Genome Assembly

26 Oct 15

Announcements

Assembly – De Bruijn

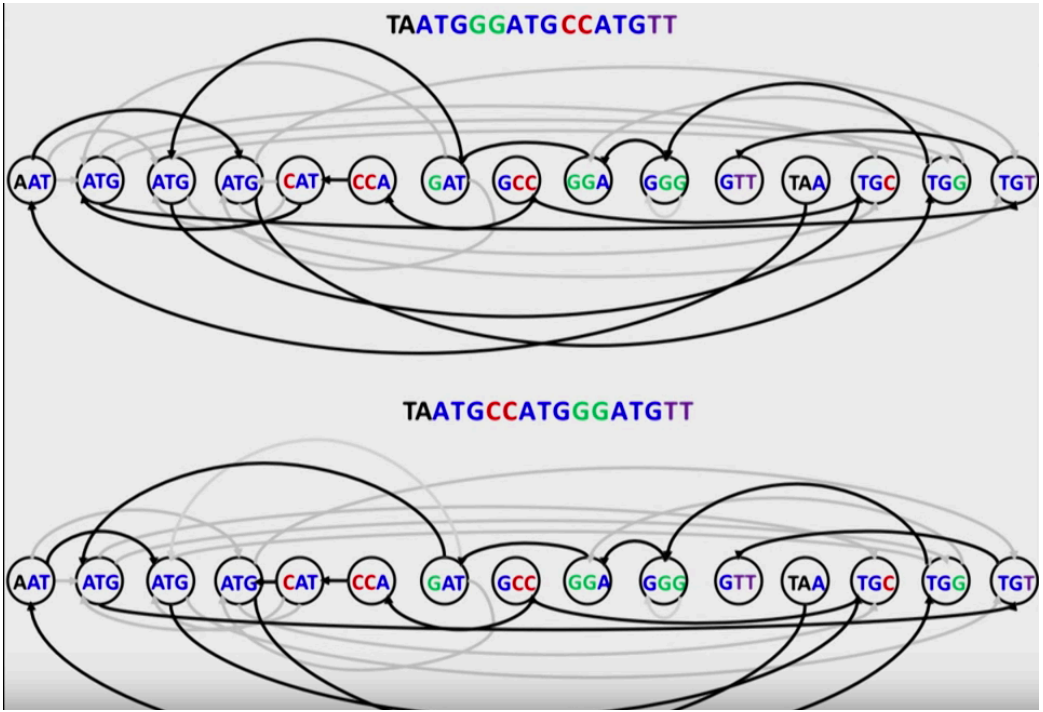
Hamiltonian Path Problem

$Composition_3$(TAATGCCATGGGATGTT) =

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

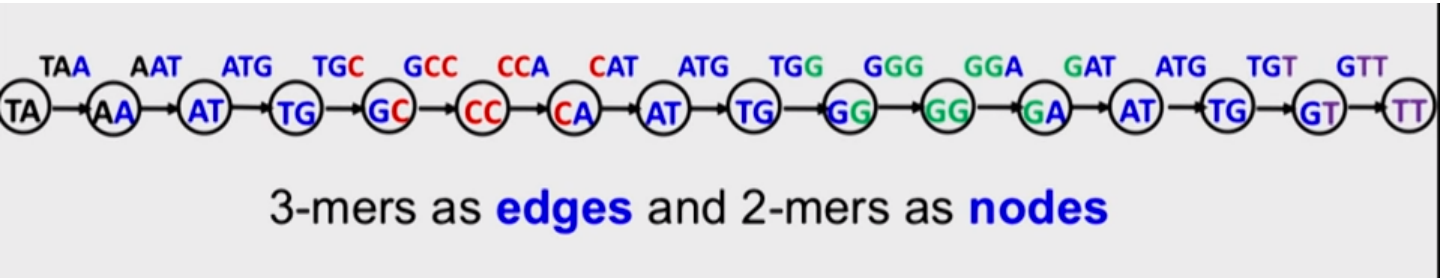
Assembly – De Bruijn

Hamiltonian Path Problem



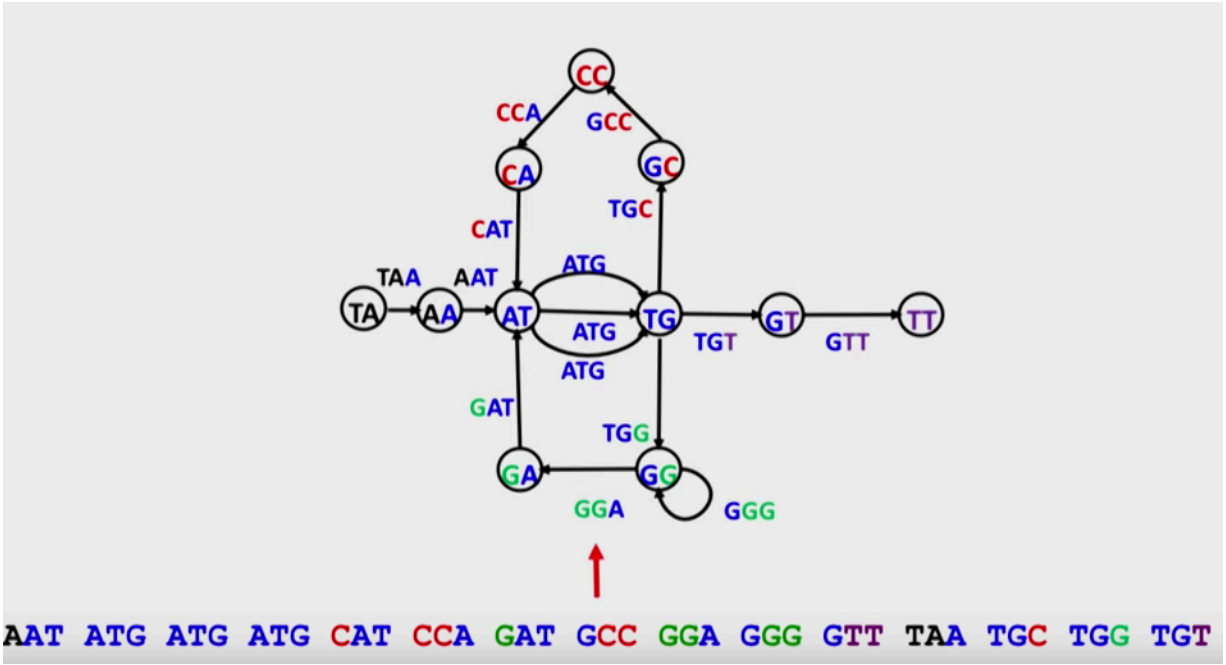
Assembly – De Bruijn

Eulerian Path Problem



Assembly – De Bruijn

Eulerian Path Problem

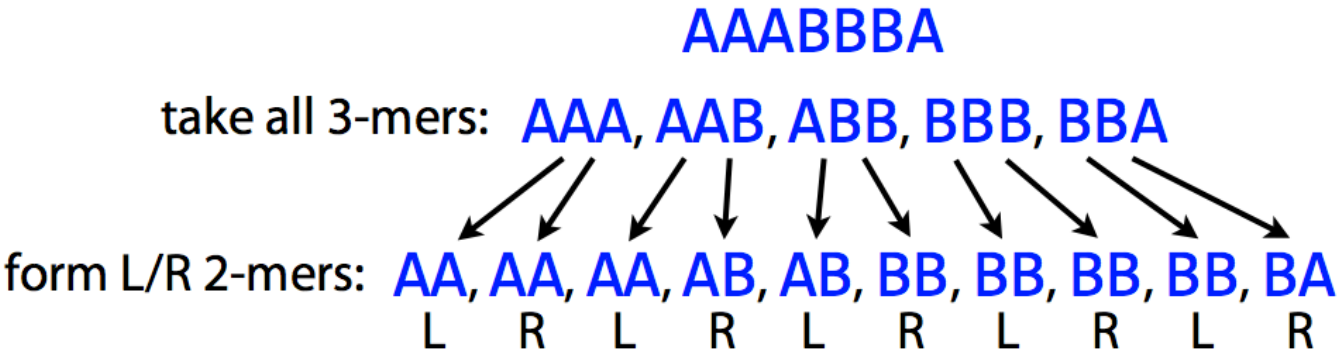


Assembly – De Bruijn

Hamiltonian vs Eulerian

Assembly – De Bruijn graph approach

Assembly – De Bruijn

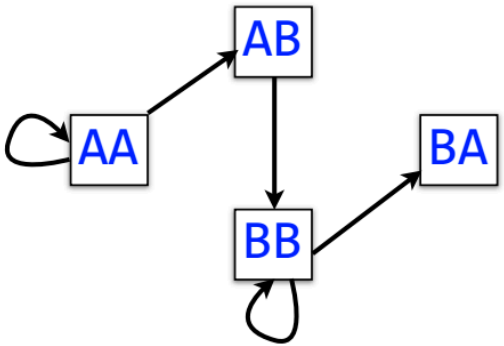


Assembly – De Bruijn

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

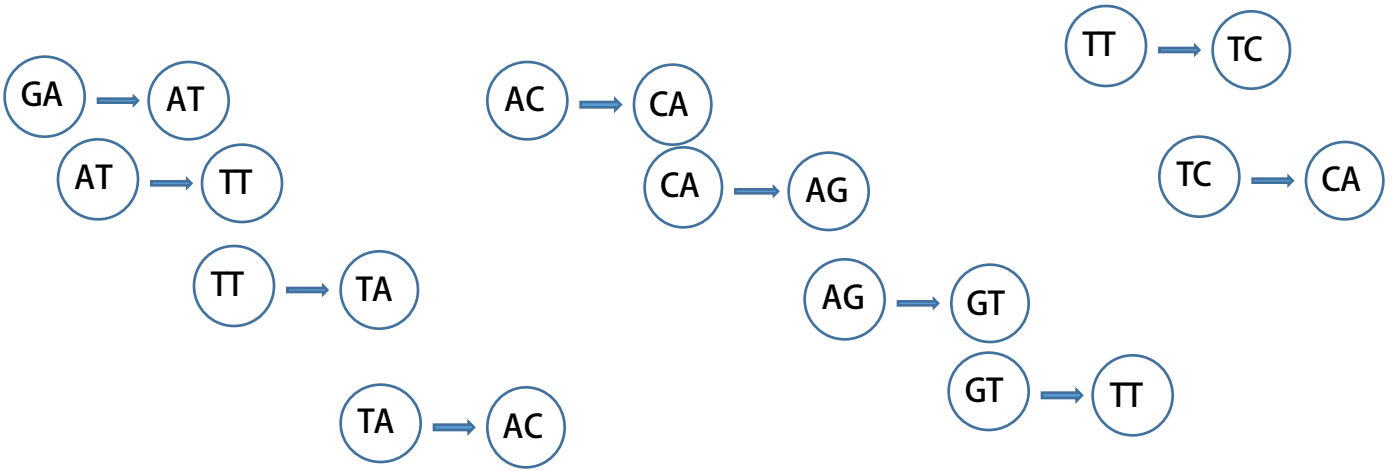
Assembly – De Bruijn

```
GATTAC
GAT
  ATT
    TTA
      TAC
```

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

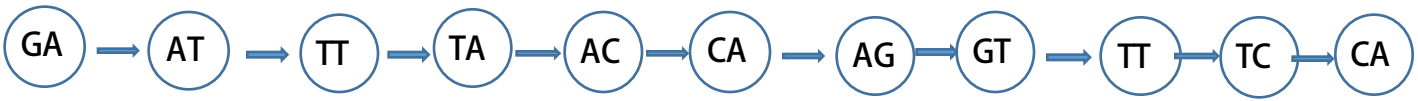
Assembly – De Bruijn

GAT ATT TTA TAC ACA CAG AGT GTT TTC TCA



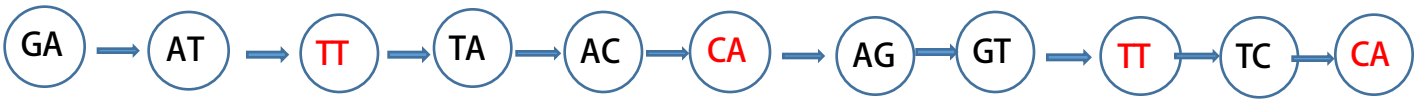
Assembly – De Bruijn

GAT ATT TTA TAC ACA CAG AGT GTT TTC TCA



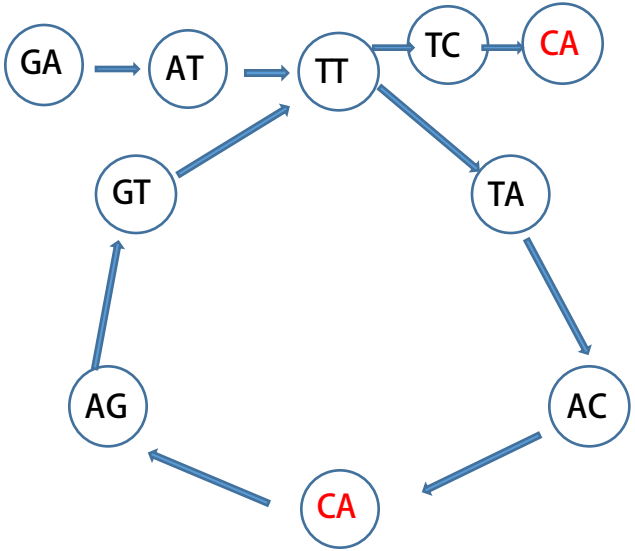
Assembly – De Bruijn

GAT ATT TTA TAC ACA CAG AGT GTT TTC TCA



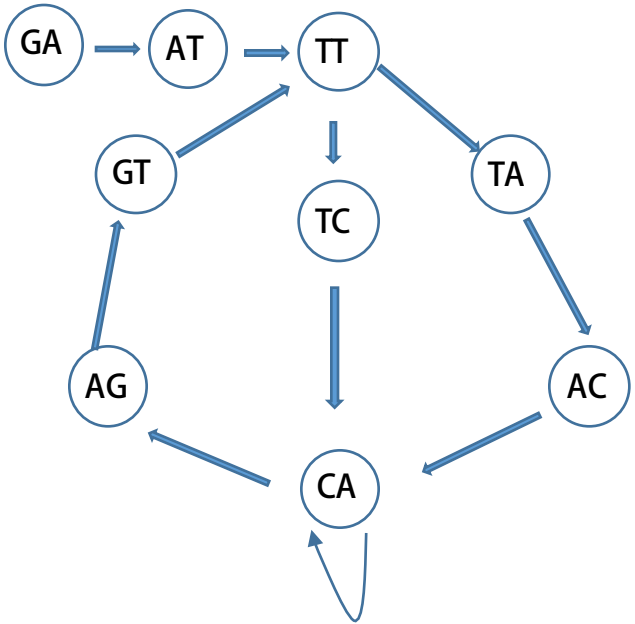
Assembly – De Bruijn

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