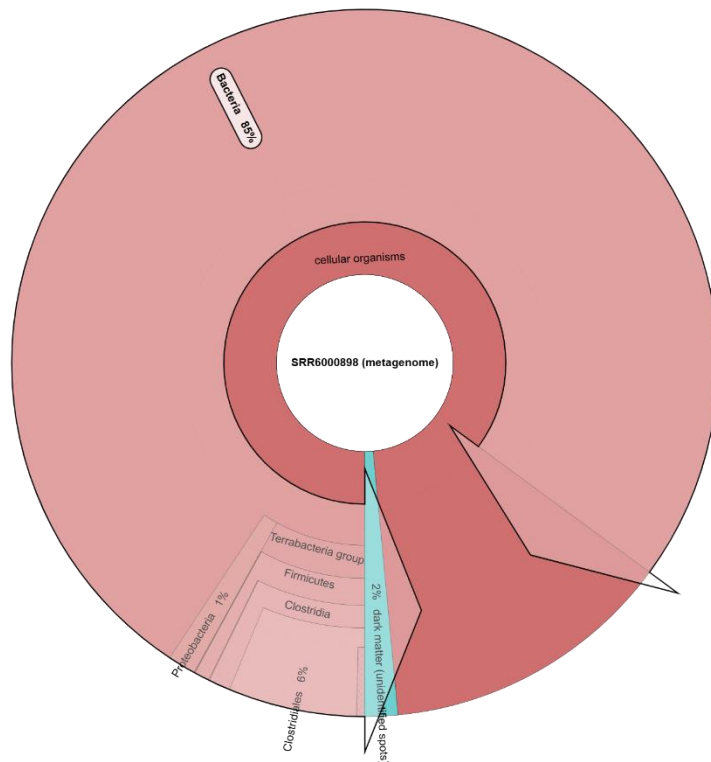


- 172 próbki, sprzęt illumina, Instrument -Illumina MiSeq, NextSeq 500

- 86% of cellular organisms
- 85% of SRR6000898 (metagenome)



- Pojawiły się 4 kolekcje, Pair-end data (fasterq-dump) zawiera dane, pozostałe 2 kolekcje są puste, a fasterq-dump log zawiera logi z przetwarzania danych.

- ### Nieznaczne różnice w kształcie dystrybucji duplikacji w sekwencji (Sequence Duplication Levels)

5. Umieść w raporcie pierwszy sklasyfikowany (Classification) odczyt oraz opisz jego klasyfikację: taksonomię do której należy oraz układ mapowanych k-merów

Column 1	Column 2	Column 3	Column 4	Column 5
C	SRR6000935.1	821	151 151	0:2 2:18 0:8 976:3 0:15 909656:3 0:3 909656:5 2:22 909656:5 0:3 821:13 0:17 0:117
C	SRR6000935.2	853	151 151	853:2 2:2 853:5 186801:1 2:14 853:34 2:3 853:18 2:4 853:2 2:5 853:5 2:22 186826:1 0:37 2:5 0:11 2:25 0:38
C	SRR6000935.3	841	151 151	186802:9 2:3 186802:7 2:3 0:33 268590:5:7 186802:1 0:5 186802:3 0:7 186803:1 0:15 841:1 0:5 186803:1 0:15 186803:1 1:10 871:5 87801:1 0:10 0:0 0:1 11067:0 0:10 0:0 1110:5 0:0 186803:0 0:14

Id taksonomii jest w 3 kolumnie - 821 i odpowiada [Phocaeicola vulgatus](#) species, CFB group bacteria. Układ mapowanych k-merów kolumna 5:

0:2 2:18 0:8 976:3 0:15 909656:3 0:3 909656:5 2:22 909656:5 0:3 821:13 0:17 |:| 0:117

Każda para takson_id:liczba_k-merów reprezentuje:

- **takson_id** – unikalny identyfikator taksonomiczny (np. gatunek, rodzaj, rodzina itp.).
 - 0 oznacza brak przypisania k-merów (nieprzypisane).
- **liczba_k-merów** – liczba k-merów przypisanych do danego taksonu.

Interpretacja:

1. 0:2 – 2 k-merów nieprzypisane.
2. 2:18 – 18 k-merów przypisanych do taksonu o ID 2.
3. 0:8 – 8 k-merów nieprzypisanych.
4. 976:3 – 3 k-merów przypisane do taksonu o ID 976.
5. 909656:3 – 3 k-merów przypisane do taksonu o ID 909656, itd.

Separator "|:|":

Dzieli szczegółowe mapowanie k-merów od statystyk podsumowujących.

Sekcja podsumowująca po "|:|":

"0:117"

- Reprezentuje ogólną liczbę k-merów nieprzypisanych.
 - **0:117** – łącznie 117 k-merów nie przypisano do żadnego konkretnego taksonu

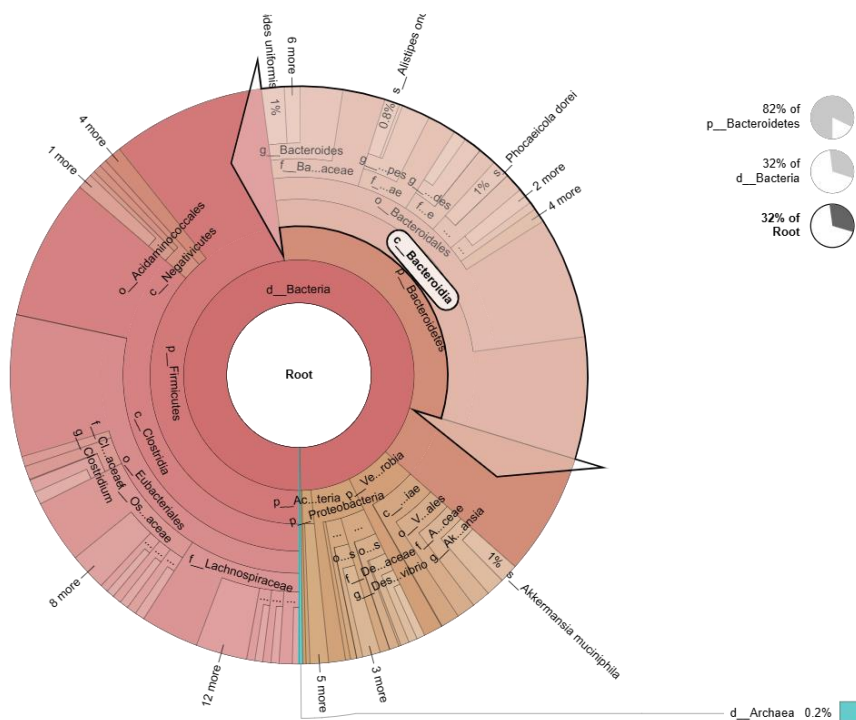
6. Opisz w raporcie zawartość plików "Report". Której klasyfikacji jest najwięcej?

W pliku raport znajdziemy szczegółową tasonomie próbki wraz z ilością poszczególnych kładów.

Najwięcej jest bakterii gatunku:

d_Bacteria p_Bacteroidetes c_Bacteroidia o_Bacteroidales g_Phocaeicola s_Phocaeicola vulgatus

7. Umieść w raporcie wykres Krona. Porównaj z uzyskanym w NCBI. Opisz różnice i ich źródło



1. Główne różnice między wykresami:

Zakres zidentyfikowanych taksonów:

- NCBI: Zawiera mniej szczegółowych informacji, skupiając się na wyższych poziomach taksonomicznych, takich jak "cellular organisms" czy ogólne "Bacteria". Większość danych przypisana jest do jednego dominującego taksonu, a niektóre fragmenty są oznaczone jako "dark matter" (niezidentyfikowane k-mery).

- Kraken2: Oferuje znacznie większą szczegółowość, rozbijając skład taksonomiczny na poziomie rzędów, rodzin i rodzajów (np. Bacteroides, Lachnospiraceae). Widać tu wyraźnie, które grupy bakteryjne dominują w próbce.

Prezentacja wyników:

- NCBI ma bardziej uproszczony układ, co sugeruje analizę o mniejszej rozdzielczości.

- Kraken2 zawiera szczegółowe dane, wskazując na zastosowanie bardziej dokładnej metody analizy - bardziej specyficznego algorytmu klasyfikacji.

Zróżnic najpewniej wynika z tego że NCBI użyło uproszczonej bazy danych a Kraken 2 wzbogacił je o dodatkowe k-mery.