Identification of bacterial virulence factors based on an integration of experimental and *in silico* transcription factor target discovery

Vychyk P.V., Nikolaichik Y.A.

Belarusian State University, Minsk, Belarus

E-mail: p.vychik@gmail.com

Key message. *In silico* transcription factors binding sites analysis in *Dickeya dadantii* genome matches published experimental results and predicts new regulators for genes inducing soft-rot disease in plants.

Keywords: virulence factors, transcriptional factor binding sites, bioinformatical analysis, LacI family, *Dickeya dadantii*

Potential novel targets involved in bacterial virulence, for transcriptional regulators LfeR and LfgR of *D. dadantii* were predicted using the motif discovery algorithm of Sigmoid.

Goal. Demonstrate the effectiveness of experimental data and *in silico* methods integration in the study of transcriptional regulation and find new virulence-related targets for LacI family regulators in *D. dadantii*.

Methods. Transcription factor binding site inference, operon prediction and database searches were all done with version 2 of Sigmoid software (1).

The *Pectobacteriaceae* family unites many species of soft rot plant pathogens. Like in most enterobacteria, their genomes encode about 300 transcription factors (TFs). According to the published transcriptomic data, about half of the TFs are either induced or repressed in planta. Despite the obvious importance for understanding plant-pathogen interaction, few of these TFs have been studied so far in most species. However, there were several important works on transcriptional regulation in *D. dadantii*, including a systematic experimental study of the LacI family regulators (2). Therefore, we have started our systematic *in silico* screen of plant pathogen transcription factor binding sites from the best-studied TF family in *D. dadantii*.

The previously reported experimental data (2) characterised 7 TFs without any proposed function at that time: LfaR, LfbR, LfcR, LfdR, LfeR, LffR, LfgR. Plants infected by bacteria with inactivated regulator genes demonstrated a change of the virulence in *lfaR*, *lfcR*, *lfdR*, *lfeR* and *lfgR* mutants. However, laborious experiments are needed to unveil a regulator's targets outside its own locus. *In silico* inference of TF binding sites could simplify 'wet' experiment design and provide a hypothesis about potential virulence mechanism.

De novo TF binding site inference algorithm implemented in Sigmoid is based on the information about specific amino acid-DNA base contacts of TF-operator complexes (available in the PDB for many TF families) and analysis of regulatory regions of the TF encoding gene and nearby operons. Full genome scan of *D. dadantii* with hidden Markov model of LacI family DNA binding domain (PF00356) has shown the presence of 19 TFs, 15 of them could be analyzed with our procedure. Plausible motifs were successively inferred for 12 of these TFs in the positions suitable for TF autoregulation and/or control of the linked target gene. Importantly, a motif was found for every target experimentally verified in (2). The newly identified motifs have been added to the growing library of the experimentally verified motifs distributed with Sigmoid.

We next looked for motifs located outside the regulator loci and have found two TFs that could control important virulence-related genes unlinked to the regulator genes. LfgR has a high scoring binding site near pyoverdine biosynthesis operon involved in iron acquisition which is known to be very important for this pathogen. A high scoring LfeR binding site is located 61 b.p. upstream from the translation start of the largest flagellar operon *fliF-fliR* right in front of the FliA-dependent promoter. This location suggests positive control of motility by LfeR. According to (2), *lfeR* mutant has reduced virulence which may indicate the importance of such regulation.

It is known that LacI family regulators most often control one or few tightly linked transcriptional units. However, our *in silico* approach can find new putative targets for uncharacterized regulators even for this TF family. We expect this method to be more efficient for other TF families and advocate its wide use to provide testable hypotheses for further experimental research of regulatory networks controlling plant-pathogen interaction.

1. Sigmoid, a cross-platform open-source application: <https://github.com/nikolaichik/Sigmoid>

2. Analysis of the LacI family regulators of *Erwinia chrysanthemi* 3937, involvement in the bacterial phytopathogenicity / F. Van Gijsegem [et al.] // MPMI. – 2008. – Vol. 21, № 11. – P. 1471-1481.