Journal of Experimental Microbiology and Immunology (JEMI) Vol. 19: 1 – 7

Subcloning and Sequencing of the Autotransporter Antigen 43 from *Escherichia coli* K-12 as an Initial Step Toward Understanding the Relationship Between Ag43 and Capsule in Biofilm Formation

Ryan Tse, Cybele Go, Stina Lindh
Department of Microbiology and Immunology, University of British Columbia

The autotransporter Antigen 43 (Ag43) is present in numerous *Escherichia coli* strains and functions in autoaggregation and adhesion leading to biofilm formation. However, biofilm formation and regulation is not well understood. We hypothesized that Ag43-mediated biofilm formation is prevented by capsule expression. In this study, we describe the construction of plasmid vector encoding the autotransporter protein Ag43 designed to test our hypothesis. Polymerase chain reaction was used to amplify Ag43 from the *E. coli* K-12 genome. The resulting PCR product was subcloned into pCR2.1-TOPO Cloning Vector and transformed into *Escherichia coli* strain OneShot TOPO. Clones were selected and isolated plasmids were screened by restriction enzyme digestion followed by agarose gel electrophoresis. The plasmid containing a putative insert, pRCS715, was sequenced and shown to contain the Ag43 sequence.

Escherichia coli (*E. coli*) biofilms have been problematic in nosocomial settings because of its increased tolerance to stress, antibiotics and host immunological defenses thus allowing for bacterial growth (1). The formation of biofilm can be mediated by various factors including surface adhesins, pili, multiple autotransporter proteins, matrix and capsular polysaccharides depending on environmental conditions (1). Antigen 43 is an autotransporter found on the outer-membrane of most *E. coli* strains. Ag43 allows the bacteria to sediment and form the initial stages of cell-cell aggregation required in biofilm formation (2, 6).

Antigen 43 (Ag43) is a 1,039 amino acid pre-protein encoded by the *flu* gene and consists of three domains; a signal peptide located on the N-terminus of the protein, a passenger (alpha) domain and a translocation (beta) domain (2) (Fig. 1). The N-terminal signal peptide directs transportation of Ag43 across the cytoplasmic membrane (2). The signal sequence is cleaved releasing the C-terminus into the periplasm. The beta domain of the cleaved protein is thought to spontaneously insert itself into the outer-membrane and form a transmembrane beta-barrel pore (3) through which the alpha domain can be transported onto the surface of the cell (2). The alpha domain of Ag43 is responsible for cell-to-cell binding, by self-recognition of Ag43 on adjacent cells. Most *E. coli* strains have multiple copies of the *flu* gene, which may have slight sequence differences (1).

The aim of this study was to generate a plasmid construct containing Antigen 43 for use in later experimentation to determine the function of this autotransporter. A plasmid copy of Ag43 is useful because it allows for the use of a regulated promoter to control gene expression levels. Furthermore, plasmids can allow for the expression of modified gene products that do not occur naturally such as the insertion of a polyhistidine tag in Ag43 to allow detection. Beloin *et al.* have discovered that capsular polysaccharides also play a role in improving cell-cell contact and that there are numerous experimental questions to be addressed regarding the ability of *E. coli* to self-aggregate (1). We hypothesized that overexpression of capsular polysaccharides in *E. coli* will mask the binding domain of Ag43 and therefore prevent Ag43-mediated cell-to-cell aggregation. In this study, we amplified the *flu* gene using primers adapted from Kjaergaard *et al.* (5). The PCR product was then subcloned into a pCR2.1-TOPO cloning vector and transformed into OneShot TOPO chemically competent cells. Putative plasmids isolated from blue-white colony screening and kanamycin selection were confirmed by gene sequencing.

MATERIALS AND METHODS

Bacterial strains and growth conditions. *E. coli* K-12 strain MG1655 was used to obtain genomic DNA for PCR amplification of the *flu* gene. The strain was obtained from the MICB 421 Culture Collection in the Microbiology and Immunology Department at University of British Columbia. Stock cultures were streaked onto Luria-Bertani (LB) agar (1.0% Tryptone, 0.5% Yeast extract, 1.0% NaCl, 1.5% agar) plates and a single colony was inoculated in 5 mL of LB broth (1.0% Tryptone, 0.5% Yeast extract, 1.0% NaCl) The culture was incubated overnight in a shaking incubator at 37˚C at 200 rpm.

Genomic DNA isolation. Genomic DNA extraction from *E. coli* K-12 MG1655 overnight culture was done with use of the PureLink Genomic DNA mini kit (Invitrogen, Cat. No. K1820-01) following the manufacturer's protocol. Genomic DNA concentration and purity was validated using Thermoscientific NanoDrop 2000c Spectrophotometer. The Sample was stored at -20˚C.

Primer construction. Primers to amplify the *flu* gene from *E. coli* K-12 MG1655 were adapted from Kjaergaard *et al*. such that the restriction sites were modified to allow ligation into the pBluescript SK(+) plasmid. The forward primer (P1: 5′-CCCGCGGCCGCGAATTCCTTTGTCAGTAACATGC-3′) binds upstream of the *flu* gene and an EcoRI restriction site (underlined) was introduced. The reverse primer (P4: 5′-CCCGCGGCCGCAAGCTTTGTGGCGTTGAAGATCCG-3′) binds downstream of the *flu* gene and a HindIII restriction site

(underlined) was introduced. All of the designed primers were obtained from Integrated DNA Technologies.

Polymerase Chain Reaction (PCR) amplification and cloning of the *flu* gene. PCR was done in 100 µl reactions with the use of *Taq* DNA Polymerase (Invitrogen, Cat. No. 18038-042). The manufacturer's protocol was followed using 2 U *Taq* Polymerase, MG1655 genomic DNA as template and a primer mix of P1 and P4.

Touchdown PCR was performed using the BioMetra Thermocycler with the following cycling conditions: 5 minutes of denaturation at 94°C, 11 amplification cycles with 94°C for 45 seconds, 61°C - 50°C with 1°C decrements for 30 seconds and 190 seconds of elongation at 72°C, followed by 25 amplification cycles with 94°C for 45 seconds, 50°C for 30 seconds and 3 min 10 seconds elongation step at 72°C, and final elongation for 10 minutes at 72°C.

Gel electrophoresis of PCR product. 10X loading buffer (Invitrogen) was added to TrackIt 1 kb Plus DNA Ladder (Invitrogen, 10488-085). 10X loading buffer (Invitrogen) was added to the PCR product. The prepared ladders and PCR samples were loaded onto a 1.0% agarose gel (containing SYBR safe). A 1 kb DNA ladder and the PCR product was loaded into wells. The gel was run in a 1X TBE buffer solution at 120 volts for 45 minutes. The gel was imaged using an Alpha Imager (Multimage™ Light Cabinet).

Transformation into OneShot TOPO chemically competent cells. The amplified *flu* gene was directly inserted into pCR2.1-TOPO cloning vectors and transformed into One Shot TOPO10 chemically competent cells using TOPO TA Cloning Kit (Invitrogen, Cat. No. K4500-40). The manufacturer's protocol was followed. The transformed cells were selected for on LB agar plates containing kanamycin and 20mg/mL of X-Gal spread on the surface.

Plasmid purification. Seven white isolated colonies were each incubated in LB media containing 50 µg/mL kanamycin. Plasmids were purified from overnight culture using PureLink Quick Plasmid Miniprep Kit (Invitrogen). The manufacturer's protocol was followed with the modification that distilled and autoclaved water was used to elute the plasmid. Plasmid concentration and purity was validated using Thermoscientific NanoDrop 2000c Spectrophotometer. The Sample was stored at -20˚C.

Confirmation of Antigen 43 insertion. Three randomly chosen plasmid samples from within the purification step were then sequenced at the NAPS unit at Michael Smith Laboratories in UBC. These samples were then aligned and verified using the NCBI BLAST database to confirm the presence of the Antigen 43 sequence.

FIG 1. Schematic of the Antigen 43 gene found in *E. coli* K12 MG1655. SP is a signal peptide that sends Ag43 into the periplasm. Alpha is the extracellular domain responsible for interaction with other Ag43. B is a beta-barrel domain that forms a pore in the outer membrane allowing the alpha domain to thread through to the extracellular space.

RESULTS

Polymerase Chain Reaction amplification of *E. coli* genome yields a 3.2kb product. The nucleotide sequence of *E. coli* K12 MG1655 *flu* gene (EcoGene Accession Number: EG12686) was compared with primer P1 and P4. In accordance with Kjaergaard *et al*., the primers were complementary to nucleotide sequences found upstream and downstream of the *flu* gene (5). PCR was used to amplify the 3.2 kBp gene from *E. coli* K12 MG1655 genome. Touchdown PCR was performed using the P1 and P4 primers, which bind upstream and downstream, respectively, of the complete *flu* gene. Using the IDT OligoAnalyzer tool, the optimal PCR annealing temperatures were determined to be 61˚ and 51˚ for P1 and P2, respectively. A Touchdown PCR technique was used to account for the difference in annealing temperature.

PCR reactions were analyzed by agarose gel electrophoresis. As shown in Fig. 2, lanes 2-4 show a discrete band between 3kb and 4kb in size for each of the PCR reactions which is consistent with the expected size of the 3.2 kBp flu gene product.

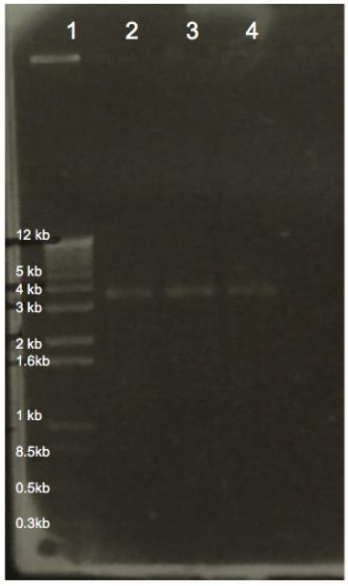

FIG 2. Gel electrophoresis of touchdown PCR amplified *flu* gene product. Lane 1: 1kb+ Track it DNA ladder; Lane 2, 3, 4: Touchdown PCR product. Bands found between 3-4kb.

Subcloning and sequencing of flu gene in plasmid pCR2.1-TOPO The 3.2 kBp PCR product was purified and ligated into plasmid pCR2.1-TOPO (Fig. 3B). Colonies were screened for inserts on X-gal and selected for plasmid transformants using kanamycin. Bacteria that contain the plasmid form colonies in the presence of kanamycin. A cloned vector has a non-functional *lacZ* gene and is a white colony in the presence of X-gal. Several white colonies were picked and grown overnight in liquid media for subsequent plasmid isolation. Three plasmids were sequenced using the M13 Forward and Reverse primers from the TOPO TA Cloning Kit (see Supplemental Data 1).

The sequences were aligned using the NCBI BLAST database. Blastn analysis showed the sequence had 98% identification as part of the *E. coli* K12 MG1655 genome (see Supplemental Data 2). Blastp analysis showed that the sequence had putative conserved autotransporter beta-barrel domains with 99% homology to an Ag43 precursor (see

Supplemental Data 3). The amplified gene is an *E. coli* K12 MG1655 gene that encodes for an autotransporter similar in structure to Ag43, the *flu* gene product.

BLAST analysis revealed a 98% identity homology between the cloned gene segment and Antigen 43 of *Escherichia coli*; however, point mutations seemed to be present with the Antigen 43 gene segment when compared to the *E. coli* K12 MG1655 genome found in the NCBI database. These point mutations were found towards the end of the cloned sequence analyzed. Sequencing and the use of the M13 reverse primer revealed the presence of the beta-fragment of Antigen 43 gene within the TOPO vector and is of a 5' to 3' orientation (Fig. 3A).

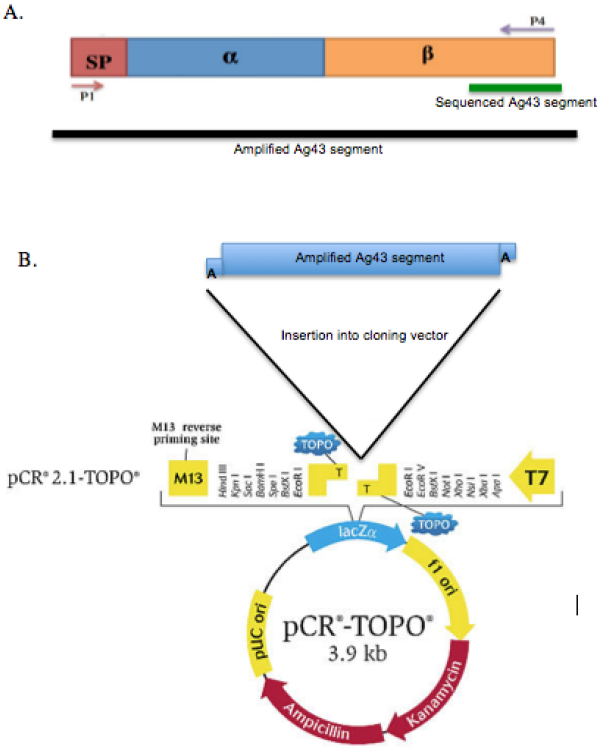

FIG 3. A) Schematic diagram of the Antigen 43 gene (*flu*) in *Escherichia coli* K-12 MG1655. The green bar located below the beta fragment is the section of the sequenced gene from NAPS. The black bar indicates the Touchdown PCR amplified region of Ag43. **B)** pRCS715 plasmid map indicating the insertion site of the amplified PCR product into pCR2.1-TOPO cloning vector.

DISCUSSION

The goal of this study was examine how overexpression of capsular polysaccharides on Antigen 43 effects autoaggregation of *Escherichia coli* K-12 MG1655. The purpose was to examine the relationship of capsule and the autotransporter in cell-cell aggregation and biofilm formation with potential implications in reducing biofilm-related infections (Fig. 4). Beloin *et al.* cited the importance of both the capsule and autotransporter in *E. coli* biofilm maturation (1); however, characterization of how these two components interact has not been well understood. Noting that type I fimbriae in *E. coli* block Antigen 43-mediated autoaggregation and that Antigen 43 functions extracellularly, the presence of overproduced capsular polysaccharides may also serve to physically block Antigen 43 and inhibit the ability of *E. coli* to aggregate (7).

To study the effects of capsule production on Ag43 cell-to-cell mediated aggregation, a cloned Ag43 plasmid was generated. The expression of Ag43 in a plasmid render both gene regulation and tagging of the protein possible. The establishment of the plasmid pRCS715 is crucial for generating capsular mutant transformants and analysis of the relationship between Antigen 43 and capsule formation. The ability to control gene expression of Ag43 enables comparison of results between different transformants and hence, the characterization of the effects capsule formation has on cell-to-cell aggregation. Constructs like pRCS715 allow potential researchers to readily transform other competent cells and experimentally test Antigen 43 in order to gain more information regarding biofilm formation and maturation of *E. coli*. In addition, the 5' to 3' orientation of the antigen 43 gene within the pCR2.1 TOPO vector allows for the autotransporter gene to be driven by the T7 promoter and can be induced independently of *E. coli* promoters thus allowing for useful future experimentation.

PCR was used to amplify Ag43 from a purified sample of the *E. coli* K12 MG1655 genome. Due to the large difference in melting temperatures of the primers used, Touchdown PCR was the most effective PCR technique to amplify the genome. Gel electrophoresis of the Touchdown PCR product demonstrated the presence of the desired Antigen 43 gene as a 3.2kb band and the success of the Touchdown PCR method. The OneShot TOPO cloning protocol allowed easy cloning of the amplified gene and screening of colonies. Cloned plasmid was sent to NAPS for sequencing.

BLAST analysis of nucleotide sequences showed significant alignment with the *E. coli* K12 MG1655 genome. This indicates that the sequence obtained was indeed from the *E. coli* K12 MG1655 strain initially isolated, and not from contaminants or other substrains of *E. coli* K12. BLASTp translated protein analysis indicated that the 1 kb sequence is potentially the beta-barrel section of an autotransporter protein. The analysis results provide evidence for the successful cloning of the Ag43 gene. Further BLAST analysis showed the presence of point mutations towards the end of the cloned gene when compared to the *E. coli* K12 MG1655 gene sequence from the NCBI database. These mutations can be caused by mutations in the *E. coli* used during the study, transcription error during PCR amplification, or from poor DNA quality.

In conclusion, the Touchdown PCR method proved to be the most effective at amplifying the *flu* gene using the primers described in this study. The plasmids were effectively cloned with the PCR gene product of Antigen 43 and transformed into a set of TOPO-competent cells for future extraction. BLAST analysis indicates that Ag43 was successfully amplified and cloned into TOPO vectors.

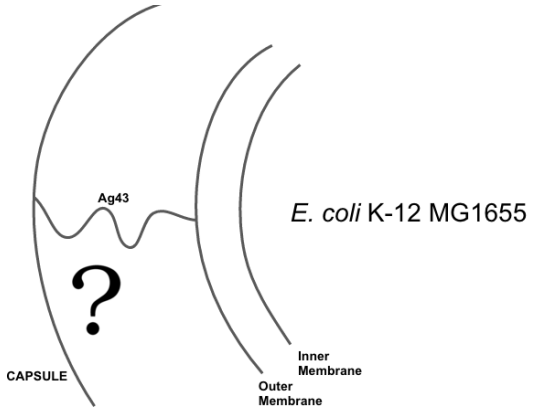

FIG 4. Diagram of possible interference between capsule and Ag43 that may affect cell-to-cell aggregation and biofilm formation in *E. coli* K12.

FUTURE DIRECTIONS

In order to study the effects of Ag43 on biofilm formation, it is vital to devise strategies to control the expression of Ag43 and to visualize Ag43 on the surface of the cell. Currently, the TOPO-vector construct will only be sufficient in characterizing the presence of Antigen 43 and not the molecular mechanisms of binding and biofilm formation. Additionally, it may be helpful to attempt to reduce the number of point mutations found in the amplified Ag43 sequence. This may be done by performing PCR and sequencing from both 5' and 3' ends of the gene or by using a polymerase with proofreading ability.

The initial goal of this study was to construct a version of Ag43 with a histidine tag inserted into the N terminus of the protein between the signal peptide and the alpha domain in order to determine if the overproduction of capsule will mask Antigen 43 binding. This polyhistidine tag allows detection of Antigen 43 on the surface of *E. coli* in order to study the effect of Ag43 on aggregation and biofilm formation. Future steps would include sequential PCR amplification of the *flu* gene to insert the histidine tag using designed primers (see Supplemental Data). This polyhistidine-tagged flu gene can then be inserted into a plasmid with an inducible lac promoter with EcoRI and HindIII restriction sites such as pBluescript SK(+) and expressed in *E. coli* K-12. This construct will allow for the assessment of the Antigen 43-related mechanisms of biofilm formation in the context of differential capsule production. Once this construct is established and transformed into *E. coli*, sedimentation assays can be performed to characterize if an overproducing capsular polysaccharide strain of *E. coli* (*lon* mutant) will mask Antigen 43 and thus inhibit sedimentation.

ACKNOWLEDGEMENTS

We would like to thank the University of British Columbia's Microbiology and Immunology department for their financial support, *Escherichia coli* strains and the use of their laboratory facilities for our experiments. We would also like to thank Dr. David C. Oliver, Christoph Deeg from the Johnson Lab at UBC, and the various MICB 447 research teams registered in the first Winter Session of 2015 for their instruction, guidance, help and the provision of reagents. We would like to specifically thank Adam Ziada for his help in optimizing our polymerase chain reaction conditions. Furthermore, we would like to acknowledge the Wesbrook media room personnel for their autoclaved equipment and general help.

REFERENCES

1. **Beloin, C., A. Roux, and J. Ghigo.** 2008. *Escherichia coli* biofilms. Curr Top Microbiol Immunol. 322:249.
2. **Heras, B., M. Totsika, K. M. Peters, J. J. Paxmana, C. L. Gee, R. J. Jarrott, M. A. Perugini, A. E. Whitten, and M. A. Schembri.** 2014. The antigen 43 structure reveals a molecular Velcrolike mechanism of autotransporter-mediated bacterial clumping. Proc. Natl. Acad. Sci. U. S. A. 111:457.
3. **Henderson, I. R., F. Navarro-Garcia, M. Desvaux, R. Fernandez, and D. Ala'Aldeen.** 2004. Type V Protein Secretion Pathway: the Autotransporter Story. Microbiol. Mol. Biol. Rev. 68:692.
4. **Zheng Zhang, Scott Schwartz, Lukas Wagner, and Webb Miller (2000),** A greedy algorithm for aligning DNA sequences. J Comput Biol 2000; 7(1-2):203-14.
5. **Kjærgaard, K, Schembri, MA, Hasman, H, Klemm, P.** 2000. Antigen 43 from *Escherichia coli* Induces Inter- and Intraspecies Cell Aggregation and Changes in Colony Morphology of Pseudomonas fluorescens. J. Bacteriol. 182:4789-4796.
6. **Schembri, MA, Hjerrild, L, Gjermansen, M, Klemm, P.** 2003. Differential Expression of the *Escherichia coli* Autoaggregation Factor Antigen 43. J. Bacteriol. 185:2236-2242.
7. **Hasman H, Chakraborty T, Klemm P.** 1999. Antigen-43-Mediated Autoaggregation of *Escherichia coli* is Blocked by Fimbriation. *Journal of Bacteriology*.181(16):4834-4841.

SUPPLEMENTAL DATA

NNNNNNNNNNNGAGCTCGGATCCNCTAGTAACGGCCGCCAGTGTGCTGGAATTCGCCCTTCCCGCGGCCGCAAGCTTTGT
GGCGTTGAAGATCCGTTGCGCCGGTCCGGGTCCCGGTCATGATGACCGGGACCACAGAGAGGCGATGGTTCTGTCAGAAG
GTCACATTCAGTGTGGCCTGACCGTTATACCCTTCAGCGCTGCTGCCGCTGACGCTGTGGGCATAACCGGCCTGAACGCC
CAGGGTGATATTTTCCCGGACACGGGCTTCCAGTCCGGCCTGCAGGTCCAGTGATGTGCCATTCTGTGAGGGAGAGAACG
TCATCCCGCTGCCTGCAGTGGAAGTCCCCACACGCATATCTCCCCGGGAGCTGAAGGTGCGGATAACAGAAGGCTGTACC
CACCAGTTCACCGGTAATTCACTCACACTGTGTTTTGCACTGTCACGCAGGGGGGCACGGGATGAGGTGCCTTCGCCAAA
GGTCATATCGTTGTGGCTGCCCAGACGGAAACCGGCACGCACATGTTGTGCACTGCCATGCCCGAACTTCACATAACCGG
CGTTGTCCTTACCGTCATCCAGGGAAAGTCCCTGCCAGGTATACTGCAGTTGTGGCTCCAGCATCAGGTTGTCAGTGATA
CTGAAGGGCAGACCGGTTTCCAGTGAGCCCAGCCAGCCCCAGCCCCGGGCGCGGAAGTCGTTATTGTCCGATGACGCTTT
CATGCTGTGGCGGGTTCCCTGTGCCACAATGTCAGCCCACAGGCCGGANGACGTGTGTACCAGATTCAGGTATCCGCCCA
GGCTGCCGGCATCATCCCGGACCGTGCCGGCACGGGAGCCGTCATCATCCTTAACATCAACGGAAGAATGGCCAGCAGCA
CCATATACCCCCGCGGTCACAGACATACCGGCAACCTCTGTTCTCATCNNTCACCCTCCAGACGGACGAATCCATAGCTG
CCGCTGCTTTNCCGGCGTGGCCCACGGGCAATACCGCCATTGTTATCGTGACCGAGATGACCGCCCTGAATGCTGAGANN
NNGCTGTGTTTCACCATTTACACCGGTCTGATGGCTGNGGNNCCTGCCNCNATCCGGTCATANTCCATTGCCTGTGTCAG
CATGNNCNTANNNGGGGNNTNCTGCANNANAGCATTTTNNNCTGNNNNNNNANNNCCAGCNNNNNCATNNNCTNNCCCNN

SUPPLEMENTAL FIG 1. Primer construction. The sequence of the *flu* gene of MG1655 was obtained from National Centre for Biotechnology Information and primers were manually designed to insert a poly-histidine tag 6 amino acids downstream of the signal peptide. The reverse primer (P2: 5′-ATG ATG ATG ATG ATG ATG GTG CAC AAC GAT GTC-3′), in P1-P2 set, encodes a poly-histidine tag (underlined). The forward primer (P3: 5′-CAT CAT CAT CAT CAT CAT CCG GGA GAA ACC GTG AAC GGC-3′), in P3-P4 set, encodes a poly-histidine tag (underlined).

Antigen 43 precursor [Escherichia coli O10:K5(L):H4 str. ATCC 23506]
Sequence ID: emb|CCP98736.1| **Length:** 495 **Number of Matches:** 2

Range 1: 239 to 495 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps	Frame
520 bits(1339)	7e-179	Compositional matrix adjust.	255/257(99%)	256/257(99%)	0/257(0%)	-3

```
Query  916  MRTEVAGMSVTAGVYGAAGHSSVDVKDDDGSRAGTVRDDAGSLGGYLNLVHTSSGLWADI  737
            MRTEVAGMSVTAGVYGAAGHSSVDVKDDDGSRAGTVRDDAGSLGGY+NL HTSSGLWADI
Sbjct  239  MRTEVAGMSVTAGVYGAAGHSSVDVKDDDGSRAGTVRDDAGSLGGYMNLTHTSSGLWADI  298

Query  736  VAQGTRHSMKASSDNNDFRARGWGWLGSLETGLPFSITDNLMLEPQLQYTWQGLSLDDGK  557
            VAQGTRHSMKASSDNNDFRARGWGWLGSLETGLPFSITDNLMLEPQLQYTWQGLSLDDGK
Sbjct  299  VAQGTRHSMKASSDNNDFRARGWGWLGSLETGLPFSITDNLMLEPQLQYTWQGLSLDDGK  358

Query  556  DNAGYVKFGHGSAQHVRAGFRLGSHNDMTFGEGTSSRAPLRDSAKHSVSELPVNWWVQPS  377
            DNAGYVKFGHGSAQHVRAGFRLGSHNDMTFGEGTSSRAPLRDSAKHSVSELPVNWWVQPS
Sbjct  359  DNAGYVKFGHGSAQHVRAGFRLGSHNDMTFGEGTSSRAPLRDSAKHSVSELPVNWWVQPS  418

Query  376  VIRTFSSRGDMRVGTSTAGSGMTFSPSQNGTSLDLQAGLEARVRENITLGVQAGYAHSVS  197
            VIRTFSSRGDMRVGTSTAGSGMTFSPSQNGTSLDLQAGLEARVRENITLGVQAGYAHSVS
Sbjct  419  VIRTFSSRGDMRVGTSTAGSGMTFSPSQNGTSLDLQAGLEARVRENITLGVQAGYAHSVS  478

Query  196  GSSAEGYNGQATLNVTF  146
            GSSAEGYNGQATLNVTF
Sbjct  479  GSSAEGYNGQATLNVTF  495
```

SUPPLEMENTAL FIG 2. Polymerase Chain Reaction (PCR) amplification and cloning of the *flu* gene. PCR was done in 100 µl reactions with the use of *Taq* DNA Polymerase (Invitrogen, Cat. No. 18038-042). The manufacturer's protocol was followed using 2 U *Taq* Polymerase, 1 µL of MG1655 genomic DNA as template and a primer mix of either P1 and P2 (10 µM each) or P3 and P4 (10 µM each).