

Assembly

Reconstruct this

CTAGGCCCTCAATTTT
CTCTAGGCCCTCAATTTT
GGCTCTAGGCCCTCATTTTT
CTCGGCTCTAGCCCCTCATTTT
TATCTCGACTCTAGGCCCTCA
TATCTCGACTCTAGGCC
TCTATATCTCGGCTCTAGG
GGCGTCTATATCTCG
GGCGTCGATATCT
GGCGTCTATATCT
→ GGCGTCTATATCTCGGCTCTAGGCCCTCATTTTT

From
these

Assembly

Reconstruct this

CTAGGCCCTCAATT
GGCGTCTATATCT
CTCTAGGCCCTCAATT
TCTATATCTCGGCTCTAGG
GGCTCTAGGCCCTCATT
CTCGGCTCTAGCCCCTCATT
TATCTCGACTCTAGGCCCTCA
GGCGTCGATATCT
TATCTCGACTCTAGGCC
GGCGTCTATATCTCG
??

From
these

Coverage

CTAGGCCCTCAATT
CTCTAGGCCCTCAATT
GGCTCTAGGCCCTCATT
CTCGGCTCTAGCCCCTCATT
TATCTCGACTCTAGGCCCTCA
TATCTCGACTCTAGGCC
TCTATATCTGGCTCTAGG
GGCGTCTATATCTCG
GGCGTCGATATCT
GGCGTCTATATCT
GGCGTCTATATCTGGCTCTAGGCCCTCATT

Coverage = 5

Coverage

CTAGGCCCTCAATT
CTCTAGGCCCTCAATT
GGCTCTAGGCCCTCATT
CTCGGCTCTAGCCCCTCATT
TATCTCGACTCTAGGCCCTCA
TATCTCGACTCTAGGCC
TCTATATCTGGCTCTAGG
GGCGTCTATATCTCG
GGCGTCGATATCT
GGCGTCTATATCT
GGCGTCTATATCTGGCTCTAGGCCCTCATT



Coverage = 5

CTAGGCCCTCAATT
CTCTAGGCCCTCAATT
GGCTCTAGGCCCTCATT
CTCGGCTCTAGCCCCTCATT
TATCTCGACTCTAGGCCCTCA
TATCTCGACTCTAGGCC 177 bases
TCTATATCTGGCTCTAGG
GGCGTCTATATCTCG
GGCGTCGATATCT
GGCGTCTATATCT 35 bases
GGCGTCTATATCTGGCTCTAGGCCCTCATT

Average coverage = 177 / 35 ≈ 7-fold

TCTATATCTGGCTCTAGG

TATCTCGACTCTAGGCC

TCTATATCTCGGCTCTAGG
| | | | | | | | | | | |
TATCTCGACTCTAGGCC

First law of assembly

If a suffix of read A is similar to a prefix of read B...

TCTATATCTGGCTCTAGG
||||||| |||||
TATCTCGACTCTAGGCC

...then A and B might *overlap* in the genome

TCTATATCTGGCTCTAGG
GGCGTCTATATCTGGCTCTAGGCCCTCATTTTT
TATCTCGACTCTAGGCC

TCTATATCTCGGCTCTAGG
| | | | | | | | | |
TATCTCGACTCTAGGCC
↑

Why the differences?

1. Sequencing errors
2. Polyploidy: e.g. humans have 2 copies of each chromosome, and copies can differ



Second law of assembly

More coverage leads to more and longer overlaps

CTAGGCCCTCAATTTT
CTCGGCTCTAGCCCCCTCATT
TCTATATCTCGGCTCTAGG
GGCGTCGATATCT
GGCGTCTATATCTCGGCTCTAGGCCCTCATT
CTAGGCCCTCAATTTT
GGCTCTAGGCCCTCATT
CTCGGCTCTAGCCCCCTCATT
TATCTCGACTCTAGGCCCTCA
TCTATATCTCGGCTCTAGG
GGCGTCTATATCTCG
GGCGTCTATATCT
less coverage
more coverage