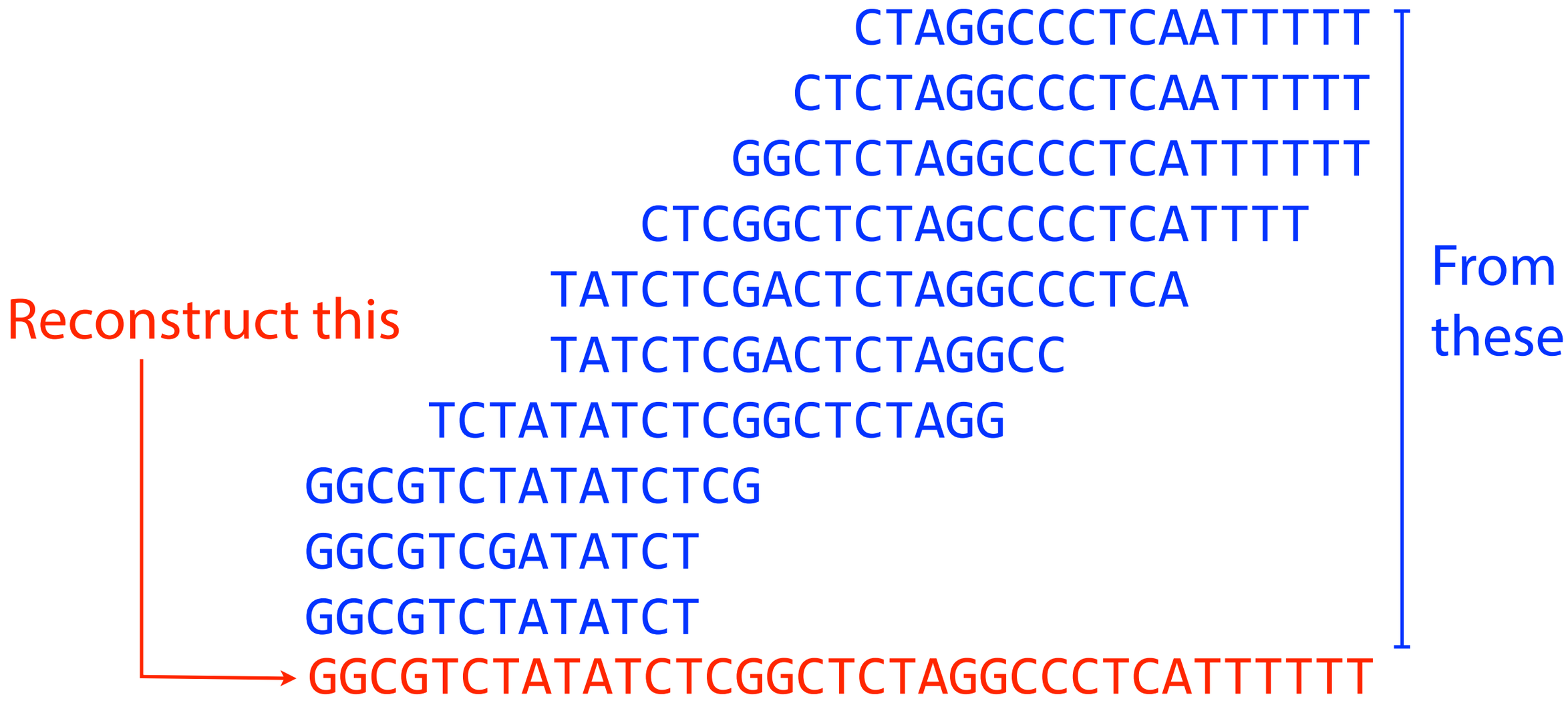
Assembly


Reconstruct this
From these
CTAGGCCCTCAATTTTT
CTCTAGGCCCTCAATTTTT
GGCTCTAGGCCCTCATTTTTT
CTCGGCTCTAGCCCCTCATTTT
TATCTCGACTCTAGGCCCTCA
TATCTCGACTCTAGGCC
TCTATATCTCGGCTCTAGG
GGCGTCTATATCTCG
GGCGTCGATATCT
GGCGTCTATATCT
GGCGTCTATATCTCGGCTCTAGGCCCTCATTTTTT

Assembly

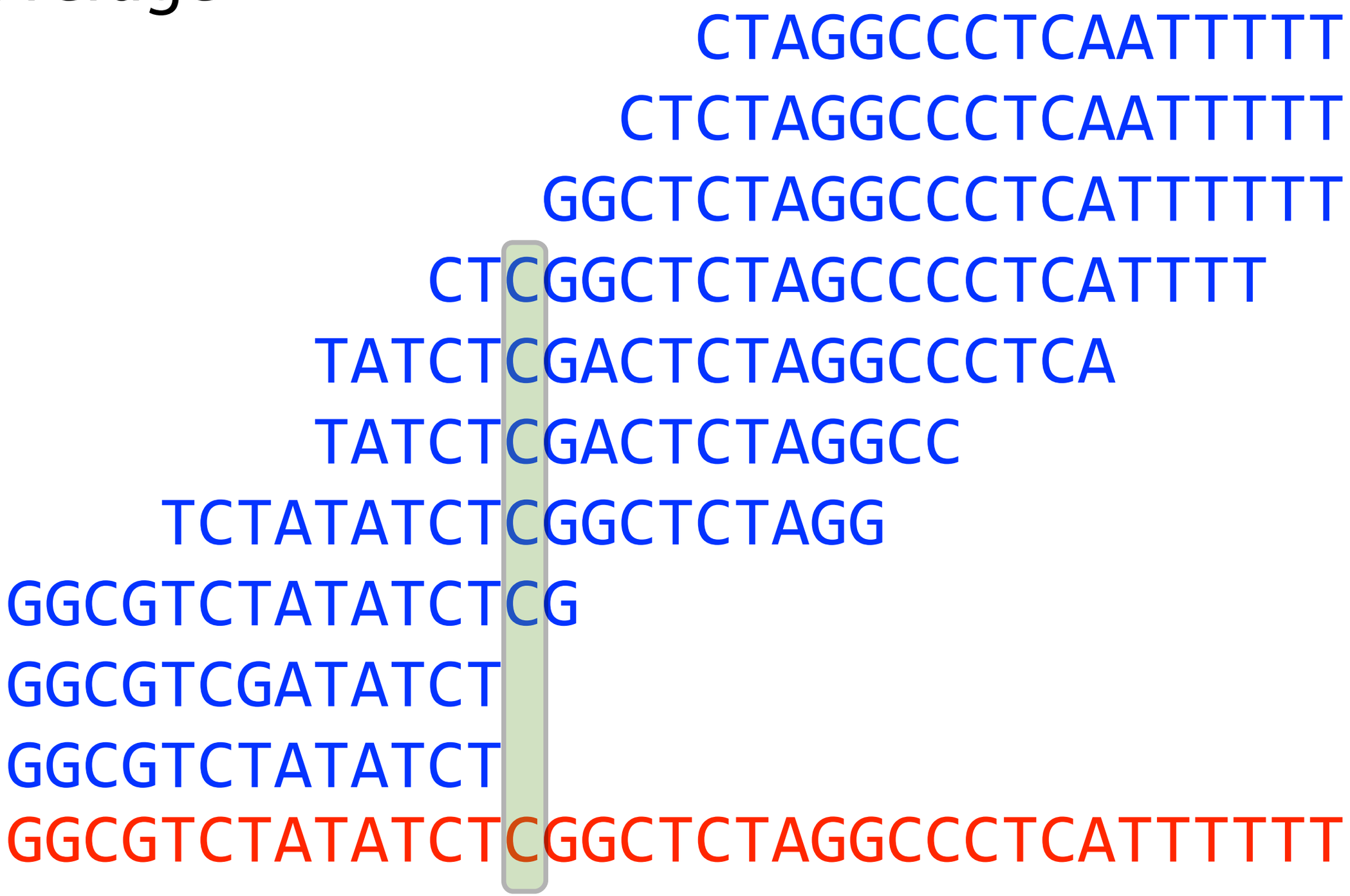
Reconstruct this

From these

CTAGGCCCTCAATTTTT
GGCGTCTATATCT
CTCTAGGCCCTCAATTTTT
TCTATATCTCGGCTCTAGG
GGCTCTAGGCCCTCATTTTTT
CTCGGCTCTAGCCCCTCATTTT
TATCTCGACTCTAGGCCCTCA
GGCGTCGATATCT
TATCTCGACTCTAGGCC
GGCGTCTATATCTCG

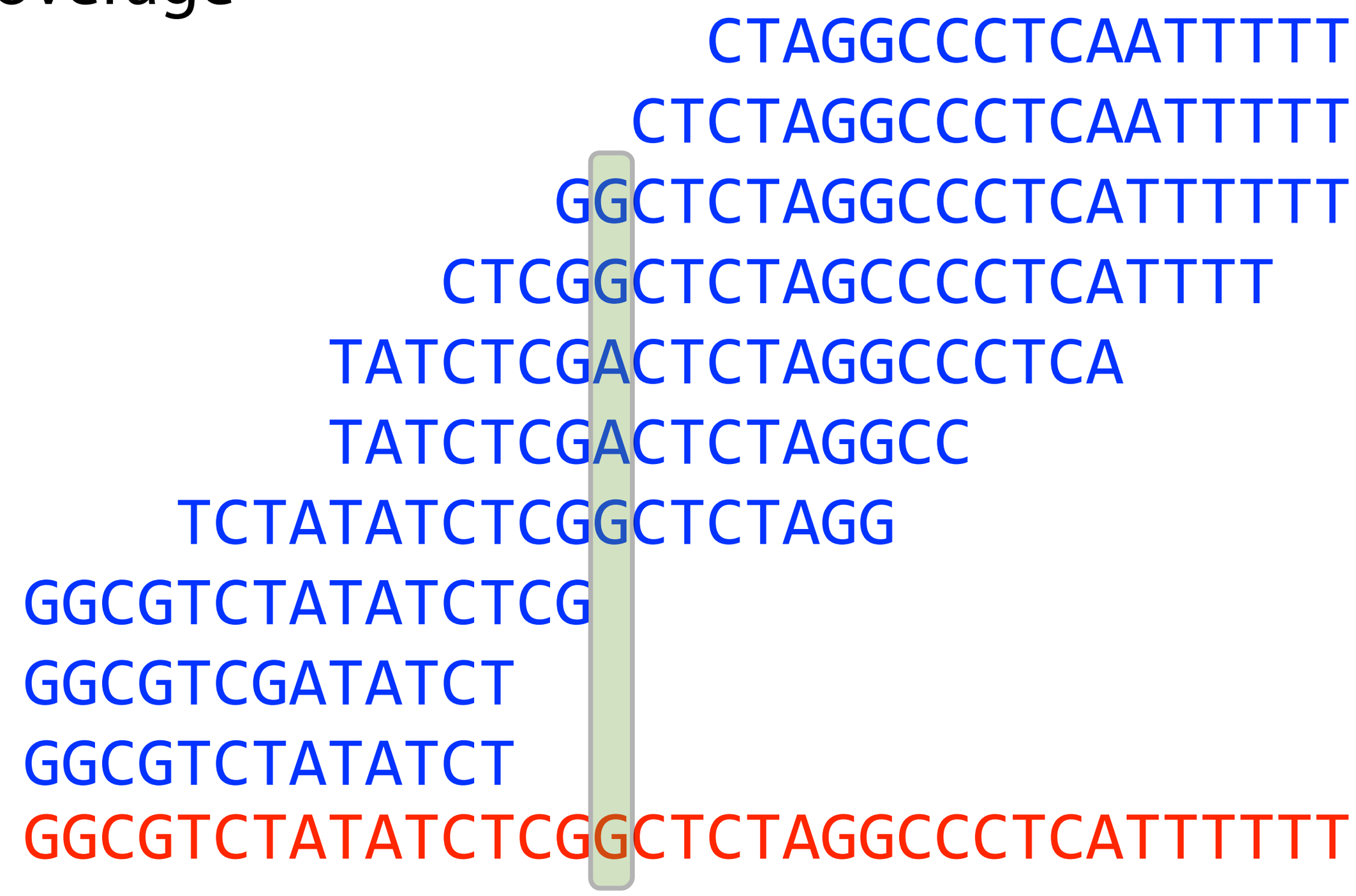
??

Coverage



Coverage = 5

Coverage



Coverage = 5

CTAGGCCCTCAATTTTT
CTCTAGGCCCTCAATTTTT
GGCTCTAGGCCCTCATTTTTT
CTCGGCTCTAGCCCCTCATTTT
TATCTCGACTCTAGGCCCTCA
TATCTCGACTCTAGGCC
TCTATATCTCGGCTCTAGG
GGCGTCTATATCTCG
GGCGTCGATATCT
GGCGTCTATATCT
GGCGTCTATATCTCGGCTCTAGGCCCTCATTTTTT

177 bases

35 bases

Average coverage = 177 / 35 ≈ 7-fold

TCTATATCTCGGCTCTAGG

TATCTCGACTCTAGGCC

TCTATATCTCGGCTCTAGG

TATCTCGACTCTAGGCC

First law of assembly

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

```
TCTATATCTCGGCTCTAGG
       ||||||| ||||||||
      TATCTCGACTCTAGGCC
```

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

```
          TCTATATCTCGGCTCTAGG
GGCGTCTATATCTCGGCTCTAGGCCCTCATTTTTTT
          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