

Autor Bartosz Czech	Opiekun naukowy dr Magda Mielczarek dr Paweł Sztromwasser
------------------------	---

**OPTIMALIZACJA SKŁADANIA GENOMÓW *DE NOVO* W OPARCIU O DANE
POCHODZĄCE Z SEKWENCJONOWANIA DRUGIEJ I TRZECIEJ GENERACJI**

**OPTIMISATION OF *DE NOVO* GENOME ASSEMBLY BASED ON SECOND- AND
THIRD-GENERATION SEQUENCING DATA**

Streszczenie

Technologie sekwencjonowania DNA nie są w stanie odczytać całych genomów za jednym zamachem, lecz produkują fragmenty DNA o długości od około kilkuset par zasad (sekwencjonowanie II generacji, np. platforma Illumina) do kilkuset tysięcy par zasad (sekwencjonowanie III generacji, np. PacBio). Składanie genomów *de novo* odnosi się do procesu ułożenia i połączenia krótkich sekwencji DNA w takiej kolejności, w jakiej występowały w genomie, z którego to DNA pochodzi. Wedle najnowszych badań, hybrydowe składanie genomów bakteryjnych, które jest oparte na różnych metodach sekwencjonowania genomów, pozwala na uzyskanie dokładnych wyników, przy zmniejszonych kosztach sekwencjonowania. W tym projekcie skupiliśmy się na porównaniu różnych metod składania genomów bakteryjnych *de novo* oraz na wyznaczeniu równowagi pomiędzy kosztami sekwencjonowania, a dokładnością złożenia genomu.

Materiał do badań stanowiły dwa szczepy *Salmonella enterica* oraz sześć szczepów *Escherichia coli*, które były sekwencjonowane metodami Illumina (krótkie i dokładne odczyty) oraz PacBio (dłuższe odczyty, z częstszymi błędami). Zastosowano dwa podejścia: (i) składanie dłuższych odczytów PacBio, a następnie skorygowanie błędów krótkimi odczytami Illumina, oraz (ii) złożenie krótkich odczytów Illumina, a następnie uzupełnienie brakujących sekwencji pomiędzy złożonymi fragmentami genomu (kontigami) odczytami PacBio. Wstępne wyniki wskazują, że składanie genomów bakteryjnych metodą hybrydową daje spójne rezultaty, jak również, że istnieje korelacja pomiędzy głębokością sekwencjonowania a dokładnością składania genomów *de novo*.