

Ćwiczenia7_Assembly-MT

Zad1

część 1

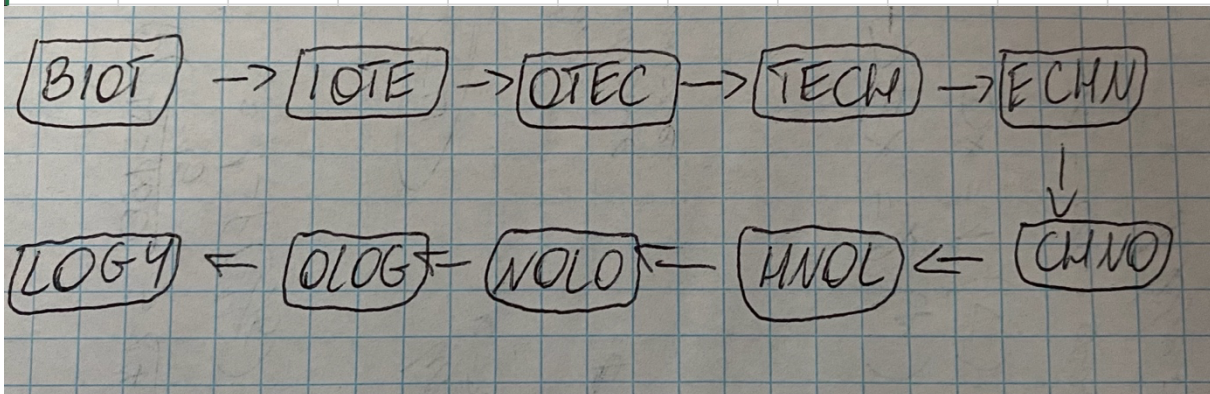
- a) najmniejszą wartością w Phred jest 0, w Illumina 1,8+ zamiast "0" stosuje się "!".
- b) jakość jest równa 1 co przekłada się na liczbę 40
- c) 99,99%

część 2

- a) Użyto Sanger/Illumina 1.9
- b) lepszą jakością charakteryzowała się sekwencja z pliku read_1. Jest to widocznie szczególnie na końcu obu sekwencji – sekwencja z pliku read_2 ma wiele powtórzeń w jakości co mogło być powodem zmniejszonej jakości. Jest to znacząca różnica

Zad2

	BIOT	IOTE	OTEC	TECH	ECHN	CHNO	HNOL	NOLO	OLOG	LOGY
BIOT		*								
IOTE			*							
OTEC				*						
TECH					*					
ECHN						*				
CHNO							*			
HNOL								*		
NOLO									*	
OLOG										*
LOGY										



Zad3

Polecenie wykonałem za pomocą kodu w pythonie. W konsekwencji udało mi się uzyskać następujące rezultaty:

sekwencja = ATGATTTCTCTCCTCGATTCCGCCAATC

kmery:

'ATGATT': 1, 'TGATTT': 1, 'GATTTCT': 1, 'ATTCTC': 1, 'TTCTCT': 1, 'TCTCTC': 1, 'CTCTCT': 1, 'CTCTCT': 1, 'CTCTCC': 1, 'TCTCCT': 1, 'CTCCTC': 1, 'TCCTCG': 1, 'CCTCGA': 1, 'CTCGAT': 1, 'TCGATT': 1, 'CGATTC': 1, 'GATTCC': 1, 'ATTCCG': 1, 'TTCCGC': 1, 'TCCGCC': 1, 'CCGCCA': 1, 'CGCCAA': 1, 'GCCAAT': 1, 'CCAATC': 1}

```
{('TTTCC', 'TTCCT'), ('ATTTC', 'TTTCC'), ('TTCCT', 'TCCTC'), ('GATTT', 'ATTTC'), ('TCCGC', 'CCGCC'), ('CTCCT', 'TCCTC'), ('CTCTC', 'TCTCC'), ('CTCGA', 'TCGAT'), ('ATTCC', 'TTCCG'), ('TGATT', 'GATTT'), ('TCGAT', 'CGATT'), ('TTCCG', 'TCCGC'), ('CCTCG', 'CTCGA'), ('CGCCA', 'GCCAA'), ('CCTCT', 'CTCTC'), ('CGATT', 'GATTC'), ('GATTC', 'ATTCC'), ('GCCAA', 'CCAAT'), ('TCTCC', 'CTCCT'), ('TCCTC', 'CCTCG'), ('ATGAT', 'TGATT'), ('TCCTC', 'CCTCT'), ('CCGCC', 'CGCCA')}
```

contig 3: TCCTCGATTCCGCCAATC

The diagram shows a blue rectangular node labeled 'contig 1' on the left. Two blue arrows originate from the right side of this node. The upper arrow points to a blue rectangular node labeled 'contig 2' on the top right. The lower arrow points to a blue rectangular node labeled 'contig 3' on the bottom right.

contig 2

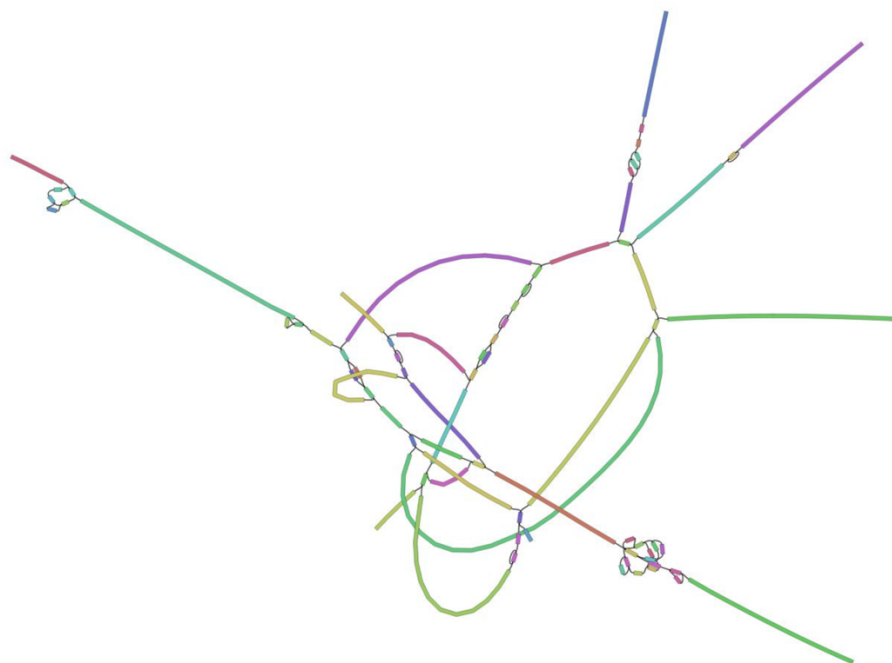
contig 3

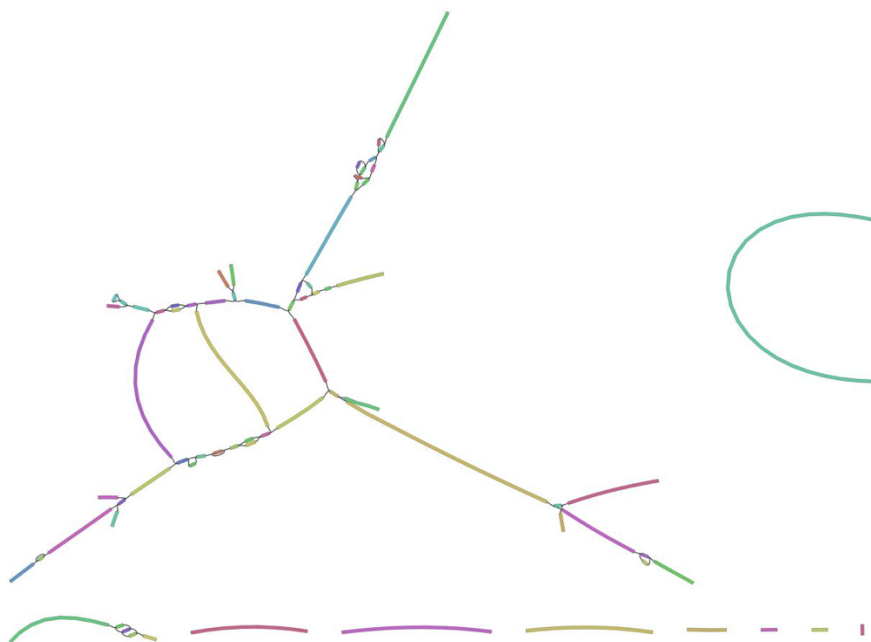
graph:

Dodatkowo sposób przedstawienia sekwencji może być umyślnie skomplikowany w celu odnalezienia miejsc z najlepszym połączeniem tak jak w przypadku grafu de Bruina.

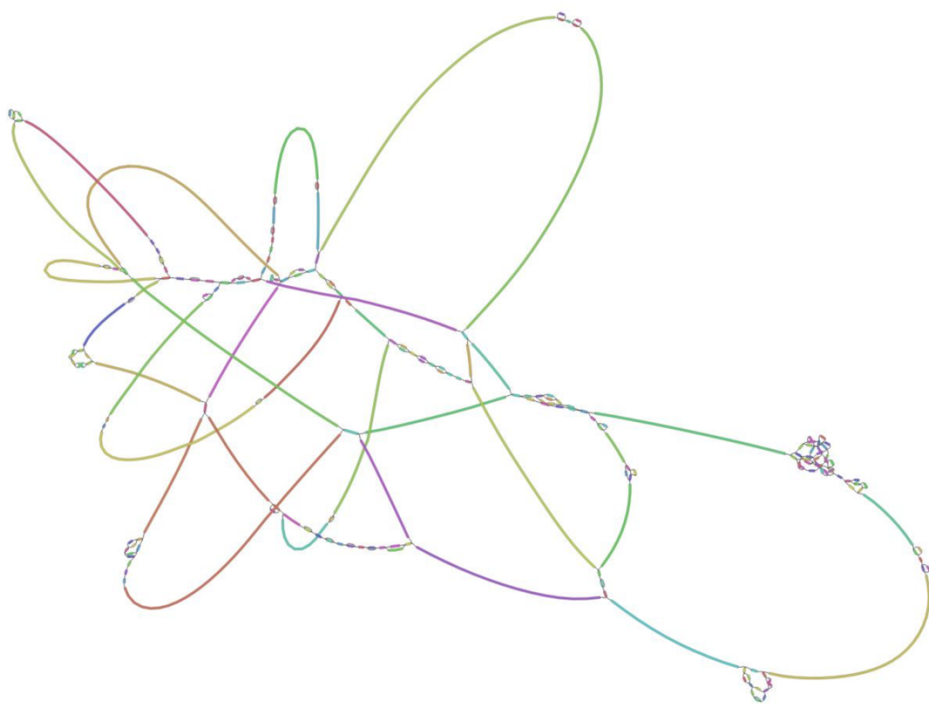
zad4.

Przedstawione poniżej zdjęcia przedstawiają grafy wytworzone za pomocą narzędzia bandage

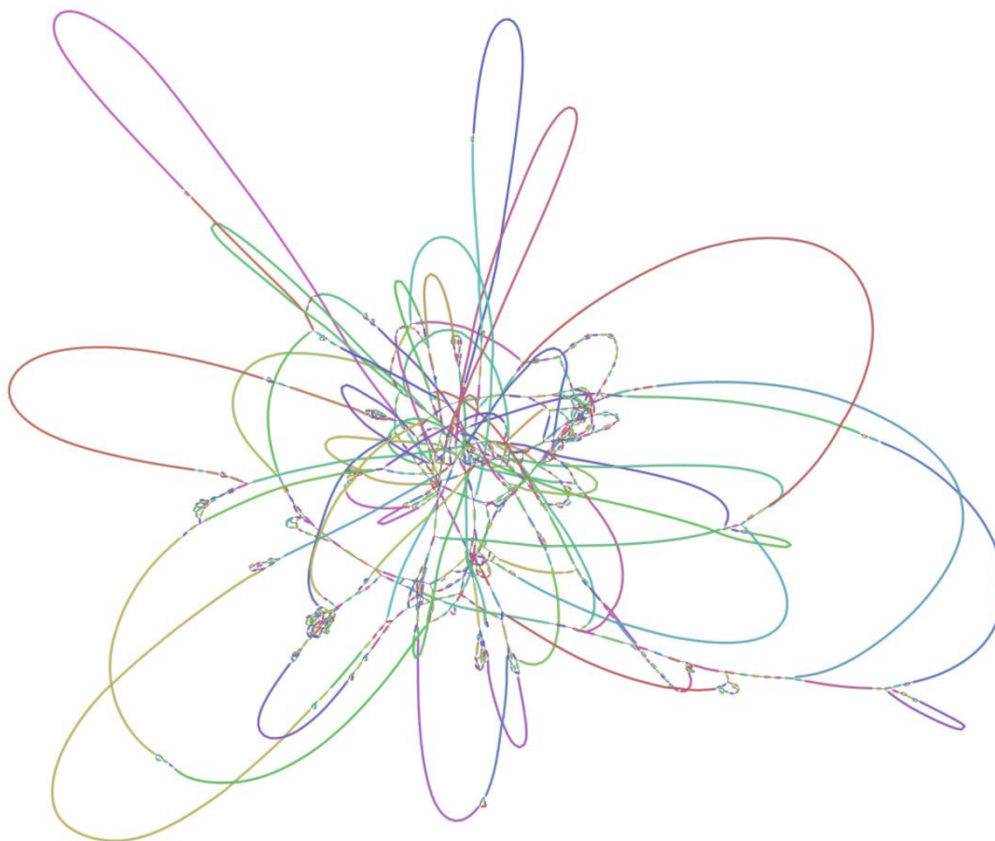




89



31



19

a)

Grafy różnią się ilością powstałych przyrównań na podstawie podobieństw sekwencji, ma to związek z parametrem k . Mniejsze k a co za tym idzie krótsze fragmenty łączą się znacznie częściej, ponieważ zwiększa się prawdopodobieństwo odnalezienia podobnych sekwencji.

b)

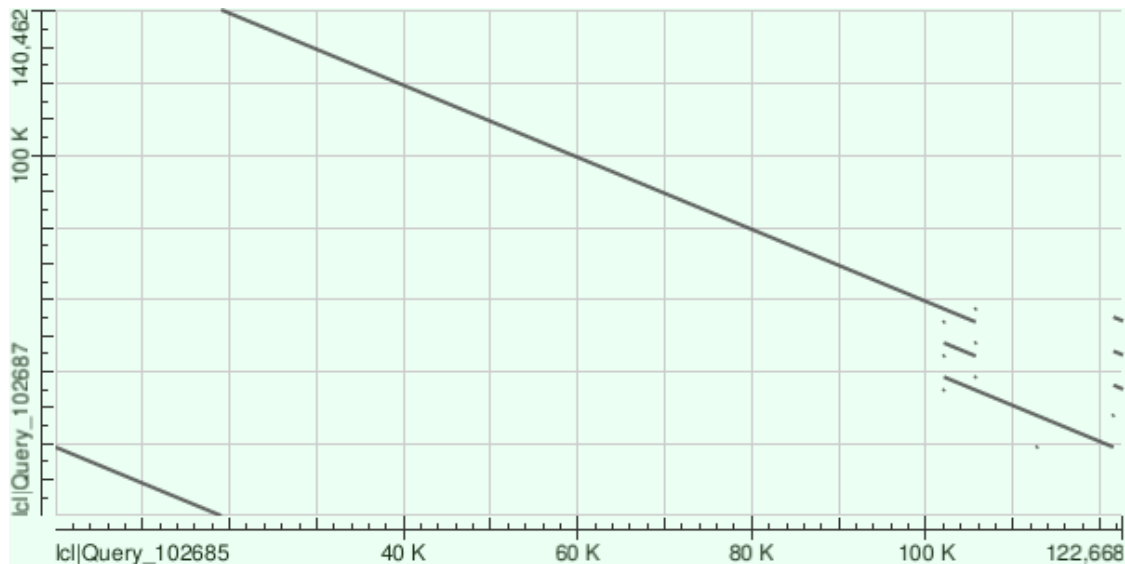
Wynika z tego, że różnice są związane z długością fragmentów (wartością k). k równe 89 czy 63 posiadają znacznie mniej połączeń, w samym grafie $k = 89$ możemy zauważyć nie połączone fragmenty co oznacza ich unikatowość w sekwencji.

zad.5

rodzaj i gatunek: *Klebsiella Pneumoniae*

kod NCBI: LR890743.1

pełna nazwa: *Klebsiella pneumoniae* isolate INF336-sc-2280147 genome assembly



Zgodnie zasadami odczytywania wykresu typu DotPlot musimy wyodrębnić pojawiające się na wykresie elementy: kropki oraz linie (linie powstają w skutek połączenia kropek). Kropki zaznacza się w miejscach, gdzie doszło do nałożenia się na siebie nukleotydów z obu sekwencji. Jak możemy zauważyć na powyższym wykresie sekwencje te są wysoce podobne. Sugeruje to pojedynczą linię biegnącą przez cały wykres. dodatkowo możemy to odczytać po niskiej wartości e-value. Jednakże na odcinku między 102K a 108K widoczne jest powtórzenie w sekwencjach, co może sugerować, że sekwencje te są do siebie odwrotnie ułożone. Najbardziej prawdopodobnym scenariuszem jest wystąpienie pewnej delekcji w tym odcinku.