Molecular Characterization of a Gene Encoding Insulin-Like Peptide 5 (ILP5) in Drosophila virilis

Nikolas R. Likourentzos
University of Tennessee, Knoxville, nlikoure@vols.utk.edu

Jae H. Park
University of Tennessee, Knoxville, jhpark@utk.edu

Follow this and additional works at: https://trace.tennessee.edu/pursuit

Part of the Bioinformatics Commons, and the Molecular Biology Commons

Molecular Characterization of a Gene Encoding Insulin-Like Peptide 5 (ILP5) in Drosophila virilis

1. INTRODUCTION

The pathway to the discovery of insulin was first recorded in 1889, when Oskar Minkowski and Joseph von Mering removed the pancreases of several animals, resulting in the animals contracting severe diabetes. Using this information, Minkowski and Mering theorized that the pancreas secreted a substance responsible for metabolic control. The substance was later named insulin because it was discovered to originate from the islet cells of the pancreas. In 1921, Charles Best, Frederick Banting, JJR Macleod, and JB Collip experimented on depancreatized animals in order to develop a cure for diabetes. By 1922, the research team had developed an acidic alcohol solution containing chilled beef pancreas. The solution was injected into Leonard Thompson, a diabetic 14-year-old boy, whose diabetic symptoms quickly diminished and blood glucose levels dropped to normal. The research team was awarded the Nobel Peace Prize in 1923 for the discovery of insulin (Joshi et. al, 2007).

Mammalian insulin is a peptide composed of an A-chain and a B-chain linked together by two disulfide bonds on cysteine residues. The precursor of insulin, preproinsulin, is synthesized in the beta cells of the pancreas and contains three amino acid chains connected by disulfide bonds and peptide bonds. Preproinsulin also contains a signal peptide connected to the A-chain that is cleaved in the rough endoplasmic reticulum. Once the signal peptide of preproinsulin is cleaved, the peptide becomes proinsulin, which is transported to the Golgi Apparatus and released in vesicles, where the C-chain between the A-chain and B-chain is cleaved off, resulting in the peptide becoming a functional insulin molecule (Joshi *et al.*, 2007).

Insulin is secreted in response to high levels of blood glucose because the peptide acts as a signal for the uptake of glucose into cells from the bloodstream. Failure to synthesize functional insulin results in numerous health issues all tied to diabetes type I, including ketoacidosis, glycemia, and glycosuria. Patients of diabetes type I require regular insulin injection to prevent excess glucose in the bloodstream and the ability to use glucose for energy (Joshi *et al.*, 2007).

Insulin-like peptides (ILP), found in *Drosophila* and other insects, are homologs to mammal insulin and insulin growth factors, showing similarity in the number and positions of cysteine residues. ILPs also mature very much like insulin, with a C-chain and signal peptide being cleaved to leave both the B-chain and A-chain after maturation (Nässel and Broeck, 2016).

Eight different ILPs have been discovered in *D. melanogaster* so far and while it is known that ILPs control various biological processes like development, growth, aging, and metabolism, there is still much to learn about ILPs. Research suggests that ILP2 regulates adult life span and blood glucose level. ILP5 regulates protein metabolism and ILP6 is suggested to regulate lipid metabolism (Post *et al.*, 2018). ILP6 and 7 have also been proposed to function similar to the mammalian

hormone Relaxin, which has various functions related to reproduction such as sperm motility, cervix softening, and relaxation of female pelvic ligaments during birthing (MacLennan, 1991). ILP6 is similar to insulin-growth factor both structurally and functionally, which regulates growth and metabolism (Nässel and Broeck, 2016). ILP3 could function to regulate growth and metabolism in larvae (Nässel and Broeck, 2016). Data suggest that ILPs are time-dependent meaning that their expression is observed at varying levels throughout the different stages of *Drosophila* development. *Drosophila* ILP1 was recorded to function mainly during pupal stage, ILPs 2-5 in the embryo stage, and ILP8 in larvae (Nässel and Broeck, 2016).

ILP5 in *D. melanogaster* is produced in the insulin-producing cells of the brain, ovaries, and Malpighian tubules (Nässel and Broeck, 2016). ILP5 is shown to directly be involved in lowering trehalose levels in *Drosophila* (Sajid et al. 2010). Research with knock-out mutants of ILP5 suggests that ILP5 decreases reproductive diapause (Schiesari et al., 2016). Inhibiting ILP5 also resulted in decreased female remating, improved sleep quality, and reduced sleep fragmentation (Wigby et al., 2011; Metaxakis et al., 2014). Peak expression of ILP5 is observed at stages throughout the larval period, during early pupal stages, and in adult male stages (Gramates et. al, 2018). In order to improve our understanding about the function of ILPs in *Drosophila*, the insulin-like peptide 5 (DvILP5) of the distantly related *Drosophila virilis,* which diverged from *D. melanogaster* about 60 million years ago, was sequenced and the structure of the peptide was predicted using bioinformatic analysis.

2. MATERIALS AND METHODS
2.1 3′ and 5′ RACE
3′ Rapid Amplification of cDNA Ends (RACE) was used to amplify the 3′ region of the *DvILP5* gene. Qt primer was used to add Q0 and Q1 primer binding locations the gene (see Figure 1A). 3′ RACE was initiated with a 40 µL mixture of 2X GoTaq PCR mix (20 µL), 2 µM f1 primer [GCTGCCGCTAATGTTGCTGT] (4 µL), 2 µM Q0 primer (4 µL), and *D. virilis* head total cDNA (1 µL). The PCR machine was set to the following protocol for each PCR: 94°C for 2 minutes, then 32 cycles of 94°C for 20 seconds, 55°C for 25 seconds, and 72°C for 40 seconds. After the cycles, the mixture was kept at 72°C for 5 minutes and then cooled to 4°C. After the initial PCR, a second nested PCR was run using a 60 µL mixture of 2X GoTaq PCR mix (30 µL), 6µM f2 primer [TGGGATTTTACAGCGCCAAAG] (3 µL), 6 µM Q1 primer (3 µL), and the mixture of the initial PCR (1 µL).

5′ RACE was employed to amplify the 5′ region of the *DvILP5* gene (Figure 2A). The procedure is the same as 3′ RACE except the initial PCR uses r1 primer [CCGCAATACTGTAACATTTC] and Qt primer. The second PCR uses r2 primer [GCTTGCGGCAACATTCATCG] and Q1 primer. In order to estimate the length

of 3′ the RACE and 5′ RACE samples, they were tested on 1 % agarose gels (Figures 1B and 2B, respectively).

2.2 DNA Purification and Ligation

To purify the DNA after PCR, the PCR samples underwent procedural gel electrophoresis and the major bands were cut out and then purified as instructed by the QIAquick gel extraction manual. After purification, the DNA was inserted into Promega pGEM-T easy vectors. The ligation reaction (15 µL) included the following materials: pGEM-T easy vector (1 µL), purified PCR sample (2.5 µL), T4 DNA Ligase (1 µL), and 2X ligation buffer (7.5 µL). The solution was vortexed gently and left overnight at room temperature.

2.3 Bacterial Transformation and PCR of Cells

To initiate bacterial transformation, an aliquot of the ligation reactions (2 µL) was added to separate vials containing 35 µL of competent DH5alpha *E. coli* cells and left on ice for 20 minutes. Meanwhile, four agar plates containing Luria Broth (LB) and Ampicillin were warmed to 37ºC and then 0.1 M IPTG (30 µL) and 25 mg/mL X-Gal (30 µL) were applied evenly on the plates. After 20 minutes, the cells were heat shocked at 37ºC for 45 seconds then put back on ice. LB (200 µL) was added to the cells and then they were kept in a shake incubator at 37ºC and 200 RPM for 50 minutes. Half of each sample of cells was spread on different agar plates and incubated overnight at 37ºC. After bacterial colonies formed on the agar plates, each white colony (positive for transformation) was inoculated to a 3 mL LB medium containing 100 µg/mL Ampicillin. The cells were grown overnight at 37ºC. To test if the cells carried desired DNA, PCR of the cells was performed as follows: a 30 µL mixture of cell culture (1 µL), T7 primer [TAATACGACTCACTATAGGG] (3 µL), Sp6 primer [ATTTAGGTGACACTATAG] (3 µL), and GoTaq PCR mix (15 µL) underwent the same conditions as the 3′RACE PCR.

2.4 Plasmid Purification and Restriction Enzyme Digestion

After the cells were harvested by centrifugation, the recombinant plasmids were purified as instructed by the QIAprep Spin Miniprep Kit. In order to test if the plasmids contained the desired DNA, EcoRI restriction enzyme was added to the plasmids as followed: a 15 µL solution containing 10X Promega Buffer H (1.5 µL), EcoRI (1 µL), and plasmids (5 µL) were mixed and kept at 37ºC for 50 minutes. Following digestion, 3 µL of 6X loading dye was added to each reaction mixture, and then 10 µL of each mixture was loaded for a gel electrophoresis (Figure 3C). Positive plasmids were sent to the UTK sequencing facility.

2.5 Genomic DNA Purification of D. virilis ILP5

A female *D. virilis* fly was added into a tube holding 50 µL of 1X squash buffer (10 mM Tris-Cl, pH 8; 1 mM EDTA; 25 mM NaCl, and 133 µg/mL Proteinase K). The fly was mashed in the solution and the solution was incubated at 37°C for 25 minutes. After incubation, the solution was heat shocked at 95°C for 5 minutes in order to terminate the Proteinase K activity. The solution was centrifuged for 30 seconds and the supernatant of *D. virilis* genomic DNA (gDNA) was extracted for PCR. Then, 1 µL of the gDNA was mixed with f1 (5 µL), r1 primer (5 µL), and 2X GoTaq mix (25 µL) to make a 50 µL solution that underwent the same conditions of PCR as the 3′ RACE. To test if there were any introns in the *DvILP5* gene, a comparative gel electrophoresis was run using 8 µL of the gDNA sample and 8 µL of a cDNA sample amplified with f1 and f2 primer (Figure 4B).

3. RESULTS
3.1 RACE

A portion of the *DvILP5* sequence was known in the genome database (Dvir\GJ13215). To determine the complete *DvILP5* sequence, the RACE protocol was employed. First, mRNA was isolated from *D. virilis* heads and then used to synthesize cDNA using reverse transcription as described previously (Scotto-Lavino et al., 2007). Figure 1A shows that the total mRNAs isolated from the fly heads was used to synthesize cDNA. The synthetic Qt primer containing oligo-T tail with Q1 and Q0 primer binding sites is designed to be complimentary of the mRNA poly-A tail, allowing the Q0 primer and nested Q1 primer to be used in PCR. After PCR, gel electrophoresis was used to estimate the length of the PCR product. As shown in Figure 1B, the number of base pairs of the 3′RACE product is between 400 and 850.

Similarly to 3′RACE, 5′RACE was used to define the portion of the *DvILP5* cDNA between the known region and the transcription start site of the gene. Figure 2A displays that cDNA was transcribed from mRNA using reverse transcription. The synthesized cDNA contained a poly-T tail with Q0 and Q1 primers at the 5′ end and a known region in the center to which r1 and r2 primers bound to amplify the 5′ region of the cDNA strand. According to the gel electrophoresis, the 5′RACE PCR product is estimated to be slightly over 400 bp (Figure 2B).

3.2 Confirmation of Recombinant Plasmid by PCR and Enzyme Digestion of Plasmids

To determine the nucleotide sequences of the 5′ and 3′ RACE, these PCR products were purified and then ligated with pGEM-T easy vector. After bacteria cells were transformed with ligation product, four colonies for each construct were screened to determine if they contain plasmids with target PCR products. As shown by

Figure 3A, plasmids have T7 and SP6 primer binding sites flanking the insert region, which is amplified using PCR. If plasmids did not contain PCR products, the amplified regions would be approximately 175 bp.

Figure 3B shows that the amplified regions of plasmids containing 5′-RACE PCR product are composed of fewer nucleotides than those containing 3′-RACE PCR product. Earlier gel electrophoresis results indicating that the 3′-RACE PCR product is larger than the 5′-RACE PCR product support that the cell plasmids tested contain the correct inserts.

In order to further confirm that the plasmids contained target DNA, EcoRI restriction enzyme was used to digest the plasmids at points flanking the insert DNA (Figure 3A). If the plasmid did not contain insert DNA, then gel electrophoresis would show only the pGEM-T easy vectors at approximately 3000 bp and the 19 bp between the two digestion points. Digestion of 3′RACE pGEM-T plasmids shows a 3 kb band and two small fragments below 500 bp, indicating that 3′RACE contains an internal EcoRI digestion site (Figure 3C). The digestion of the 5′RACE pGEM-T plasmids shows one 3 kb and a ~500 bp band (Figure 3C). These results suggest that all purified plasmids contained the desired inserts.

3.3 Determining the Sequence and Structure of *DvILP5* Gene

After confirming that the plasmids contain the target DNA, the plasmids were sequenced. Two 5′RACE and two 3′RACE sequences were considered clean enough to piece together to make a full cDNA sequence of the *DvILP5* gene. From the analysis of Clustal Omega (Embl-Ebi), a full-length cDNA sequence from 5′ and 3′RACE sequences was constructed.

To identify the protein coding sequence, i.e. open reading frame (ORF), within the cDNA the full-length cDNA sequence was analyzed by a web-based ORF Finder software (Sequence Manipulation Suite). As a result, the ORF is 366 bp in length and the polypeptide produced is determined to be 121 amino acid residues. The remaining nucleotides constitute the 5′UTR and 3′UTR at 67 and 150 bp, respectively (Figure 4A).

Next, because most eukaryotic genes contain introns, we wanted to determine if the *DvILP5* gene also contained any introns. To do so, PCR was used to amplify segments of cDNA and gDNA using f1 and R1 primers. It is reasonable that any introns between the two primers would be amplified in the PCR of genomic DNA but not cDNA. As seen in Figure 4B, the gDNA PCR product was larger than the cDNA PCR product, suggesting that there is at least one intron in the *DvILP5* gene. To verify this result, the sequence of genomic DNA was compared to that of cDNA, revealing 123-bp sequence that shows only in the genomic DNA (Figure 4C). In summary, *DvILP5* gene consists of two exons separated by an intron, and the ORF encodes 121 amino acid long polypeptide.

3.4 Bioinformatic analyses of the prepro-DvILP5 Peptide

The ExPASy ProtParam determined the molecular weight of prepro-DvILP5 to be 13865.13 g/mol (Swiss Institute of Bioinformatics). As demonstrated in mammalian insulin, nascent polypeptide produced from insulin mRNA, called preproinsulin, undergoes post-translational maturation process to produce functional insulin. The first step of this processing is to enter ER lumen via N-terminal signal peptide region of the nascent polypeptide.

To determine the signal peptide in the prepro-DvILP5, a web-based SignalP program was employed (Center for Biological Sequence Analysis). According to this analysis, the signal peptide was predicted to be the first 29 amino acid residues, resulting in the production of 92-amino acid long pro-DvILP5.

The second step of processing is the conventional proteolytic cleavage at double basic amino acid residues (K and R) within the polypeptide (Figure 5A and B). Visual inspection of the sequence found that amino acid residues 56 and 57 as well as residues 95 and 96 are typically cleaved residues, resulting in the separation of 39 residues (C-Chain) between the A-chain and the B-chain (Figure 5A and B).

Another important structural feature of mature insulin is the triple disulfide (SS) bonds. In mammalian insulin, there are two interchain SS bonds, and one intra-B chain SS bond (Figure 5D). Pro-DvILP5 also contains six Cys residues, which are likely to involve the formation of three SS bonds. To predict which pair of Cys makes a SS bond, I ran a web-based software DiANNA (Boston College). With this, we determined that there are most likely three SS bonds, which are represented in Figure 5C.

In summary, prepro-DvILP5 is likely processed to mature to DvILP5 in a manner similar to mammalian insulin precursors, suggesting that the molecular mechanisms underlying hormone maturation process are highly conserved between mammals and flies.

4. DISCUSSION
4.1 Sequenced Gene vs. FlyBase Genome Project

The results of the *DvILP5* sequencing can be used to compare against the sequence predicted from genome project algorithms (Flybase). The genome project uses algorithms to predict which segments of *Drosophila* genomes are protein-coding sequences. The algorithm did correctly determine the *DvILP5* gene, titled *Dvir\GJ13215*; however, we found a slight difference between our experimental

data and genomic project. *Dvir\GJ13215* is missing 19 nucleotides from the 5′UTR and 9 nucleotides from the 3′UTR (see Figure 6). Testing against the algorithm shows that it is capable of determining the location of a gene but needs improvement with determining where the gene begins and ends. Wet lab work is still necessary to accurately sequence the entirety of certain genes. Determining all of the nucleotides of a gene is crucial in order to define the functional and regulatory regions of the gene.

4.2 Characteristics of the DvILP5 Peptide

The DvILP5 peptide shows conservation in structure through evolution, justifying the use of DvILP5 as a model for human insulin. Just like in mammal insulin, DvILP5 precursor consists of a signal peptide and peptides A, B, and C. During maturation, the signal peptide and peptide C are cleaved. The function of the C-peptide in insulin is yet to be well known, however, it is known that the C-peptide of preproinsulin holds the A- and B-peptide in a conformation that allows disulfide bonds to form between the A- and B-peptide and promotes proper folding of the peptides. Cleavage of the C-peptide exposes the C-terminal of the B-peptide, allowing the B-peptide to assume the conformation necessary to interact with the insulin receptor.

Interestingly, it has also been discovered that C-peptide shows stimulating effects on Na(+)-K(+)-ATPase and endothelial nitric oxide synthase activities in rats (Warren et al., 2000). Research also shows that the administration of C-peptide to patients with type-1 diabetes leads to increased blood flow in skeletal muscle and skin, diminished glomerular hyperfiltration, reduced urinary albumin excretion, and improved nerve function (Warren et al., 2000). Further research could entail the use of C-peptide along with insulin to treat symptoms of type-1 diabetes.

4.3 Comparison of the *Drosophila virilis* ILP5 to all of the *Drosophila melanogaster* ILPs.

When a comparison of all DmILPs to DvILP5 is performed using web-based multiple sequence alignment software (Embl-Ebi), the results show that DvILP5 shares more similarity in structure with DmILP5 than the other seven DmILPs with a match of 36% (Figure 7). After what is predicted to be 50 million years of divergence, a 36% similarity in structure of DvILP5 and DmILP5 suggests that the structure of the peptide is essential for *Drosophila* life. The similarity in structure of DvILP5 and DmILP5 compared to the other DmILPs, especially with the same number of cysteine residues being shared between the two peptides, also suggests that DvILP5 shares the most similarity in function with DmILP5 because structure and function are closely related.

REFERENCES

Boston College. (n.d.). DiANNA 1.1 web server.

Center for Biological Sequence Analysis. (n.d.). SignalP-5.0 Server.

Embl-Ebi. (n.d.). Clustal Omega.

Gramates L.S., Marygold S.J., dos Santos G., Urbano J-M., Antonazzo G., Matthews B.B., Rey A.J., Tabone C.J., Crosby M.A., Emmert D.B., Falls K., Goodman J.L., Hu Y., Ponting L., Schroeder A.J., Strelets VB, Thurmond J, Zhou P and the FlyBase Consortium. (2018). *FlyBase*.

Joshi, S., Parikh, R., Das, AH. (2007). Insulin -- history, biochemistry, physiology and Pharmacology. *J Assoc Physicians India.* 5, 19-25.

MacLennan, A. (1991). The role of the hormone relaxin in human reproduction and pelvic girdle relaxation. *Scand J Rheumatol Supp.* 88, 7-15.

Metaxakis A., Tain L. S., Grönke S., Hendrich O., Hinze Y., Birras U., Partridge L. (2014). Lowered insulin signalling ameliorates age-related sleep fragmentation in *Drosophila. PLoS Biol.* 1-13.

Nässel, D., Broeck, J. (2016). Insulin/IGF signaling in *Drosophila* and other insects: factors that regulate production, release and post-release action of the insulin-like peptides. *Cell Mol Life Sci.* 73, 271-290.

Post, S., Karashchuk, G., Wade, J. D., Sajid, W., De Meyts, P., Tatar, M. (2018). *Drosophila* insulin-like peptides DILP2 and DILP5 differentially stimulate cell signaling and glycogen phosphorylase to regulate longevity. *Front Endocrinol*. 9, 1-16.

Sajid, W., Kulahin, N., Schluckebier, G., Ribel, U., Henderson, H. R., et al. (2010). Structural and biological properties of the *Drosophila* insulin-like peptide 5 show evolutionary conservation. *J Biol Chem.* 286, 661-673.

Schiesari, L., Andreatta, G., Kyriacou, C. P., O'Connor, M. B., & Costa, R. (2016). The insulin-like proteins dILPs-2/5 determine diapause inducibility in *Drosophila*. *Plos One.* 11, 1-16.

Scotto-Lavino, E., Du, G., & Frohman, M. (2006). Amplification of 5' end cDNA with 'new RACE'. *Nat Protoc.* 1, 3056-61.

Sequence Manipulation Suite. (2017). ORF Finder.

Swiss Institute of Bioinformatics. (n.d.). ProtParam tool.

Warren, J., Ekberg, K., Johansson, J., Henriksson, M., Pramanik, A., Johansson, B., Jörnvall, H. (2000). Role of C-peptide in human physiology. *Am J Physiol Endocrinol Metab.* 278, E759-E768.

Wigby, S., Slack, C., Grönke, S., Martinez, P., Calboli, F. C., Chapman, T., & Partridge, L. (2011). Insulin signalling regulates remating in female *Drosophila*. *Proc Biol Sci.* 278, 424-431.

FIGURE CAPTIONS

Figure 1. 3′RACE
- **(A)** Schematic of 3′ RACE. f1 and f2 are *DvILP5* gene specific primers.
- **(B)** Gel electrophoresis results of 3′RACE. Lane 1: 3′RACE product; Lane 2: DNA size marker

Figure 2. 5′RACE
- **(A)** Schematic of 5′ RACE. The r1 and r2 primers are *DvILP5* gene specific primers.
- **(B)** Gel electrophoresis results of 5′RACE. Lane 1: DNA size marker; Lane 2: 5′RACE PCR product.

Figure 3. Plasmid PCR and EcoRI Digestion
- **(A)** Schematic of plasmid PCR and EcoRI digestion of plasmids. T7 and SP6 are primer binding sites located on plasmid regions flanking insert DNA. EcoRI digestion sites also flank insert DNA. 3′RACE product has one internal EcoRI digestion site.
- **(B)** Gel electrophoresis results of PCR of E. coli containing plasmids using T7 and SP6 primers. Lanes 1-4: 5′RACE insert DNA in plasmids; Lane 5: DNA size marker; Lanes 6-9: 3′RACE insert DNA in plasmids.
- **(C)** Gel Electrophoresis of PCR of plasmids containing 3′RACE PCR product after EcoRI digestion. Lane 5: DNA size marker. Large bands are approximately 3-kb long and represent pGEM-T easy vector. Two <500 bp bands indicate single digestion site within 3′RACE PCR product.
- **(D)** Gel Electrophoresis of PCR of plasmids containing 5′RACE PCR product after EcoRI digestion. Lane 5: DNA size marker. Large bands are approximately 3-kb long and represent pGEM-T easy vector. One ~500 bp band indicates no digestion site within 5′RACE PCR product.

Figure 4. Gene Structure
- **(A)** Full cDNA sequence including open reading frame (ORF) and amino acid code. The f1 primer binding site is highlighted in cyan and f2 primer binding site in yellow. The R1 primer binding site is highlighted with orange and R2 primer binding site in green. The ORF codes for 121 amino acid residues. The asterisk represents the stop codon. A black arrow points to the location of an intron. The EcoRI digestion site is underlined.
- **(B)** Gel electrophoresis of cDNA and gDNA after PCR using f1 and r1 primers. Lane 1: DNA size marker; Lane 2: cDNA PCR product; Lane 3: gDNA PCR product.
- **(C)** *DvILP5* intron highlighted in blue.

(D) Actual sequence data of intron (highlighted in yellow).

(E) *DvILP5* gene structure with lengths of exons (grey boxes), intron (line), and UTR (blue boxes) in base pairs (bp).

Figure 5. Comparison of Preproinsulin Structure

(A) Boxed structure of DvILP5. Each box represents a segment of the peptide. The numbers in the box indicate how many amino acids are in each segment. The letters between the boxes indicate the amino acid residues that are predicted to be cleaved.

(B) Human preproinsulin for comparison.

(C) Predicted peptide structure of DvILP5. Signal peptide in green, A-Chain in red, C-Chain in yellow, B-Chain in blue. White circles represent points of cleavage. Lines marked with S indicate disulfide bonds. The peptide is 121 amino acids in length.

(D) Human preproinsulin for comparison at 110 amino acids in length.

Figure 6. Genome Project vs Sequenced Gene.

Gene sequence of *DvILP5* reported by the Genome project on top. Gene sequence obtained from this research on bottom.

Asterisks indicate identical nucleotides. Dashed lines indicate nucleotides experimentally found to be part of *DvILP5* gene that genome project did not include.

Figure 7. DvILP5 vs DmILP 1-8

Sequence alignment of the DvILP5 with DmILP1-8 shows that DvILP5 is most similar to DmILP5 in structure with a 36% match. The most similar peptides are closer to each other on the list. Asterisks indicate amino acid residues conserved in all peptides. A period indicates similarly among some of the peptides. A colon indicates a similarity among most peptides.

FIGURES

Figure 1.

A.
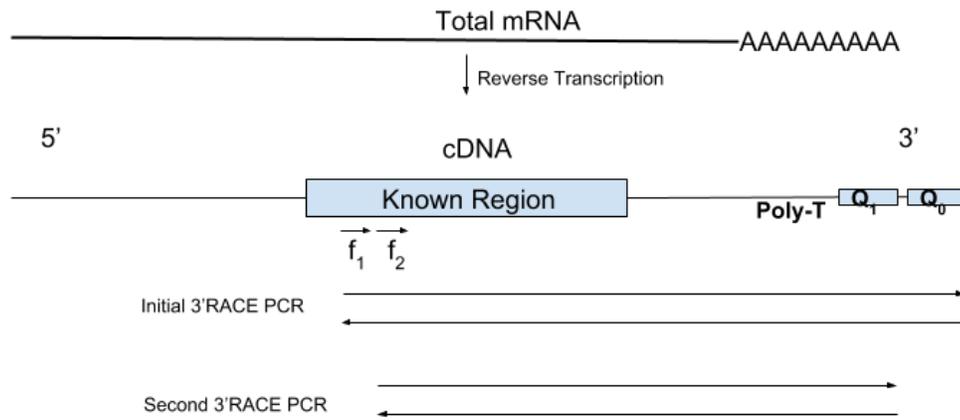
B.
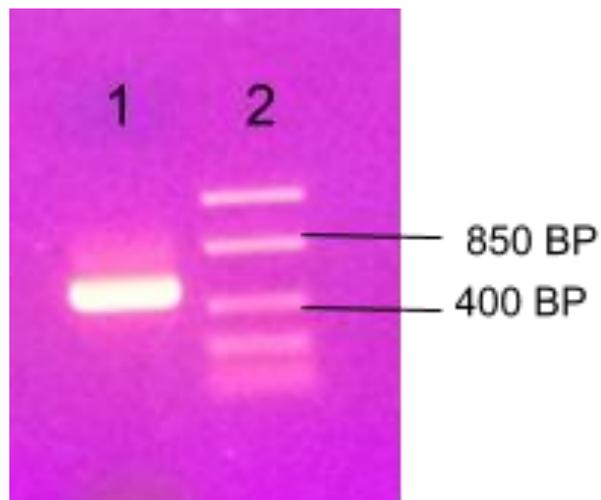
Figure 2.

A.

B.

Figure 3.

A.

B.

C.

D.
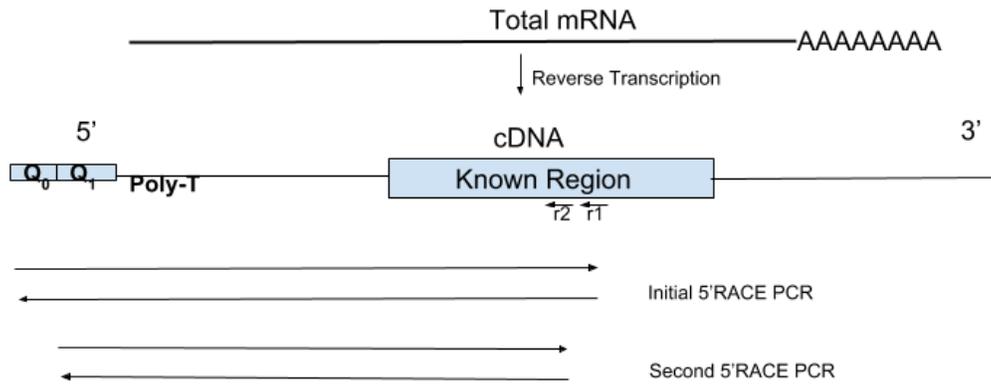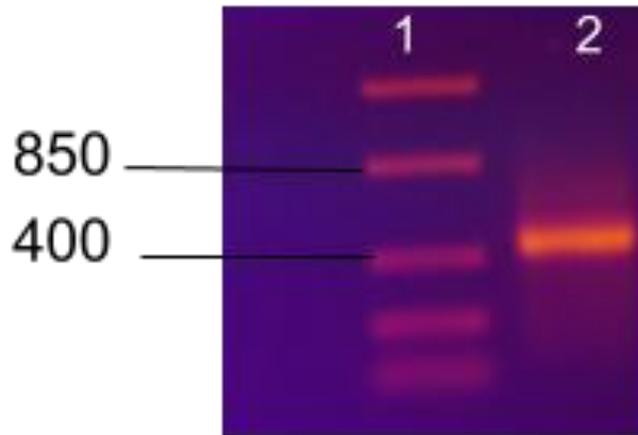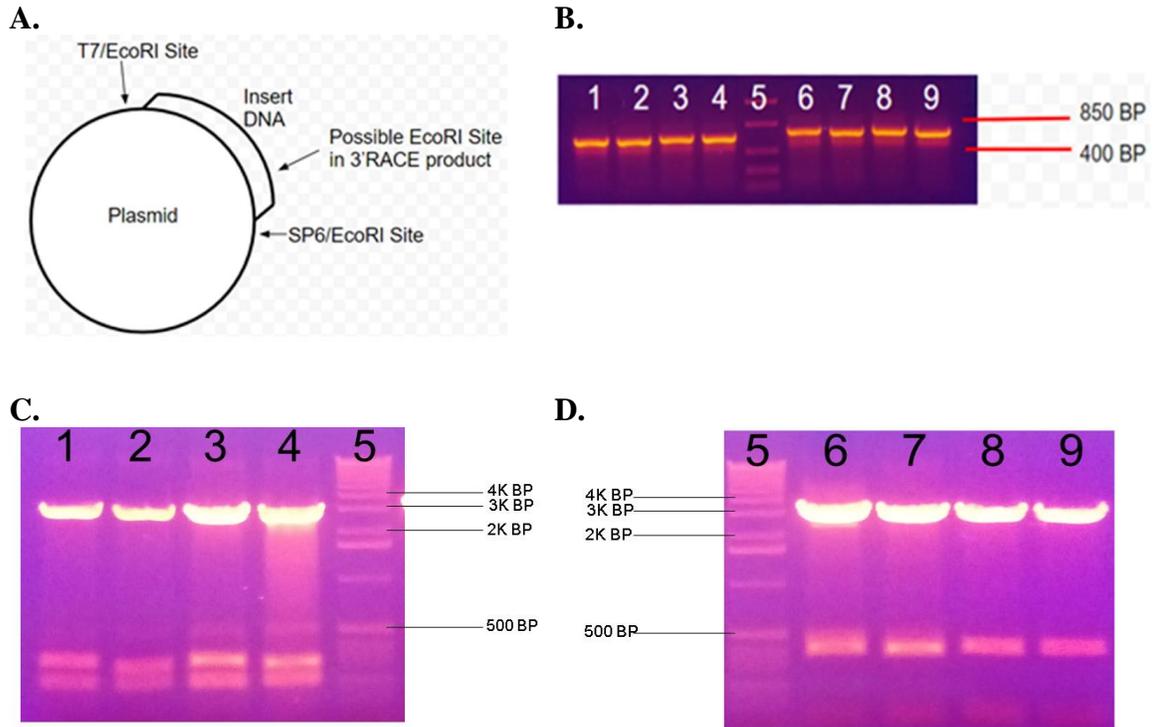
Figure 4.

A.

```
GATCACATCTTGTGTGGCATCTAACCAGGCACAATATCCAGCATCGATAAA
ATTCCAGCAGCACAAA

ATGGCGAAAACCATACAAATCTTCTTGCTGCCGCTAATGTTGCTGTTGCTG
 M   A   K   T   I   Q   I   F   L   L   P   L   M   L   L   L
ATGGGATTTTACAGCGCCAAAGCCGCGAACAGCTATAAGGTTTGTGGTTCG
 M   G   F   Y   S   A   K   A   A   N   S   Y   K   V   C   G   S
GAACTGACCGAAGCACTTTACATAGTCTGCAAGGGCAAATTCGGCTCCTTT
 E   L   T   E   A   L   Y   I   V   C   K   G   K   F   G   S   F
CAACACCACAAACGCGGCTCGTTGGATCTCTTTGACTATGTGGACCAGCAG
 Q   H   H   K   R   G   S   L   D   L   F   D   Y   V   D   Q   Q
TACGAGACGAACGATGTGGAGGACAGGGATACGCTCTGGGGATTGCCGCCA
 Y   E   T   N   D   V   E   D   R   D   T   L   W   G   L   P   P
AGGAGTCAGAATTCCCTGGTGGCCACCAGGCGGCTAATGCGCGGCGTTGTC
 R   S   Q   N   S   L   V   A   T   R   R   L   M   R   G   V   V
GATGAATGTTGCCGCAAGCCCTGCACGAGATTGGAAATGTTACAGTATTGC
 D   E   C   C   R   K   P   C   T   R   L   E   M   L   Q   Y   C
GGAAACTAA
 G   N   *

GAATTGCCAAAAGTAAACACTCATGTATTCCAACTCCAGTTTGTATTCCCC
TATATATAATGTATATGCATGTATTCCTGTTAAAGCAAGTGTACAAAAACT
GGTTCGCCTAATATAACGGAGGAAATAAATAACCTACAATGAAACCAG
```

B.
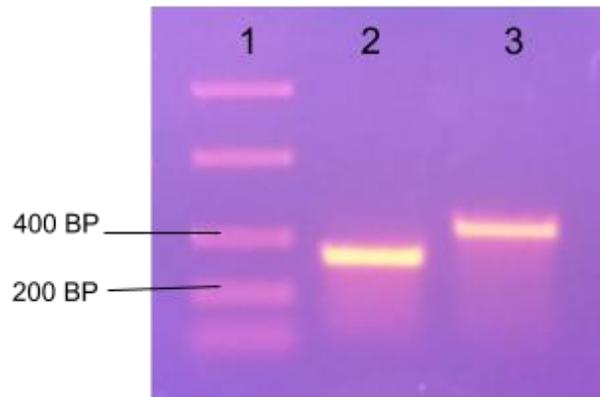
C.

```
CGTGTAGGTATCTCATACTCTAGCTGTGGTTGCAGCTGCGCCT
GCCTGTAACTGCAACTGCTGCTGCTTCTGCTGATTGGCTTTGA
TTCGAACAGCAATTTGCATTAACGCATTTCTGCATTGCAGTGG
```

D.

CGGCTCGTGTAGGTATCTCATACTCTAGCTGTGGTTGCAGCTGCGCCTGCCTGTAACTGCAACTGCTGCTGCTTCTGCTGATTGGCTTTGATTCGAACAGCAATTTGCATTAACGCATTTCTGCATTGCAGTGGATCTC

E.

ATG ... TAA

| 67 bp | 175 bp | 123 bp | 191 bp | 154 bp |

Figure 5.

A.

| SP | B-PEPTIDE | | C-PEPTIDE | | A-PEPTIDE |
| 29 | 25 | KR | 39 | RR | 25 |

B.

| SP | B-PEPTIDE | | C-PEPTIDE | | A-PEPTIDE |
| 24 | 30 | RR | 31 | KR | 21 |

C.

SIGNAL PEPTIDE

B-CHAIN

A-CHAIN

C-CHAIN

D.

SIGNAL PEPTIDE

B-CHAIN

A-CHAIN

C-CHAIN
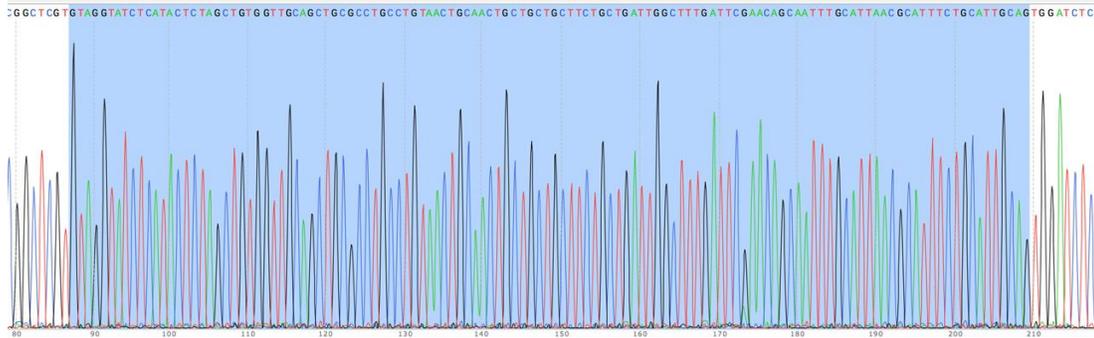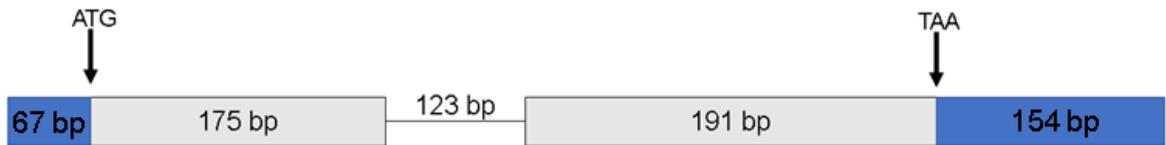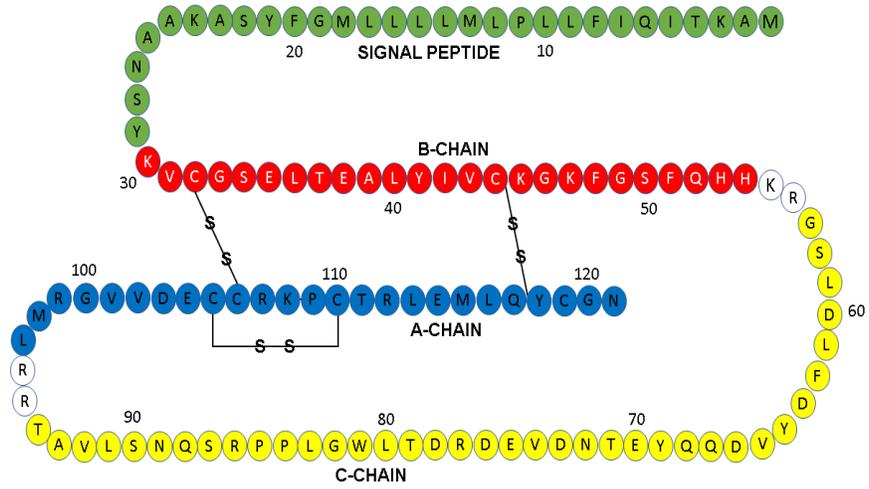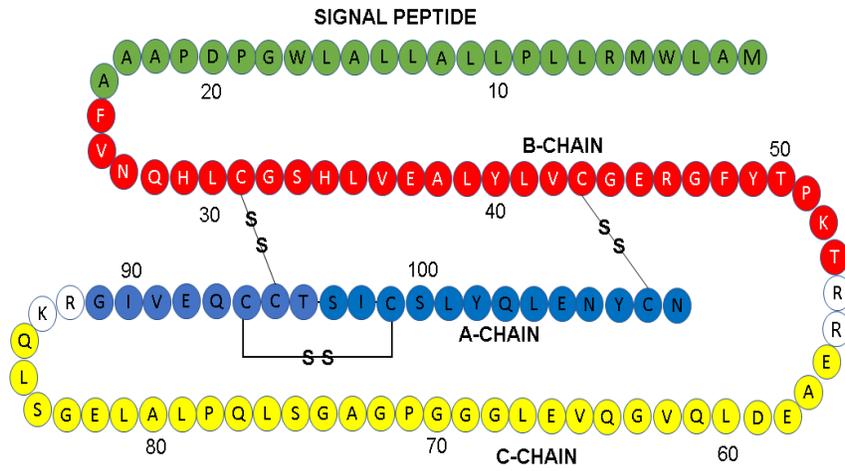
Figure 6.

```
DmILP8    -------------MSSKLHMCRWMLLVIGVCC---LM----------------------  21
DmILP7    MTRMIIQNSGS-----------WTLCGAVLLFVLPLIPTPEALQHTEEGLEMLFRERSQS  49
DmILP6    -------------MVLKVPTS-KVLLVLATLFAVAAMISSWMPQVAASPL----------  36
DmILP4    --------------------MSLIRLGLALLLLAT----VSQL---------------  20
DmILP2    -----------------MS---KPLSFISMVAVILLASS--------------------  19
DmILP1    ---MFSQHNGAAVHGLRL----QSLLIAAMLTAAMAMVTPTGSGH--------------  38
DmILP3    -------------MGIEMRCQDRRILLPSLLLLILMI----------------------  24
DvILP5    ----------------MAKTIQIFLLPLMLLLLMGFYS--------------------  22
DmILP5    -------------------MMFRSVIPVLLFLIPLLLS--------------------  19


DmILP8    ------GSSSGSFCSLERMKKFAMEACEHLFQADEGARRDRRSIEFAHHHLNRLGSG---  72
DmILP7    DWENVWHQETHSRCRDKLVR-QLYWACEKDIYRLTRRNKKRTGNDEA--WIKK-----TT 101
DmILP6    --APTEYEQRRMMCSTGLSD-VIQKICVSGTVALGD----------------------  69
DmILP4    ---LQPVQGRRKMCGEALIQ-ALDVICVNGFTRRVRRSSASKDARV-RDLIRKLQQPDED  75
DmILP2    ----TVKLAQGTLCSEKLNE-VLSMVCEEYNPVIPHKRAM-PGADSDLDALNPLQFV-QE  72
DmILP1    ---QLLPPGNHKLCGPALSD-AMDVVCPHGFNTLPRKRESLLGNSDD--DEDTEQ----E  88
DmILP3    ----GGVQATMKLCGRKLPE-TLSKLCVYGFNAMTKRTLDPVN---F----NQID----G  68
DvILP5    ----AKAANSYKVCGSELTE-ALYIVCKGKFGSFQHHKRGSLDLFDYVDQQYETN----D  73
DmILP5    ----AQAANSLRACGPALMD-MLRVACPNGFNSMF-AKRGTLGLFDYEDHLADLD----S  69
                        *  *


DmILP8    --KTHNKHHYISRSSYPMGGYLKVTREHFNRLSELDIFPR-YKPIKPHHEKKHRFKRDHS 129
DmILP7    TEPDGSTWLHVN----------------------------YANM----FLRSR-RSD-- 125
DmILP6    ------------------------------------VF-------P--NSFGK-RRK-- 80
DmILP4    IEQETETGRLKQKH----------------------TDADTEKGVPPAVGSGR-KLR-- 109
DmILP2    FEEEDNSISEPLRS-----A---------------LFPGSYLGGVLNSLAEVR-RRT-- 108
DmILP1    VQDDSSMWQTLDGAGYSFSP---------------LLTNLYGSEVL--IKMRR-HRR-- 127
DmILP3    FEDRSLLERLLSDSSVQ--------------------------M--L-----KTR-- 90
DvILP5    VEDRDTLWGLPPRSQNS--------------------------L--V-----ATR-- 95
DmILP5    SE---------SHHMNS--------------------------L--S-----SIR-- 82


DmILP8    SRSYNNIPYCCLN-QCEEEF---FC--------- 150
DmILP7    GNTPSISNECCTKAGCTWEEYAEYCPSNKRRNHY 159
DmILP6    RDLQNVTDLCCKSGGCTYRELLQYCKG------- 107
DmILP4    RHRRRIAHECCKE-GCTYDDILDYCA-------- 134
DmILP2    RQRQGIVERCCKK-SCDMKALREYCSVVRN---- 137
DmILP1    HLTGGVYDECCVK-TCSYLELAIYCLPK------ 154
DmILP3    RLRDGVFDECCLK-SCTMDEVLRYCAAKPRT--- 120
DvILP5    RLMRGVVDECCRK-PCTRLEMLQYCGN------- 121
DmILP5    RDFRGVVDSCCRK-SCSFSTLRAYCDS------- 108
            **  .  *  :*
```

Figure 7.

```
Dvir\GJ13215    -------------------TCTAACCAGGCACAATATCCAGCATCGATAAAATTCCAGCA    41
DvILP5          GATCACATCTTGTGTGGCATCTAACCAGGCACAATATCCAGCATCGATAAAATTCCAGCA    60
                                   *****************************************


Dvir\GJ13215    GCACAAAATGGCGAAAACCATACAAATCTTCTTGCTGCCGCTAATGTTGCTGTTGCTGAT    101
DvILP5          GCACAAAATGGCGAAAACCATACAAATCTTCTTGCTGCCGCTAATGTTGCTGTTGCTGAT    120
                ************************************************************


Dvir\GJ13215    GGGATTTTACAGCGCCAAAGCCGCGAACAGCTATAAGGTTTGTGGTTCGGAACTGACCGA    161
DvILP5          GGGATTTTACAGCGCCAAAGCCGCGAACAGCTATAAGGTTTGTGGTTCGGAACTGACCGA    180
                ************************************************************


Dvir\GJ13215    AGCACTTTACATAGTCTGCAAGGGCAAATTCGGCTCCTTTCAACACCACAAACGCGGCTC    221
DvILP5          AGCACTTTACATAGTCTGCAAGGGCAAATTCGGCTCCTTTCAACACCACAAACGCGGCTC    240
                ************************************************************


Dvir\GJ13215    GTGTAGGTATCTCATACTCTAGCTGTGGTTGCAGCTGCGCCTGCCTGTAACTGCAACTGC    281
DvILP5          GTGTAGGTATCTCATACTCTAGCTGTGGTTGCAGCTGCGCCTGCCTGTAACTGCAACTGC    300
                ************************************************************


Dvir\GJ13215    TGCTGCTTCTGCTGATTGGCTTTGATTCGAACAGCAATTTGCATTAACGCATTTCTGCAT    341
DvILP5          TGCTGCTTCTGCTGATTGGCTTTGATTCGAACAGCAATTTGCATTAACGCATTTCTGCAT    360
                ************************************************************


Dvir\GJ13215    TGCAGTGGATCTCTTTGACTATGTGGACCAGCAGTACGAGACGAACGATGTGGAGGACAG    401
DvILP5          TGCAGTGGATCTCTTTGACTATGTGGACCAGCAGTACGAGACGAACGATGTGGAGGACAG    420
                ************************************************************


Dvir\GJ13215    GGATACGCTCTGGGGATTGCCGCCAAGGAGTCAGAATTCCCTGGTGGCCACCAGGCGGCT    461
DvILP5          GGATACGCTCTGGGGATTGCCGCCAAGGAGTCAGAATTCCCTGGTGGCCACCAGGCGGCT    480
                ************************************************************


Dvir\GJ13215    AATGCGCGGCGTTGTCGATGAATGTTGCCGCAAGCCCTGCACGAGATTGGAAATGTTACA    521
DvILP5          AATGCGCGGCGTTGTCGATGAATGTTGCCGCAAGCCCTGCACGAGATTGGAAATGTTACA    540
                ************************************************************


Dvir\GJ13215    GTATTGCGGAAACTAAACCTGAATTGCCAAAAGTAAACACTCATGTATTCCAACTCCAGT    581
DvILP5          GTATTGCGGAAACTAAACCTGAATTGCCAAAAGTAAACACTCATGTATTCCAACTCCAGT    600
                ************************************************************


Dvir\GJ13215    TTGTATTCCCCTATATATAATGTATATGCATGTATTCCTGTTAAAGCAAGTGTACAAAAA    641
DvILP5          TTGTATTCCCCTATATATAATGTATATGCATGTATTCCTGTTAAAGCAAGTGTACAAAAA    660
                ************************************************************


Dvir\GJ13215    CTGGTTCGCCTAATATAACGGAGGAAATAAATAACCTACAA--------- 682
DvILP5          CTGGTTCGCCTAATATAACGGAGGAAATAAATAACCTACAATGAAACCAG 710
                *****************************************
```