Final project report

Oxford Nanopore technology: enzyme optimization and genomic data analysis for commercial applications

Anna Czmil[a], Małgorzata Semik[b], Marta Sochacka-Piętal[b], Magdalena Totoń[a], Michał Ćmil[a], Michał Piętal[a], Dominik Strzałka[a*], Michał Wroński[a]

[a] Department of Complex Systems, Rzeszów University of Technology, Al. Powstańców Warszawy 12, 35-959 Rzeszów, POLAND

[b] Department of Biotechnology and Bioinformatics, Rzeszów University of Technology, Al. Powstańców Warszawy 12, 35-959 Rzeszów, POLAND

[*] strzalka@prz.edu.pl

The project was done between January 2 and September 30, 2020. Originally, the project should last 6 months (the schedule assumed two stages, until June 30, 2020), but due to the COVID-19 epidemic, the second stage was extended by 3 months. In the first phase of the project, the organization of the project team was started, due to the fact that the project also assumed active participation of students and a doctoral candidate. As a result of the announced competitions ((https://zsz.prz.edu.pl/aktualnosci/konkurs-dla-studentow-i-doktorantow-na-udzial-w-projekcie-naukowym-88.html)), two students were selected (MSc Eng Anna Czmil and Eng. Michał Ćmil) who were engaged to work in the project from January 22, 2020. Bearing in mind the fact that the project involved people representing both IT and biology specialists, in February 2020, a scientific seminar was organized on "Bioinformatic analysis of bacterial genomes for commercial applications", which was presented by Ph.D. Marta Sochacka-Piętal from the Department of Biotechnology and Bioinformatics, Faculty of Chemistry, Rzeszów University of Technology (Fig. 1). In addition, all necessary administrative procedures for the purchase of the services and equipment were started. The COVID-19 epidemic caused a delay of about two months in the project implementation, in particular the impossibility of the first sequencing experiment. But finally, at the end of May 2020, the necessary tender procedure was completed and all tasks planned in the project were carried out successively. The access to knowledge on next-generation sequencing and data analysis was obtained by carrying out the experiments of sequencing by the Laboratory of DNA

Sequencing and Oligonucleotide Synthesis, Institute of Biochemistry and Biophysics of the Polish Academy of Sciences, Warsaw. Sequencing took place in two stages, all done in Rzeszów, thanks to the personal involvement of Jan Gawor, the main person carrying out the experiment. It was the first sequencing with the Oxford Nanopore method at the Rzeszów University of Technology, performed on 17-23.06.2020 On. On June 8, a meeting to prepare this experiment was held (Fig. 2). The meeting was attended by Ph.D. Marta Sochacka-Piętal and MSc Małgorzata Semik (Department of Biotechnology and Bioinformatics/Faculty of Chemistry), MSc Eng. Michał Wroński (Department of Complex Systems/Faculty of Electrical and Computer Engineering, PRz), MSc Jan Gawor, Laboratory of DNA Sequencing and Oligonucleotide Synthesis, Institute of Biochemistry and Biophysics of the Polish Academy of Sciences, Warsaw. The experiment was divided into two parts - experimental and bioinformatics (Fig. 3), carried out, respectively, in the Department of Biotechnology and Bioinformatics of the Faculty of Chemistry and in the Department of Complex Systems of the Faculty of Electrical and Computer Engineering. Efficient sequencing as well as the correct bio-information analysis of the obtained data was supervised by J. Gawor. The necessary organizational and technical assistance was also provided by prof. Mirosław Tyrka from Faculty of Chemistry. The second part, bioinformatics, required the involvement of the rest of the project team, including PhD. Dominik Strzałka, PhD Michał Piętal, MSc Magdalena Totoń, MSc Michał Ćmil and MSc. Anna Czmil. The experiment lasted 5 working days: the first two days concerned work in a wet laboratory, the remaining days were devoted to bioinformatics work, including data analysis and data processing on Department of Complex Systems servers.

The series of sequencing experiments would not be possible without the purchase of the necessary equipment, i.e. the MinION Starter Pack (7 pcs in total, Fig. 4) with the additional equipment in the form kits: CN1398 First Release Rapid barcoding Kit, Native Barcoding Expansion 112, Native Barcoding Expansion 1324, RAP Top Up Kit, Flow Cell Priming Kit and SPOT ON FLOW CELL MK 1 R9 VERSION, Field Sequencing Kit, Flow Cell Priming Kit, Blue Pippin cassette, DNAse enzyme, and the rental of the Blue Pippin device (Fig. 5).

During the project, some new mutants of the thermotolerant Bacillus subtlis MSP4 strain were obtained by chemical mutagenesis using ethidium bromide (Fig. 6). The ONT sequencing was performed for the obtained mutants. A pool of a dozen mutants with altered gene sequence coding enzymes with potential commercial application, such as proteases, fatty

acid desaturase, epoxide hydrolase or subtilisin E, was obtained. On the basis of bioinformatics analyzes (annotation of the obtained genomic sequences of the wild strain and the obtained mutants using RAST server tools - Fig. 7) some strains of B.subtilis MSP4 mutants producing mutated enzymes were selected for patent.

A number of necessary IT works were also carried out, allowing for the implementation of both the software and hardware solutions of the NanoForms server for the processing and analysis of raw bioinformatics data. The prepared server (available at http://nanoforms.tech) is able to handle small genomes (up to 50 MB). The server user loads the archived, single sequence file (fastq), then the data is pre-processed and the user selects the available options on an ongoing basis, which allows obtaining the DNA/RNA sequence in the form of a fasta file. For the NanoForms server, a proprietary pipeline algorithm for data processing from ONT was developed (Fig. 8). During the construction of the server, its functionality was extended by the so-called hybrid assembly (together with Illumina data), which significantly improves the final sequencing quality. The description of the entire server will be presented in the publication with the working title “NanoForms: an integrated server for processing, analysis and assembly of raw genomic data of prokaryotic species, from Oxford Nanopore technology”.

One of the students working in the project, MSc. Michał Ćmil, after completing the project and defending his diploma thesis, was employed from October 2020 at the Rzeszów University of Technology in the Department of Complex Systems.

The project team believes that the grant was successful.

The project was financed by the Subcarpatian Innovation Center (PCI) under the grant no. F3_116 (contract no. 05/PRZ/1/DG/PCI/2019).

Fig. 1 Seminar entitled Bioinformatic analyzes of bacterial genomes for commercial applications

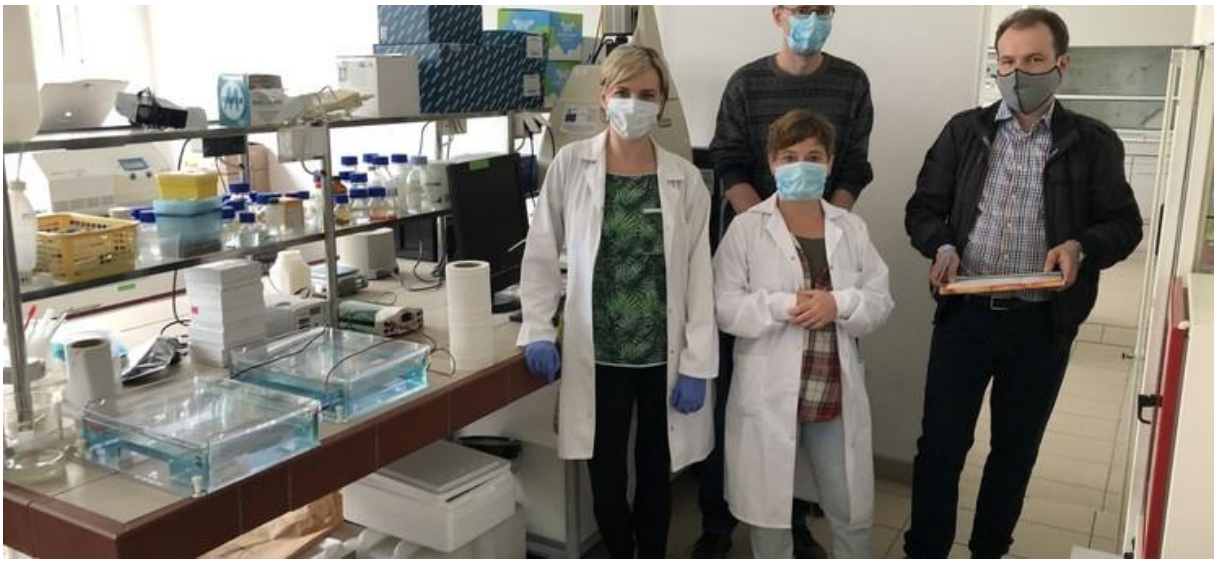


Fig. 2 Preparation of the first sequencing experiment.

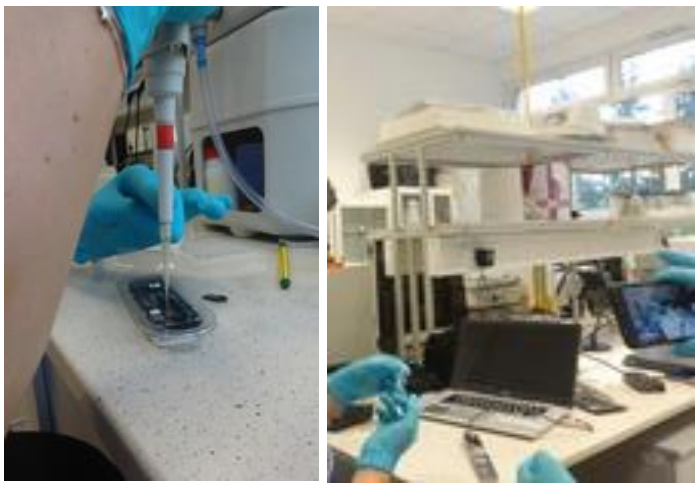


Fig. 3 Work in a wet laboratory - preparation of a sample for sequencing

Fig. 4 MinION devices purchased in the project

Fig. 5 The Blue Pippin device

Fig. 6 A culture of new bacterial strain of 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

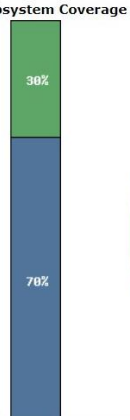
Browse through the features of bacillus sp. both graphically and through a table. Both allow quick navigation and filtering for features of your interest. Each feature is linked to its own detail page.

Click here to get to the Genome Browser

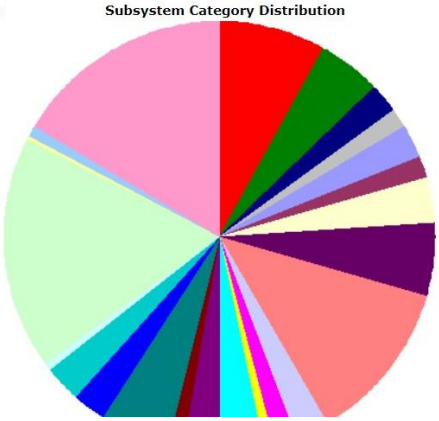
Subsystem Information

Subsystem Statistics | Features in Subsystems

Subsystem Coverage



Subsystem Category Distribution



Subsystem Feature Counts

- Cofactors, Vitamins, Prosthetic Groups, Pigments (143)
- Cell Wall and Capsule (86)
- Virulence, Disease and Defense (37)
- Potassium metabolism (3)
- Photosynthesis (0)
- Miscellaneous (26)
- Phages, Prophages, Transposable elements, Plasmids (4)
- Membrane Transport (44)
- Iron acquisition and metabolism (31)
- RNA Metabolism (57)
- Nucleosides and Nucleotides (98)
- Protein Metabolism (216)
- Cell Division and Cell Cycle (4)
- Motility and Chemotaxis (45)
- Regulation and Cell signaling (31)
- Secondary Metabolism (12)
- DNA Metabolism (61)
- Fatty Acids, Lipids, and Isoprenoids (48)
- Nitrogen Metabolism (20)
- Dormancy and Sporulation (97)
- Respiration (42)
- Stress Response (47)
- Metabolism of Aromatic Compounds (12)
- Amino Acids and Derivatives (309)
- Sulfur Metabolism (8)

Fig. 7 Comparative analysis of the obtained mutants with the use of RAST server tools

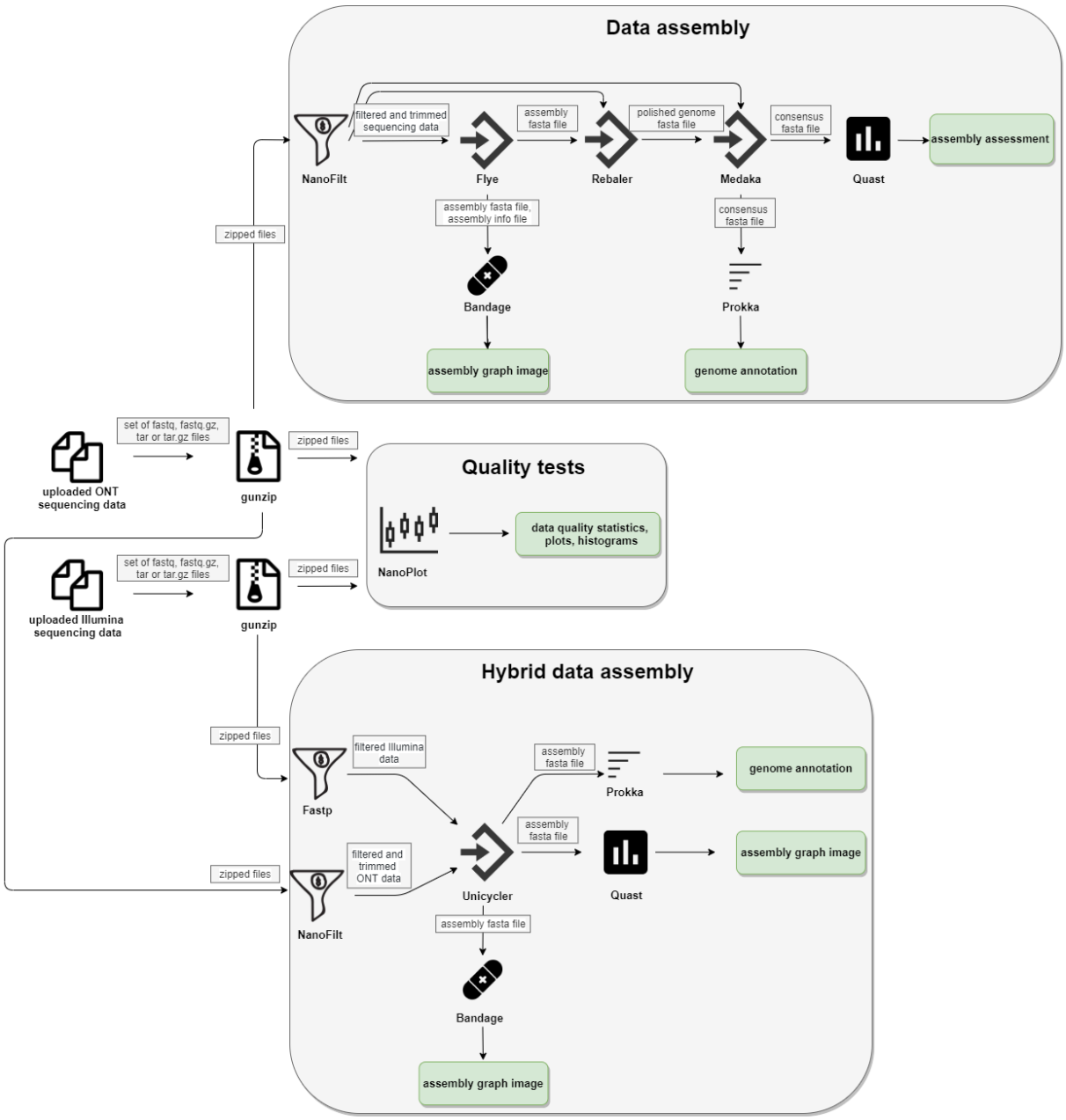


Fig. 8 A processing pipeline on the NanoForms server