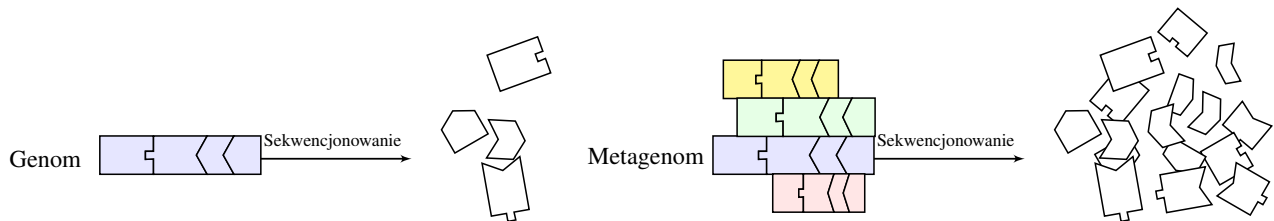


Procesy reprodukcji i regulacji funkcji życiowych są w przyrodzie koordynowane poprzez system informacji, w którym kluczową rolę odgrywa *genom* kodowany za pomocą kwasu dezoksyrybonukleinowego (DNA). Z uwagi na to, że obecnie nie istnieją rozwiązania technologiczne umożliwiające poprawny odczyt długich sekwencji DNA w całości, są one najpierw powielane, a następnie losowo cięte na odpowiednio krótkie fragmenty, które mogą zostać odczytane — proces ten nazywany jest *sekwencjonowaniem* DNA. Pojedynczy fragment sekwencji nazywany jest odczytem (ang. *read*). Metody sekwencjonowania DNA są obecnie na tyle zautomatyzowane i wydajne, że z dnia na dzień rośnie liczba poznanych sekwencji.

Genom to cała informacja genetyczna pojedynczego organizmu. Możliwa jest jednak również analiza naraz całego zbioru genomów pochodzących od wszystkich organizmów zamieszkujących dane środowisko. Zbiór takich genomów zwany jest *metagenomem*.

W wyniku sekwencjonowania metagenomu otrzymywany jest zbiór pomieszanych odczytów pochodzących z sekwencji DNA wszystkich mikroorganizmów znajdujących się w badanej próbce środowiskowej. Różnice pomiędzy sekwencjonowaniem pojedynczego genomu oraz metagenomu są zilustrowane na rysunku poniżej.



Największą zaletą analizy metagenomu jest brak konieczności izolacji i hodowania mikroorganizmów w laboratorium. Jest to o tyle ważne, że większości mikroorganizmów nie da się wyhodować w warunkach sztucznych. Analiza metagenomu pomaga rozwiązać wiele praktycznych problemów w medycynie, inżynierii, rolnictwie, jak i w ekologii. W ostatnich latach, metody analizy odczytów metagenomowych były intensywnie badaniem zagadnieniem bioinformatycznym, w wyniku czego opracowanych zostało wiele metod oraz narzędzi służących do przetwarzania danych metagenomowych. Większość opublikowanych prac dotyczy klasyfikacji badanych fragmentów metagenomu do określonych kategorii taksonomicznych lub grupowaniu ich ze względu na wzajemne podobieństwo.

Warto odnotować, że całkowita liczba genów człowieka stanowi tylko 1% genów mikrobów zamieszkujących ludzki organizm. Istnieje wiele badań pokazujących zależność stanu zdrowotnego (fizycznego oraz psychicznego) człowieka od występowania danego mikrobiomu (głównie w jelitach). Przykładowo, wyniki niektórych badań wskazują na powiązanie ryzyka wystąpienia otyłości u dzieci ze zmianami w mikroflorze jelit w okresie niemowlęcym. Należy przy tym zaznaczyć, że metagenomika jest młodą dziedziną nauki i do tej pory prowadzonych było niewiele badań mających na celu umożliwienie wykorzystania tej wiedzy w praktyce. Diagnostyka uwzględniająca analizę mikroflory biologicznej wymaga obecnie przeprowadzania czasochłonnych posiewów, natomiast rozwój metod analizy metagenomowej pozwoli w przyszłości znacznie przyspieszyć ten proces. To z kolei może przełożyć się na zwiększenie skuteczności leczenia.

Celem naukowym projektu jest opracowanie nowych metod służących do efektywnej nadzorowanej i nienadzorowanej klasyfikacji odczytów uzyskanych w wyniku sekwencjonowania metagenomu. Metody te mogą okazać się niezwykle użyteczne w wielu dziedzinach, w których jest lub może być wykorzystywana analiza metagenomów, takich jak inżynieria, rolnictwo, ekologia, czy medycyna. Dokładniej, stworzone algorytmy będą mogły zostać zastosowane do rozwiązywania szeregu zadań bioinformatycznych, takich jak:

- ∞ określanie składu jakościowego i ilościowego próbki pobranej z badanego środowiska,
- ∞ badanie korelacji w występowaniu różnych gatunków mikrobów,
- ∞ wykrywanie nowych gatunków mikrobów (w szczególności występujących w środowiskach cechujących się ekstremalnymi warunkami do życia),
- ∞ badanie stanu zdrowia pacjenta na podstawie pobranej od niego próbki metagenomowej.