Dominic Colter
5/3/2021

Gene annotation uncovers the evolutionary significance of STUB1 in *D. melanogaster*

Abstract

STUB1 is known to participate in protein degradation and regulation of the insulin signaling pathway. The purpose of this experiment was to annotate STUB1 in *D. melanogaster*, *D. simulans*, *D. virilis*, and *D. willistoni* as well as discover its evolutionary significance. To accomplish this, the GEP UCSC Genome Browser Gateway and NCBI Blast were utilized in the analysis and annotation of STUB1 in each species. This analysis showed that STUB1 is mostly conserved in these species and that the genomic neighborhood of STUB1 changed drastically across time. The conservation of STUB1 in each target species demonstrates that STUB1 is evolutionarily significant and is a critical player in the insulin pathway.

Introduction

STUB1 participates in a significant cellular process known as protein degradation or proteolysis. Protein degradation is defined as the breakdown of proteins into polypeptides or amino acids. Protein degradation is significant for a variety of reasons. For example, damaged proteins are targeted and degraded within cells to prevent mistakes during protein synthesis (Cooper). The ubiquitin-proteasome pathway is a very common pathway that cells undergo for targeted protein degradation. This pathway uses ubiquitin, a small protein, as a marker for protein targets to be degraded and destroyed (Cooper). In this process, ubiquitin is passed on to its target protein through a cascade of enzymes called E1, E2, and E3. First, ubiquitin binds to an E1 enzyme. Then, this E1 enzyme transfers the ubiquitin to an E2 enzyme. Lastly, the E2 enzyme transfers the ubiquitin to an E3 enzyme which then attaches the ubiquitin to the target protein. The ubiquitin tagged protein can then be recognized by a proteasome and be degraded. STUB1 codes for an E3 ubiquitin ligase. Protein degradation is also important for regulating many complex pathways in the body. In this project, STUB1 serves as a regulator of the insulin signaling pathway and participates in the ubiquitin-proteasome pathway. Since the insulin signaling pathway is a crucial biochemical process in organisms, the STUB1 gene is evolutionarily conserved in *D. melanogaster* to ensure proper regulation of the insulin pathway. In order to test this hypothesis, STUB1 was annotated in *D. melanogaster*, *D. simulans*, *D. virilis*, and *D. willistoni*. STUB1 in *D. melanogaster* served as a reference for the other species. *D. simulans*, *D. virilis*, and *D. willistoni* are farther away from *D. melanogaster* in the evolutionary tree. Therefore, if STUB1 is evolutionarily conserved, there should be minimal differences in the STUB1 gene between each species. By observing the similarities and differences between each gene annotation in each species, a conclusion can be drawn about the evolutionary significance of STUB1. Ultimately, the results verified our hypothesis. Across the different species, STUB1 was highly conserved. However, the genes around STUB1 changed greatly between the species. STUB1 is composed of six exons and the most variation occurred in exon four. Overall, STUB1 is greatly conserved and this gene annotation approach revealed that STUB1 is evolutionary significant in *D. melanogaster*.

Materials and Methods

1. Examination of the genomic neighborhood surrounding STUB1 in *D. melanogaster*

   By using the GEP UCSC Genome Browser Gateway, STUB1 was searched for in the *D. melanogaster* genome. The genes around STUB1 were also observed in the genome. This examination was done so that we could compare STUB1 and the genomic neighborhood of STUB1 to other species. The genes around STUB1 are important to note because they could be used as evidence for identification of STUB1 orthologs in the other species.
2. Identify the location of the STUB1 ortholog in *D. simulans*, *D. virilis*, and *D. willistoni*.

   The STUB1 protein sequence was obtained from *D. melanogaster* using the Genome Browser. Using this protein sequence, a tblastn search was performed with each species using NCBI Blast to identify regions of local similarity. The best match was determined based on the E-value and percent identity statistics provided by the blast search. These blastn searches were performed in order to determine the regions that had the best similarity in each of the target species to the STUB1 region in *D. melanogaster*.

3. Examination of the genomic neighborhood of the STUB1 ortholog in *D. simulans*, *D. virilis*, and *D. willistoni*.

   Using the matches identified in step 2, the regions of similarity were searched for in each target species using the Genome Browser. In the Genome Browser, alignment prediction tracks were observed for STUB1 and for the genes in the genomic neighborhood. The computationally predicted protein sequence for each gene was searched for in the *D. melanogaster* genome using NCBI Blast. Blastp searches were performed for each species in order to identify similarities between STUB1 and the genomic neighborhood of STUB1 with the *D. melanogaster* genome. This step was performed in order to compare and contrast the genomic regions surrounding STUB1 in each species with the STUB1 region in *D. melanogaster*.

4. Determination of the STUB1 gene structure in *D. melanogaster*

   The Gene Record Finder was used to identify the exons in STUB1 and retrieve their DNA sequences in *D.melanogaster*. These exon coding sequences were used to determine the approximate coordinates of each coding exon (CDS) of STUB1 in each target species. Each CDS was compared to the matches obtained in step 2 by performing tblastn searches. These searches found the coordinates of each CDS in the orthologs identified in step 2.

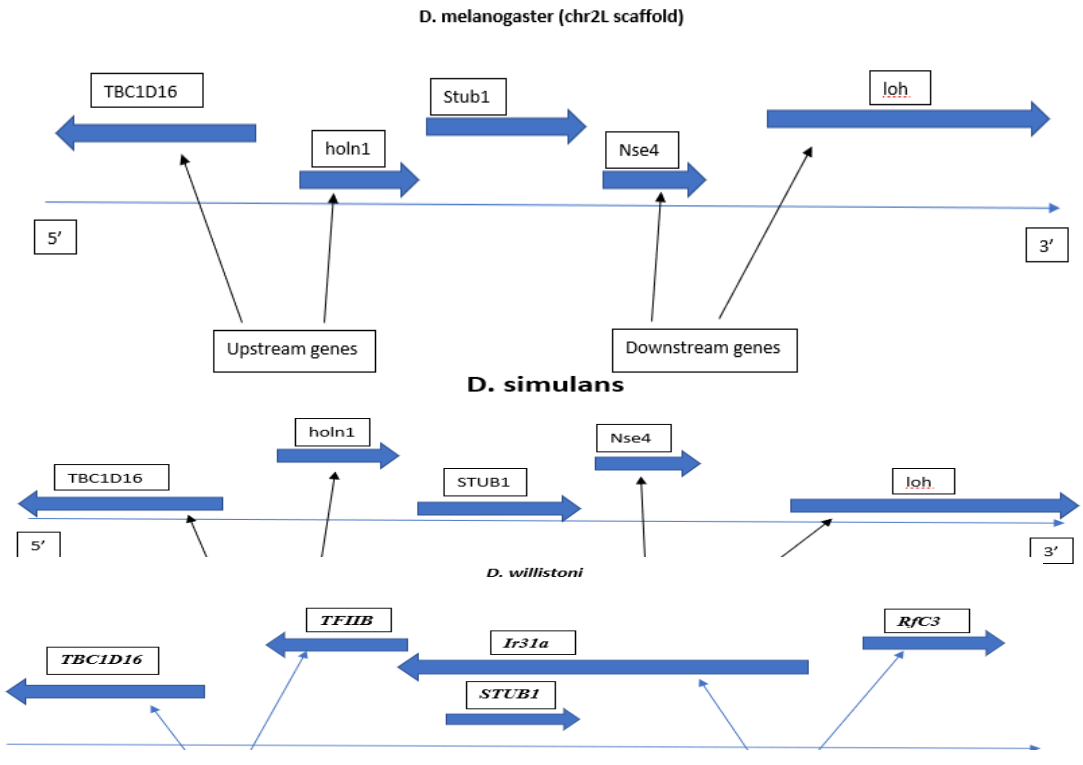
5. Refining coordinates of coding exons

   After discovering the boundaries of each coding exon in STUB1 for each species, splice donor and acceptor sites were identified by visual inspection using the Genome Browser. RNA-Seq data was observed for each coding exon to determine if each CDS alignment was supported. The location of the stop and start codons were verified using RefSeq data and alignment prediction tracks. Splice donor and acceptor sites for each CDS were determined by observing the Genome Browser in order to further verify the predicted coordinates of each coding exon. Splice junction prediction tracks were examined in order to verify the coordinates of each intron in STUB1 for each species.

6. Verification of gene models

   To verify each proposed gene model in each species, the Gene Model Checker was utilized to determine if each model satisfied the basic biological constraints. Dot plots and alignment screenshots were produced by the Gene Model Checker. This alignment information was observed and examined for each proposed model. From this information, conclusions can be made about the evolutionary significance of STUB1 in *D. melanogaster*.

Results



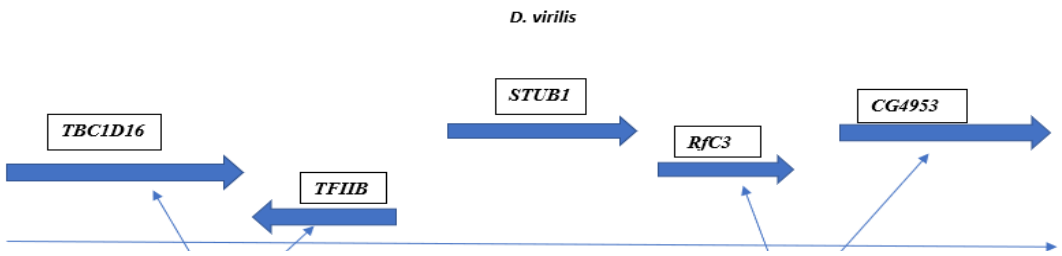


Figure 1. Genomic neighborhood of STUB1 in each species.

Each solid blue arrow represents the direction in which each gene is read. The genomic neighborhood of each species is composed of the two nearest upstream and two nearest downstream genes in relation to STUB1. The name of each gene is illustrated above each solid blue arrow. The name of each species is also shown for each genomic neighborhood at the top of each image.

Identity: 285/289 (98.6%), **Similarity:** 287/289 (99.3%), **Gaps:** 0/289 (0.0%)

```
Dmel_STUB1-PA         1 MTTKHIYSTTNLSDLQLKEQGNCLFAARKYDDAINCYSKAIIKNPTNATYFTNRALCNLK  60
                        ************** ******************************************
DsimGB2_STUB1-PA      1 MTTKHIYSTTNLSDQQLKEQGNCLFAARKYDDAINCYSKAIIKNPTNATYFTNRALCNLK  60

Dmel_STUB1-PA        61 LKRWELCCQDSRRALDIDGNLLKGHFFLGQGLMEIDNFDEAIKHLQRAYDLSKEQKQNFG 120
                        **********.************************************************
DsimGB2_STUB1-PA     61 LKRWELCCQDCRRALDIDGNLLKGHFFLGQGLMEIDNFDEAIKHLQRAYDLSKEQKQNFG 120

Dmel_STUB1-PA       121 DDITLQLRLARKKRWNVMEEKRIQQEIELQSYLNGLIKGDMESRLANLKLNGNVHDEQLK 180
                        ************************************************.*******
DsimGB2_STUB1-PA    121 DDITLQLRLARKKRWNVMEEKRIQQEIELQSYLNGLIKGDMESRLANLKLNGSVHDEQLK 180

Dmel_STUB1-PA       181 DKQQEIEQECDDHIKELNNIFSKVDERRKXREVPDFLCGKISFEILTDPVITPSGITYER 240
                        ****************************:*****************************
DsimGB2_STUB1-PA    181 DKQQEIEQECDDHIKELNNIFSKVDERRKXRDVPDFLCGKISFEILTDPVITPSGITYER 240

Dmel_STUB1-PA       241 KDIEEHLQRVGHFDPVTRVKLTQDQLIPNFSMKEVVDSFIAENEWSLDY 289
                        ************************************************
DsimGB2_STUB1-PA    241 KDIEEHLQRVGHFDPVTRVKLTQDQLIPNFSMKEVVDSFIAENEWSLDY 289
```

Identity: 271/289 (93.8%), **Similarity:** 283/289 (97.9%), **Gaps:** 0/289 (0.0%)

```
Dmel_STUB1-PA         1 MTTKHIYSTTNLSDLQLKEQGNCLFAARKYDDAINCYSKAIIKNPTNATYFTNRALCNLK  60
                        **:*****************************.**********************
DwilCAF1_STUB1-PA     1 MTSKHIYSTTNLSDLQLKEQGNCLFAARKYDDAISCYSKAIIKNPTNATYFTNRALCNLK  60

Dmel_STUB1-PA        61 LKRWELCCQDSRRALDIDGNLLKGHFFLGQGLMEIDNFDEAIKHLQRAYDLSKEQKQNFG 120
                        ****:*****.************************.**********************
DwilCAF1_STUB1-PA    61 LKRWDLCCQDCRRALDIDGNLLKGHFFLGQGLMEIDSFDEAIKHLQRAYDLSKEQKQNFG 120

Dmel_STUB1-PA       121 DDITLQLRLARKKRWNVMEEKRIQQEIELQSYLNGLIKGDMESRLANLKLNGNVHDEQLK 180
                        ************************:********** ***.***.******:* *:::*:**
DwilCAF1_STUB1-PA   121 DDITLQLRLARKKRWNVMEEKRIHQEIELQSYLNRLIKDDMENRLANLKINENLNEEHLK 180

Dmel_STUB1-PA       181 DKQQEIEQECDDHIKELNNIFSKVDERRKXREVPDFLCGKISFEILTDPVITPSGITYER 240
                        ***********************.******:**:****************************
DwilCAF1_STUB1-PA   181 DKQQEIEQECDDHIKELNNIFAKVDERRKXRDVPDFLCGKISFEILTDPVITPSGITYER 240

Dmel_STUB1-PA       241 KDIEEHLQRVGHFDPVTRVKLTQDQLIPNFSMKEVVDSFIAENEWSLDY 289
                        ************************************************
DwilCAF1_STUB1-PA   241 KDIEEHLQRVGHFDPVTRVKLTQDQLIPNFSMKEVVDSFIAENEWSLDY 289
```

Identity: 268/289 (92.7%), **Similarity:** 282/289 (97.6%), **Gaps:** 0/289 (0.0%)

```
Dmel_STUB1-PA         1 MTTKHIYSTTNLSDLQLKEQGNCLFAARKYDDAINCYSKAIIKNPTNATYFTNRALCNLK  60
                        **:****************************:***  **********************
DvirCAF1_STUB1-PA     1 MTSKHIYSTTNLSDLQLKEQGNCLFAARKYEDAIICYSKAIIKNPTNATYFTNRALCNLK  60

Dmel_STUB1-PA        61 LKRWELCCQDSRRALDIDGNLLKGHFFLGQGLMEIDNFDEAIKHLQRAYDLSKEQKQNFG 120
                        **********.************************:**********************
DvirCAF1_STUB1-PA    61 LKRWELCCQDCRRALDIDGNLLKGHFFLGQGLMEIDNYDEAIKHLQRAYDLSKEQKQNFG 120

Dmel_STUB1-PA       121 DDITLQLRLARKKRWNVMEEKRIQQEIELQSYLNGLIKGDMESRLANLKLNGNVHDEQLK 180
                        ****************:*****:********** ***.***.***.**** *:::*:**
DvirCAF1_STUB1-PA   121 DDITLQLRLARKKRWNVLEEKRIHQEIELQSYLNRLIKDDMENRLASLKLNENLNEEHLK 180

Dmel_STUB1-PA       181 DKQQEIEQECDDHIKELNNIFSKVDERRKXREVPDFLCGKISFEILTDPVITPSGITYER 240
                        **************:******:******:**:****************************
DvirCAF1_STUB1-PA   181 DKQQEIEQECDDHIQELNNIFAKVDERRKXRDVPDFLCGKISFEILTDPVITPSGITYER 240

Dmel_STUB1-PA       241 KDIEEHLQRVGHFDPVTRVKLTQDQLIPNFSMKEVVDSFIAENEWSLDY 289
                        ****************************************:********
DvirCAF1_STUB1-PA   241 KDIEEHLQRVGHFDPVTRVKLTQDQLIPNFSMKEVVDSFIADNEWSLDY 289
```

Figure 2. Alignment of STUB1 in each target species to STUB1 in *D. melanogaster*.

Each alignment is accompanied by percent identity, percent similarity, and percent gaps. These statistics describe how closely related STUB1 is within each species to STUB1 in *D. melanogaster*. The STUB1 protein sequence in *D. melanogaster* is placed above the STUB1 protein sequence from each target species, depending on the alignment. Dmel_STUB1-PA represents the STUB1 sequence from *D. melanogaster* while the other strand is the STUB1 sequence from the target species. The colored regions of the sequences represent each exon starting from position one. Differences in the sequences are indicated by "." or ":" in the alignment and "*" indicates no change.

In order to examine STUB1 and its genomic neighborhood in each species, the GEP UCSC Genome Browser Gateway and NCBI Blast were used to examine each genome. STUB1 and its genomic neighborhood appeared to stay the same between *D. simulans* and *D. melanogaster*. Each gene in these neighborhoods matched in orientation as well. The genomic neighborhood of STUB1 in *D. virilis* and *D. willistoni* changed drastically compared to the genomic neighborhood of STUB1 in *D. melanogaster*. Different genes appeared with different lengths and orientations. The *D. virilis* and *D. willistoni* STUB1 genomic neighborhoods both contain genes TBC1D16, TFIIB, and RfC3. TBC1D16 was present in every species. However, its orientation was different in *D. virilis*. STUB1 was present in every species, maintaining its orientation. The alignments show most of the variation occurring in the fourth exon of STUB1 but STUB1 in each species is very similar to STUB1 in *D. melanogaster*.

Discussion

The results obtained support our hypothesis stated in the introduction. The high percent similarity, the absence of gaps, and the minimal mutations across the various target species indicate that STUB1 is highly conserved evolutionarily. The lasting presence of TBC1D16 shows that the correct STUB1 orthologs were identified in each species and that TBC1D16 is evolutionarily significant. Likewise, the other genes in the genomic neighborhood of STUB1 changed frequently between species. This finding suggests that STUB1 may have translocated to a different area of the genome multiple times across time. Another reason could be that the surrounding genes were not evolutionarily selected for, replacing them with new genes. The mutations seen in the alignments were expected, as *D. virilis* and *D. willistoni* are evolutionarily distant from *D. melanogaster*. The most variation occurred in the fourth exon. This observation suggests that this region of STUB1 possibly changed overtime due to external pressures. However, this region was highly conserved between *D. simulans* and *D. melanogaster*, suggesting that the mutations that occurred across time became favored and were kept in *D. melanogaster*. The conservation of STUB1 in each target species supports our hypothesis that STUB1 is evolutionarily significant and is a critical player in the insulin pathway.

The goal of this experiment was to annotate the STUB1 gene in *D. melanogaster, D. simulans*, *D. virilis*, and *D. willistoni*. Another aspect of the goal was to determine the evolutionary significance of STUB1 in *D. melanogaster*. After completing this investigation, the goal was reached. STUB1 was annotated in each species and evidence was obtained that supported our evolutionary hypothesis of STUB1. This investigation could have several implications. For instance, STUB1 in *D. melanogaster* is orthologous to the STUB1 gene in humans. This correlation could prove important for human experiments. This investigation provides evidence that studying STUB1 in *D. melanogaster* could be useful for understanding STUB1 in human beings. For example, one study found that the overexpression of STUB1 was linked to decreased neurodegeneration in patients diagnosed with spinocerebellar ataxia type 1 **(**Al-Ramahi et al.**)**. Studying STUB1 in *D. melanogaster* could further advance this type of research. Further experiments could be done in order to discover the function of STUB1 in each of the species studied in this investigation. For instance, STUB1 could be mutated in each species to examine their phenotypes. Based on the phenotypes discovered, a conclusion can be made about the preservation or destruction of STUB1's function over time.

Literature Cited

1. Cooper GM. The Cell: A Molecular Approach. 2nd edition. Sunderland (MA): Sinauer Associates; 2000. Protein Degradation. Available from: https://www.ncbi.nlm.nih.gov/books/NBK9957/
2. Al-Ramahi, Ismael et al. "CHIP protects from the neurotoxicity of expanded and wild-type ataxin-1 and promotes their ubiquitination and degradation." The Journal of biological chemistry vol. 281,36 (2006): 26714-24. doi:10.1074/jbc.M601603200