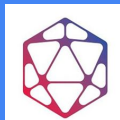


Identification and comparison of somatic antigen structures of symbiotic and pathogenic bacteria from Morganeliaceae family



Students:

Anna Churkina, Anna Rybina



Supervisors:

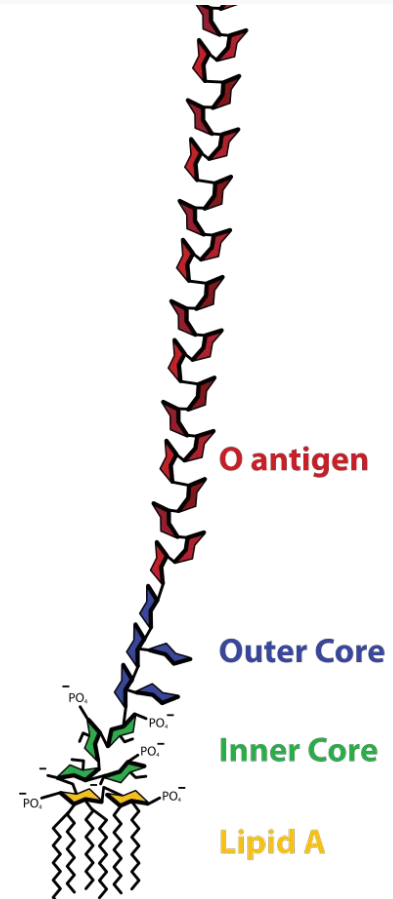
Polina Kuchur, Aleksey Komissarov

Applied Genomics Laboratory, SCAMT Institute, ITMO University

Background

Somatic O-antigen:

- a repetitive glycan polymer contained within an lipopolysaccharides (LPS);
- important virulence factor;
- its diversity provides bacteria with selective advantage in their specific niche and serves as a basis for serotyping schemes for many Gram-negative bacteria, including *Morganellaceae*;



Aim and objectives

Aim:

- to identify and compare candidate O-antigen operons in bacteria of the Morganellaceae family with different levels of virulence

Objectives:

- to find genes of O-antigen biosynthesis in the *Providencia* and *Xenorhabdus* species analyzing the literature;
- to analyze the quality of all *Providencia* assemblies available at NCBI (204 assemblies) and select the best one for each species;
- to create and test the pipeline for identifying, visualizing O-antigen operons;
- to compare candidate O-antigen operons in selected *Providencia* and *Xenorhabdus* species;

Materials and methods

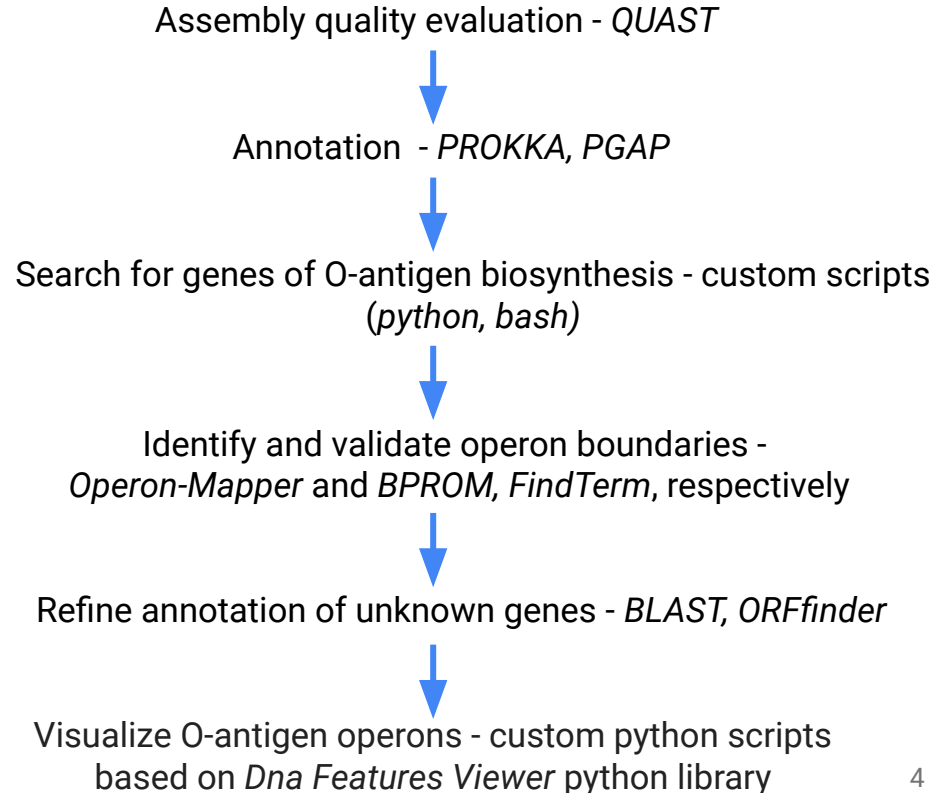
Objects of study

Taxonomy

Phylum: Proteobacteria
Class: Gammaproteobacteria
Order: Enterobacterales
Family: **Morganellaceae**

Genus: *Providencia* (mainly, opportunistic pathogens)
Xenorhabdus (mainly, symbionts)

Pipeline



Results. Quality of *Providencia* assemblies selected for analysis

Out of 204 *Providencia* assemblies currently available at NCBI, we selected only 7 complete genomes and 2 genomes of contig/scaffold assembly level that demonstrated the best quality. In total, we analyzed 6 distinct *Providencia*

species. Assembly ID	Organism Name	Total sequence length	# contigs	N50	L50
GCF_001874625.1	<i>Providencia rettgeri</i> strain RB151	4889093	2	4780676	1
GCF_010669105.1	<i>Providencia stuartii</i> strain MF1	4528609	2	4525926	1
GCF_000259175.1	<i>Providencia stuartii</i> MRSN 2154	4402109	1	4402109	1
GCA_016618195.1	<i>Providencia vermicola</i> strain LLDRA6	4342370	1	4342370	1
GCF_013702185.1	<i>Providencia alcalifaciens</i> strain 1701003	4033976	1	4033976	1
GCF_900455075.1	<i>Providencia rustigianii</i> strain NCTC11802	3969308	3	3926179	1
GCF_900635875.1	<i>Providencia rustigianii</i> strain NCTC6933	3913850	1	3913850	1
GCF_014652175.1	<i>Providencia thailandensis</i> KCTC 23281	4342585	39	484316	3
GCA_013702025.1	<i>Providencia rettgeri</i> strain 2055	4674814	1	4674814	1

Results. Genes of O-antigen biosynthesis defined from the literature for searching O-antigen operons in *Providencia* and *Xenorhabdus* species

Gene name	Product
<i>galE</i>	UDP-galactose 4-epimerase
<i>gne</i>	SDR family oxidoreductase
<i>ugd</i>	UDP-glucose 6-dehydrogenase
<i>wpaC</i>	Acetyltransferase
<i>qdtA</i>	dTDP-6-deoxy-3,4-keto-hexulose isomerase
<i>qdtB</i>	Aminotransferase
<i>rmlA</i>	glucose-1-phosphate thymidyltransferase
<i>qdtf</i>	N-formyltransferase
<i>wecA</i>	Undecaprenyl-phosphate alpha-N-acetylglucosaminyl 1-phosphate transferase
<i>wpaB</i>	Glycosyltransferase
<i>wpaA</i>	Glycosyltransferase

Gene name	Product
<i>wzx</i>	Lipid III flippase
<i>wza</i>	Polysaccharide export protein
<i>cpxA</i>	envelope stress sensor histidine kinase
<i>yibK</i>	tRNA (cytidine(34)-2'-O)-methyltransferase
<i>wzc</i>	Tyrosine-protein kinase wzc
<i>wzb</i>	Protein-tyrosine-phosphatase

Nucleotide sugar biosynthesis genes

Glycosyl transferases and O-unit formation

O-antigen processing genes

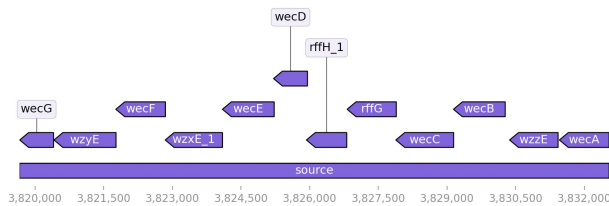
Housekeeping genes flanking the O-antigen cluster in *Providencia alcalifaciens* O40 (Ovchinnikova et al. 2012)

Results. Identified genes fall into two categories: one from conserved and other from variable O-antigen operons

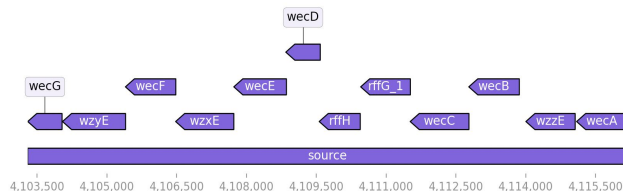
Assembly / Organism Name	Genes from the conserved operon	Genes from variable operon(s)
GCA_000259175 / <i>Providencia stuartii</i> MRSN 2154	<i>wecA, wzzE, wecB, wecC, wecD, wecE, wecF, wecG, rffG, rffH, wzxE, wzyE</i>	<i>galE, rfaQ, ugd, wzc, wzb, mshA, rfbM</i>
GCA_001874625 / <i>Providencia rettgeri</i> strain RB151	<i>wecA, wzzE, wecB, wecC, wecD, wecE, wecF, wecG, rffG, rffH, wzxE, wzyE</i>	<i>galE, rfaQ, ugd, rfaP, rfaG, wzb, wzc, rfbA, rfbB, wcaG, wza, waaO, wcaA</i>
GCA_013702025 / <i>Providencia rettgeri</i> strain 2055	<i>wecA, wzzE, wecB, wecC, wecD, wecE, wecF, wecG, rffG, rffH, wzxE, wzyE</i>	<i>galE, rfaQ, ugd, wzb, wzc, mshA, wbgU, rfbM, gmm, gmd, glmM, arnC, fcl</i>
GCA_010669105 / <i>Providencia stuartii</i> strain MF1	<i>wecA, wzzE, wecB, wecC, wecD, wecE, wecF, wecG, rffG, rffH, wzxE, wzyE</i>	<i>galE, rfaQ, ugd, rfaP, rfaG, wzb, wzc, mshA, gspA, wbpB, wbpD, wbpE</i>
GCA_013702185 / <i>Providencia alcalifaciens</i> strain 1701003	<i>wecA, wzzE, wecB, wecC, wecD, wecE, wecF, wecG, rffG, rffH, wzxE, wzyE</i>	<i>galE, rfaQ, ugd, wzb, wzc, rfbA, wcaG</i>
GCA_014652175 / <i>Providencia thailandensis</i> KCTC 23281	<i>wecA, wzzE, wecB, wecC, wecD, wecE, wecF, wecG, rffG, rffH, wzxE, wzyE</i>	<i>galE, rfaQ, ugd, rfaP, rfaG, wzc, wzb, mshA, gspA</i>
GCA_016618195 / <i>Providencia vermicola</i> strain LLDRA6	<i>wecA, wzzE, wecB, wecC, wecD, wecE, wecF, wecG, rffG, rffH, wzxE, wzyE</i>	<i>galE, rfaQ, ugd, rfaP, rfaG, wzb, wzc, rmlA (rfbA), rfbB, rfbD, rfbC, wza, waaO</i>
GCA_900455075 / <i>Providencia rustigianii</i> strain NCTC11802	<i>wecA, wzzE, wecB, wecC, wecD, wecE, wecF, wecG, rffG, rffH, wzxE, wzyE</i>	<i>galE, rfaQ, ugd, rfaP, rfaG, mshA</i>
GCA_900635875 / <i>Providencia rustigianii</i> strain NCTC6933	<i>wecA, wzzE, wecB, wecC, wecD, wecE, wecF, wecG, rffG, rffH, wzxE, wzyE</i>	<i>galE, rfaQ, ugd, rfaP, rfaG</i>
GCA_014295015 / <i>Xenorhabdus nematophila</i>	<i>wecA, wzzE, wecB, wecC, wecD, wecE, wecF, wecG, rffG, rffH, wzxE, wzyE</i>	<i>galE, rfaQ, rfaP, rfaG, mshA, wbgU</i>
GCF_000973125 / <i>Xenorhabdus bovienii</i> CS03	<i>wecA, wzzE, wecB, wecC, wecD, wecE, wecF, wecG, rffG, rffH, wzxE, wzyE</i>	<i>galE, rfaQ, rfaP, rfaG, wbgU, rfaD, rfaF, rfaC</i>

Results. Gene order and genetic content in the conserved O-antigen operon is maintained across all *Providencia* bacteria used in the study

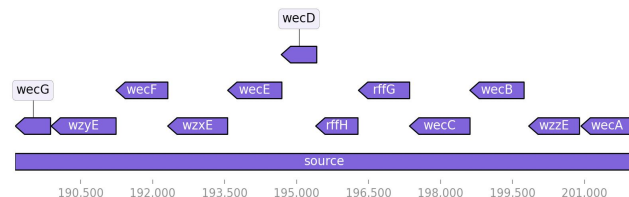
Providencia alcalifaciens strain 1701003



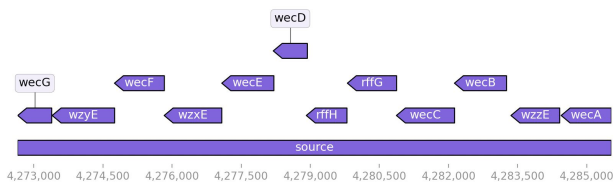
Providencia vermicola strain LLDRA6



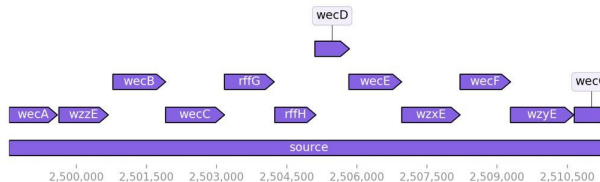
Providencia thailandensis KCTC 23281



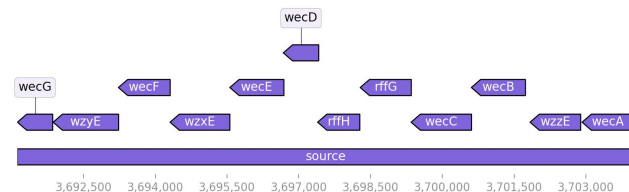
Providencia stuartii strain MF1



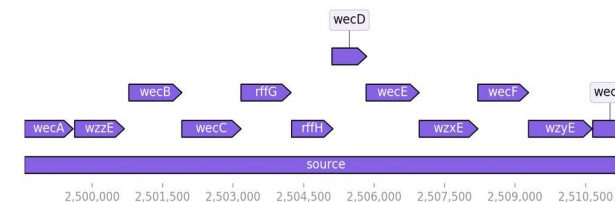
Providencia rettgeri strain 2055



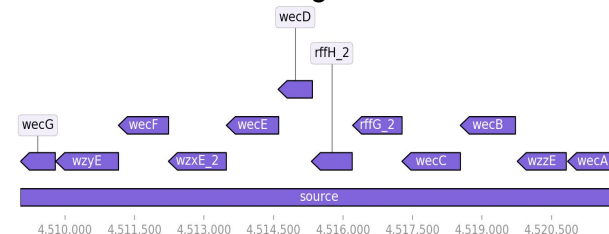
Providencia rustigianii NCTC11802



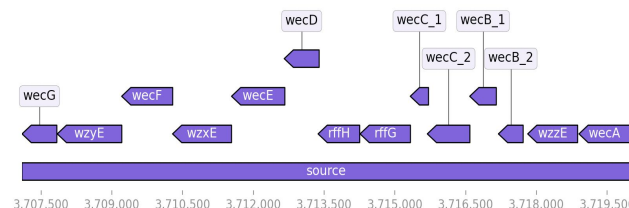
Providencia stuartii MRSN 2154



Providencia rettgeri strain RB151



Providencia rustigianii NCTC6933



Results. Conserved O-antigen operon in *Providencia*: gene and operon length differ slightly across analyzed organisms

Length (bp) of the conserved O-antigen operon and its genes from the *Providencia* species used in the study.

<i>Providencia</i> Assembly	<i>wecA</i>	<i>wzzE</i>	<i>wecB</i>	<i>wecC</i>	<i>wecD</i>	<i>wecE</i>	<i>wecF</i>	<i>wecG</i>	<i>rffG</i>	<i>rffH</i>	<i>wzxE</i>	<i>wzyE</i>	Whole operon
<i>Providencia stuartii</i> MRSN 2154	1047	1059	1131	1263	741	1131	1083	738	1068	882	1251	1356	12828
<i>Providencia rettgeri</i> strain RB151	1086	1059	1131	1263	741	1131	1083	753	1068	882	1251	1356	12878
<i>Providencia rettgeri</i> strain 2055	1086	1059	1131	1264	741	1131	1083	753	1068	882	1251	1356	12882
<i>Providencia stuartii</i> strain MF1	1089	1059	1131	1263	741	1131	1083	738	1068	882	1251	1356	12882
<i>Providencia alcalifaciens</i> strain 1701003	1068	1059	1131	1263	741	1131	1083	738	1074	882	1251	1356	12880
<i>Providencia thailandensis</i> KCTC 23281	1089	1059	1131	1263	741	1131	1083	738	1068	882	1251	1356	12888
<i>Providencia vermicola</i> strain LLDRA6	1086	1059	1131	1263	741	1131	1083	738	1068	882	1251	1356	12895
<i>Providencia rustigianii</i> strain NCTC11802	1086	1059	1131	1263	741	1131	1083	738	1068	882	1251	1368	12891
<i>Providencia rustigianii</i> strain NCTC6933	1086	1059	1133	1264	741	1131	1083	738	1068	882	1251	1368	12894

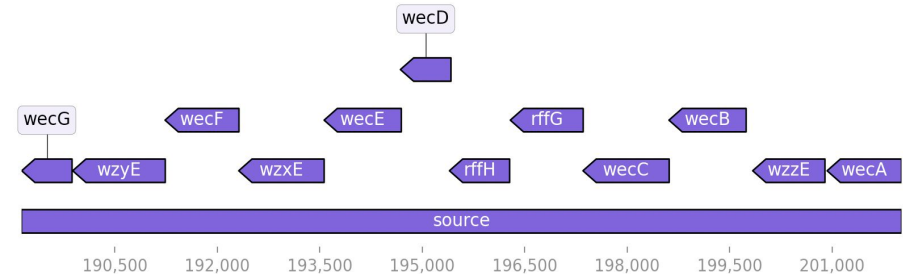
wzzE, *wecD*, *wecE*, *wecF*, *rffH*, and *wzxE* genes from conserved O-antigen operon have the same length for each of the studied species (marked green), while the rest genes are of varying length.

To test if a negative selection is preserving each gene from the conserved operon, we would compare the abundance of non-synonymous and synonymous substitutions in a gene using the Z-test of neutral evolution.

Results. Genes from the conserved O-antigen operon might be under negative selection which preserves encoded amino acid sequences from spread of mutations

Codon-based Test of Neutrality for analysis averaging over all sequence pairs.

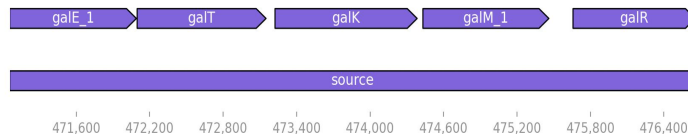
Gene	P-value	Z-test statistics of Neutral Evolution, $Z(dN-dS)$	Selection
wecA	0.00	-8.95	negative
wzzE	0.00	-9.25	negative
wecB	0.00	-9.35	negative
wecC	0.00	-10.32	negative
rffG	0.00	-10.47	negative
rffH	0.00	-6.84	negative
wecD	0.00	-6.47	negative
wecE	0.00	-9.71	negative
wzxE	0.00	-10.59	negative
wecF	0.00	-10.18	negative
wzyE	0.00	-5.88	negative
wecG	0.00	-6.83	negative



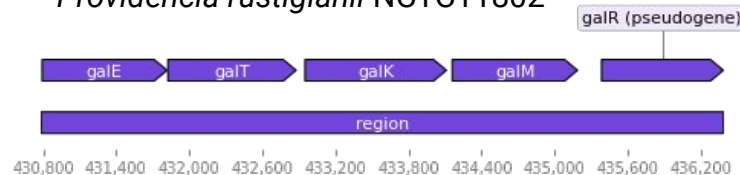
- $Z(dN-dS)$
- Nei-Gojobori (Jukes-Cantor) model
- 1000 bootstraps

Results. Almost all *Providencia* used in the study have *galETKMR* operon (except for *Providencia alcalifaciens* strain 1701003) with the same genetic order and gene content

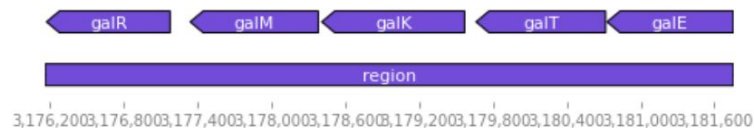
Providencia rustigianii NCTC6933



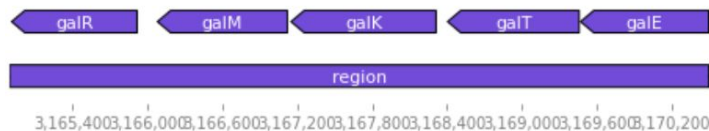
Providencia rustigianii NCTC11802



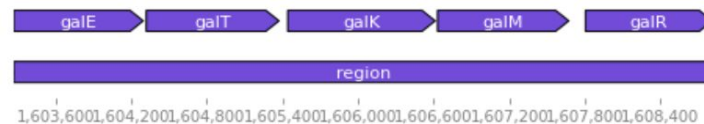
Providencia rettgeri strain RB151



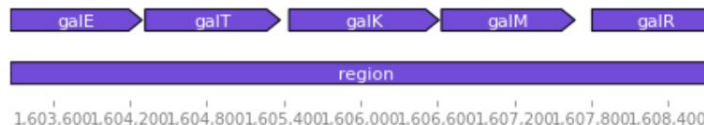
Providencia rettgeri strain 2055



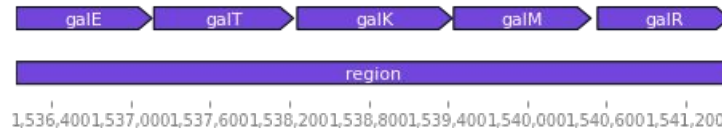
Providencia stuartii MRSN 2154



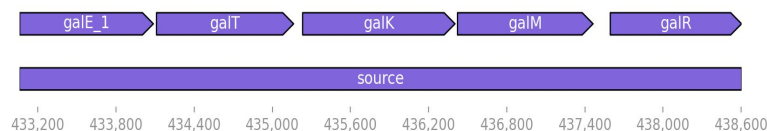
Providencia stuartii strain MF1



Providencia vermicola strain LLDRA6



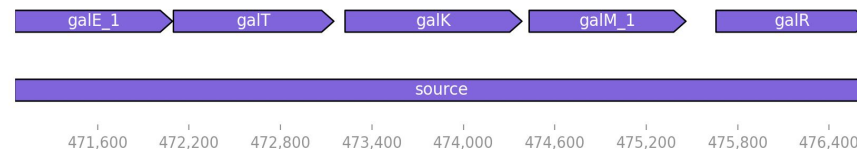
Providencia thailandensis KCTC 23281



Results. Gene sequences from the *galETKMR* operon are preserved from non-synonymous mutations being under negative selection

Codon-based Test of Neutrality for analysis averaging over all sequence pairs.

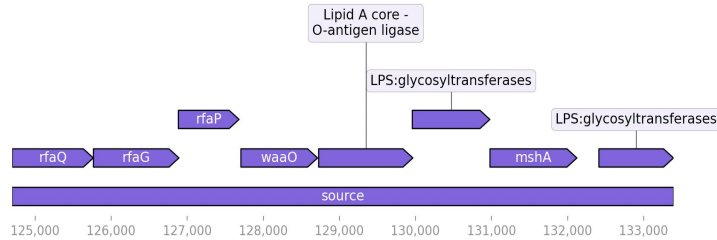
Gene	P-value	Z-test statistics of Neutral Evolution, $Z(dN-dS)$	Selection
<i>galE</i>	0.00	-8.15	negative
<i>galT</i>	0.00	-6.20	negative
<i>galK</i>	0.00	-6.33	negative
<i>galM</i>	0.00	-8.22	negative
<i>galR</i>	0.00	-7.53	negative



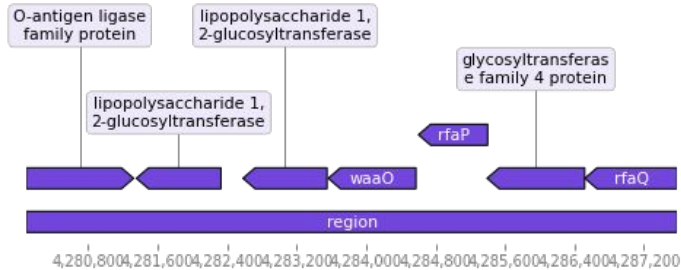
- $Z(dN-dS)$
- Nei-Gojobori (Jukes-Cantor) model
- 1000 bootstraps

Results. Variable operons in all studied species mainly contained genes encoding glycosyltransferase and O-antigen processing proteins

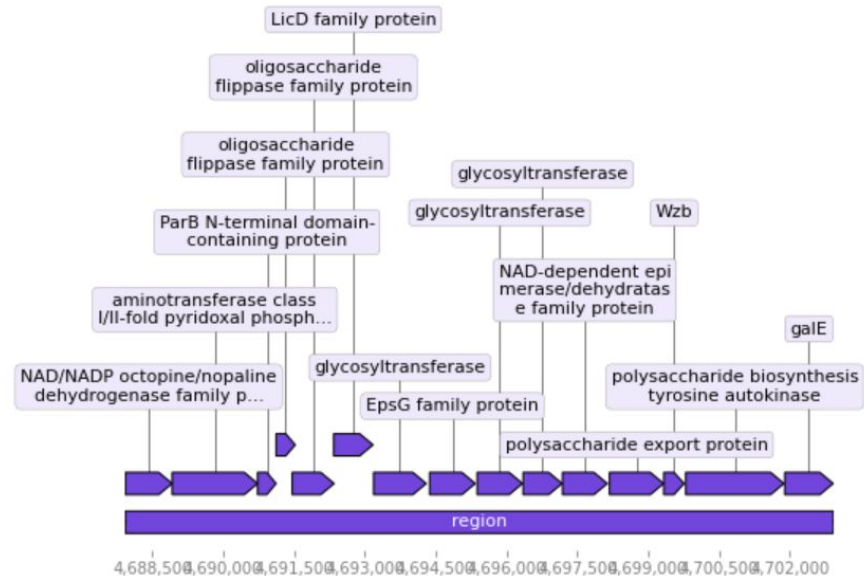
Providencia thailandensis KCTC 23281



Providencia vermicola strain LLDRA6

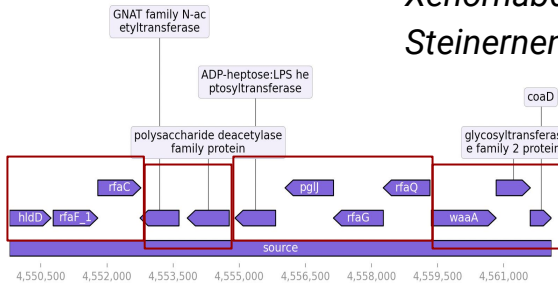
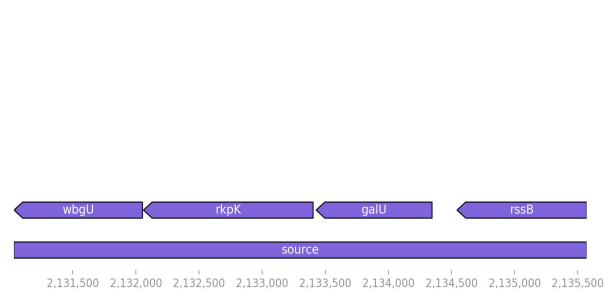


Providencia rettgeri RB151



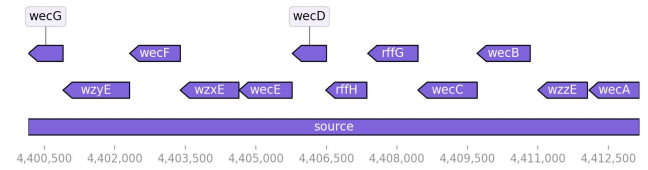
The *Providencia* bacteria studied in the project have a similar lifestyle, they are opportunistic microorganisms, while their O-antigen operons differ significantly in structure.

Results. Structural organization of candidate O-antigen operons is similar between *Xenorhabdus nematophila* and *Xenorhabdus bovienii* CS03 despite the difference in lifestyle

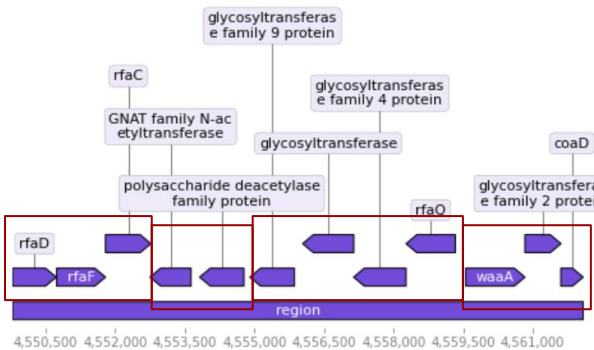
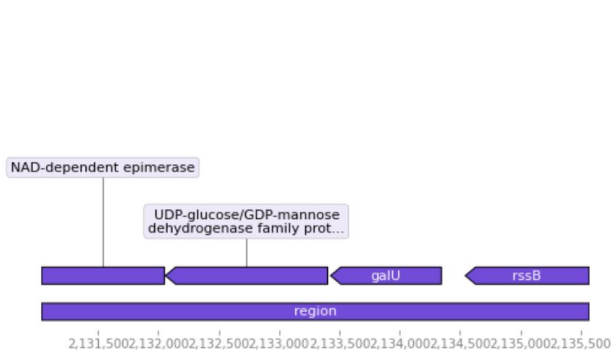


4 operons

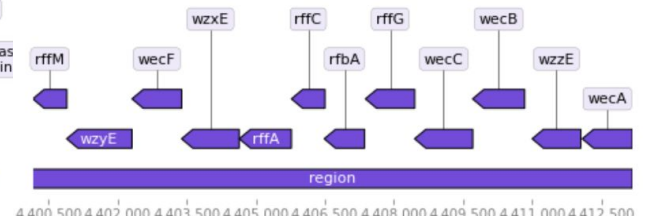
Xenorhabdus nematophila is mutualist of Nematode *Steinernema carpocapsae* and **pathogen** of insects
*PROKKA



Xenorhabdus bovienii CS03 - mutualist of Nematode *Steinernema nematodes* and **non-pathogen** of insects
*PGAP

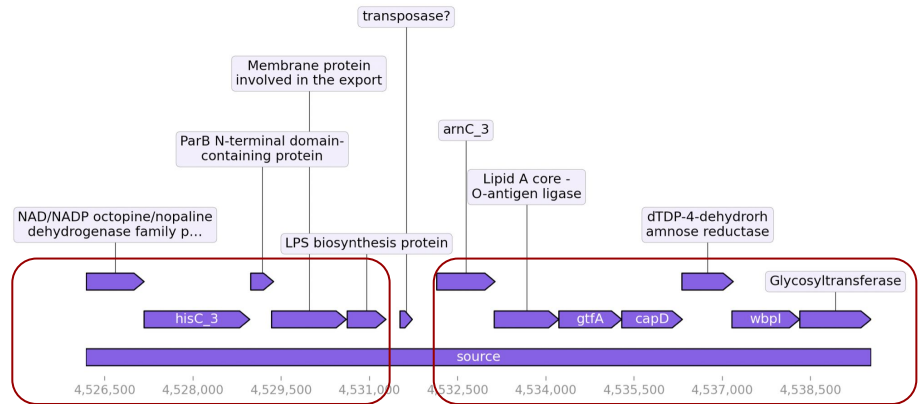


4 operons

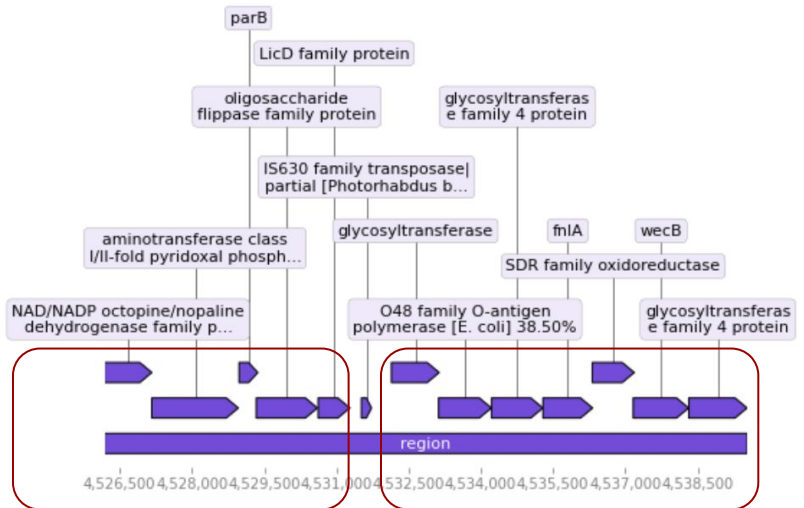


Results. Structural organization of candidate O-antigen operons is similar between *Xenorhabdus nematophila* and *Xenorhabdus bovienii* CS03 despite the difference in lifestyle

Xenorhabdus nematophila



Xenorhabdus bovienii CS03



The studied species of *Xenorhabdus* are characterized by different lifestyles:

- *Xenorhabdus nematophila* is mutualist of Nematode *Steinernema carpocapsae* and **pathogen** of insects
- *Xenorhabdus bovienii* CS03 - mutualist of Nematode *Steinernema nematodes* and **non-pathogen** of insects

But despite these differences, they have a similar structure of operons

Conclusion

- As a result of the literature analysis, we found the following O-antigen operon genes reported for *Providencia* sp.: 8 genes of nucleotide sugar biosynthesis, three genes involved in glycosyl transferring, and O-unit formation, two O-antigen processing genes, and four housekeeping genes flanking the O-antigen cluster.
- Out of 204 *Providencia* assemblies currently available at NCBI, we selected only seven complete genomes and two genomes of contig/scaffold assembly level that demonstrated the best quality. In total, we analyzed six distinct *Providencia* species.
- We have designed a unique pipeline for identifying candidate O-antigen operons. The key features of our pipeline are: employing several genome annotation tools, additional prediction and validation of operon boundaries, and manual curation and annotation of unknown genes. Our pipeline allowed us to detect 23 and 12 genes in variable and conserved candidate O-antigen operons, respectively. Among these genes, only seven (*galE*, *wxz*, *wzc*, *wza*, *ugd*, *rmlA*, and *wecA*) were previously described as O-antigen ones in *Providencia* and *Xenorhabdus* species.
- Both *Providencia* and *Xenorhabdus* species have O-antigen conserved operon involved in the nucleotide (UDP- or dTDP-) sugar biosynthesis (*wecB*, *wecC*, *rffG*, *rfbA*, *rffC*, *rffA*), glucosyl to lipid transfer (*wecA*, *wecF*, *wecG*), and O-antigen processing (*wzzE*, *wzxE*, *wzyE*).
- We did not observe a correlation between lifestyle and O-antigen operon organization as we initially expected. Although *Providencia* bacteria used in the study share similar virulence levels being opportunistic pathogens, their O-antigen operon significantly varies by structure. Despite different relationship towards insects (pathogenic or non-pathogenic ones), *Xenorhabdus* exhibit high similarity in O-antigen organization.

References

1. Bisch, Gaëlle, Sylvie Pagès, John G. McMullen 2nd, S. Patricia Stock, Bernard Duvic, Alain Givaudan, and Sophie Gaudriault. 2015. "Xenorhabdus Bovienii CS03, the Bacterial Symbiont of the Entomopathogenic Nematode *Steinernema Weiseri*, Is a Non-Virulent Strain against Lepidopteran Insects." *Journal of Invertebrate Pathology* 124 (January): 15–22.
2. Herbert, Erin E., and Heidi Goodrich-Blair. 2007. "Friend and Foe: The Two Faces of *Xenorhabdus Nematophila*." *Nature Reviews. Microbiology* 5 (8): 634–46.
3. Ovchinnikova, Olga G., Bin Liu, Dan Guo, Nina A. Kocharova, Magdalena Bialczak-Kokot, Alexander S. Shashkov, Lu Feng, Antoni Rozalski, Lei Wang, and Yuriy A. Knirel. 2012. "Structural, Serological, and Genetic Characterization of the O-Antigen of *Providencia Alcalifaciens* O40." *FEMS Immunology and Medical Microbiology* 66 (3): 382–92.
4. Samuel, Gabrielle, and Peter Reeves. 2003. "Biosynthesis of O-Antigens: Genes and Pathways Involved in Nucleotide Sugar Precursor Synthesis and O-Antigen Assembly." *Carbohydrate Research* 338 (23): 2503–19.
5. Wang, Lei, Quan Wang, and Peter R. Reeves. 2010. "The Variation of O Antigens in Gram-Negative Bacteria." *Sub-Cellular Biochemistry* 53: 123–52.
6. **GitHub** <https://github.com/rybinaanya/O-antigens>