

Sequenzassemblierung

ASE Übung

21.06.2019

Shotgun-Sequenzierung

ACGTGTGCGCTTCTT
GCTTCCGCCGACGTGTGCG
TTCGATCGCTTCCGC
GCTACGTTTCGATCGCT
CGACTAAGCTA

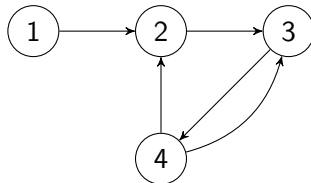
CGACTAAGCTACGTTTCGATCGCTTCCGCCGACGTGTGCGCTTCTT

Shotgun-Sequenzierung (cont.)

ACGTGTGCGCTTCTT
GCTTCCGCCGACGTGTGCG
TTCGATCGCTTCCGC
GCTACGTTCGATCGCT
CGACTAAGCTA

????????????????????????????????????

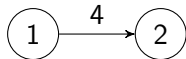
Graph



Overlap graph

1: CGACTAAGCTA
2: GCTACGTTTCGATCGCT

1: CGACTAAGCTA
2: GCTACGTTTCGATCGCT



De Bruijn Graph

1: CGACTAA...

$k = 6$

CGACTA

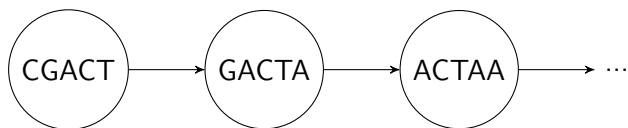
GACTAA

⋮

De Bruijn Graph (cont.)

CGACTA
/ \
CGACT GACTA

GACTAA
/ \
GACTA ACTAA



Frage 1

Bei der Assemblierung der Sequenz mittels des *overlap*-Graphs wurde ein Problem festgestellt.

Worum handelt es sich?

Wie würde das selbe Problem den *de Bruijn*-Graph Ansatz beeinflussen?

Was wären mögliche Lösungsansätze?

Frage 1 (cont.)

```
AAGGCCATTCTT
      TTCTTATTCTTA
      TTCTTATTCTTAT
      | CTTATTCTTATT
      |           TATTCACGCGG
      |           |
AAGGCCATTCTTATTCTTATTCACGCGG
```

```
AAGGCCATTCTT
      TTCTTATTCTTAT
      |       TTCTTATTCTTA
      |       CTTATTCTTATT
      |           TATTCACGCGG
      |           |
AAGGCCATTCTTATTCTTATTCTTATTCACGCGG
```

Repeats

.....xxxxxxAAAABBBBAAAABBBBAAAABBBByyyyyy.....

|

.....xxxxxxAAAABBBByyyyyy.....

.....AAAABBBBxxxxxxAAAABBBByyyyyyAAAABBBB.....

oder

.....AAAABBBByyyyyyAAAABBBBxxxxxxAAAABBBB.....

?

Frage 2

Welche weiteren Probleme können bei *Shotgun*-Sequenzierung und Graph-Assemblierung auftreten und wie könnte man mit diesen umgehen?

Fehler und Coverage

1: CGACTAAGCTA

|||

2: GCCACGTTCGATCGCT

CCATTCTT

TTCTTATTCT**T**A

TTCTTATTCT**T**AT

CTTATTCT**C**ATT

TATTCACGCGG

|

CCATTCTTATTCT**T**ATTCACGCGG

Frage 3

Welcher der graphbasierten Ansätze ist komplexer?

Welche Komplexität hat die Erstellung der Graphen (grob) in Abhängigkeit von der Anzahl der Fragmente N ?