

Whole genome of novel *Lactobacillus fermentum* *HFD1* strain producing various antimicrobial metabolites

Georgii Ozhegov¹, Monyr Nait Yahia¹, Alexey Vasilchenko², Natalya Gogoleva¹, Dina Yarullina¹, Airat Kayumov¹

1 – Kazan Federal University, Kazan, Russia

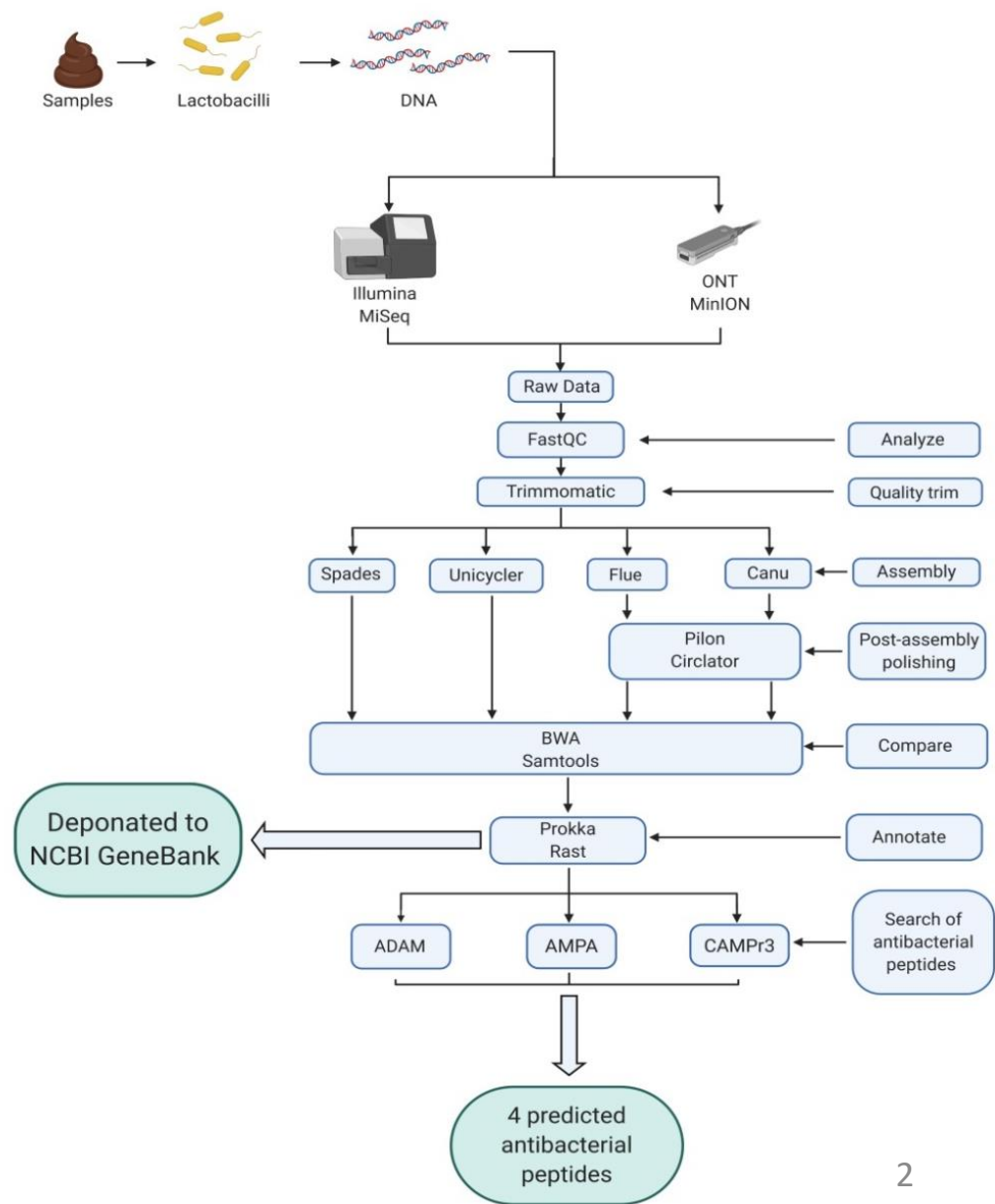
2 – University of Tyumen, Tyumen, Russia

Lactobacilli, generally recognized as safe for humans, are characterized with high antagonistic properties. Among 40 strains of Lactobacilli isolated from the faeces of healthy humans, *Lactobacillus fermentum* *HFD1* exhibited significant antimicrobial activity against various nosocomial pathogens.

From the culture liquid of *L. fermentum* *HFD1* several antimicrobial compounds repressing growth of *P. aureginosa* were isolated by using reverse-phase chromatography. The chemical structure and nature of these antimicrobial metabolites remain to be identified and for this purpose the whole genome sequence is strictly required.

Materials and methods

DNA was extracted by phenol-chloform method. Sequencing was performed on Illumina MiSeq and ONT MinION instruments. Hybrid assemblers Unicycler v0.4.8-beta and SPAdes v3.13.0, and long-read assembler Canu 1.8 combined with short-read polishing with Pilon v. 1.23 have been used. Genome has been annotated by using PROKKA 1.14.5 software. Putative genes and genes clusters responsible for antimicrobial peptides synthesis were predicted by using ADAM, AMPA and CAMPr3 web-services.



Results

The genome was read with 575× coverage for Illumina data and 454× coverage for MinION data. By comparing different strategies, Unicycler exhibited better performance. After assembly 2 circular contigs with respective sizes of 2101878 bp and 5386 bp has been obtained. The BLAST search identified the short contig as similar to bacteriophage phiX174. Full annotated genome deposited in NCBI GenBank database with accession number CP050919.1

	Unicycler conservative	Unicycler bold	Canu	Canu + pilon
Number of contigs	2	2	2	2
Max contig length	2.01 Mbp	2.01 Mbp	2.11 Mbp	2.15 Mbp
Circular	+	+	-	-
Error rate (Illumina)	8×10^{-4}	8×10^{-4}	7×10^{-3}	2×10^{-3}
Error rate (MinIon)	1×10^{-1}	1×10^{-1}	1×10^{-1}	1×10^{-1}

Genome annotation was performed by using PROKKA software. In total 2120 coding sequences (CDS) were predicted. By using CAMPr3 (protocols SVM, RF, Ann DAC), ADAM and AMPA services 4 CDSs with highest summary prediction score were identified as coding for antimicrobial peptides with length from 56 to 120 amino acids. Further validation of their antimicrobial activity in vitro is required.

GenBank protein ID	ADAM	CAMPR3			AMPA	Number of AA
		SVM	RF	DAC		
QIX58482.1	1.52	1.0	0.566	0.527	0.24	84
QIX58771.1	0.82	0.889	0.6715	0.81	0.2226	56
QIX58909.1	0.09	1.00	0.9855	1.0	0.248	120
QIX58994.1	1.32	1.00	0.7745	1.0	0.237	110

This research was funded by the Russian Foundation for Basic Research (grant number 17-00-00456) and by the subsidy allocated to Kazan Federal University for the state assignment in the sphere of scientific activities, project No. 13556.2019/13.1