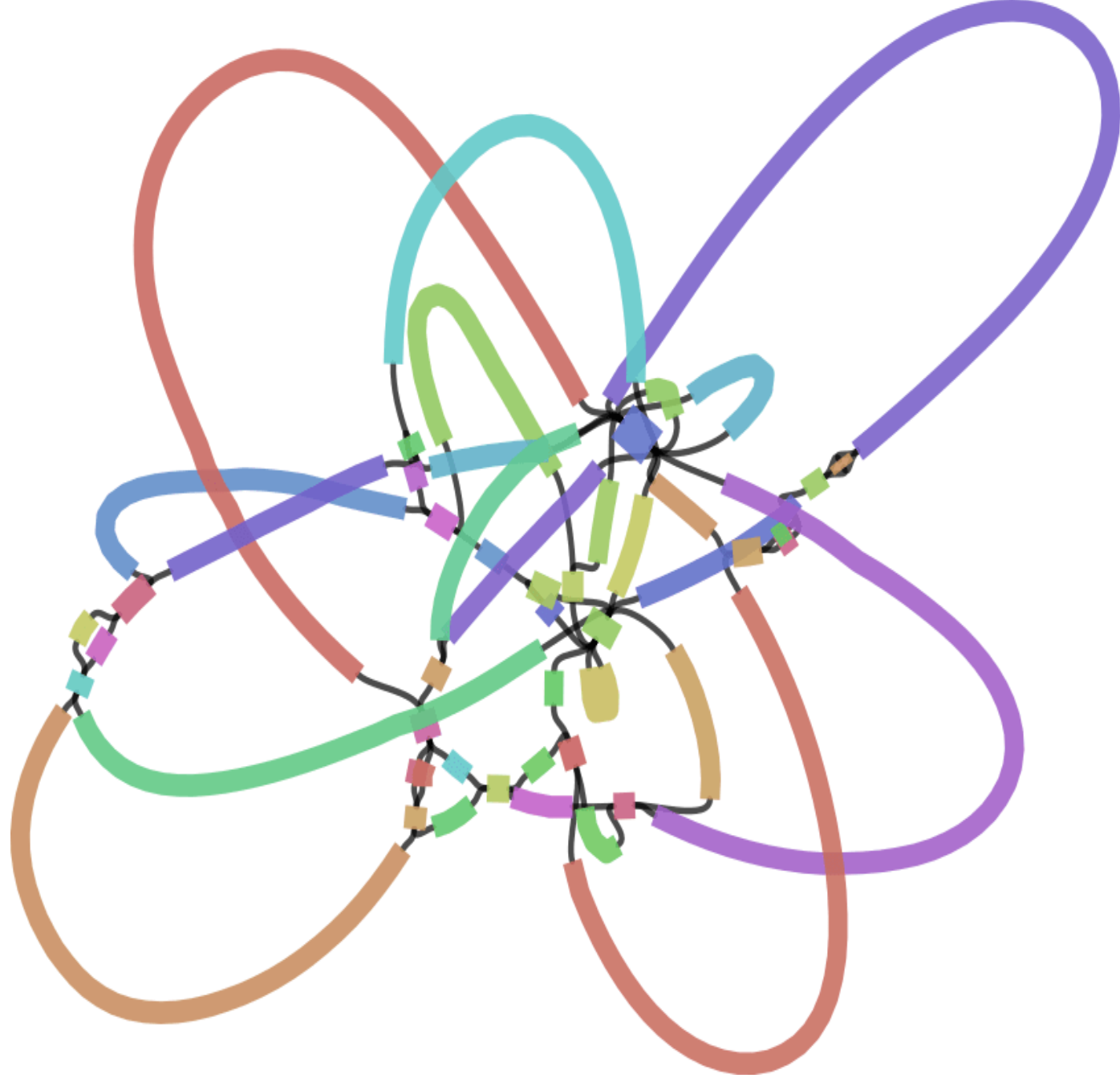


# Genome Assembly

Collins Kigen  
18-July-2025

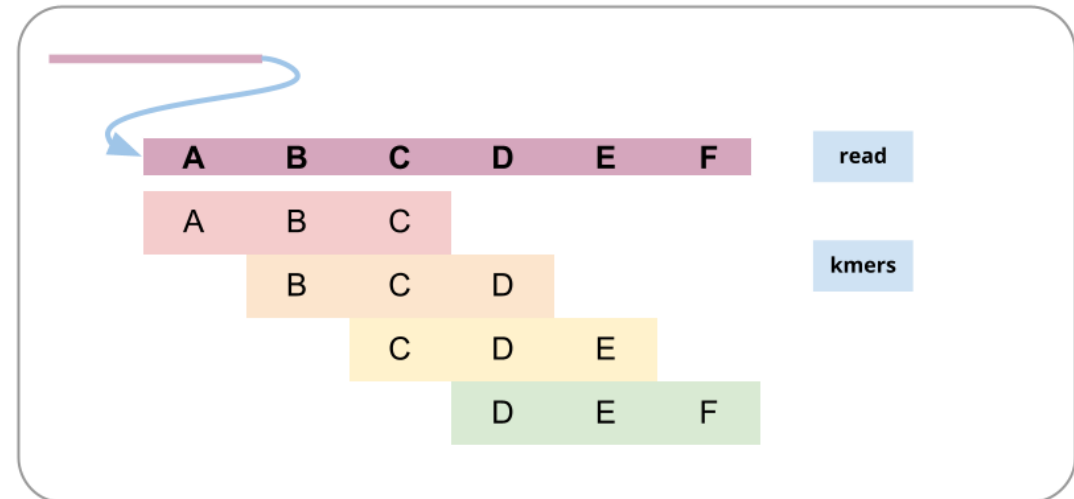


# Learning outcomes

- Perform de novo genome assembly using both short-read and long-read sequencing data
- Evaluate the quality of genome assemblies using QUAST

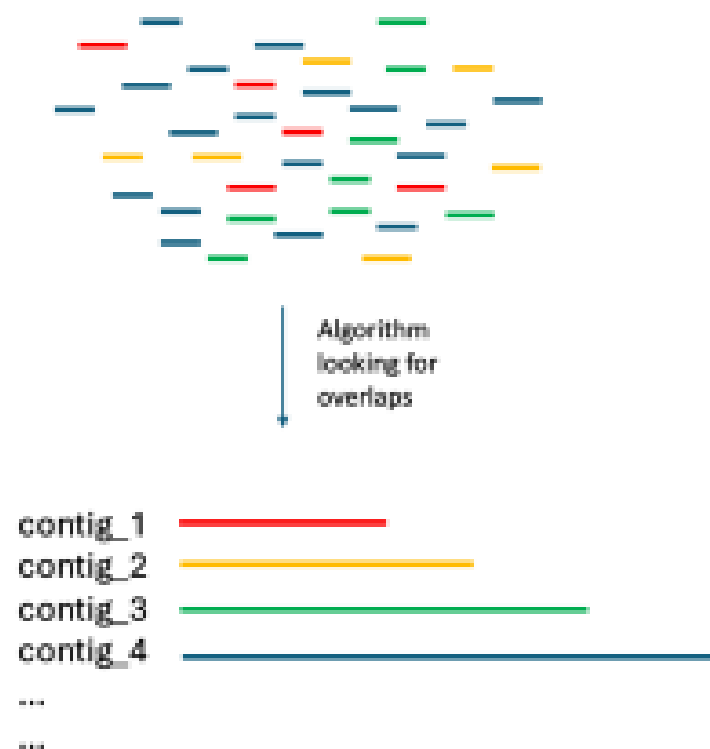
# Genome Assembly

- **Genome assembly** is the process of reconstructing an organism's genome from short or long sequencing reads generated by high-throughput sequencing technologies. The goal is to arrange these reads into longer continuous sequences called **contigs** and further link them into **scaffolds**, which represent larger portions of the genome.

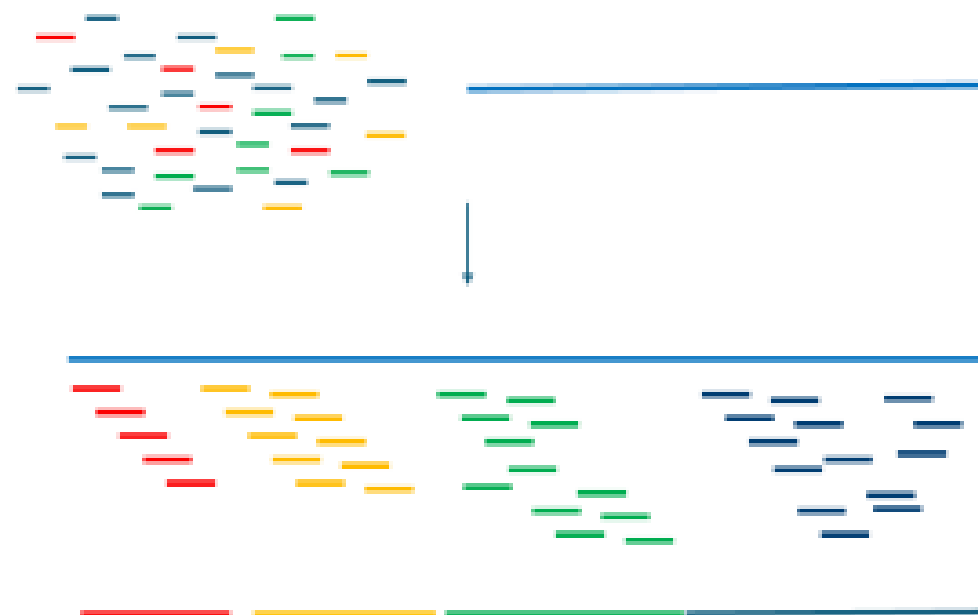


# Approaches

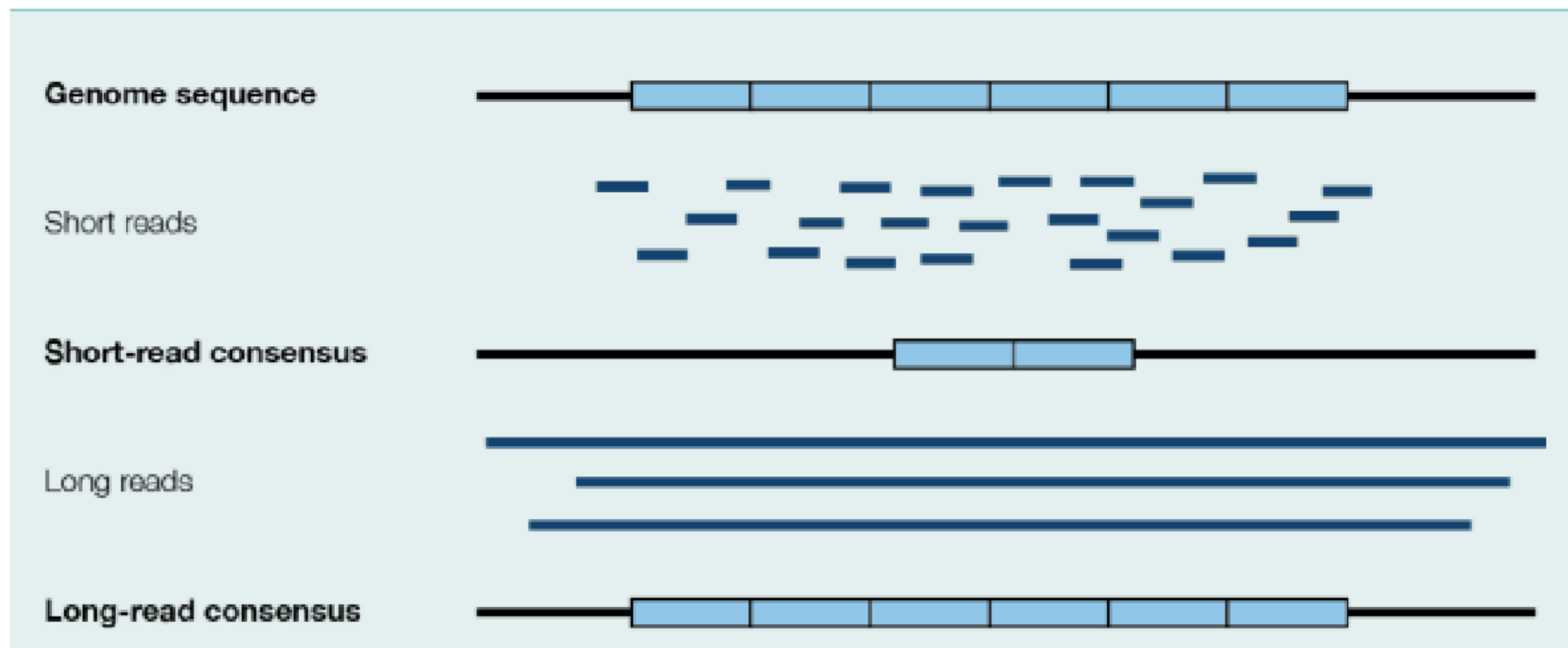
*de novo* assembly



Reference-based assembly

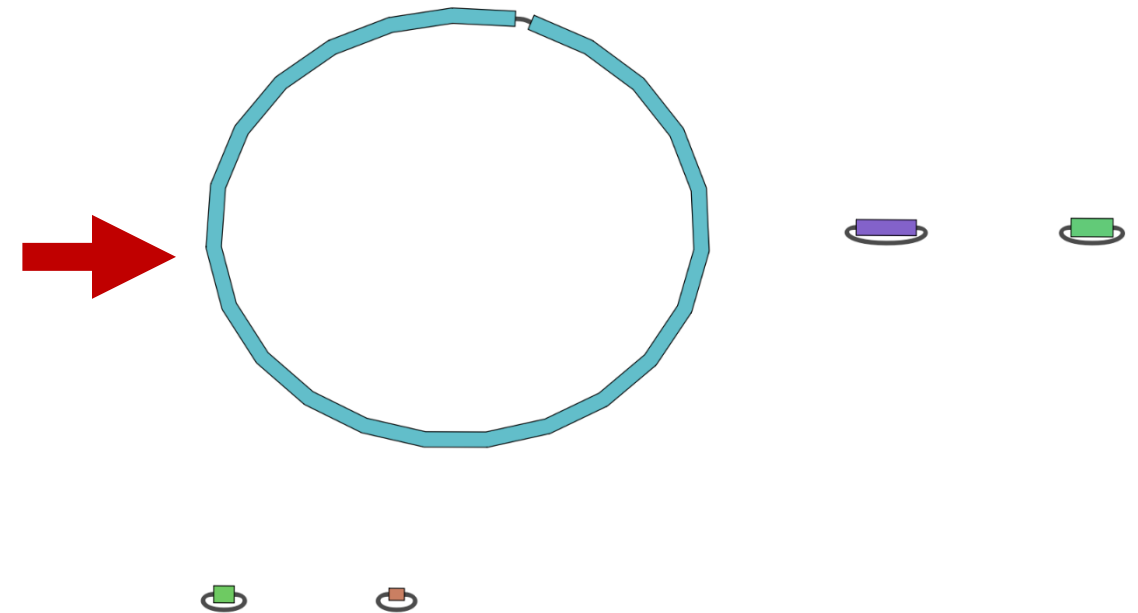


# Short-read vs long-read consensus

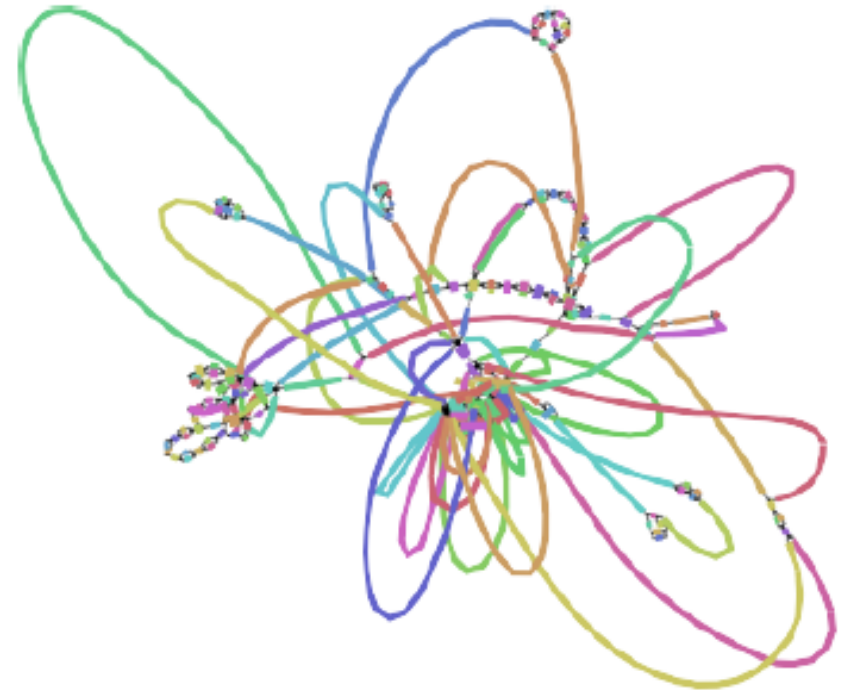
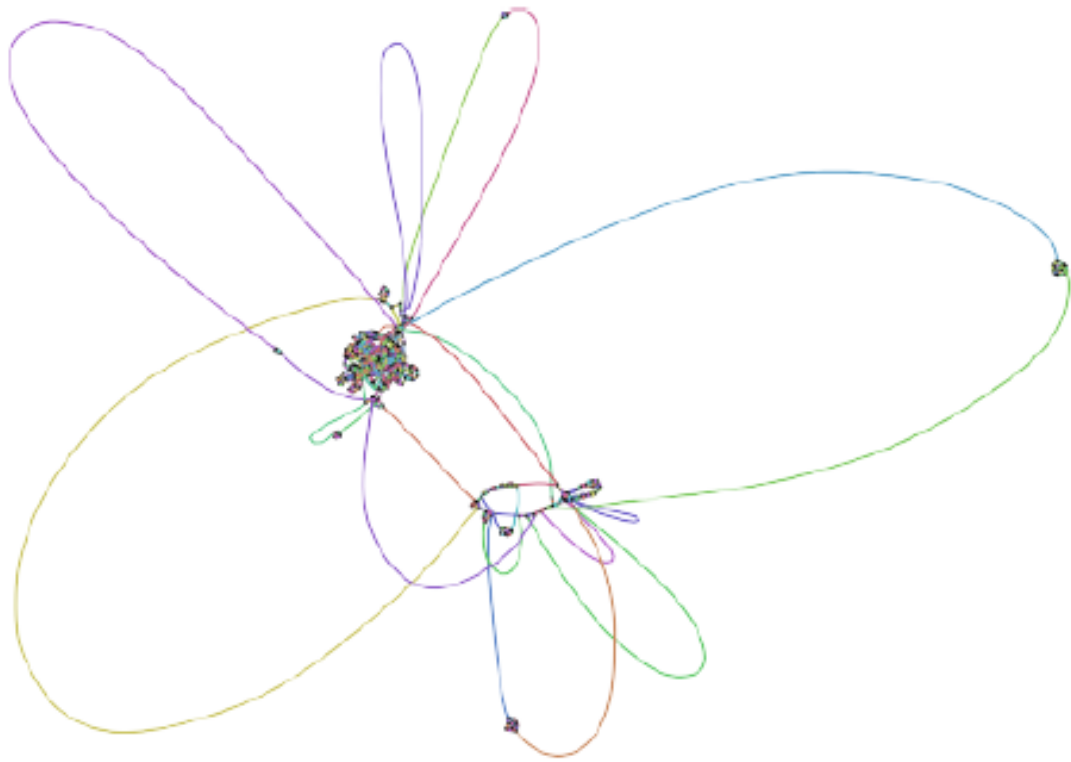


# Reconstruct bacterial Genomes

```
@NB501745:156:HJNNMBGXJ:1:11101:12420:1066 1:N:0:TCGCCTTA+CTAGTCGA
GTTTGCATCACGCCGTGCCTATGACTTTATTTGTATGGCAATTGCAGAGGATAACCTGNACGTCAAATTTGTGGCGGCACT
+
A/AAAAEEEEEEAEEEEEEEEEEEEEEEEEEEEEEEE/EEEEAE/EEEEEEE/6AEA#AEAEAAAAEEEEEEEE/EE<E
@NB501745:156:HJNNMBGXJ:1:11101:21970:1076 1:N:0:TCGCCTTA+CTAGTCGA
TTTAACTTGATGGTAGAAATACCTTTTCTTAATAGGCTTCATAAAAAATTCAAGTGCTTCCACCCGGTCTATTCCATA
+
AAAAAEAAAAAAAAAAAAAAAAEEEEAEAEAAAAEEEE66AAAAAAAA6<EAEA/EEEEAE/EEEEEEEE6A/
@NB501745:156:HJNNMBGXJ:1:11101:20441:1078 1:N:0:TCGCCTTA+CTAGTCGA
TGGAGATTTGGTCTATGTTGAACTCCTGAAGTGGGTAGCCAAGTAACAGCTGGCGAACAAGCCGGAGTTGTTGAATCAGT
+
AAAAEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE/EEAAAAEE
@NB501745:156:HJNNMBGXJ:1:11101:16251:1078 1:N:0:TCGCCTTA+CTAGTCGA
GTTTAAAGACGGCAGTTTCTGTTCAAGTGAAGGCTGACGCGGCGAAACCGTGATGCTGATATTGCTGCTTCATTCCAAC
+
AAAAEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE
@NB501745:156:HJNNMBGXJ:1:11101:17513:1079 1:N:0:TCGCCTTA+CTAGTCGA
CCCTGAATCAGGTATTTAACTATCTTGAATAAGTATAATAGCTTAGCGCTAAAAATAATGTTTTGATAGAGAAAGTTAC
+
AAAAEEEEEEEEEEEE/EEEEEEEEEEEEEEEEAAAAEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE/
@NB501745:156:HJNNMBGXJ:1:11101:20034:1080 1:N:0:TCGCCTTA+CTAGTCGA
CAGCGTACTTGTCAAATCAAACATACTGCTACAAGGCTACGCCAAGAAAAATTTAAAAACGTAGTTGCATATGCA
+
AAAAEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE
@NB501745:156:HJNNMBGXJ:1:11101:16704:1082 1:N:0:TCGCCTTA+CTAGTCGA
```



# Short Read Graph



# Tools

**Flye**

**SPAdes**



**Unicycler**

**Trycycler**



**Polypolish**



**Quast**



**Bandage**

a Bioinformatics Application for Navigating *De novo* Assembly Graphs Easily