Sequencing *Salmonella*, *hilA*, HilA, A, A, A...

A sequence analysis of two mutant *hilA* genes

Slater Sharp and Becca Lensing
Cornell College
9/22/2014

INTRODUCTION

Salmonella, a facultative intracellular parasite, is one of the most common causes of foodborne bacterial disease in the world. It becomes incorporated into the human gastrointestinal system primarily through the ingestion of infected food-producing animals (Rychlik and Barrow, 2005). Virulence, the capacity of *Salmonella* to invade and infect host epithelial cells, is dependent on both environmental and genetic factors. Genes encoding pathways for intracellular pathogenesis are located on "pathogenicity islands" in the bacterial chromosome (Bäumler et al., 2000). Although twelve pathogenicity islands have been identified, the roles of specific genes in virulence are not well understood (Hensel, 2004). Understanding the genetic controls and mechanisms responsible for the invasive properties of *Salmonella* may enable the prevention of health endemics, economic suffering, and productivity losses caused by the pathogen. The *hilA* gene encodes an important transcriptional regulator of *Salmonella* pathogenicity island 1 (SPI1). Mutations in *hilA* can lead to the reduced expression of SPI1 genes, resulting in avirulence (Lostroh and Lee, 2001). In our study the *hilA* genes of two avirulent *Salmonella* strains, mutated in-vitro, were sequenced and compared to that of wild-type *Salmonella*. Base-pair mutations conserved across the two avirulent strains indicated crucial sites for *Salmonella* pathogenicity. Through this retrospective approach, we were able to provide insight into the nature of *Salmonella* virulence by drawing conclusions about the genetic causes of avirulence.

The primary role of *hilA*-regulated genes on SPI1 is the expression of the type III secretion system 1 (T3SS-1), a known determinant of *Salmonella* virulence. T3SS-1 is organized into a supramolecular structure known as the "needle complex", responsible for the transfer of bacterial proteins into eukaryotic epithelial host cells (Lostroh and Lee, 2001). Discrete T3SS-1 substructures include a needle-like projection that extends outwards from the bacterial envelope and a base that spans

the width of the double membrane (Kubori et al., 2000). The base is composed of two pairs of rings joined by a hollow cylindrical structure. Structural protein InvG comprises the outer rings while PrgH and PrgK form the inner rings and cylinder. The needle is made up of a single protein, PrgI, stabilized by PrgJ (Kubori et al., 2000). InvG is encoded on the *invF* operon at one end of SPI1, while PrgH, PrgK, PrgI, and PrgJ are encoded on the *prgH* operon at the opposite end of the 40 Kbp pathogenicity island (Lostroh and Lee, 2001).

Transcription of T3SS-1 structural genes *invF* and *prgH* is activated by HilA, a transcription factor coded for by the gene *hilA*. HilA recognizes specific 17-nucleotide-long sequences in the *invF* and *prgH* promoters on SPI1. Each of these 17-nucleotide sequences contains two copies of the hexameric repeat TTXYAT, where X represents an adenine or thymine base and Y represents an adenine, thymine, or cytosine base. This repeated motif makes up the HilA protein binding site, or the "HilA box." Binding of HilA to the HilA box directly activates the promoters P_{invF} and P_{prgH}, which in turn activate the *invF-1* and *prgH* genes, producing the proteins required for T3SS-1 biosynthesis (Lostroh and Lee, 2001).

Salmonella virulence requires a HilA-dependent cascade of transcriptional activation, and through the expression of *invF* and *prgH*, initiates the construction of the T3SS-1 "needle complex" (Lostroh and Lee, 2001). Mutations in the *hilA* gene structurally alter the protein HilA, potentially affecting its ability to bind to the HilA box and activate the *invF* and *prgH* promoters, compromising the cascade. The bacteria, unable to assemble T3SS-1, is rendered avirulent.

Here we report the *hilA* sequence analysis of two avirulent *Salmonella* strains (A5 and B3) and characterize the mutant sequences. Select mutations were conserved across the strains, indicating crucial

base pair sequences necessary for the translation of functional HilA protein. Origins of avirulence were derived at the nucleotide level.

METHODS AND MATERIALS

Plasmid Isolation

C. Phoebe Lostroh introduced random mutations into pHilA and isolated mutant plasmids unable to activate the *invF* and *prgH* genes. *E. coli* containing pHilA strains A5 and B3 were grown overnight in LB Broth containing 100 µg/mL ampicillin. Plasmid DNA was isolated using the IBI Fast Ion Plasmid Isolation Kit. Cells were pelleted at 7,000 rpm for 15 minutes. The bacterial pellets were resuspended in buffer PM1 (50 mM Tris-Cl, pH 8.0, 10mM EDTA, 100 µg/mL RNase A) and incubated at room temperature for five minutes in buffer PM2 (100 mM NaOH, 1% w/v SDS). Buffer PM3 (3.0 M potassium acetate, pH 5.5) was added, and the lysate was centrifuged at 11,000 rpm for 20 minutes. The supernatants were spun at 11,000 rpm for 15 minutes and transferred into IBI anion-exchange columns equilibrated with PEQ buffer (750 mM NaCl, 50 mM MOPS pH 7.0, 15% v/v isopropanol, 0.15% Triton X-100). The columns were washed with PW buffer (1 M NaCl, 50 mM MOPS pH 7.0, 15% v/v isopropanol) and the plasmids were eluted with PEL buffer (50 mM Tris-Cl pH 8.5, 1.25M NaOH, 15% v/v isopropanol). Plasmid DNA was precipitated with 0.75 volumes of room-temperature isopropanol and centrifuged at 12,000 rpm for 30 minutes at 4°C. The DNA was ethanol (75%) precipitated, centrifuged at 12,000 rpm for 10 minutes, and resuspended in sterile dH_20. Plasmid mass was determined by scanning the DNA from 230-350 nm on a Nanodrop spectrophotometer and recording peak absorbance at 260 nm.

Plasmid Characterization

Plasmids pHilA(A5) and pHilA(B3) (0.5 μg) were digested with Nco I and Xba I to remove the *hilA* insert, Xba I to linearize plasmid the construct, or left undigested to function as negative controls. Lambda (0.5g) digested with Hind III and Eco RI and a 100 b.p. ladder served as molecular weight markers. Samples were digested overnight at 37°C. Digested samples containing tracking dye (50% v/v glycerol, 0.4% w/v bromophenol blue) ran on a 0.7% (w/v) agarose gel made in Tris acetate buffer (0.04M Tris-acetate, 0.35M glacial acetic acid, 0.01M EDTA, pH 8.0) containing 0.5μL Midori Green Advance DNA Stain, a non-mutagenic alternative to Ethidium bromide. Samples were run at 65 volts in Tris acetate buffer (0.04M Tris-acetate, 0.35M glacial acetic acid, 0.01M EDTA, pH 8.0) and were photographed under ultraviolet light.

PCR Amplification and Sequencing

Plasmids pHilA(A5) and pHilA-B3 were sequenced using dideoxy DNA sequencing. The PCR reaction contained pHilA plasmid template (1,2,3μg), USB Thermo Sequenase Reaction Buffer (11.76% v/v), USB Thermo Sequenase DNA Polymerase (11.76% v/v), M13 Forward primer ATGCCATAGCATTTTTATCC (88nM), M13 Reverse primer CCTGATACAGATTAAATC (88nM), and dNTP nucleotide mix (5.88% v/v). The final master mix volume was 17μL. 4μL of the mix was aliquoted into thin-walled PCR tubes and 4μL dideoxynucleoside triphosphate (A,T,G, or C) was added. The tubes were run in a thermocycler at 92°C for two minutes, then cycled at 92°C for 30 seconds, 47°C for 30 seconds, and 70°C for 1 minute for a total of 30 cycles. The reactions were denatured at 92°C for 3 minutes and run on 5.5% (w/v) denaturing polyacrylamide gels (7.0M urea) in 0.8X TBE (0.072M Tris, 0.072M boric acid, 0.0016M EDTA) at 2,000 volts at 45°C and read under

infrared by a LI-COR 4300 DNA Analyzer. Computer program LI-COR e-Seq translated the resulting banding pattern into a base pair sequence.

RESULTS AND DISCUSSION

Gel Electrophoresis of *hilA*

Prior to our investigation, two in-vitro mutated 1,600 base pair (bp) *hilA* sequences were inserted into 4,100bp pBAD plasmids. The resulting plasmid pHilA, measuring around 5,700 bp, was inserted into an *E. coli* host to allow for its replication. Following plasmid isolation, pHilA(A5) and pHilA(B3) were characterized using restriction enzymes Nco I and Xba I and run on an agarose gel (*Figure 1*) in order to confirm the identity of the pHilA plasmids and the isolated *hilA* sequence. Undigested plasmid functioned as a negative control in order to confirm that the linearized plasmids were properly digested.

Lambda (λ)/EcoRI presented 21,226 bp, 7,421 bp, 5,804-5,645 bp and 4,878 bp molecular weight markers and Lambda (λ)/HindIII displayed weight markers at the 23,130 bp, 9,416 bp, 6,557 bp, 2,322 bp, and 2,027 bp regions. A faint 4,361 bp marker was also present. The 1,500 bp and 500 bp markers were the most distinguished bands of the 100 bp ladder.Ṗrominent bands of undigested plasmid presented near the 7,421 bp marker of λ/EcoRI. Plasmids linearized with only Xba I expressed multiple banding patterns, most of which were conserved across the linearized pHilA(A5) and pHilA(B3). The band of interest appeared at the 5,804-5,643 bp marker of λ/EcoRI. Other bands presented at the 20,000 bp, 18,500 bp, 14,000 bp, and 7,500 bp regions, approximately. Plasmids digested with Nco I and Xba I exhibited three consistent bands near the 5,804-5,643 bp and slightly below the 4,878 bp markers of λ/EcoRI and just above the 1,500 bp marker of 100 bp ladder.

Both undigested plasmids appeared near the 7,421 bp marker of λ/EcoRI, despite their 5.7 Kbp length. This is most likely due to the plasmids' supercoiled conformations, altering their abilities to travel through the gel. The undigested plasmids and Xba I-linearized pHilA exhibited dissimilar banding patterns, confirming that the cut plasmids were properly digested. The 5,804-5,643 bp-localized bands of the Xba I-linearized pHilA indicated the length of the linearized plasmids to be approximately 5.7 Kbp long. This model is consistent with pHilA. The extraneous bands of Xba I-linearized pHilA were unaccounted for, and were likely the result of sample contamination. Nco I and Xba I-digested plasmids each exhibited three linearized fragments 5.7 Kbp, 4.1 Kbp, and 1.6 Kbp in length. The 4,100 bp fragments were characteristic of linearized pBAD, and the 1,600 bp fragments were presumptively the linearized *hilA* inserts. The 5.7 Kbp fragments are likely partially digested pHilA plasmids that have been linearized by a single restriction enzyme. This assumption is supported by their banding similarity with the Xba I-linearized pHilA fragments. These final results strongly support the conclusion that the plasmids were indeed pHilA, and that both plasmids included the *hilA* insert.

Dideoxy DNA Sequencing of *hilA*

hilA template DNA was amplified through polymerase chain reaction (PCR) using pBAD-specific M13 forward (ATGCCATAGCATTTTTATCC) and reverse (CCTGATACAGATTAAATC) primers. Elongation of the templates was terminated in 1-bp increments by the addition of a specific dideoxynucleoside triphosphate (ddNTP). The four reactions (ddATP, ddCTP, ddGTP, and ddTCP) were run separately on a 5.5% denaturing polyacrylamide gel where variations in chain-length as small as one nucleotide could be differentiated. Each band on the gel marked a specific termination point in the daughter strand; therefore, by running the four reactions in

parallel, the sequence of the *hilA* gene could be read off the gel by the LI-COR e-Seq computer program. Multiple four-reaction samples of different volumes (1.0 μL, 0.75μL, and 0.5μL) were run simultaneously for each of the pHilA(A5) and pHilA(B3) plasmids.

The resulting sequences were tested against the full *Salmonella* genome of *Salmonella enterica* subsp. *enterica serovar Enteritidis* strain SEJ, as released by the Los Alamos National Laboratory (Bishop-Lilly et. al, 2014), and the National Center for Biotechnology Information. In order to verify that the conserved non-homology between the avirulent strains was not due to natural variations between the SEJ subspecies strain and the original strain mutated by Lostroh, the conserved regions were compared to *Salmonella enterica* subsp. *enterica serovar Enteritidis* strains OLF-SE6-00219-16 (SE6), OLF-SE1-1019-1 (SE1), and OLF-SE5-1104-2 (SE5), also from the National Center for Biotechnology Information.

Mutated *hilA*(A5) was sequenced 720 bp in the forward direction and 562 bp in the reverse direction with 99% homology (714/720 and 559/562). These 1,282 nucleotides comprise 77.4% of the 1,656 bp *hilA* gene. The unsequenced middle region measures 374 bp in length. Mutated *hilA*(B3) was sequenced 511 bp in the forward direction and 312 bp in the reverse direction with 99% homology (505/511 and 310/312). These 820 nucleotides comprise 49.5% of the *hilA* gene. The unsequenced middle region measures 836 bp in length. Sequenced regions displaying higher rates of ambiguity were disregarded; the probability of a non-homologous base in the regions being the result of a mutation was decidedly higher. However, these dissimilarities between the SEJ *hilA* sequence and the mutated *hilA* sequences may have been the result of experimental error and not random mutation.

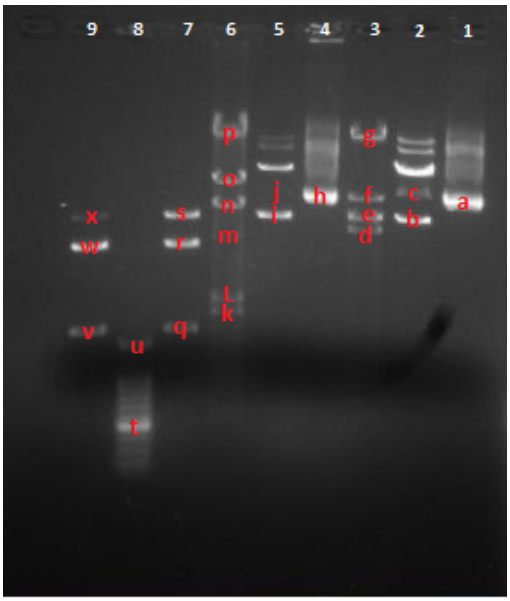
Comparison of *hilA*(A5) to the *Salmonella enterica* SEJ strain genome revealed two point mutations, two insertions, and five deletions. Comparison of *hilA*(B3) to the *Salmonella enterica* SEJ

strain genome revealed two point mutations, one insertion, and five deletions. Of these, two point mutations (C to G and A to T) and two deletions (C and T) on the forward strands and one deletion (A) on the reverse strand are conserved between the mutated *hilA*(A5) and *hilA*(B3) sequences. When *hilA*(A5) and *hilA*(B3) were compared to the SE6, SE1, and SE5 strains, these mutations remained conserved.

One possible explanation for the avirulence of *Salmonella* strains A5 and B3 is that the induced mutations in *hilA* modified the amino acid sequence of protein HilA. These changes may have rendered the protein's active site inert, by altering either protein folding or the active site itself; consequently, the ability of protein HilA to bind to the HilA box may have been compromised. Without a functional transcription factor for *invF* and *prgH*, the T3SS-1 structural proteins were not expressed. These *Salmonella* bacteria were unable to physically pierce, and therefore infect, eukaryotic epithelial host cells. Conservation of these mutations between the avirulent *Salmonella* strains A5 and B3 suggests that *Salmonella enterica* subsp. *enterica serovar Enteritidis* strain SEJ base pairs 1201733, 1201736, 1201809, 1201813, and 1203340 may be necessary for proper *hilA* function, and by extension *Salmonella* virulence.

APPENDIX

Figure 1: Agarose gel electrophoresis



We confirmed the presence of *hilA* in plasmid pHilA by agarose gel electrophoresis: **a)** supercoiled pHilA(A5) (5700bp); **b)** linearized pHilA(A5) (5700bp); **c)** supercoiled pHilA(A5) (5700bp); **d)** 4878bp; **e)** 5804bp + 5643bp; **f)** 7421bp; **g)** 21,226bp; **h)** supercoiled pHilA(B3) (5700bp); **i)** linearized pHilA(B3) (5700bp); **j)** supercoiled pHilA(B3) (5700bp); **k)** 2027bp; **L)** 2322bp; **m)** 4361bp; **n)** 6557bp; **o)** 9416bp; **p)** 23,130bp; **q)** lineazed *hilA*(A5) (1600bp); **r)** linearized pBAD (4100bp); **s)** linearized pHilA(A5) (5700bp); **t)** 500bp; **u)**1500bp; **v)** lineazed *hilA*(B3) (1600bp); **w)** linearized pBAD (4100bp); **x)** linearized pHilA(B3) (5700bp)

Table 1: Gel electrophoresis loads with restriction enzymes

Lane #	Sample	Enzyme
1	pHilA(A5)	-
2	pHilA(A5)	XBA I
3	λ/EcoRI	-
4	pHilA(B3)	-
5	pHilA(B3)	XBA I
6	λ/HindIII	-
7	pHilA(A5)	NCO I + XBA I
8	100bp Ladder	-
9	pHilA(B3)	NCO I + XBA I

REFERENCES

Bäumler, A.J., R.M. Tsolis and F. Heffron. 2000. Virulence mechanisms of *Salmonella* and their genetic basis: In *Salmonella* in domestic animals. C. Wray & A. Wray, eds. CABI Publishing, Wallingford, Oxfordshire, pp. 57-72.

Bishop-Lilly, K.A., S.M. Broomall, P.S. Chain, O. Chertkov, S.R. Coyne, H.E. Daligault, K.W. Davenport, T. Erkkila, K.G. Frey, H.S. Gibbons, W. Gu, J. Jaissle, S.L. Johnson, G.I. Koroleva, J.T. Ladner, C.-C. Lo, T.D. Minogue, C. Munk, G.F. Palacios, C.L. Redden, C.N. Rosenzweig, M.B. Scholz, H. Teshima, and Y. Xu. 2014. Salmonella enterica subsp. enterica serovar Enteritidis strain SEJ, complete genome. National Center for Biotechnology Information. CP008928.1.

Fahlen, T.F., N. Mathur, and B.D. Jones. 2000. Identification and characterization of mutants with increased expression of hilA, the invasion gene transcriptional activator of Salmonella typhimurium. Federation of European Microbiological Societies Immunology and Medical Microbiology. 28: no. 1: pp. 25-35.

Hensel, M. 2004. Evolution of pathogenicity islands of Salmonella enterica. International Journal of Medical Microbiology, no. 294: pp. 95-102.

Kubori, T., A. Sukhan, S. Aizawa, and J. Galán. 2000. Molecular characterization and assembly of the needle complex of the *Salmonella typhimurium* type III protein secretion system. Proceeding of the National Academy of Sciences of the United States of America. 9: no. 18: pp. 10225–10230.

Lostroh, C.P and C.A. Lee. 2001. The HilA and sequences outside it determine the magnitude of the HilA-dependent activation of P_{prgH} from *Salmonella* pathogenicity island 1. Journal of Bacteriology, American Society for Microbiology. pp. 4876-4885.

Rychlik, I. and P.A. Barrow. 2005. Salmonella stress management and its relevance to behaviour during intestinal colonisation and infection. FEMS Microbiology Review, no. 29: pp. 1021-1040.