Comparative genomics and quantitative proteomics reveal differentially produced proteins underlying virulence and host specificity in *Bacillus thuringiensis*

Yury Malovichko[1], Maria Belousova[1], Elena Lukasheva[2], Daria Gorbach[2], Ekaterina Romanovskaya[2], Christian Ihling[3], Andrej Frolov[2,3], Anton Nizhnikov[1,4], Kirill Antonets[1,4*]

1 Laboratory for Proteomics of Supra-Organismal Systems, All-Russia Research Institute of Agricultural Microbiology, Saint Petersburg, Russia.
2 Department of Biochemistry, St. Petersburg State University Saint Petersburg, Russia.
3 Department of Pharmaceutical Chemistry and Bioanalytics Institute of Pharmacy, Martin-Luther Universität Halle-Wittenberg, Halle, Germany.
4 Department of Genetics and Biotechnology, St. Petersburg State University, Saint Petersburg, Russia.
*to whom correspondence should be addressed: k.antonets@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

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

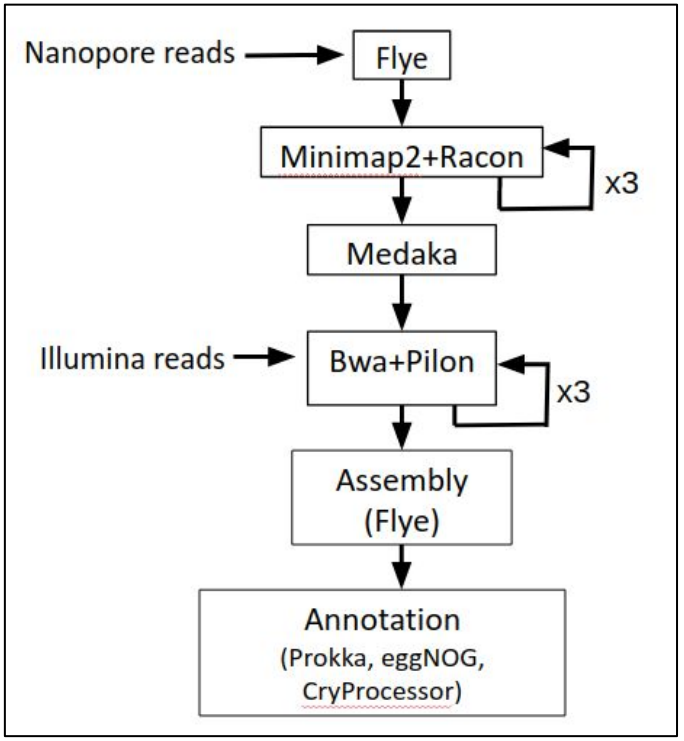

Figure 1. Genome assembly and annotation pipeline

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:

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

Comparative genomics approach

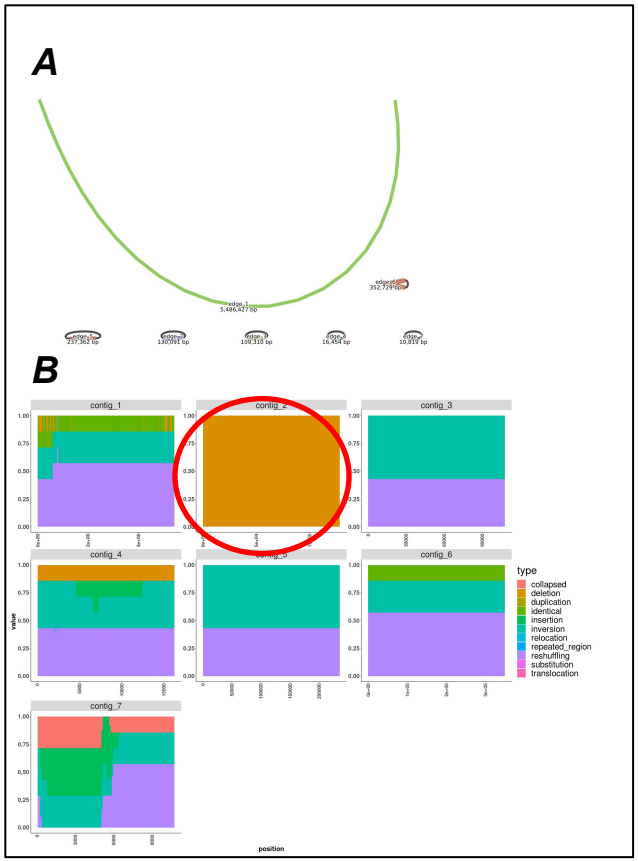

Figure 2. Results of genome assembly. ***A.*** A representation of strain 800/3 genome assembly. Linear contig represents chromosome, small circular contigs represent plasmids. ***B.*** A MUMmer plot of strain 800/3-15 genome aligned to genome of virulent ser. *israeliensis* strains. Highlighted contig absent in strain 800/3-15 bears 120 genes, of which 3 were found to encode for Cry toxins

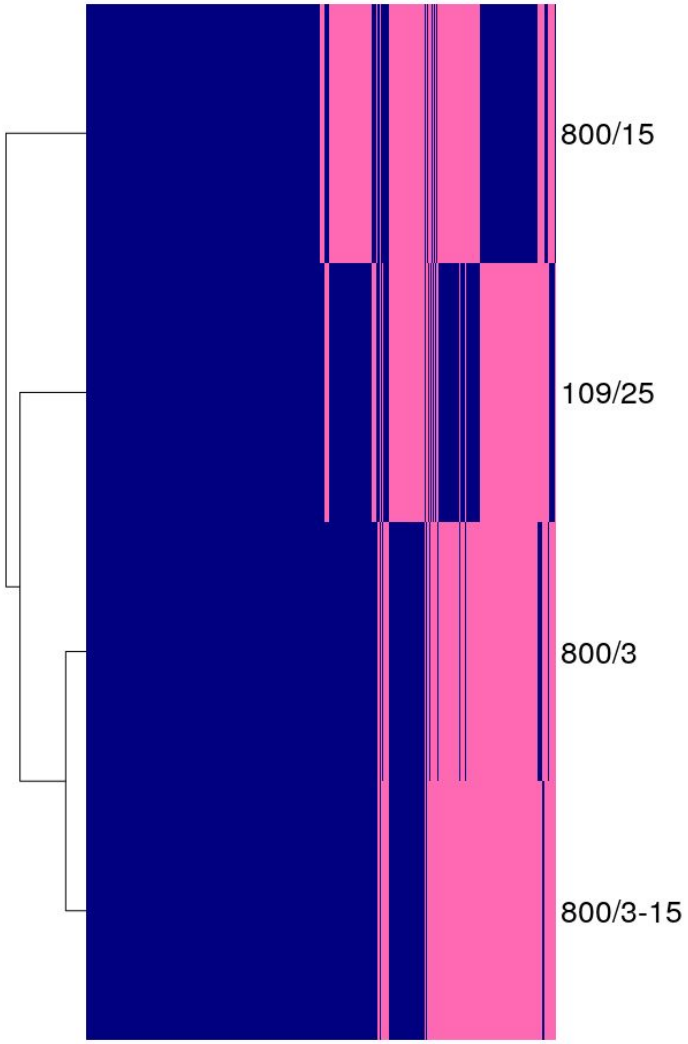

Figure 3. A pan-genome reconstruction of the studied strains according to Roary. Blue lines indicate present orthologues, pink lines indicate absent ones. Similarity threshold was set to 90%

Table 2. 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 3. *Cry* genes found with CryProcessor

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

Quantitative proteomics approach

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 4. Differential protein production in sporulating cultures (predictor) compared to vegetative cultures (intercept)

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

Table 5. 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

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 at protein level with q-value threshold equal 0.01.

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 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.