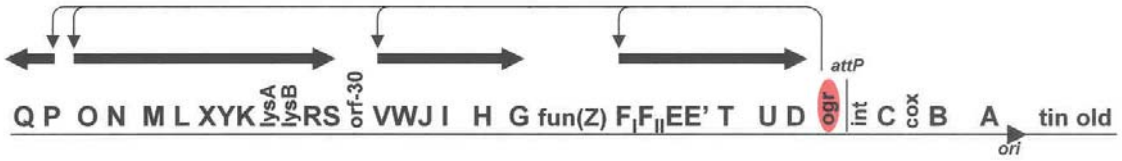
Scenario: Exploitation of a phage transcription factor by a bacterium: Identification of NucC-regulated genes in *Serratia marcescens*

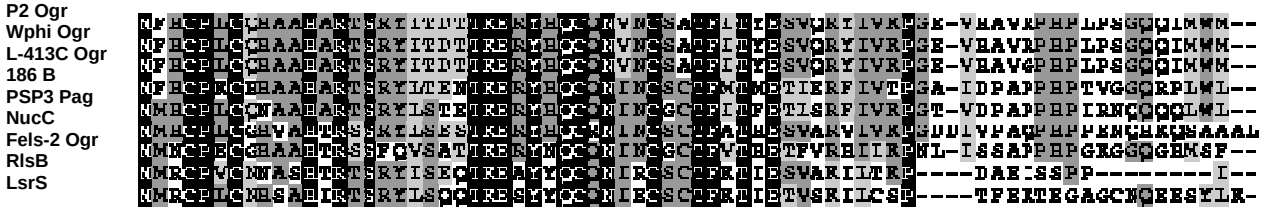
Background:

Bacteriophage P2 is the best characterized member of a family of temperate bacteriophages that are widespread among the Enterobacteriaceae. As with all phages, gene regulation during infection follows a precise temporal order. Expression of the late genes of P2 (which encode the phage virion proteins and the functions that lyse the bacterial host) requires a phage-encoded transcription factor, Ogr, which recognizes specific sequences upstream of the phage late promoters and interacts with the host RNA polymerase to stimulate transcription. All P2-related phages have similar late gene control by an Ogr homolog, and all family members tested to date are functionally interchangeable (i.e. will work to activate the late promoters of the other P2-related phages).



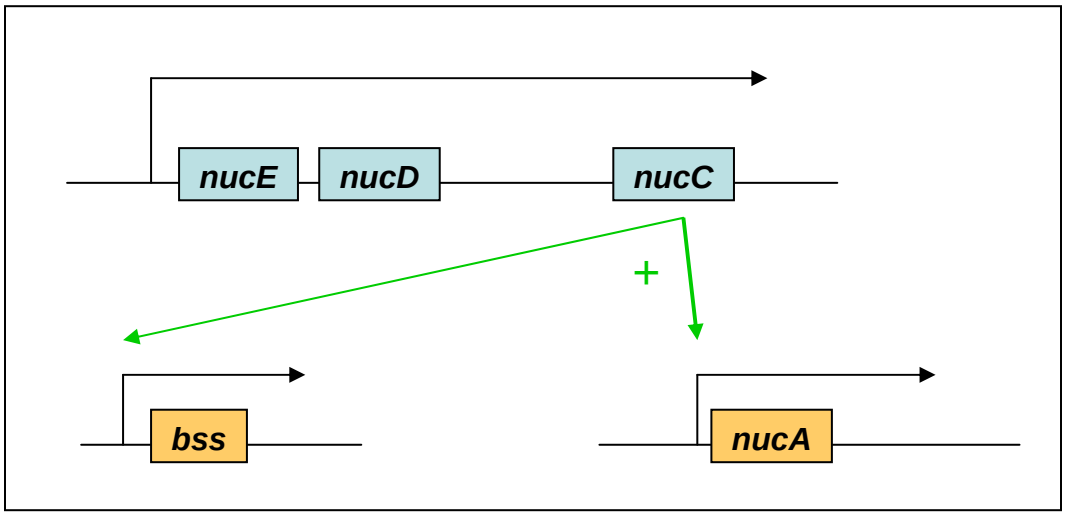
Late transcription of phage P2. Heavy arrows indicate the four late transcription units, each of which is under the positive control of the P2 *ogr* gene product.

The P2 Ogr family represents a unique class of transcription factors. They are extremely small (generally 72-75 amino acids) highly conserved proteins containing 4 invariant Cys residues that coordinate an atom of zinc. Zinc-binding transcription factors are common in eukaryotes, but this is the only prokaryotic example thus far and is much smaller than any of the eukaryotic proteins. A comparison of Ogr-related proteins reveals that the amino acid conservation is highest in the amino-terminal 2/3 of the protein. Deletion analysis of several of these proteins has demonstrated that this conserved amino-terminal region (to residue 51) is all that is required for function.



Comparison of Ogr homologs. NucC, RcsB and LsrA are regulators of bacterial genes (see discussion below); the remainder are the "Ogr" equivalents from a number of P2-related phages. Shading of amino acid residues corresponds to the degree of conservation, ranging from invariant (black) to nonconserved (white).

Ogr-dependent transcription activation ceased to be a phenomenon of just the bacteriophage world about a decade ago, when two groups independently identified an Ogr homolog in two different strains of *Serratia marcescens* that was required for the expression of two different secreted proteins.
(doi:10.1006/jmbi.1996.0084 , http://jb.asm.org/cgi/reprint/178/4/951?view=reprint&pmid=8576068
The gene for this protein, called *nucC* or *regC* by the two groups, was located in an operon that included two additional genes that resembled phage lysis functions, leading to the conclusion that this operon represented the remnant of a prophage in the *S. marcescens* genome. The two genes shown to be regulated by NucC encode an extracellular nuclease (the product of the *nucA* gene) and a small bacteriocidal peptide called bacteriocin 28b (the product of the *bss* gene). There is no evidence suggesting that *nucA* or *bss* are phage-derived genes; rather, it appears that *S. marcescens* may have co-opted the phage regulatory system for its own use.

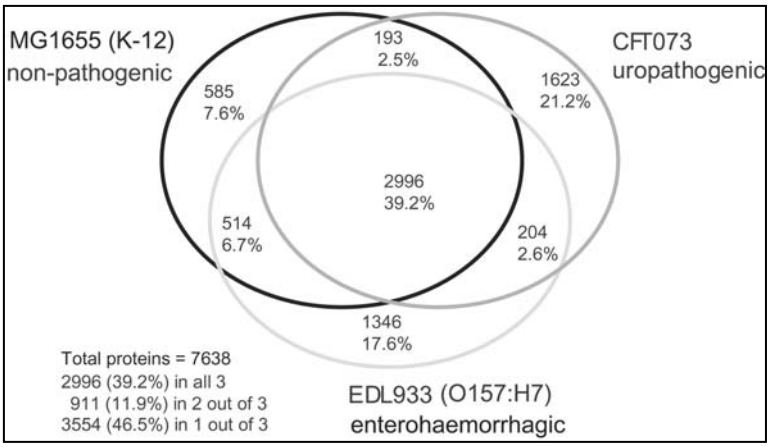


Subsequently, another Ogr homolog was identified in yet another bacterial species, *Rahnella aquatilis*, as a regulator of a secreted levansucrase
http://www.pubmedcentral.gov/articlerender.fcgi?tool=pubmed&pubmedid=12374819 .
A similar protein appears to regulate levansucrase in *Erwinia amylovora* (unpublished; accession number AJ344351 http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=nucleotide&val=18074038.

These observations raise interesting questions. We have examples of secreted proteins in 3 different bacterial species that are regulated by Ogr homologs. Are there others? What is the complete spectrum of NucC-regulated proteins in *Serratia marcescens*? Are they all secreted proteins? Are the Ogr homologs in *R. aquatilis* and *E. amylovora* also of apparent phage origin?...or did the phage steal this regulator from the bacteria? What kinds of approaches would one take to answering some of these questions?

Given the available experimental data, identification of additional NucC-regulated genes in *S. marcescens* should now be possible. The complete genome sequence of one strain of *S. marcescens* has recently been determined at the Sanger Centre. This strain, Db11, has been used in France for studies of virulence in a *C. elegans* model. It is not the same strain as the one(s) in which the *nucC*-defective mutations have been isolated. Can we use the genome sequence of Db11 to look for NucC-regulated proteins encoded by our strain?

While only one *S. marcescens* genome has been sequenced, there are now many examples in the database of multiple sequences of a particular bacterial species. If we look at *E. coli* as an example, it suggests that different strains of the same species may in fact be quite different.



Comparison of the predicted proteins encoded by three sequenced *E. coli* strains. EDL933 is a pathogenic strain that causes intestinal infection, CFT073 causes urinary tract infections, and MG1655 is a commonly used non-pathogenic laboratory strain. Only 39.2% of proteins are encoded by all three strains.

In order to determine whether the Db11 genome sequence might serve as a suitable model for analysis of NucC-regulated genes, the goal of this exercise is to determine whether the known regulated genes from *S. marcescens*, as well as the genes in the *nucC* operon itself, are encoded by this strain. The Db11 sequence is not yet annotated, so it is not yet in the GenBank database, but can be accessed at the Sanger Center http://www.sanger.ac.uk/Projects/S_marcescens/