

Progress report

Project title: **Sequencing the genome of a useful bacteria: *Paenarthrobacter nicotinovorans* – next step in extending it's biotechnological applications**

Contract no: **PCE 152 / 2021**

Proposal code: **PN-III-P4-ID-PCE-2020-0656**

Project Manager: **Marius MIHĂȘAN , Prof. Dr.**

Timeframe: **04 JANUARY - 31 DECEMBER 2021**

Abstract

Genomic DNA from the soil bacteria *Paenarthrobacter nicotinovorans* was isolated and sequenced using two different technologies. A draft genome containing 43 contigs was assembled and deposited in GenBank with accession number: JAESVM000000000. The final genome is currently on-going.

Results

Whole Genome Sequencing of any bacterial strains involves 3 different steps: genomic DNA isolation, DNA sequencing and finally genome assembly and functional characterization. We used a hybrid approach for sequencing the *Paenarthrobacter nicotinovorans* genome that relies on two sequencing technologies. The precise *short reads sequencing* of 150 pb fragments is done using the Illumina platform. The long reads sequencing is not so accurate, but it is useful in the final assembly and was performed using and ONT MinION device. This approach requires the isolation and preparation of two different DNA libraries to be sequenced: one containing short sequences and one containing high molecular weight (HMW) DNA.

As in our lab we also have a pAO1 cured strain derived from the wild type (wt) *Paenarthrobacter nicotinovorans*, we decided to also sequence this strain as it would work perfectly as a positive control for the whole experiment.

A total of 6 samples for the wt and 3 samples for the cured strain were used for sequencing using the Illumina platform and the amount of data generate is presented in table 1.

Table 1. Amount of nucleotide acquired by Illumina as shown by the raw data files.

Strain	Sample ID	Total read bases (bp)	Total reads	GC(%)	AT(%)	Q20(%)	Q30(%)
P. nicotino vorans	Pnic1	2 091 872 762	13 853 462	63,05	36,95	96,91	92,34
	Pnic2	2 137 039 278	14 152 578	63,00	37,00	97,08	92,63

Strain	Sample ID	Total read bases (bp)	Total reads	GC(%)	AT(%)	Q20(%)	Q30(%)
	Pnic3	2 139 191 632	14 166 832	63,03	36,97	96,82	92,23
	Pnic4	2 080 032 550	13 775 050	63,14	36,86	98,02	94,56
	Pnic5	2 078 919 378	13 767 678	63,23	36,77	97,76	94,03
	Pnic6	2 116 302 448	14 015 248	63,14	36,86	98,10	94,86
	Totals	12 643 358 048	83 730 848				
	Mean			63,10	36,90	97,45	93,44
P. nicotovorans pAO1-	Pnic7	2083608834	13798734	63,36	36,64	97,95	94,44
	Pnic8	2128301814	14094714	63,35	36,65	97,87	94,38
	Pnic9	2087609126	13825226	63,33	36,67	97,87	94,43
	Totals	6 299 519 774	41 718 674,				
	Mean			63,35	36,65	97,90	94,42

A total of 3 samples for the wt and 3 samples for the cured strain were used for sequencing using the ONT MinIon platform and the amount of data generate is presented in tabel 2.

Table 2. Amount of nucleotide acquired by MinION as shown by the raw data files.

Strain	Sample ID	Total read bases (bp)	Total reads	Min seq length	Min seq length	GC(%)	AT(%)
P. nicotovorans pAO1+	AES_BR_Minlon_7_10_2021	3373803981	288121	23	138891	62,70	37,30
P. nicotovorans pAO1-	Amada_Flongle_21_10_2021_#1	113946981	12705	94	72770	62,90	37,10
	Amada_Flongle_22_10_2021_#2	82995497	9427	110	70901	62,90	37,10
	Amada_Flongle_26_10_2021_#3	82918870	8831	135	75094	62,90	37,10
	Amada_Flongle_27_10_2021_#4	77258991	9240	95	183730	62,90	37,10
	Amada_Flongle_28_10_2021_#5	22284449	3830	116	251951	62,50	37,50
	Amada_Flongle_01_11_2021_#6	68335149	6599	98	367753	62,70	37,30
	Amada_Flongle_02_11_2021_#7	76493552	8114	124	84775	63,00	37,00
	Totals	3898037470	58746				
	Min/Max/Means			94	367753	62,83	37,17

The short reads data was processed, assembled, and deposited in NCBI Genome with the following coordinates BioProject: PRJNA693273, BioSample: SAMN17383832 și SAMN23521210. Datele au fost asamblate în 2 versiuni ale genomului în draft accesibile în baza de date NCBI Genome cu numerele de identificare **JAESVM000000000** <https://dataview.ncbi.nlm.nih.gov/object/PRJNA693273?reviewer=guidbohippi2s7egc7rnncfrsv>

Both the short and the long reads were used in a hybrid assembly approach. Several “in the lab” versions of the genome are available, an overview of their completeness and assembly characteristics is presented in table 3.

Table 3. Analytical data on various version of the genome assembled in the lab.

Strain	version	No of contigs	No of scaffolds	Min sequence length scaffolds	Max sequence length scaffolds	No nucleotides in scaffolds	No scaffold > 10 000 bp	No nucleotides in scaffolds > 10 000	No scaffolds > 100 000 bp	No nucleotides in scaffolds > 100 000	No scaffolds > 1 000 000 bp	No nucleotides in scaffolds > 1 000 000	GC content
P. nictotinovorans pAO1+	BBDuk+Spades_Amada_Genious_v1	4 636	4 614	78	4 315 603	6 148 331	3	4 542 002	2	4 531 876	1	4 315 603	62,40
	BBDuk+Spades_Amada_Genious_v2	4 637	4 615	78	4 325 566	6 097 417	2	4 490 784	2	4 490 784	1	4 325 566	62,50
	BBDuk+Geneious_v1												
	Fastp+Spades_Amada_CLI_v1	8 693	8 230		1 543 639	10 434 475	26	4 704 891	6	4 441 387	2	2 897 421	61,07
	Fastp+Unicycler_Amada_CLI_v1	2	2		4 316 184	4 481 325	2	4 481 325	2	4 481 325	1	4 316 184	63,1
P. nictotinovorans pAO1-	BBDuk+Spades_Amada_Genious_v1	392	391	78	4 321 433	4 430 699	1	4 321 433	1	4 321 433	1	4 321 433	63,20
	BBDuk+Spades_Amada_Genious_v2	392	391	78	4 321 433	4 430 699	1	4 321 433	1	4 321 433	1	4 321 433	63,20
	BBDuk+Spades_Amada_Genious_v3	1 140	1 133	78	3 615 095	4 658 160	2	4 321 870	2	4 321 870	2	3 615 095	63,10
	BBDuk+Geneious_v1												
	Fastp+Spades_Amada_CLI_v1	3 985	3 906	78	4 331 185	5 613 472	1	4 331 185	1	4 331 185	1	4 331 185	63,05
	Fastp+Unicycler_Amada_CLI_v1	1	1	4 323 902	4 323 902	4 323 902	1	4 323 902	1	4 323 902	1	4 323 902	63,24

Publications and results disseminations

At this point (12 month after the projects debut) the results were presented at 5 conferences (3 international and 2 national) as 2 lectures and 7 poster presentations. Some results were included in 2 articles (1 in ACS Omega, Q2 in Chemistry and one in Journal of Molecular and Experimental Biology).

News on results, publications, acquisition of novel equipment or key materials were posted in a timely manner on the project webpage (<http://cercetare.bio.uaic.ro/grupuri/bioactive/content/grants/PCE2021.html>), and group Facebook page: <https://www.facebook.com/bioactive.bio.uaic/>