

Raport końcowy realizacji projektu

Technologia Oxford Nanopore: optymalizacja enzymów oraz analizy danych genomicznych pod kątem zastosowań komercyjnych

Anna Czmil^a, Małgorzata Semik^b, Marta Sochacka-Piętał^b, Magdalena Toton^a, Michał Ćmil^a, Michał Piętał^b, Dominik Strzałka^{a*}, Michał Wroński^a

^a Zakład Systemów Złożonych, Politechnika Rzeszowska, Al. Powstańców Warszawy 12, 35-959 Rzeszów, Polska

^b Zakład Biotechnologii i Bioinformatyki, Politechnika Rzeszowska, Al. Powstańców Warszawy 12, 35-959 Rzeszów, Polska

* strzalka@prz.edu.pl

Projekt był realizowany w okresie 02 stycznia - 30 września 2020 r. Pierwotnie realizacja projektu miała trwać 6 miesięcy (harmonogram zakładał dwa etapy, do 30 czerwca 2020 r.), ale w związku z wystąpieniem epidemii COVID-19, etap drugi został wydłużony o 3 miesiące. W pierwszej fazie projektu przystąpiono do organizacji zespołu projektowego w związku z tym, iż a projekcie przewidziano także udział studentów oraz doktoranta. W wyniku ogłoszonych konkursów (<https://zsz.prz.edu.pl/aktualnosc/konkurs-dla-studentow-i-doktorantow-na-udzial-w-projekcie-naukowym-88.html>) wyłoniono dwie osoby (mgr inż. Anna Czmil i inż. Michał Ćmil), które zostały począwszy od 22 stycznia 2020 r. zaangażowane do pracy w projekcie. Mając na uwadze fakt, iż w projekcie zaangażowane były osoby reprezentujące zarówno specjalistów z zakresu informatyki jak i biologii w lutym 2020 r. zorganizowane seminarium naukowe na temat „Analizy bioinformatyczne genomów bakteryjnych pod kątem zastosowań komercyjnych”, które wygłosiła dr Marta Sochacka-Piętał z Zakładu Biotechnologii i Bioinformatyki Wydziału Chemicznego PRz (Rys. 1). Ponadto rozpoczęto realizację wszystkich niezbędnych procedur przetargowych w celu zakupu przewidzianych usług i wyposażenia. Wystąpienie epidemii COVID-19 spowodowało około dwumiesięczne opóźnienie w realizacji projektu w szczególności brak możliwości realizacji pierwszego eksperymentu sekwencjonowania. Niemniej jednak pod koniec maja 2020 r. udało się zrealizować niezbędne postępowanie przetargowe i w ramach projektu wykonano kolejno wszystkie przewidziane do realizacji zadania. W szczególności uzyskano dostęp do wiedzy z sekwencjonowania nowej generacji i analizy danych poprzez wykonanie eksperymentu sekwencjonowania, przez

Pracownię Sekwencjonowania DNA i Syntezy Oligonukleotydów, Instytut Biochemii i Biofizyki PAN, Warszawa. Sekwencjonowanie odbyło się w dwóch etapach, na miejscu w PRz w Rzeszowie dzięki osobistemu zaangażowaniu Pana mgr Jana Gawora, głównej osoby realizującej eksperyment. Było to pierwsze sekwencjonowanie metodą Oxford Nanopore na Politechnice Rzeszowskiej wykonane w dn. 17-23.06.2020 r. W dn. 08.06.20 odbyło się spotkanie przygotowujące ten eksperyment (Rys. 2). W spotkaniu wzięli udział dr Marta Sochacka-Piętał oraz mgr Małgorzata Semik (Zakład Biotechnologii i Bioinformatyki/WCh PRz), mgr inż. Michał Wroński (Zakład Systemów Złożonych/WEil PRz), mgr Jan Gawor, Pracownia Sekwencjonowania DNA i Syntezy Oligonukleotydów, Instytut Biochemii i Biofizyki PAN, Warszawa. Eksperyment podzielono na dwie części – doświadczalną oraz bioinformatyczną (Rys. 3), realizowane odpowiednio, w Zakładzie Biotechnologii i Bioinformatyki Wydziału Chemicznego oraz w Zakładzie Systemów Złożonych Wydziału Elektrotechniki i Informatyki. Sprawne przeprowadzenie sekwencjonowania, jak również prawidłową analizę bioinformatyczną uzyskanych danych, nadzorował mgr J. Gawor. Niezbędnej pomocy organizacyjno-technicznej udzielił także prof. dr hab. inż. Mirosław Tyrka z WCh. W części drugiej, bioinformatycznej, wymagane było zaangażowanie pozostałej części zespołu projektowego, w tym dr inż. Dominik Strzałka, dr Michał Piętał, mgr Magdalena Totoń, mgr inż. Michał Ćmil i mgr inż. Anna Czmił. Eksperyment trwał 5 dni roboczych: pierwsze dwa dni dotyczyły pracy w laboratorium mokrym, pozostałe dni przeznaczono na prace o charakterze bioinformatycznym w tym na analizy danych oraz ich przetwarzanie na serwerach Zakładu Systemów Złożonych.

Realizacja serii eksperymentów sekwencjonowania nie byłaby możliwa bez zakupów niezbędnego wyposażenia tj. MinION Starter Pack (łącznie 7 szt., Rys. 4) wraz z wyposażeniem dodatkowym w postaci tzw. kits.: CN1398 First Release Rapid barcoding Kit, Native Barcoding Expansion 112, Native Barcoding Expansion 1324, RAP Top Up Kit, Flow Cell Priming Kit oraz SPOT ON FLOW CELL MK 1 R9 VERSION, Field Sequencing Kit, Flow Cell Priming Kit, użycia kaset Blue Pippin, enzymu DNAza oraz wynajmu urządzenia Blue Pippin (Rys. 5).

W czasie realizacji projektu wyhodowano nowe mutanty termotolerancyjnego szczepu *Bacillus subtilis* MSP4 otrzymane na drodze mutagenезy chemicznej z wykorzystaniem bromku etyldyny (Rys. 6). Dla otrzymanych mutantów wykonano proces sekwencjonowania metodą ONT. Uzyskano pulę kilkunastu mutantów o zmienionej sekwencji genów, kodujących enzymy

o potencjalnym zastosowaniu komercyjnym, takich jak proteazy, desaturaza kwasów tłuszczowych, hydrolaza epoksydowa czy subtylizyna E. Na podstawie przeprowadzanych analiz bioinformatycznych (adnotacja uzyskanych sekwencji genomowych szczepu dzikiego oraz otrzymanych mutantów z wykorzystaniem narzędzi serwera RAST – Rys. 7) wytypowano do opatentowania szczepy mutantów *B.subtilis* MSP4, wytwarzających zmutowane enzymy.

Wykonano także szereg niezbędnych prac informatycznych pozwalających na implementację zarówno od strony programowej jak i sprzętowej serwera NanoForms do przetwarzania i analizy surowych danych bioinformatycznych. Przygotowany serwer (dostępny pod adresem <http://nanoforms.tech>) jest w stanie obsłużyć małe genomy (do 50 MB). Użytkownik serwera ładuje zarchiwizowany, pojedynczy plik sekwencji (fastq), następnie dane są wstępnie przetwarzane a użytkownik na bieżąco wybiera dostępne opcje, co pozwala na uzyskanie sekwencji DNA/RNA w postaci pliku fasta. Dla serwera NanoForms opracowano autorski algorytm pipeline przetwarzania danych z ONT (Rys. 8). W czasie budowy serwera, ponadplanowo rozszerzono jego funkcjonalność o tzw. złożenie hybrydowe (wraz z danymi z urządzenia Illumina), co znacząco poprawia ostateczną jakość sekwencjonowania. Opis całego serwera zostanie przedstawiony w publikacji (do zgłoszenia w czasopiśmie z IF) pod roboczym tytułem „NanoForms: an integrated server for processing, analysis and assembly of raw genomic data of prokaryotic species, from Oxford Nanopore technology”.

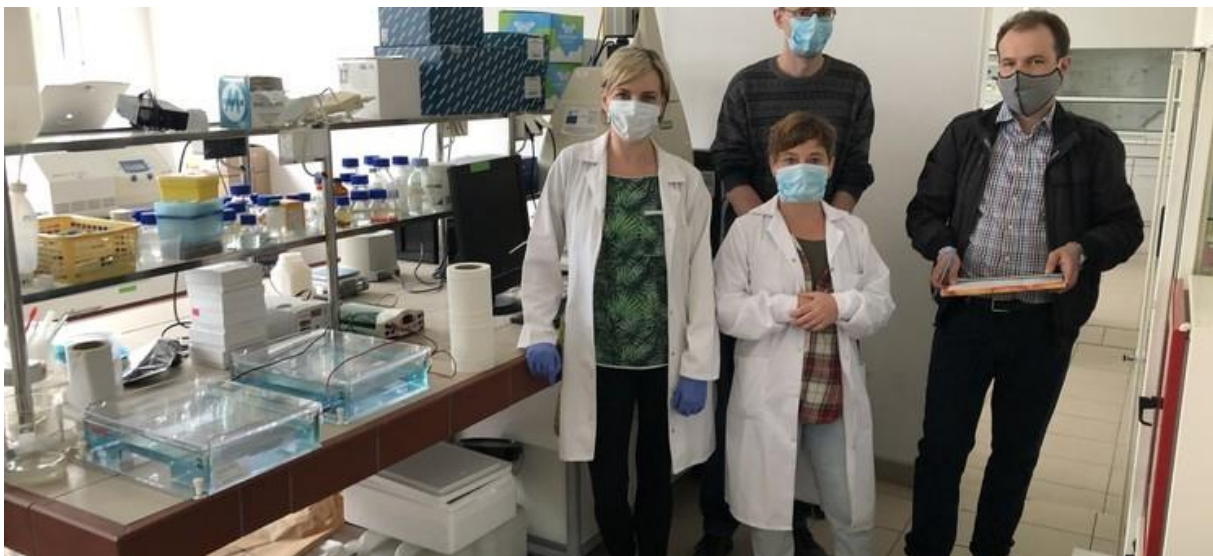
Jeden ze studentów pracujących w projekcie, Pan mgr inż. Michał Ćmil, po zakończeniu projektu oraz obronie pracy dyplomowej został zatrudniony od października 2020 w Politechnice Rzeszowskiej w Zakładzie Systemów Złożonych.

Zespół projektowy uważa, że grant zakończył się powodzeniem.

Projekt był finansowany przez Podkarpackie Centrum Innowacyjności w ramach grantu nr F3_116 (umowa nr 05/PRZ/1/DG/PCI/2019).



Rys. 1 Seminarium pt. Analizy bioinformatyczne genomów bakteryjnych pod kątem zastosowań komercyjnych



Rys. 2 Przygotowanie eksperymentu pierwszego sekwencjonowania.



Rys. 3 Realizacja prac w laboratorium mokrym – przygotowanie próbki do sekwencjonowania



Rys. 4 Zakupione w projekcie urządzenia MinION



Rys. 5 Urządzenie Blue Pippin



Rys. 6 Wyhodowany nowy szczep bakterii *Bacillus subtilis* MSP4

Genome	bacillus sp.
Domain	Bacteria
Taxonomy	Bacteria; bacillus sp.
Neighbors	View closest neighbors
Size	4,096,133
GC Content	43.8
L50	1
Number of Contigs (with PEGs)	1
Number of Subsystems	336
Number of Coding Sequences	4190
Number of RNAs	116

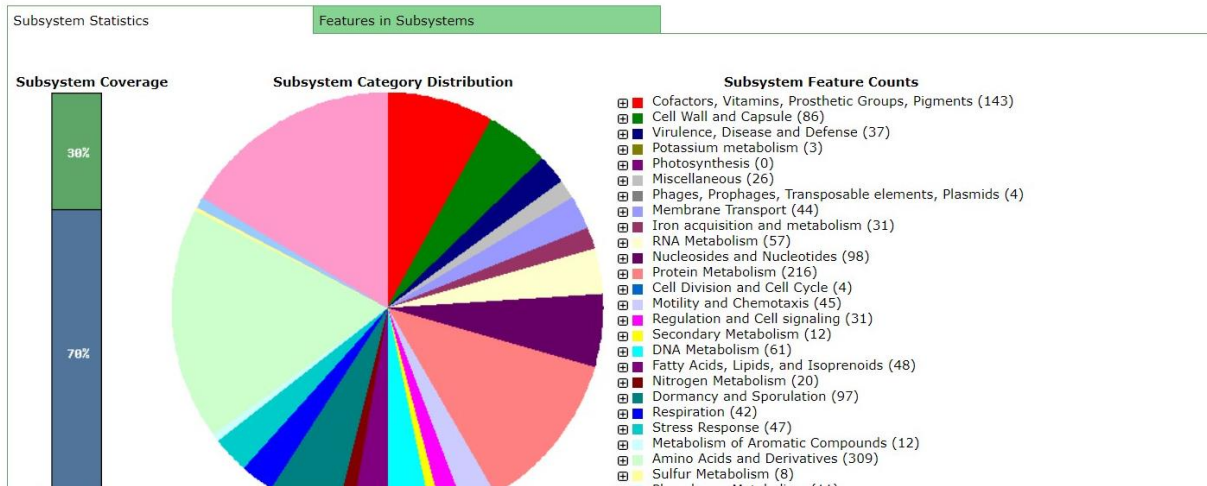
For each genome we offer a wide set of information to browse, compare and download.

Browse [Compare](#) [Download](#) [Annotate](#)

Browse through the features of [bacillus sp.](#), both graphically and through a table. Both allow quick navigation and filtering for features of your interest. Each feature is linked to its own detail page.

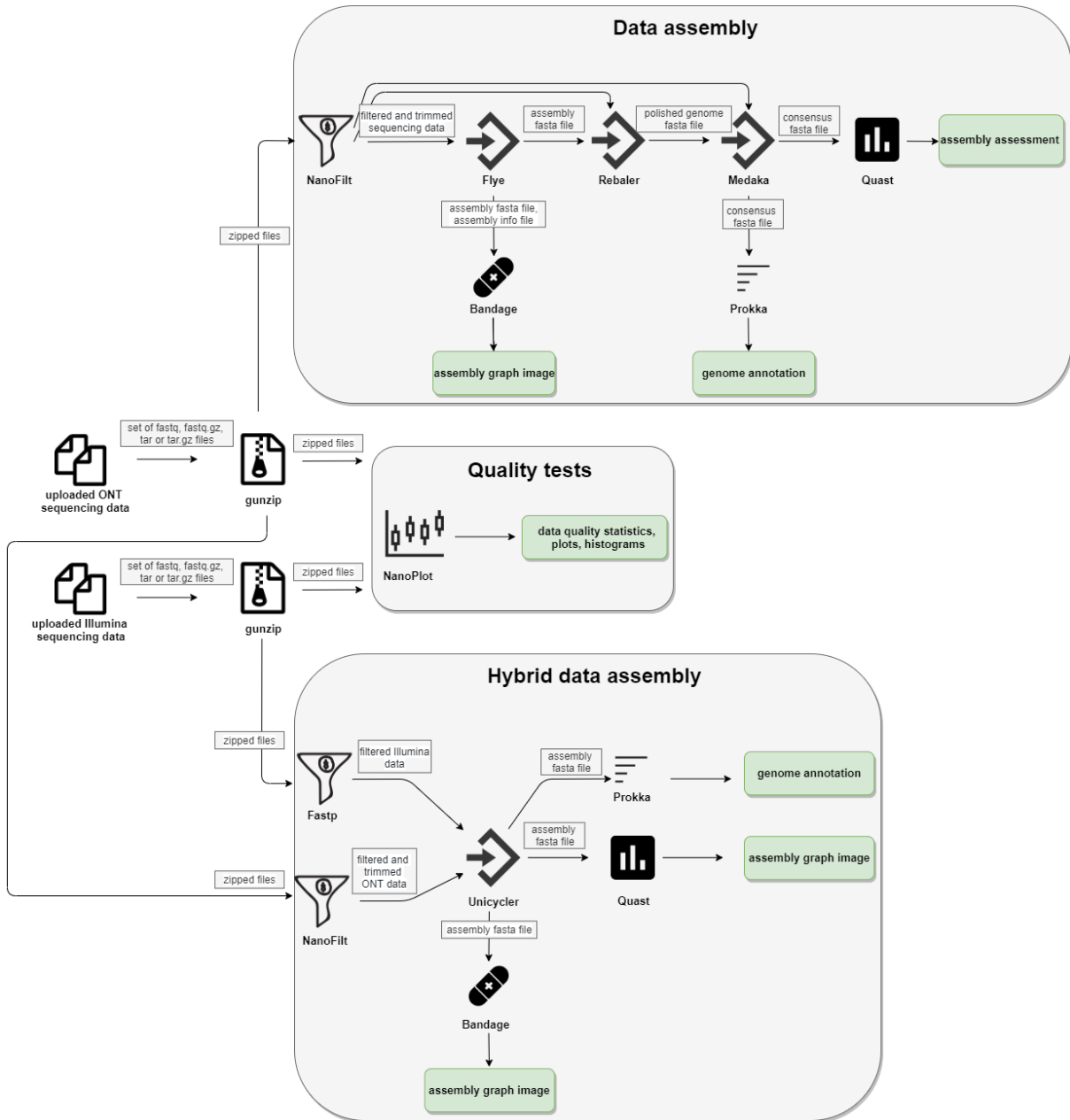
Click [here](#) to get to the Genome Browser

Subsystem Information



Rys. 7 Analiza porównawcza otrzymanych mutantów z wykorzystaniem narzędzi serwera

RAST



Rys. 8 Autorski pipeline przetwarzania na serwerze NanoForms