Dissecting the bases of *Bacillus thuringiensis* host specificity: genomic and proteomic approaches

Yury Malovichko[12*], Daria Gorbach[2], Maria Belousova[1], Ekaterina Romanovskaya[2], Elena Lukasheva[2], Arseniy Lobov[23], Christian Ihling[4], Andrej Frolov[24], Anton Nizhnikov[12], Kirill Antonets[12]

[1] *Laboratory for Proteomics of SupraOrganismal Systems, All-Russian Research Institute of Agricultural Microbiology, Saint Petersburg, Russia* [2] *Department of Biology, St. Petersburg State University, Saint Petersburg, Russia* [3] *Laboratory of Regenerative Biomedicine, Institute of Cytology, Russian Academy of Sciences, Saint Petersburg, Russia* [4] *Department of Pharmaceutical Chemistry and Bioanalytics, Institute of Pharmacy, Martin-Luther Universität Halle-Wittenberg, Halle, Germany*

* *e-mail: yu.malovichko@arriam.ru*

Background. *Bacillus thuringiensis* (*Bt*) is a spore-forming bacterium affecting a wide range of invertebrate hosts. *Bt* is renowned for its high specificity towards the hosts; however, the mechanisms underlying this specificity remain unclear. In the present work we combine whole genome sequencing (WGS), qualitative and quantitative proteomics approaches to study the differences between four strains of three different serovars at vegetative and sporulating states and assess their virulence factors in respect to their host range.

Table 1. The list of strains used in this study. Vegetative cultures were grown on LB medium, and sporulating cultures grown on T3 medium for all four strains at 30^{0}C for 72h

Strain	Serovar	Host range
109/25	*darmstadiensis*	Lepidoptera, Coleoptera, mites
800/15	*thuringiensis*	Coleoptera, Hemiptera
800/3	*israeliensis*	Diptera
800/3-15	*israeliensis*	Avirulent

Acknowledgements. Proteomic assays were supported by the Russian Foundation for Basic Research (grant No 20-316-70020) Genomic part of this work was previously accomplished under support of the Russian Science Foundation (grant No 18-76-00028).

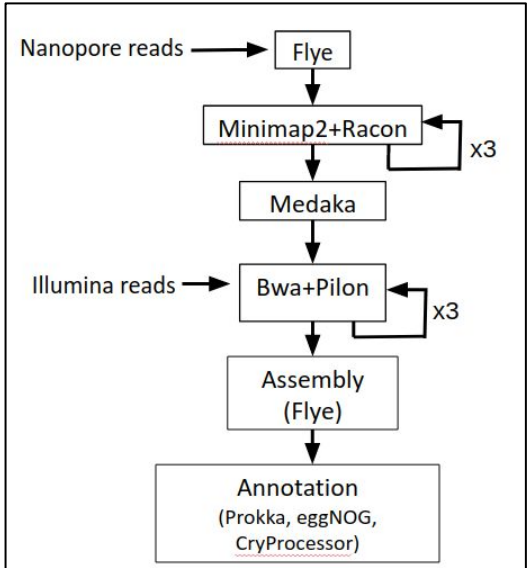

Figure 1. Genome assembly and annotation pipeline

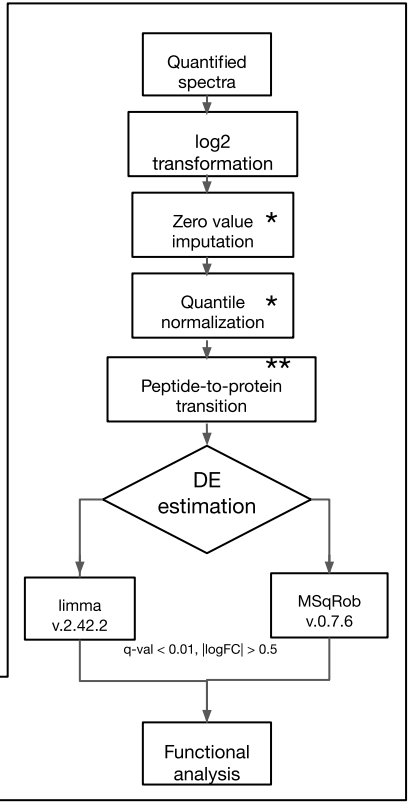

Figure 2. Quantitative proteomic analysis pipeline. Mass-spectrometry (MS) step was performed using HPLC/ESI-Orbitrap MS approach. * Performed using MSnbase v.2.12.0 R package. ** Performed using MSqRob v.0.7.6 R package.

Comparative genomics approach

Table 2. Genomes ON-based assembly statistics

Strain	ON read number	Genome length, bp	N50, bp	Coverage, x
109/25	58289	273589223	9080	49.74349509
800/15	96747	443435819	8193	80.62469436
800/3	71602	365644310	9198	66.48078364
800/3-15	444850	863249530	4674	156.95446

Table 3. Genomes annotation summary

Strain	No of contigs	No of genes
109/25	10	6525
800/15	8	5951
800/3	7	6412
800/3-15	10	6307

Table 4. *Cry* genes found with CryProcessor

Strain	*Cry* genes found
109/25	Cry1Ea10
800/15	Cry1Ba1 Cry1Ab12
800/3	Cry4Ba1 Cry4Aa2 Cry10Aa3
800/3-15	N/A

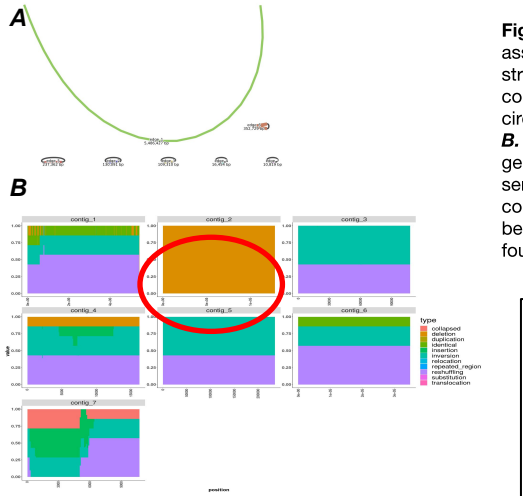

Figure 3. Results of genome assembly. ***A.*** A representation of strain 800/3 genome assembly. Linear contig represents chromosome, small circular contigs represent plasmids. ***B.*** A MUMmer plot of strain 800/3-15 genome aligned to genome of virulent ser. *israeliensis* strains. Highlighted contig absent in strain 800/3-15 bears 120 genes, of which 3 were found to encode for Cry toxins

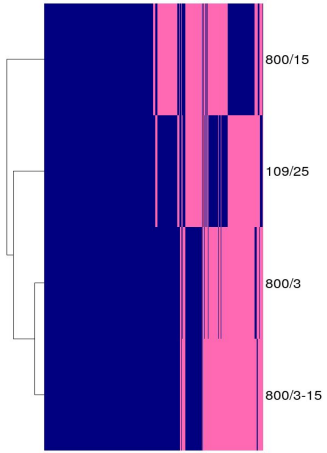

Figure 4. A pan-genome reconstruction of the studied strains according to Roary. Blue lines indicate present orthologues, pink lines indicate absent ones. Similarity threshold was set to 90%

CryProcessor (Shikov *et al., 2020*).
To mine for novel *Bt* insecticidal toxin genes, we elaborated a fast HMM-based tool, which can be found at the following link:

Minor virulence factors of *Bt* (Malovichko *et al*., 2019)
Currently we are working on a project dedicated to distribution of virulence factors genes unrelated to Cry toxins among *Bt* strains. A literature review on this topic and some preliminary results can be found at the following link:

Qualitative/quantitative proteomics approach

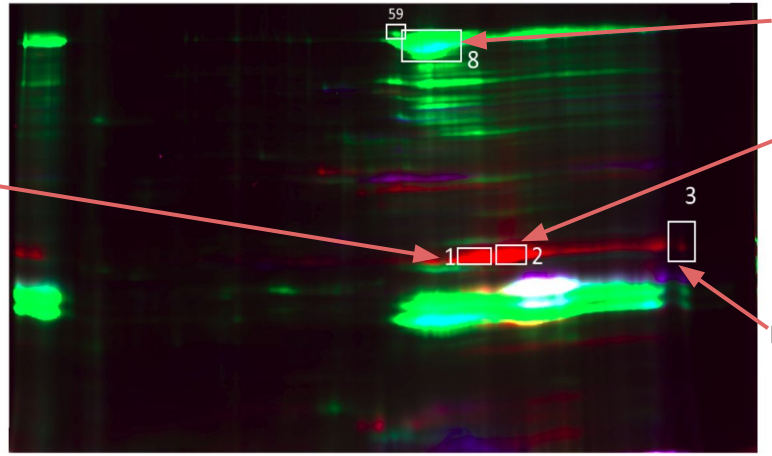

Figure 5. Two-dimensional difference gel electrophoresis (2D-DIGE) of sporulating culture proteins. Red channel denotes strain 800/3, blue channel denotes strain 109/25, and green channel denotes strain 800/15 proteins.

Table 5. A total number of proteins identified from 59 differential spots using FPLC/ESI-quadrupole MS

Serovar	109/25	800/15	800/3
No of identified proteins	15	14	11

Quantitative proteomics assays were performed using a non-labeled protocol for HPLC/ESI-Orbitrap mass-spectrometry. Annotated and quantified spectra underwent MLE/MinDet imputation and quantile-dependent normalization with MSnBase. Differential production was assessed with MSqRob and limma packages at protein level with q-value threshold equal 0.01.

Table 6. Differential protein production in sporulating cultures (predictor) compared to vegetative cultures (intercept) according to MSqRob

Differentially produced protein, No	Strain			
	109/25	800/15	800/3	800/3-15
Promoted	16	24	24	55
Repressed	25	50	55	80

Table 7. Differential protein production in sporulating cultures compared measured between the studied strains. In each pair the first strain indicates intercept (baseline) and the latter one indicate predictor

Differentially produced protein, No	Strain			
	109/25 vs 800/15	109/25 vs 800/3	800/15 vs 800/3	800/3 vs 800/3-15
Promoted	24	0	21	3
Repressed	16	0	24	8

Highlights. Several functional groups of proteins were enriched in the obtained datasets for interstrain comparisons. While for vegetative cells differentially produced proteins comprised mostly cellular metabolism enzymes, those differentially produced between sporulating cultures enclosed certain major virulence determinants (e.g. Cry and Cyt toxins and InhA/ColB metalloproteases) as well as exosporium maturation proteins. These assumptions, however, need further elucidation with the statistically correct assessment of Gene Ontology and/or KEGG term representation. Taken together, the presented data clearly reflect the differences between the strains studied, but their association with the observed phenotypes and host range is yet to be accomplished.