Identification of possible virulence factors in the broad-spectrum pathogen *Serratia marcescens*

Honors Thesis

Presented to the College of Agriculture and Life Sciences, Department of Entomology

of Cornell University

in Partial Fulfillment of the Requirement for the

Research Honors Program

by

Gabriel T. LaHue

May 2010

conducted under the supervision of Brian P. Lazzaro

Abstract

The identification and characterization of bacterial virulence factors is important for effective treatment of bacterial infections, especially in the case of opportunistic human pathogens like *Serratia marcescens*. In this study I assess the virulence level of five *S. marcescens* strains by infecting *Drosophila melanogaster*, a natural host of the pathogen. I had the genomes of these strains sequenced using high-throughput Illumina-Solexa technology, and compared the whole genomes of a highly virulent *S. marcescens* strain and a non-virulent one to find genes or regions that were absent in the non-virulent strain. Forty-nine putative virulence genes were identified by sequence homology to genes involved in virulence in other pathogens. These putative virulence genes were located in fourteen genetic regions that were missing in the non-virulent strain. These potential virulence factors may prove to be important in the pathogenicity of *Serratia marcescens* against a broad spectrum of natural hosts.

Introduction

Virulence factors are molecules that promote pathogenicity by enabling the pathogen to enter into the host, evade host defenses and proliferate, cause local damage at the infection site, and disseminate itself or its products (Weiss & Hewlett 1986). Identification and characterization of virulence factors is central toward understanding how a pathogen causes disease and to devising new strategies for preventing disease. The quantity of research on bacterial virulence factors and the breadth of its scope illustrate the importance of determining mechanisms of pathogenicity (e.g. Chain *et al.* 2004, Read *et al.* 2003, Fleischmann *et al.* 2002). This research has been aided by the advent of next-generation sequencing technologies, such as Illumina-Solexa, Roche 454, and Applied Biosystem's SOLiD, which has significantly reduced the cost and time of sequencing, allowing whole genome comparisons between different species or strains to identify putative virulence factors in many pathogenic bacteria (Ahmed 2009).

Serratia marcescens is a Gram-negative bacterium that causes disease in a wide range of hosts including plants, insects, and vertebrates (Grimont & Grimont 1978). The species is an

opportunistic human pathogen and is responsible for increasing nosocomial infections in hospitalized patients, which can be life-threatening (Hejazi & Falkiner 1997). In addition, isolates of *S. marcescens* have been discovered that are resistant to several antibiotics, including ampicillin and cephalosporins (Hejazi & Falkiner 1997). Research into the virulence mechanisms of the bacterium may be critical in the development of new antibiotics or the treatment of disease symptoms caused by *S. marcescens* infections. Several *S. marcescens* virulence factors have been identified, such as mannose-resistant and mannose-sensitive pili, lipopolysaccharide, chitinase, lipase, chloroperoxidase, and the extracellular protein HasA (Hejazi & Falkiner 1997). However, many of the genes responsible for the synthesis of these virulence factors have not been characterized.

Model systems are often used for the study of pathogen-host interactions because many virulence factors are essential for pathogenicity regardless of the specific host (Kurz *et al.* 2003). Not only is *Drosophila melanogaster* is a natural host of *Serratia marcescens* (Flyg *et al.* 1980), but the fruit fly has also been extensively studied, both for identification of bacterial virulence mechanisms (e.g. Dionne *et al.* 2003) and for elucidation of the insect's innate immunity (Lemaitre & Hoffman 2007). In this study I have characterized the degree of pathogenicity of five *S. marcescens* strains (Type, Db11, BPL, PJ, and Unckless) on *D. melanogaster*, and then used whole-genome comparison to identify genes or gene regions with possible involvement in virulence. I have also developed a transformation protocol for *Serratia marcescens*, which can be used in future research to test these putative virulence factors for their role in pathogenicity.

Materials and Methods

Strains Used

The Type strain of *Serratia marcescens* was the first isolate of the species, and was identified by Bizio in 1819 (Anía 2009). The strain Db11 was originally isolated by from *Drosophila melanogaster* hemolymph as a spontaneous streptomycin-resistant mutant (Flyg *et al.* 1980). The *S. marcescens* strain BPL is a spontaneous mutant from the Type strain, the PJ strain was isolated from the hemolymph of wild *D. melanogaster* (unpublished P. Juneja 2005),

and the Unckless strain was isolated from the wild *D. melanogaster* environment (unpublished R. Unckless 2009). All bacterial culture growth was done on lysogeny broth (LB) media at 37°C with the addition of antibiotics when appropriate.

Assessing Differential Virulence

All five *S. marcescens* strains were grown overnight and the concentration of the bacterial cultures was standardized to an optical density at 600 nm of 1.00 using a *SmartSpec*™ 3000 (BIORAD). *Drosophila melanogaster* flies from the OregonR line were reared at room temperature on yeast-glucose media containing 12 g agar, 100 g glucose, 100 g Brewer's yeast, and 10 mL of a 4.2% phosphoric acid / 41.8% propionic acid mixture per liter of media prepared by the Cornell University Fly Food Kitchen. Flies that were 3-6 days old were anesthetized with carbon dioxide, and pricked in the thorax with a 0.1 mm base diameter stainless steel minutien pin (Fine Science Tools, Inc.) that had either been dipped in a standardized bacterial culture of one of the five strains or had been left sterile as a control. A group of *D. melanogaster* flies were anesthetized with carbon dioxide (CO_2) and not pricked with an infecting pin as another control. I infected 25 male flies and 25 female flies with each bacterial strain treatment, pricked 15 male and 15 female flies with sterile needles, and anesthetized ten male and ten female flies without pricking or infecting them during every infection period. Both male and female flies were used for the infections, since during preliminary trial infections I did not observe a large difference in mortality between sexes. In addition, each treatment had the same number of male and female flies, so any differences in response to infection based on gender would not have affected the analysis differing virulence levels among the strains. I performed 20 independent infections on different days, with a total of 1000 flies per bacterial infection treatment, 330 flies for the sterile needle control, and 220 flies for the anesthesia (CO_2) control. Flies were monitored for ten days, and mortality was recorded at 0 and 4 hours, every 2 hours from 12-28 hours, every 4 hours from 40-48 hours, and every 24 hours for the remainder of the 240-hour period. The proportion of flies surviving over time for each treatment was analyzed with the proc lifetest function using SAS® 9.1.3 software (SAS Institute, Inc. 2002). Since 21 pair-wise comparisons are made, I applied a Bonferoni correction to lower the probability of a Type 1 error, reducing the alpha level from $\alpha = 0.05$ to $\alpha = 0.0024$. This correction is conservative since

the Bonferoni correction assumes that all tests are independent, but the tests are not independent in my analysis.

Complete Genome Sequencing and Analysis

The genome of the Db11 strain has already been sequenced, and the annotated genome is publicly available (http://www.sanger.ac.uk/Projects/S_marcescens/). The complete genomes of the *Serratia marcescens* strains Type, BPL, PJ, and Unckless were sequenced using Illumina-Solexa high-throughput sequencing technology with 85-base pair reads by the Cornell University Life Sciences Sequencing and Genotyping Core Facility. A reference-based assembly was done to the Db11 genome for each strain using the Burrows-Wheeler Alignment (BWA) Tool (Li & Durbin 2009). Since the Db11 strain was found to be highly virulent (*D. melanogaster* mortality > 90%), it was first compared to the least virulent strain, BPL. Gene sequences in the Db11 genome that were completely absent in the BPL genome were identified using the Integrative Genomics Viewer (Broad Institute 2009). Due to the high frequency of single nucleotide polymorphisms (SNPs) and small gaps in the assembly, only genes to which no BPL sequence aligned were considered to be missing, as only completely absent genes could be considered non-functional with certainty. The sequences of open-reading frames (ORFs) that were present in the Db11 genome and absent in the BPL genome were obtained using the public Db11 genome and the Artemis genome viewer program (Rutherford et al. 2000). These sequences were then searched in the National Center for Biotechnology Information (NCBI) GenBank database using BLAST to identify the genes' putative functions. The list of genes that varied between the two strains was narrowed to those likely involved in virulence based on what has been shown to be important for virulence in other pathogens (Alm & Mattick 1996, Kurz *et al*. 2003).

Verification of Deleted Regions

In order to verify that the regions of the Db11 genome to which no sequence from the BPL genome aligns were actually absent in the BPL strain, I designed primers using Primer3 (http://frodo.wi.mit.edu/primer3/input.htm) with sequences complementary to the Db11

genome sequence surrounding the regions of interest. I used polymerase chain reaction (PCR) to amplify the gene region from Db11 DNA, and to amplify the shorter product from BPL DNA. DNA was extracted from both strains using the QIAprep® Spin Miniprep Kit (QIAGEN). I tested the primers used for amplification of the regions of interest with Db11 DNA at several different annealing temperatures, using the following protocol: 94°C for 2 minutes; 30 cycles of 94°C for 30 seconds, a gradient of annealing temperatures for 30 seconds, and 72°C for 1.5 minutes; 72°C for five minutes. The results of these reactions were visualized by gel electrophoresis. In addition, the absence of these regions was also verified computationally. A database of the Illumina-Solexa reads from the BPL strain was made, and NCBI BLAST+2.2.20 was used to search for the gene regions in that database.

Protocol for Knocking Out Putative Virulence Genes

To test the involvement of putative virulence genes in pathogenicity, I attempted to construct a plasmid vector with an antibiotic resistance gene flanked by sequences that were homologous to those on either side of the gene regions of interest. This would allow homologous recombination to knock out the putative virulence genes when the plasmid was transformed into *S. marcescens* Db11. The *Escherichia coli* strain DH5α pir+ was used as a vector for the cloning of sequences on either side of regions of interest. I amplified approximately 300 base pairs of sequence on either side of the region containing these genes with PCR. I designed primers to amplify these homologous regions with Primer3, and added restriction sites to the 5' end of each primer (Table 1). Recognition sites for the restriction enzymes *NotI* and *SacI* or *SpeI* and *SwaI* were incorporated into the primers to facilitate cloning into the pGPS2.1 plasmid vector.

The pGPS2.1 plasmid (New England Biolabs) was used as the plasmid vector for the cloning steps, since it is able to replicate in the *E. coli* DH5α pir+ strain, but not in any of the *S. marcescens* strains because the former bacterium has the *pi* gene necessary for replication while the latter doesn't. This ensures that only bacteria in which homologous recombination takes place with the Db11 bacterial chromosome will be able to replicate the antibiotic resistance gene and form colonies. The pGPS2.1 plasmid also contains a chloramphenicol

resistance gene, and selection for the plasmid was done on media with 25 µg/ml chloramphenicol. I used sequential ligations with T4 DNA ligase (Invitrogen) to insert the amplified homologous regions on either side of the chloramphenicol resistance gene in the plasmid, and the plasmid was transformed into competent *Escherichia coli* DH5α pir+ cells.

E. coli DH5α cells were made competent by centrifuging 8 ml of culture in the logarithmic growth phase at 4°C and 7000 X g for ten minutes, re-suspending them in 333 µl ice cold 0.1M CaCl2, centrifuging again at 4°C and 7000 X g for ten minutes, re-suspending again in 333 µl ice cold 0.1M CaCl2, adding 12 µl DMSO, incubating on ice for 15 minutes, adding 11 µl DMSO, incubating on ice for 15 minutes, and snap-freezing the cells in dry ice and ethanol. To transform the competent cells with the recombinant pGPS2.1 plasmid, the cells were thawed on ice, mixed with 50 ng of purified plasmid DNA, incubated on ice for 30 minutes, heat-shocked at 42°C for 30 seconds, mixed with 200 µl of SOC medium, incubated with shaking for one hour, and then plated on media containing the antibiotic chloramphenicol. If transformed colonies were recovered, the recombinant plasmid would have been extracted from the *E. coli* DH5α cells and transformed into the *Serratia marcescens* strain Db11 using the protocol described in the next section. Homologous recombination would replace the gene region of interest with the chloramphenicol resistance gene, and the transformed Db11 strain would then be used to infect *Drosophila melanogaster* flies.

Transformation Protocol Development

Since the success of a transformation protocol can vary with different strains of a bacterium, I adapted a protocol from Reid *et al.* 1982 and Cohen *et al.* 1972 and tested it with the *Serratia marcescens* strains Type and Db11. I constructed growth curves for all five strains to determine what the optical density was for the strains in the mid-logarithmic phase. Liquid LB media was spiked with overnight cultures and grown until the proper optical density, and split into micro-centrifuge tubes. These tubes were centrifuged at 12,000 X g for 10 minutes at 4°C and the pellet was re-suspended in 1 ml of 10 mM ice cold NaCl. The tubes were heat-shocked at 65°C for one minute, centrifuged again at 12,000 X g for 10 minutes at 4°C, and the pellet was re-suspended in 1 ml of 30 mM ice cold $CaCl_2$. The tubes were then centrifuged at 3000 X g for 5 minutes at 4°C and the pellet was re-suspended in 0.8 ml of ice cold 30 mM $CaCl_2$

and 15% glycerol. The competent *S. marcescens* cells were then divided into aliquots in sterile micro-centrifuge tubes, snap-frozen in dry ice and ethanol, and stored at -80°C. Competent cells could then be thawed, mixed with either 100 ng or 500 ng of plasmid DNA, incubated on ice for 60 minutes, heat-shocked for two minutes at 42°C, and then plated on LB media with the appropriate antibiotic. To test the competency of the treated *S. marcescens* cells, I used the pEGFP plasmid (BD Biosciences) and the pBCSK+ plasmid (Stratagene), since both plasmids contain the pUC replication of origin, which has previously been shown to maintain the plasmid in *S. marcescens* cells (Ball *et al.* 1990). The pEGFP plasmid has an ampicillin resistance gene (100 µg/ml), as well as the green fluorescent protein (GFP) reporter gene. The pBCSK+ plasmid has a chloramphenicol resistance gene (25 µg/ml).

Results

Assessing Differential Virulence

The five *Serratia marcescens* strains were found to differ significantly in their level of virulence, as assessed by mortality of infected *Drosophila melanogaster* flies (Figure 1). The five strains largely fell into two groups: the highly virulent category in which strains caused > 90% mortality on average (Db11, PJ, and Unckless) and the non-virulent category in which strains caused < 20% mortality (Type and BPL). The strains Type, BPL, and Db11 each had virulence levels that were significantly different from those of all the other strains (Table 2). The two most virulent strains, PJ and Unckless, did not differ in their virulence level, but were significantly more virulent than the next most virulent strain, Db11. The BPL strain did not differ significantly from either the sterile needle or the carbon dioxide anesthesia control. Flies that were infected with the pigmented Unckless strain turned pink after death, indicating that the bacterium was proliferating inside of the host (Figure 2).

Identification of Candidate Virulence Genes

The entire genomes of all five *Serratia marcescens* strains were sequenced with Illumina-Solexa technology, which yielded 8-12 million 85-base pair reads per strain (Table 3). This translates to an average coverage of 160X to 220X for each strain. The genomes of the *S.*

marcescens strains BPL and Db11 were compared, and Db11 genes that were completely absent in the BPL genome were analyzed to find sequence homology with known genes. Approximately five-hundred genes that were present in the Db11 genome were absent in the BPL genome, which equates to about 10% of Db11's 4,763 genes. The sizes of the deletions ranged from many with only a single gene missing (e.g. Figure 3) to a deletion that encompassed 53 genes. Missing genes that had significant sequence homology with known genes involved in pathogenicity in other organisms or that had a role in virulence-related processes, such as iron-uptake and fimbria production, were identified as candidate virulence genes. Forty-nine Db11 genes that were absent in BPL had significant sequence homology with virulence-related genes in other organisms, and these genes were predominately in fourteen gene regions that were completely missing in the BPL strain (Table 4). Due to time constraints, the entire genomes of the *S. marcescens* strains PJ, Type, and Unckless were not compared to the Db11 genome, but the presence or absence of the gene regions identified by the BPL-Db11 comparison to have possible virulence involvement was noted (Table 4). All of these gene regions were also absent in the non-virulent Type strain, but while some were present in the virulent strain Unckless or the virulent strain PJ, none were present in both, and many regions were absent from both.

Verification of Deleted Regions

In order to confirm that apparent deletions observed in the sequencing were real, I used PCR to amplify gene regions possibly involved in virulence from the Db11 strain and to amplify a shorter fragment containing the deletion from the BPL strain. Due to time constraints, three gene regions that had several putative virulence genes (SMAD1051-1057 and SMAD2299-2302) or that had one gene that was highly likely to be involved in virulence (SMAD4212-4216) were chosen for further study, in order to maximize the chance of identifying a gene involved in virulence. The SMAD1051-1057 region was successfully amplified from Db11 DNA, but no shorter fragment was amplified from BPL DNA. This may indicate that the region absent in the BPL genome also includes non-coding sequence on either side of the SMAD1051-1057 genes, and if the complementary sequences to the primers were in this region, then no product would be expected. Unfortunately the reference-based assembly of the BPL genome does not have high enough resolution to conclusively determine whether the complementary sequence of these primers is present or not. No product was obtained with PCRs to amplify the SMAD2299-2302 and SMAD4212-4216 regions with either Db11 or BPL DNA, even after multiple primers

were designed for the region, and the reason for this is currently unknown. While the absence of these two gene regions could not be verified by PCR, the absence of all three regions was verified computationally. A search for the sequence of all three regions in a database of Illumina-Solexa reads for the BPL genome using BLAST resulted in no significant alignments of reads to the sequences of the Db11 regions, which indicates that these regions are not present in the BPL genome.

Knocking-Out Putative Virulence Genes

I am currently in the process of attempting to knock out the three gene regions previously mentioned in the *Serratia marcescens* strain Db11. I have successfully amplified the homologous regions between the BPL and the Db11 strains on either side of the SMAD1051-1057, SMAD2299-2302, and SMAD4212-4216 regions using PCR. I have not, however, been able to successfully insert these amplicons into the pGPS2.1 plasmid. Transformation experiments with the unaltered plasmid into the *Escherichia coli* cloning strain DH5α pir+ have been successful, indicating that the problem is in the insertion of the homologous regions into the plasmid. Gel electrophoresis has shown that the restriction enzymes used have been cutting the plasmid at two sites instead of one, rendering it useless for transformation. These old restriction enzymes have been replaced with new enzymes of the same type, which gel electrophoresis has shown to cut at only the one site. I am therefore hopeful that the homologous regions will be successfully inserted before long, and given that we have developed a successful transformation protocol for the *S. marcescens* strain Db11, *Drosophila melanogaster* infections and virulence studies could follow soon after. If the virulence of the knock-out Db11 strain was significantly attenuated, this would support the involvement of these genes in virulence. To confirm their involvement, the gene region could be amplified by PCR from wild type Db11 DNA, inserted into a plasmid vector, and transformed into the knock-out Db11 strain. If the virulence of this strain was rescued by this transformation, as determined by infection of *D. melanogaster* flies, this would confirm the role of the gene region in pathogenicity.

Transformation Protocol Development

Two plasmids, pEGFP and pBCSK+, were transformed into *S. marcescens* Type and Db11 using the transformation protocol. Both the Type and Db11 strains transformed with the pEGFP plasmid formed bacterial lawns on selective media, indicating that the wild type strains likely had resistance to the low concentration of ampicillin used (40 μg/ml). Transgenic colonies were therefore identified by microscopic visualization of green fluorescent protein, the gene for which is contained on the pEGFP plasmid (Figure 4). The Type strain plates had a high density of transformed colonies, while the Db11 plates had sparse transformed colonies, but the transformation protocol with the pEGFP plasmid was clearly successful for both strains. The Type strain transformed with the pBCSK+ plasmid and plated on LB media with chloramphenicol grew approximately 10-60 colonies for each μl of transformation mixture plated, depending on the concentration of plasmid used in the transformation. The Db11 strain transformed with the pBCSK+ plasmid grew one colony for each 5-25 μl of transformation mixture plated. The presence of the plasmid was confirmed by amplifying a 200 base-pair region of the plasmid from whole bacterial colonies using PCR with the M13 primers (Heidecker *et al.* 1980). The presence of Type and Db11 colonies transformed with the pEGFP and the pBCSK+ plasmids indicates that these plasmids are maintained in both strains, and that the pUC replication of origin common to both of these plasmids is functional in *S. marcescens*.

Discussion

In this study I used whole genome comparison of two *S. marcescens* strains with significantly different virulence phenotypes to identify candidate virulence genes, and analyzed the distribution of these putative virulence genes in three other *S. marcescens* strains with differing virulence levels. I have initially focused only on Db11 genes that were completely absent in the BPL genome for the identification of possible virulence genes. However, a single nucleotide polymorphism (SNP) or a small deletion can still make a gene non-functional, and exploring the functionality and roles of BPL genes interrupted by SNPs or small deletions is an area for future research. Fourteen gene regions that were present in the highly virulent Db11 strain and absent in the non-virulent BPL strain contained at least one gene with significant homology to a gene involved in virulence in another organism. None of these fourteen gene

regions were present in all three highly virulent strains (Db11, PJ, and Unckless) and absent in both non-virulent strains (Type and BPL), suggesting that there might be several genetically independent mechanisms by which a bacterium can become virulent or benign.

Several genes involved in iron uptake or fimbrial biosynthesis were identified as present in Db11 but absent in BPL. Both of these processes are related to virulence in other bacteria. Genes for the production of iron-chelating siderophores have been shown to be involved in virulence in the *Yersinia* genus of bacterial pathogens (Carniel 2001). Hosts often sequester iron in proteins, leaving too low a level of free iron for bacterial growth (Carniel 2001). Iron uptake is therefore central to proliferation of the bacterium and iron uptake genes are necessary for virulence. Adhesion of pathogenic bacteria to host mucosal surfaces is essential for the development of an infection, and in Gram-negative bacteria (such as *S. marcescens*) adhesion is usually mediated by filamentous organelles on the cell surface called fimbriae (Struve *et al.* 2008). This is the case for *Klebsiella pneumoniae*, in which fimbriae were shown to be an important virulence factor (Struve *et al.* 2008). I hypothesize that iron transport and fimbrial biosynthesis genes may contribute to the virulence of *S. marcescens* Db11, but testing this prediction will require further experimentation.

Given that the phylogenetic relationship of most of the *S. marcescens* strains used in this study is unknown (although we do know that BPL is derived from Type), and considering the ease of gaining virulence genes by horizontal transfer or losing them through deletions (Guttman 2009), it is not surprising that none of the putative virulence genes are uniformly present in all the virulent strains and absent in the non-virulent strains. For example, none of the iron uptake genes are present in all three virulent strains, but each virulent strain has at least one gene region involved in iron uptake (Table 2). In addition, the only phenotype assessed for the *S. marcescens* strains in this study was the mortality they cause in *Drosophila melanogaster* flies. Therefore, the mechanism behind this pathogenicity is unknown in all of the strains and could vary widely for each virulent strain. In this case we would not expect to find a single cassette of virulence genes that is absent in all of the non-virulent strains and present in the virulent strains, but pair-wise comparisons of strain genomes might still uncover genetic differences responsible for varying virulence levels.

This study has characterized the differences in virulence between five *Serratia marcescens* strains, identified genes and gene regions with possible involvement in virulence, and developed a successful transformation protocol for some of these *S. marcescens* strains. In future research it will be important to functionally validate the role of the putative virulence genes identified in this study by using homologous recombination in a highly virulent strain to replace each gene with an antibiotic resistance gene, and assess whether a drop in virulence occurs. Reciprocally, the gene sequence of a putative virulence gene in a highly virulent strain can be isolated and transformed into a non-virulent strain to see whether the virulence phenotype can be rescued. The characterization of the virulence factors of *Serratia marcescens* will contribute to an understanding of the biology of the bacterium and may aid in the development of new antibiotics or disease-prevention strategies.

Acknowledgements

I would like to thank Dr. Brian P. Lazzaro and Madeline Galac for help with the experimental design and their critiques of this paper, as well as Punita Juneja and Rob Unckless for providing the PJ and Unckless strains respectively. This work was funded in part by a summer research fellowship awarded to G.T.L. by the Howard Hughes Medical Institute.

References

Ahmed, N. (2009) A flood of microbial genomes – do we need more? *PLoS ONE* **4(6):** e5831.

Alm, R.A. and Mattick, J.S. (1996) Identification of two genes with prepilin-like leader sequences involved in Type 4 fimbrial biogenesis in *Pseudomonas aeruginosa*. *Journal of Bacteriology* **178(13):** 3809-3817.

Anía, B.J. (2009) *Serratia*. *eMedicine from WebMD*. http://emedicine.medscape.com/article/228495-overview

Ball, T.K., Wasmuth, C.R., Braunagel, S.C., and Benedik, M.J. (1990) Expression of *Serratia marcescens* extracellular proteins requires *recA*. *Journal of Bacteriology* **172(1):** 342-349.

Broad Institute (2009) Integrative Genomics Viewer beta. http://www.broadinstitute.org/igv/

Carniel, E. (2001) The *Yersinia* high-pathogenicity island: an iron uptake island. *Microbes and Infection* **3(7):** 561-569.

Chain, P.S.G., Carniel, E., Larimer, F.W., Lamerdin, J., Stoutland, P.O., Regala, W.M., Georgescu, A.M., Vergez, L.M., Land, M.L., Motin, V.L., Brubaker, R.R., Fowler, J., Hinnebusch, J., Marceau, M., Medigue, C., Simonet, M., Chenal-Francisque, V., Souza, B., Dacheux, D., Elliott, J.M., Derbise, A., Hauser, L.J., and Garcia, E. (2004) Insights into the evolution of *Yersinia pestis* through whole-genome comparison with *Yersinia pseudotuberculosis*. *PNAS* **101(38):** 13826-13831.

Cohen, S.N., Chang, A.C.Y., and Hsu, L. (1972) Non-chromosomal antibiotic resistance in bacteria: genetic transformation of *Escherichia coli* by R-Factor DNA. *PNAS* **69(8):** 2110-2114.

Dionne, M.S., Ghori, N., and Schneider, D.S. (2003) *Drosophila melanogaster* is a genetically tractable model host for *Mycobacterium marinum*. *Infection and Immunity* **71(6):** 3540-3550.

Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A., Bishai, W., Jacobs, Jr., W.R., Venter, J.C., and Fraser, C.M. (2002) Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. *Journal of Bacteriology* **184(19)**: 5479-5490.

Flyg, C., Kenne, K., and Boman, H.G. (1980) Insect pathogenic properties of *Serratia marcescens*: phage-resistant mutants with a decreased resistance to *Cecropia* immunity and a decreased virulence to *Drosophila*. *Journal of General Microbiology* **120:** 173-181.

Grimont, P.A.D. and Grimont, F. (1978) The Genus *Serratia*. *Annual Review of Microbiology* **32:** 221-248.

Guttman, D.S. (2009) Bacterial evolution: dynamic genomes and the power of transformation. *Current Biology* **19(18):** r857.

Heidecker, G., Messing, J., and Gronenborn, B. (1980) A versatile primer for DNA sequencing in the M13mp2 cloning system. *Gene* **10:** 69-73.

Hejazi, A. and Falkiner, F.R. (1997) *Serratia marcescens*. *Journal of Medical Microbiology* **46(11):** 903-912.

Kurz, C.L., Chauvet, S., Andrés, E., Aurouze, M., Vallet, I., Michel, G.P.F., Uh, M., Celli, J., Filloux, A., de Bentzmann, S., Steinmetz, I., Hoffmann, J.A., Finlay, B.B., Gorvel, J.P., Ferrandon, D., and Ewbank, J.J. (2003) Virulence factors of the human opportunistic pathogen *Serratia marcescens* identified by *in vivo* screening. *The EMBO Journal* **22(7):** 1451-1460.

Lemaitre, B. and Hoffmann, J. (2007) The host defense of *Drosophila melanogaster*. *Annual Review of Immunology* **25:** 697-743.

Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25:** 1754-60. [PMID: 19451168]

Read, T.D., Peterson, S.N., Tourasse, N., Baillie, L.W., Paulsen, I.T., Nelson, K.E., Tettelin, H., Fouts, D.E., Eisen, J.A., Gill, S.R., Holtzapple, E.K., Økstad, O.A., Helgason, E., Rilstone, J., Wu, M., Kolonay, J.F., Beanan, M.J., Dodson, R.J., Brinkac, L.M., Gwinn, M., DeBoy, R.T., Madpu, R., Daugherty, S.C., Durkin, A.S., Haft, D.H., Nelson, W.C., Peterson, J.D., Pop, M., Khouri, H.M., Radune, D., Benton, J.L., Mahamoud, Y., Jiang, L., Hance, I.R., Weidman, J.F., Berry, K.J., Plaut, R.D., Wolf, A.M., Watkins, K.L., Nierman, W.C., Hazen, A., Cline, R., Redmond, C., Thwaite, J.E., White, O., Salzberg, S.L., Thomason, B., Friedlander, A.M., Koehler, T.M., Hanna, P.C., Kolstø, A.B., and Fraser, C.M. (2003) The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. *Nature* **423:** 81-86.

Reid, J.D., Stoufer, S.D., and Ogrydziak, D.M. (1982) Efficient transformation of *Serratia marcescens* with pBR322 plasmid DNA. *Gene* **17:** 107-112.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., and Barrell, B. (2000) Artemis: sequence visualization and annotation. *Bioinformatics Applications* **16(10):** 944-945.

Struve, C., Bojer, M., and Krogfelt, K.A. (2008) Characterization of *Klebsiella pneumoniae* Type 1 fimbriae by detection of phase variation during colonization and infection and impact on virulence. *Infection and Immunity* **76(9):** 4055-4065.

Weiss, A.A. and Hewlett, E.L. (1986) Virulence factors of *Bordetella pertussis*. *Annual Review of Microbiology* **40:** 661-686.

Figures and Tables

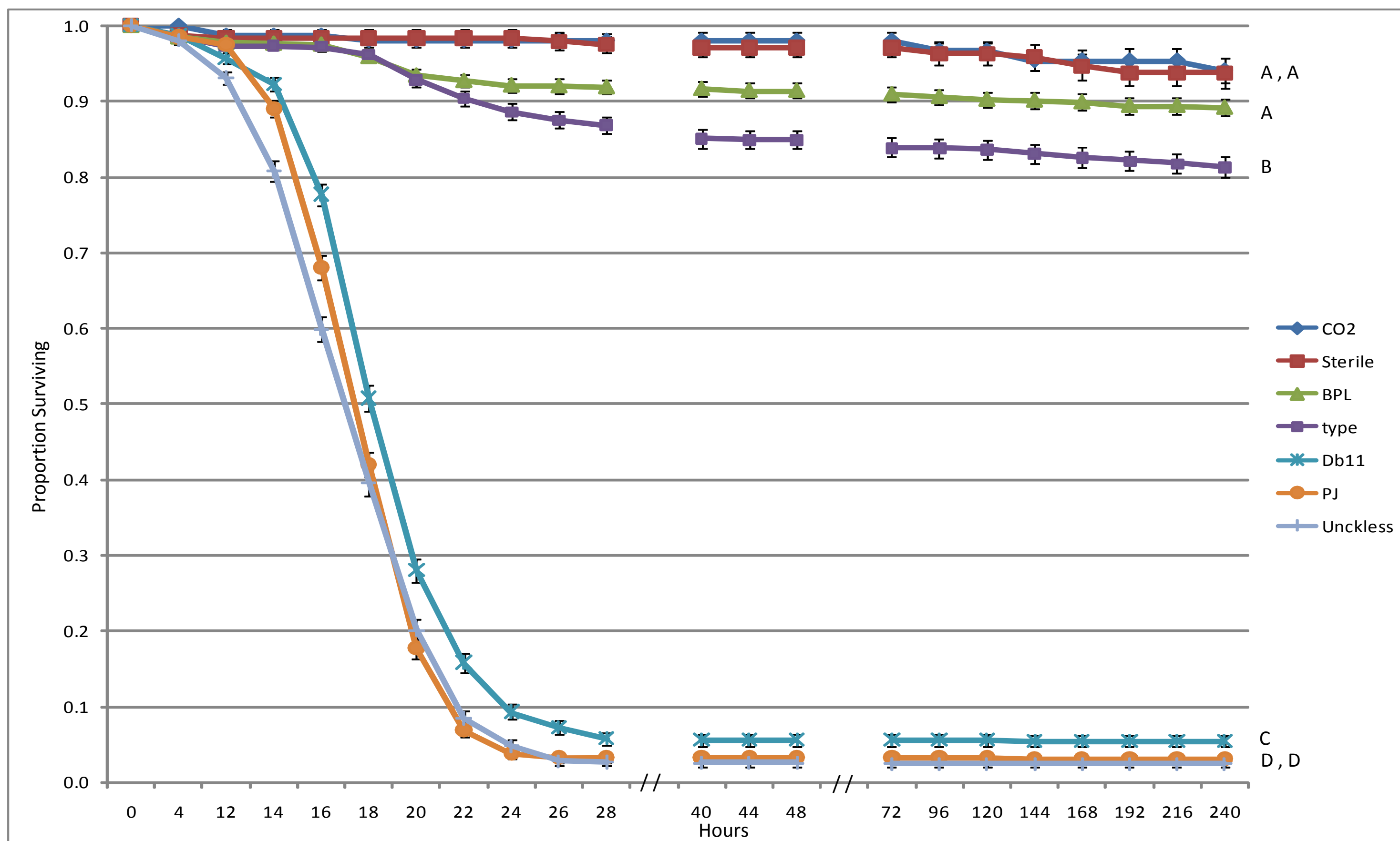

Figure 1: The proportion of *Drosophila melanogaster* flies surviving for ten days after being infected with one of five *Serratia marcescens* strains (BPL, Type, Db11, PJ, and Unckless). Lines with the same letter next to them are not significantly different from each other.

Figure 2: A *Drosophila melanogaster* fly infected with the pink-pigmented *Serratia marcescens* strain Unckless (bottom) and an uninfected fly (top).

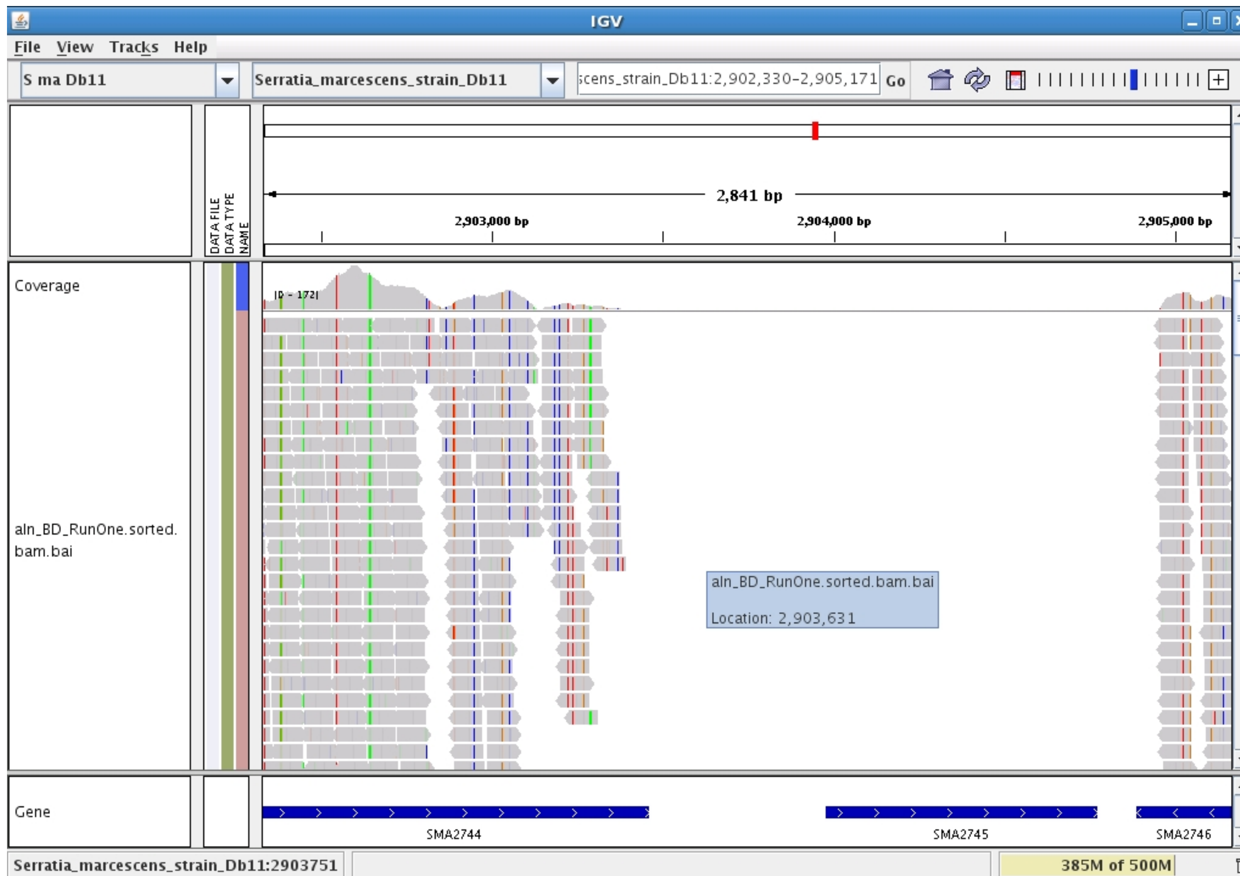

Figure 3: An example of a single gene deletion in the *Serratia marcescens* BPL genome, as shown by a screenshot of the Integrative Genomics Viewer (IGV) software interface. The blue lines at the bottom of the screen are the open-reading frames (ORFs) in the Db11 genome, and the gray blocks each represent an individual Illumina-Solexa read of the BPL genome. Colored lines in the gray blocks show single nucleotide polymorphisms in the reads. The absence of any reads in the region of the Db11 gene SMAD2745 indicates that this gene is absent in the BPL genome.

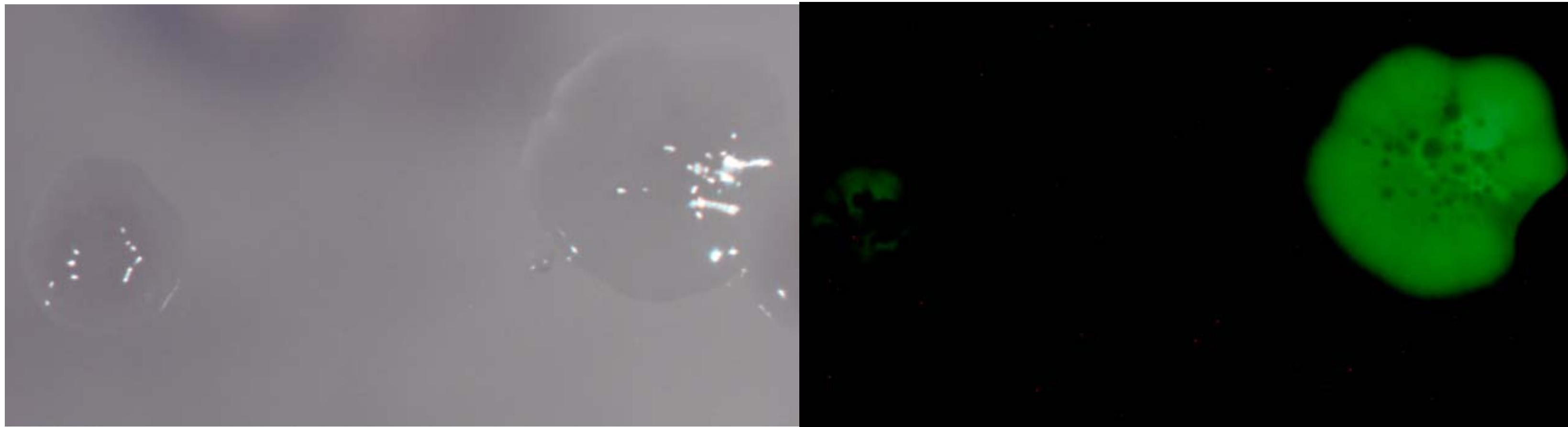

Figure 4: *Serratia marcescens* Type colonies transformed with the pEGFP plasmid, which contains a gene for green fluorescent protein (GFP). The colonies are shown in white light (left) and in light that makes the GFP fluoresce (right).

Target Region	Orientation	Primer Sequence	Purpose
SMAD1051-1057	Forward	GGACTAGTCAGGTGTAACACGCGATCAG	Confirm presence of region in Db11/absence in BPL
	Reverse	GGGAGCTCAGAAACACGGTATCGCCGTA	
Upstream of SMAD1051	Forward	GGACTAGTCAGGTGTAACACGCGATCAG	Amplify 300 bp of homologous sequence for cloning
	Reverse	GGATTTAAATAATAGATAAGGCGGCGTTGA	
Downstream of SMAD1057	Forward	GGGCGGCCGCCCAAATATCGTTTGGGTTGC	Amplify 300 bp of homologous sequence for cloning
	Reverse	GGGAGCTCTTTTACAGCCAAAGCCATCC	
Upstream of SMAD2299	Forward	GGACTAGTACCACGCTGAAGGTTTTACG	Amplify 300 bp of homologous sequence for cloning
	Reverse	GGATTTAAATTCTACTTGATCGTCGCCAGA	
Downstream of SMAD2302	Forward	GGGCGGCCGCTGAAGGATTCCCCTGAGTGT	Amplify 300 bp of homologous sequence for cloning
	Reverse	GGGAGCTCGCCAGTCGCAATTTTATTGG	
Upstream of SMAD4212	Forward	GGACTAGTGCAGTACGTGCTGGAAGACA	Amplify 300 bp of homologous sequence for cloning
	Reverse	GGATTTAAATAATTTGGCGGAAGATCACAG	
Downstream of SMAD4216	Forward	GGGCGGCCGCCCAATCATCGTCTCCCTTGT	Amplify 300 bp of homologous sequence for cloning
	Reverse	GGGAGCTCCACGTTCTAAGCCAGTGCAA	

Table 1: List of primers used for verification of deleted regions and in the amplification of flanking homologous sequence to the gene regions for use in cloning. Attempts to amplify the SMAD2299-2302 and SMAD4212-4216 regions in both Db11 and BPL were unsuccessful, and for this reason none of the primers used in these attempts are included above. Restriction sites are included in the primers to allow digestion of the amplified product and incorporation into a plasmid vector. Restriction sites in the primers are highlighted, with the site for *NotI* (GCGGCCGC) in blue, for *SpeI* (ACTAGT) in yellow, for *SwaI* (ATTTAAAT) in red, and for *SacI* (GAGCTC) in green.

	BPL	Db11	PJ	Type	Unckless	CO2
Db11	< 0.0001					
PJ	< 0.0001	< 0.0001				
Type	< 0.0001	< 0.0001	< 0.0001			
Unckless	< 0.0001	< 0.0001	0.0134	< 0.0001		
CO2	0.0576	< 0.0001	< 0.0001	0.0001	< 0.0001	
Sterile	0.0296	< 0.0001	< 0.0001	< 0.0001	< 0.0001	0.8848

Table 2: *P*-values of pair-wise comparisons using the proc lifetest function of SAS software between the five *S. marcescens* and the two controls.

Strain	**# of Reads**	**Average Coverage**
Type	8,679,350	160.38X
BPL	10,487,765	193.79X
PJ	11,952,491	220.86X
Unckless	11,723,470	160.38X

Table 3: Number of 85-base pair Illumina-Solexa reads and average coverage for the four *Serratia marcescens* strains whose genomes were sequenced. Average coverage was calculated using the size of the Db11 genome (4,600 Kb).

Gene Region	Db11	BPL	Type	PJ	Unckless	Genes of Interest
SMAD0789-0792	P	A	A	A	A	Fimbrial Proteins
SMAD1051-1057	P	A	A	A	P	Mannose-sensitive Fimbrial Proteins
SMAD1443	P	A	A	A	A/P	Iron Uptake Gene
SMAD2299-2302	P	A	A	P	A	HasB Protein / Iron Uptake Genes
SMAD2473-2479	P	A	A	P	A	Iron Uptake Genes
SMAD2745	P	A	A	A	A	Fimbrial Protein
SMAD2862	P	A	A	A	P	Iron Uptake Gene
SMAD3224-3226	P	A	A	A	A	Fimbrial Proteins
SMAD3379-3385	P	A	A	A	A	Fimbrial Proteins
SMAD3392	P	A	A	A	P	Iron Uptake Gene
SMAD4212-4216	P	A	A	P	A	Virulence Protein
SMAD4418-4437	P	A	A	A	A	Fimbrial Proteins
SMAD4440-4444	P	A	A	A	A	Fimbrial Proteins
SMAD4446-4449	P	A	A	A	A	Fimbrial Proteins

Table 4: List of *S. marcescens* Db11 gene sets that are absent in the BPL genome, and in which at least one of the genes shared sequence homology with genes involved in virulence in other bacteria. "A" indicates that the genes listed are absent in that strain, whereas "P" indicates the presence of the genes in a strain. The "A/P" listed for SMAD1443 in the Unckless strain indicates that part of the gene was absent, and the functionality of the gene or any protein product is unknown.