1. This study examines the effect of *k*-mer length (*k*) and read-pair distance (*d*) on assembling the *Carsonella ruddii* genome using the de Bruijn graph approach. Testing k=10,100,1000 across d=[1,100] revealed that k=10 failed at all d values due to high ambiguity. In contrast, k=100,1000 successfully reconstructed the genome even at d=0, showing that larger *k*-mers provide sufficient context for assembly without paired-read constraints. These results emphasize that choosing a sufficiently large *k* is key to accurate genome reconstruction, while *d* is only useful when *k* is too small.

```
● gafur@gafur-HP-Spectre-x360-Convertible-13-ap0xxx:~/Documents/computational_biology$ /bin/python3 /home/gafur/
  hw3.py
  (k,d)-mer composition Results saved to /home/gafur/Documents/computational_biology/hw3/k_d_mers.txt
  Testing with k= 10
  Incorrectly constructed sequence
● gafur@gafur-HP-Spectre-x360-Convertible-13-ap0xxx:~/Documents/computational_biology$ /bin/python3 /home/gafur/
  hw3.py
  (k,d)-mer composition Results saved to /home/gafur/Documents/computational_biology/hw3/k_d_mers.txt
  Testing with k= 100
  Correctly constructed sequence
● gafur@gafur-HP-Spectre-x360-Convertible-13-ap0xxx:~/Documents/computational_biology$ /bin/python3 /home/gafur/
  hw3.py
  (k,d)-mer composition Results saved to /home/gafur/Documents/computational_biology/hw3/k_d_mers.txt
  Testing with k= 1000
  Correctly constructed sequence
○ gafur@gafur-HP-Spectre-x360-Convertible-13-ap0xxx:~/Documents/computational_biology$ ▯
```

```
  (k,d)-mer composition Results saved to /home/gafur/Documents/computational_biology/hw3/k_d_mers.txt
  Testing with k= 10 d= 99
  Incorrectly constructed sequence
  (k,d)-mer composition Results saved to /home/gafur/Documents/computational_biology/hw3/k_d_mers.txt
  Testing with k= 100 d= 0
  Correctly constructed sequence
  k=100|d=0
  (k,d)-mer composition Results saved to /home/gafur/Documents/computational_biology/hw3/k_d_mers.txt
  Testing with k= 1000 d= 0
  Correctly constructed sequence
  k=1000|d=0
  k=100|d=0
  k=1000|d=0
⊗ gafur@gafur-HP-Spectre-x360-Convertible-13-ap0xxx:~/Documents/computational_biology$ /bin/python3 /home/gafur/
  hw3.py
  (k,d)-mer composition Results saved to /home/gafur/Documents/computational_biology/hw3/k_d_mers.txt
```

2. Comparison of Standard and Paired de Bruijn Graph Approaches

This study evaluates genome assembly of *Carsonella ruddii* using both standard and paired de Bruijn graphs with varying k-mer sizes (k=10,100,1000). In both approaches, k=10 failed to reconstruct the genome due to high ambiguity and repetitive sequences. However, for k=100 and k=1000, the graphs were successfully constructed, demonstrating that a sufficiently large k ensures accurate genome assembly even without paired-end constraints. These findings confirm that **choosing an appropriate k is crucial, while read-pair information (d) primarily aids reconstruction when k is too small.**

```
144
145
146   def main():
147       file_path = "/home/gafur/Documents/computational_biology/hw3/sequence.fasta"
148       genome_sequence = read_fasta(file_path)
149
150       dna = kmer_composition(genome_sequence, 1000)
151       reconstructed = reconstruct_from_kmer(dna)
152       if reconstructed == genome_sequence:
153           print("correct")
154       else:
155           print("incorrect")
156
157       # k_values = [10, 100, 1000]
```

PROBLEMS   OUTPUT   DEBUG CONSOLE   TERMINAL   PORTS

```
● gafur@gafur-HP-Spectre-x360-Convertible-13-ap0xxx:~/Documents/computational_biology$ /bin/pytho
  hw3.py
● gafur@gafur-HP-Spectre-x360-Convertible-13-ap0xxx:~/Documents/computational_biology$ /bin/pytho
  hw3.py
  incorrect
● gafur@gafur-HP-Spectre-x360-Convertible-13-ap0xxx:~/Documents/computational_biology$ /bin/pytho
  hw3.py
  correct
● gafur@gafur-HP-Spectre-x360-Convertible-13-ap0xxx:~/Documents/computational_biology$ /bin/pytho
  hw3.py
  correct
○ gafur@gafur-HP-Spectre-x360-Convertible-13-ap0xxx:~/Documents/computational_biology$ ▯
```