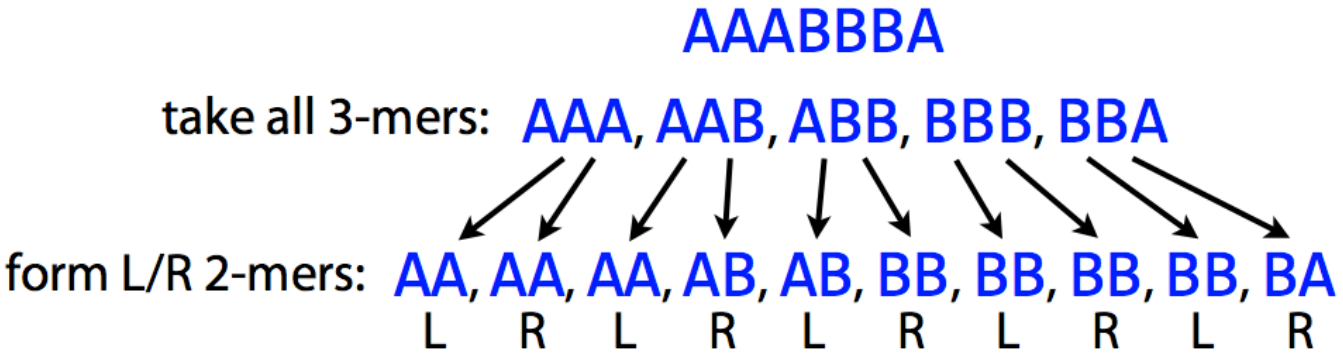
Assembly – De Bruijn

Eulerian Path Problem



Assembly – De Bruijn

Assembly – De Bruijn

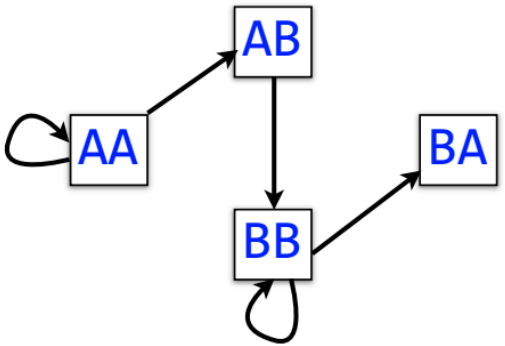


Assembly – De Bruijn

form L/R 2-mers: AA, AA, AA, AB, AB, BB, BB, BB, BB, BA

L R L R L R L R L R

Let 2-mers be nodes in a new graph. Draw a directed edge from each left 2-mer to corresponding right 2-mer:



Each *edge* in this graph corresponds to a length-3 input string

Assembly – De Bruijn

GATTACAGTTCA

Assembly – De Bruijn

GATTACAGTTCA
GATTAC
ACAGTTCA

Assembly – De Bruijn

GATTAC

ACAGTTCA

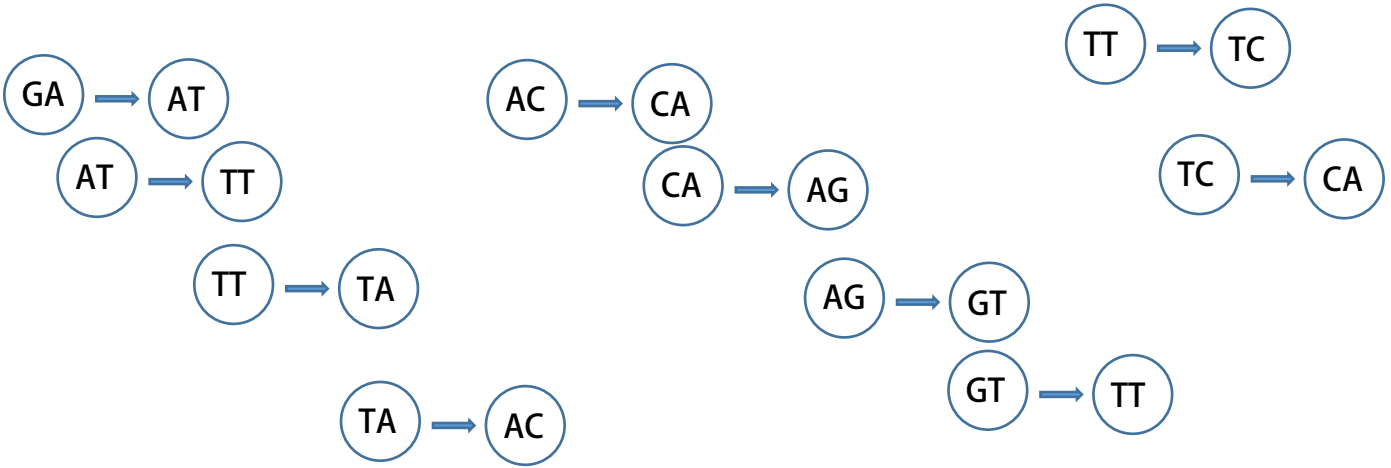
Assembly – De Bruijn

```
GATTAC
GAT
  ATT
    TTA
      TAC
```

```
ACAGTTCA
ACA
  CAG
    AGT
      GTT
        TTC
         TCA
```

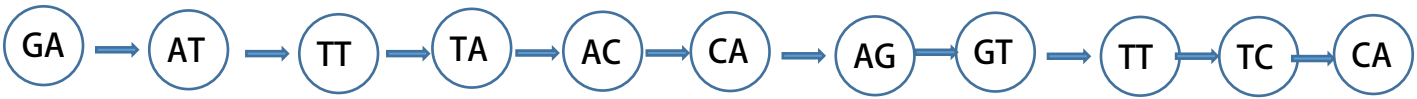
Assembly – De Bruijn

GAT ATT TTA TAC ACA CAG AGT GTT TTC TCA



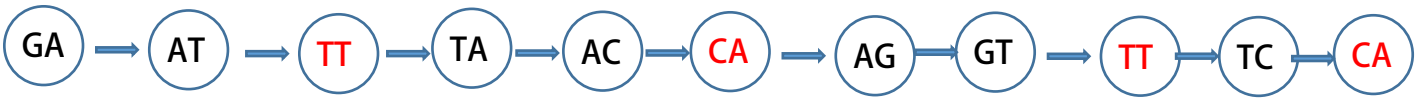
Assembly – De Bruijn

GAT ATT TTA TAC ACA CAG AGT GTT TTC TCA



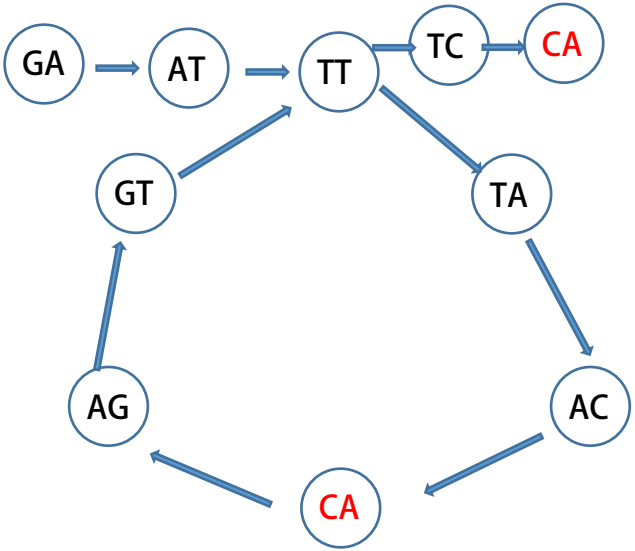
Assembly – De Bruijn

GAT ATT TTA TAC ACA CAG AGT GTT TTC TCA



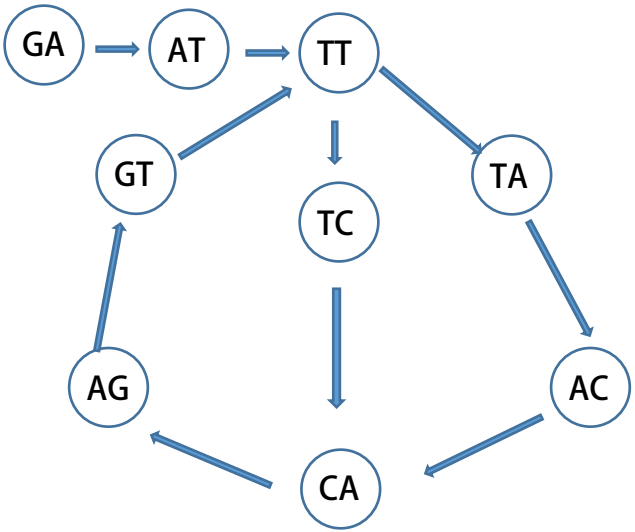
Assembly – De Bruijn

GAT ATT TTA TAC ACA CAG AGT GTT TTC TCA



Assembly – De Bruijn

GAT ATT TTA TAC ACA CAG AGT GTT TTC TCA



Assembly – De Bruijn

GAT ATT TTA TAC ACA CAG AGT GTT TTC TCA