Annotation Example:
Alpha catenin in the *Leptinotarsa decemlineata* genome

For training purposes, we have implemented special instances of our BLAST and Web Apollo tools specifically to allow for training or practice annotations. If you want to take time to familiarize yourself with the various tools before annotating on an active project, feel free to use our training resources.

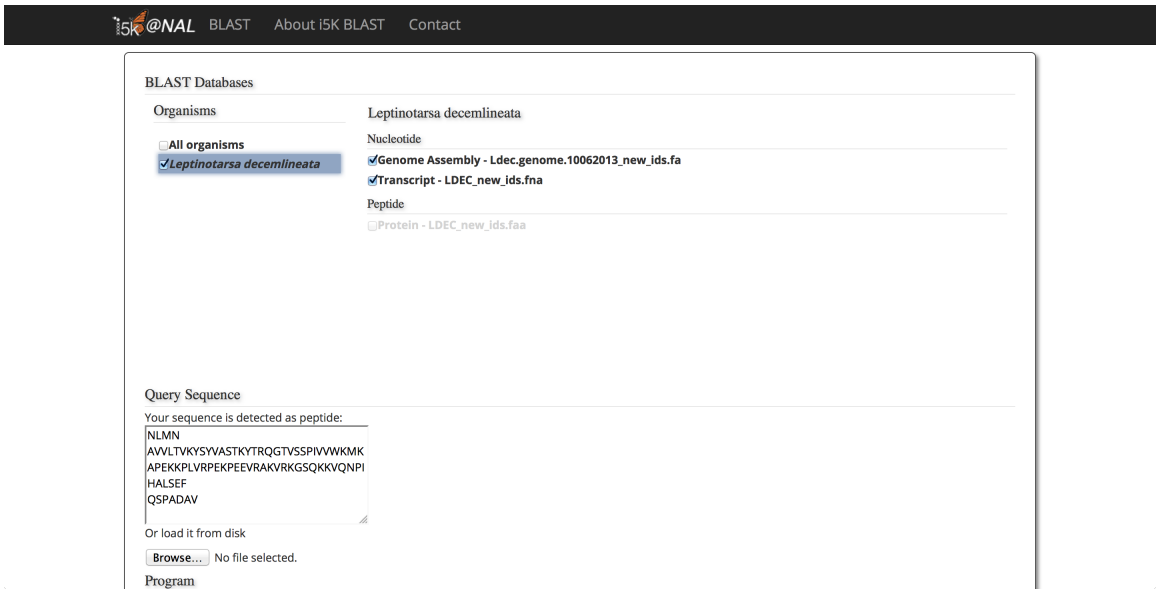
The training BLAST instance is located here: i5k.nal.usda.gov/training/blast
The training Web Apollo instance created for this workshop is for *Leptinotarsa decemlineata* (lepdec), the Colorado Potato Beetle. You can access the training Web Apollo instance here: apollo.nal.usda.gov/lepdec_training
Username: demo
Password: demo

Retrieve the sequence for alpha catenin for fruit fly, and download the protein sequence (available at http://www.ncbi.nlm.nih.gov/protein/NP_524219.1).

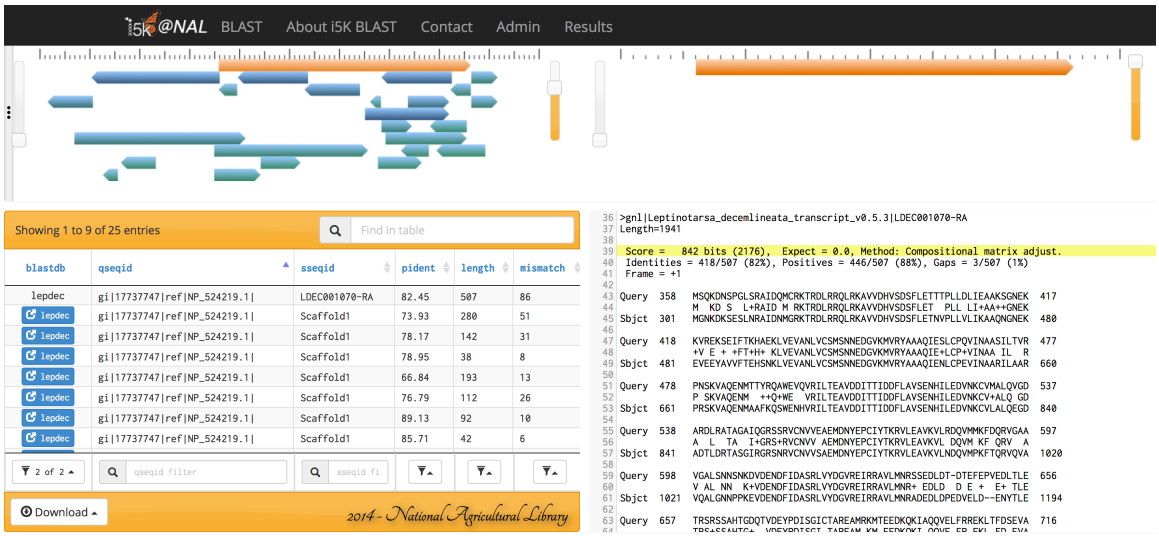
```
>gi|17737747|ref|NP_524219.1| alpha catenin, isoform A [Drosophila melanogaster]
MLKPDKMGTLTDFGQIALKWDPKNLEIRTMSVEKTLEPLVLQVTTLVNTKGPSKKKKGKSKRASALVAAV
EKATENFIQKGEQIAYENPDITQEMLTAVDEVKKTGDAMSIAAREFSEDPCSSLKRGNMVRAARNLLSAV
TRLLILADMVDVHLLLKSLHIVEDDLNKLKNASSQDELMDNMRQFGRNAGELIKQAAKRQQELKDPQLRD
DLAAARAMLKKHSTMLLTASKVYVRHPELDLAKVNRDFILKQVCDAVNTISDVAQGKSSQPTDIYSGAGE
LAAALDDFDEGIVMDPMTYSEKRSRQLLEERLESIISAAALMADADCTRDERRERIVAECNAVRQALQDL
LSEYMSNMSQKDNSPGLSRAIDQMCRKTRDLRRQLRKAVVDHVSDSFLETTTPLLDLIEAAKSGNEKKVR
EKSEIFTKHAEKLVEVANLVCSMSNNEDGVKMVRYAAAQIESLCPQVINAASILTVRPNSKVAQENMTTY
RQAWEVQVRILTEAVDDITTIDDFLAVSENHILEDVNKCVMALQVGDARDLRATAGAIQGRSSRVCNVVE
AEMDNYEPCIYTKRVLEAVKVLRDQVMMKFDQRVGAAVGALSNNSNKDVDENDFIDASRLVYDGVREIRR
AVLMNRSSEDLDTDTEFEPVEDLTLETRSRSSAHTGDQTVDEYPDISGICTAREAMRKMTEEDKQKIAQQ
VELFRREKLTFDSEVAKWDDTGNDIIFLAKHMCMIMMEMTDFTRGRGPLKTTMDVINAAKKISEAGTKLD
KLTREIAEQCPESSTKKDLLAYLQRIALYCHQIQITSKVKADVQNISGELIVSGLDSATSLIQAAKNLMN
AVVLTVKYSYVASTKYTRQGTVSSPIVVWKMKAPEKKPLVRPEKPEEVRAKVRKGSQKKVQNPIHALSEF
QSPADAV
```

Use the training BLAST tool (i5k.nal.usda.gov/training/blast) to map this gene back to the *Leptinotarsa decemlineata* (Colorado Potato Beetle) genome.

- In the BLAST page, check the *Leptinotarsa decemlineata* box on the left.
- In the right panel, check both the "Genome Assembly" and "Transcript" boxes.
- Paste the fruit fly sequence from NCBI into the Query Sequence box and click Submit.

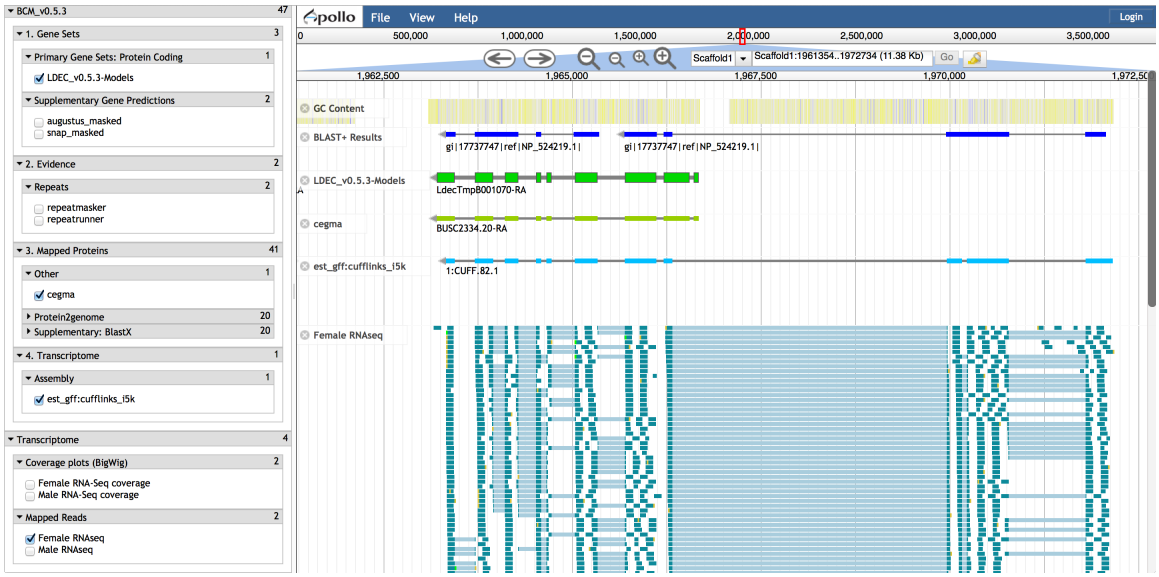


Blast shows a hit to one gene model (LDEC001070-RA) and eight HSPs on Scaffold1. By inspecting the HSPs in the Subject and Query panels along the top, you can verify that the HSPs are consistent in order and organization. Click on the lepdec button to open the HSPs for Scaffold1 in the genome browser.



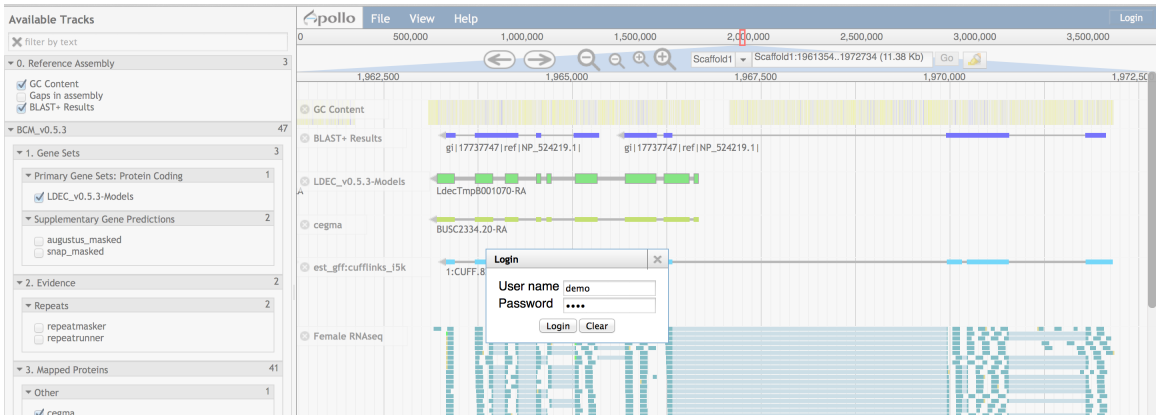
The BLAST+ Results track should be visible in the initial view of the genome browser. The initial view only shows some of the hits; Zoom out until all eight HSPs are visible.

Turn on the LDEC_v0.5.3-Models, cegma, est_gff:cufflinks and Female RNAseq tracks. Also take some time to examine the "protein2genome_Arthropoda" track to see how proteins from other arthropod species aligned to this region.

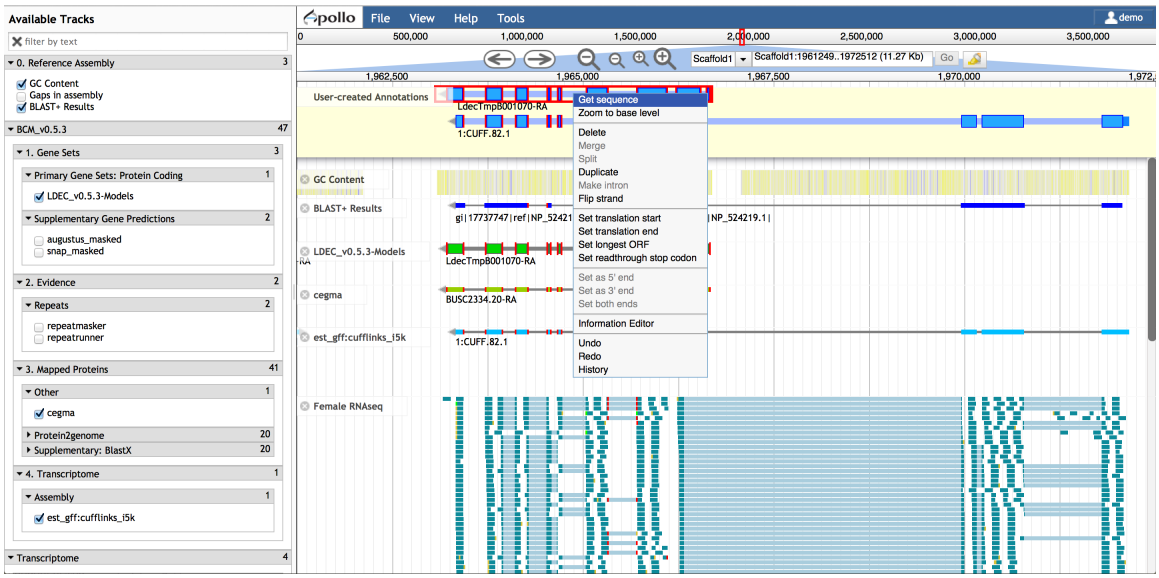


Look at the LDEC_v0.5.3 model alongside the other tracks. Clicking on an intron will highlight the entire gene model. Exons with boundaries identical to the selected model will have red edges. The LDEC_v0.5.3 model agrees with the cegma protein alignment. The cufflinks and RNAseq data support a different CDS structure for the first three exons. The BLAST results also indicate that the five-prime end of the gene may be further upstream than predicted by the LDEC_v0.5.3 model.

Log in to enable edit mode.



Click and drag both the LDEC model and the Cufflinks model into the yellow "User-created Annotations" pane at the top of the browser. Next, right-click an intron to select the entire model and click "Get sequence". Save the translated sequence in a text file and repeat for the other gene model. To make the multiple alignment easier to follow, prepend each sequence name with a short descriptive tag.



Open a multiple sequence aligner like Clustal Omega
http://www.ebi.ac.uk/Tools/msa/clustalo/
Perform a multiple sequence alignment with the sequences saved from NCBI (fruit fly) and the gene models saved from Web Apollo (LDEC_v0.5.3 and Cufflinks).
Paste the three sequences into the input box and click submit.
The results show that there are differences in the five prime and three prime ends of the gene. This is more pronounced in the LDEC_v0.5.3 model (See attached Clustal report LDEC_clustal_1).
To try to get a better alignment, remove the LDEC model and rerun Clustal (see LDEC_clustal_2). The new alignment has a shorter area of disagreement at the five prime end.

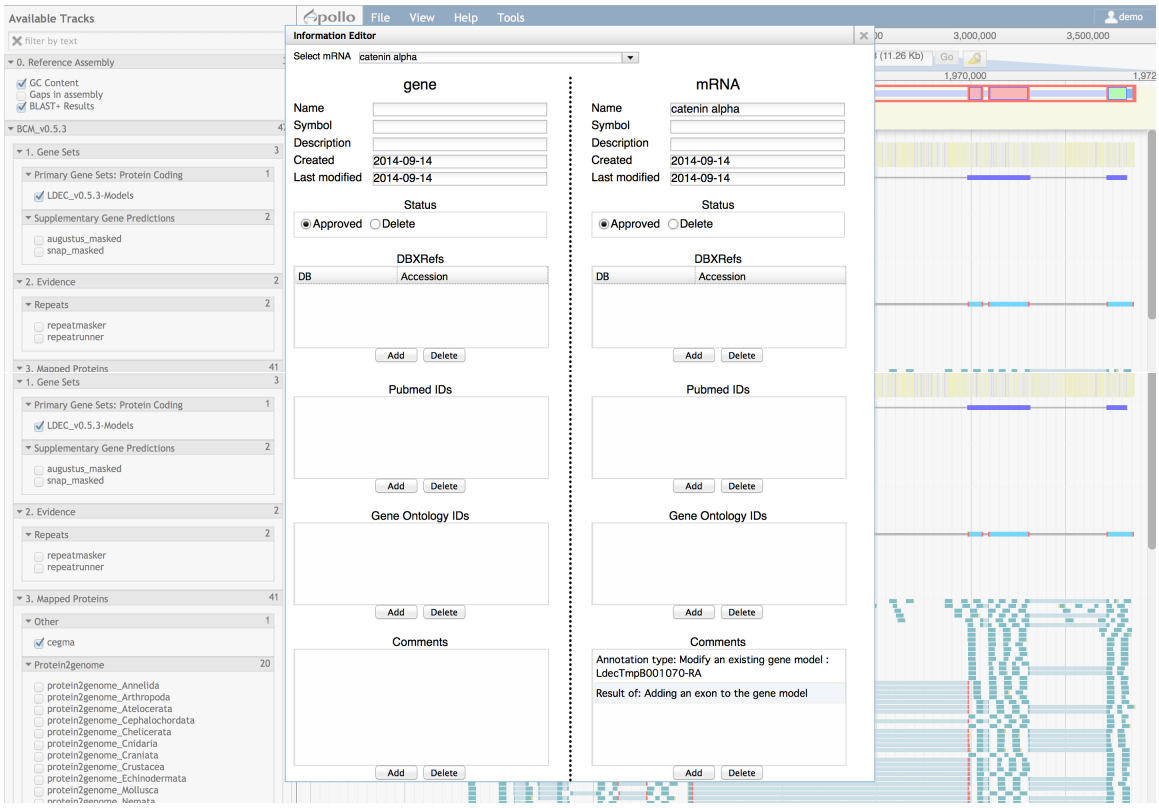
In Web Apollo, zoom in on the start of the longer model. You can see that the model runs into a gap in the assembly, limiting our ability to annotate further. Zoom back out to view the entire gene and then zoom in to the three prime end of the gene. The cufflinks model does not end with a stop codon. You should

extend the exon until it includes a stop codon. You can do this by dragging the end of the exon until it includes a stop codon OR by merging the last exon from the LDEC model with the active annotation. To merge them, click on the last exon in the LDEC model and drag it up to the Cufflinks model until the box around the active annotation turns green then release the mouse button. The end of the two models should now be the same, which you can see through the red edge highlighting.

Get the new peptide sequence and use NCBIs BLASTP to search it back against the non-redundant (nr) database. The top two hits are to alpha catenin genes in *Tribolium castaneum*, the red flour beetle, and *Bombus impatiens*, a bumblebee. Use these two protein sequences along with the sequence from the annotation in another multiple-alignment (see LDEC_clustal_3). These models have better agreement at the five prime end of the gene, but there is still some disagreement with the annotation at the three prime end. Unfortunately, there is also a gap in the assembly at the three prime end, so further refinement of this model is not possible at this time. The annotation does have in-frame start and stop codons. This is the best that I can annotate this model.

Now that our new candidate gene model is annotated, delete the shorter model, leaving the Cufflinks-based model. Now we need to add information to describe this model and the actions we took to make it. Right click on the model and select "Information Editor".

- For the Name fields, enter "catenin alpha".
- On the mRNA side, click Add to add a comment, then use the drop down option to select "Annotation type: Modify an existing gene model ". Add more information to clearly state what the model is: "Annotation type: Modify an existing gene model: LdecTmpB001070-RA
- You also need to add additional comments here as applicable to your work. In this case I also added "Result of: Adding an exon to the gene model".
- Finally click the "Approved" option in the Status section. This lets others know that you feel the annotation is complete. Even if you do not change the status the annotation will remain so you can save the URL and return to complete it later.



Once this is done, simply click the X to close the Information Editor. Your modifications are already saved back to the server.

Web Apollo Workshop Example Supplemental Materials

Web Resources used:
BLAST: https://i5k.nal.usda.gov/training/blast
JBrowse : https://apollo.nal.usda.gov/lepdec_training/jbrowse
Web Apollo: https://apollo.nal.usda.gov/lepdec_training/
Clustal Omega: http://www.ebi.ac.uk/Tools/msa/clustalo/

Additional Resources
Annotation guide

LDEC_clustal_1

```
CLUSTAL O(1.2.1) multiple sequence alignment


LDEC_009905F2F303802AED9B6891B7FFF088           ------------------------------------------------------------
gi|17737747|ref|NP_524219.1|                   MLKPDKMGTLTDFGQIALKWDPKNLEIRTMSVEKTLEPLVLQVTTLVNTKGPSKKKKGKS
Cufflinks_92BA0B7F01612DBCD26CE6DAEA73897C      -----------MTDRFGLKWDPKNLEIRTMSVEKTLEPLVLQVTTLVNTKGPSKKKKGKS


LDEC_009905F2F303802AED9B6891B7FFF088           ------------------------------------------------------------
gi|17737747|ref|NP_524219.1|                   KRASALVAAVEKATENFIQKGEQIAYENPDITQEMLTAVDEVKKTGDAMSIAAREFSEDP
Cufflinks_92BA0B7F01612DBCD26CE6DAEA73897C      KRANALVSTVEKATENFIEKGEQIAYENPDITDEMLAAVEEVKKTGTAMSVAAREFSEDP


LDEC_009905F2F303802AED9B6891B7FFF088           ------------------------------------------------------------
gi|17737747|ref|NP_524219.1|                   CSSLKRGNMVRAARNLLSAVTRLLILADMVDVHLLLKSLHIVEDDLNKLKNASSQDELMD
Cufflinks_92BA0B7F01612DBCD26CE6DAEA73897C      CSSLKRGNMVRAARNLLSAVTRLLILADMVDVHLLLKSLHVVEDDLEKLKNASSHGELLD


LDEC_009905F2F303802AED9B6891B7FFF088           -------------------------------------MLSSLKSVLLCI-------QQL-
gi|17737747|ref|NP_524219.1|                   NMRQFGRNAGELIKQAAKRQQELKDPQLRDDLAAARAMLKKHSTMLLTASKVYVRHPELD
Cufflinks_92BA0B7F01612DBCD26CE6DAEA73897C      NIKAFGQNANELMNQAAKRQQELKDPQLRDDLAAARAVLKKHSTMLLTASKVYVRHPELA
                                                                                      :*.. .::**          :*

LDEC_009905F2F303802AED9B6891B7FFF088           ---------------------EVLRGVT--RPLSIYSLK--ITA--SNQDMHSTMSQLTK
gi|17737747|ref|NP_524219.1|                   LAKVNRDFILKQVCDAVNTISDVAQGKSSQPT-DIYSGAGELAAALDDFDEGIVMDPMTY
Cufflinks_92BA0B7F01612DBCD26CE6DAEA73897C      AAKANRDYVLKQVCEAVHTINDVAQGRTPQPACGPYDGPGELAAALDDFDDHMVMEPLAY
                                                                     :* :* :       *.    ::*  .: *   .*. ::

LDEC_009905F2F303802AED9B6891B7FFF088           ------------NQSYIFCISFFNDVALK-WAAISEIPIECSLFLFLSVKICFKPHLQQF
gi|17737747|ref|NP_524219.1|                   SEKRSRQLLEERLESIISAAALMADADCTRDERRERIVAECNA-----VRQALQDLLSE-
Cufflinks_92BA0B7F01612DBCD26CE6DAEA73897C      NEVHTRPSLEERLESIISGAALMADSSCTRDERRERIVAECNA-----VRQALQDLLSE-
                                                             :* *   ::: *   .     ..*  **.      *: .::  *.:

LDEC_009905F2F303802AED9B6891B7FFF088           FICLQMGNKDKSESLNRAIDNMGRKTRDLRRQLRKAVVDHVSDSFLETNVPLLVLIKAAQ
gi|17737747|ref|NP_524219.1|                   -YMSNMSQKDNSPGLSRAIDQMCRKTRDLRRQLRKAVVDHVSDSFLETTTPLLDLIEAAK
Cufflinks_92BA0B7F01612DBCD26CE6DAEA73897C      -YMTNMGNKDKSESLNRAIDNMGRKTRDLRRQLRKAVVDHVSDSFLETNVPLLVLIKAAQ
                                                    :*.:**:* .*.****:* *************************..*** **:**:

LDEC_009905F2F303802AED9B6891B7FFF088           NGNEKEVEEYAVVFTEHSNKLVEVANLVCSMSNNEDGVKMVRYAAAQIENLCPEVINAAR
gi|17737747|ref|NP_524219.1|                   SGNEKKVREKSEIFTKHAEKLVEVANLVCSMSNNEDGVKMVRYAAAQIESLCPQVINAAS
Cufflinks_92BA0B7F01612DBCD26CE6DAEA73897C      NGNEKEVEEYAVVFTEHSNKLVEVANLVCSMSNNEDGVKMVRYAAAQIENLCPEVINAAR
                                                .****:*.* : :**:*::******************************.***:*****

LDEC_009905F2F303802AED9B6891B7FFF088           ILAARPRSKVAQENMAAFKQSWENHVRILTEAVDDITTIDDFLAVSENHILEDVNKCVLA
gi|17737747|ref|NP_524219.1|                   ILTVRPNSKVAQENMTTYRQAWEVQVRILTEAVDDITTIDDFLAVSENHILEDVNKCVMA
Cufflinks_92BA0B7F01612DBCD26CE6DAEA73897C      ILAARPRSKVAQENMAAFKQSWENHVRILTEAVDDITTIDDFLAVSENHILEDVNKCVLA
                                                **:.**.********::::*:** :********************************:*

LDEC_009905F2F303802AED9B6891B7FFF088           LQEGDADTLDRTASGIRGRSNRVCNVVSAEMDNYEPCIYTKRVLEAVKVLNDQVMPKFTQ
gi|17737747|ref|NP_524219.1|                   LQVGDARDLRATAGAIQGRSSRVCNVVEAEMDNYEPCIYTKRVLEAVKVLRDQVMMKFDQ
Cufflinks_92BA0B7F01612DBCD26CE6DAEA73897C      LQEGDADTLDRTASGIRGRSNRVCNVVSAEMDNYEPCIYTKRVLEAVKVLNDQVMPKFTQ
                                                ** ***  *  **..*:***.******.**********************.**** ** *

LDEC_009905F2F303802AED9B6891B7FFF088           RVQVAVQALGNNPPKEVDENDFIDASRLVYDGVREIRRAVLMNRADEDLDPE-DVELDEN
gi|17737747|ref|NP_524219.1|                   RVGAAVGALSNNSNKDVDENDFIDASRLVYDGVREIRRAVLMNRSSEDLDTDTEFEPVED
Cufflinks_92BA0B7F01612DBCD26CE6DAEA73897C      RVQVAVQALGNNPPKEVDENDFIDASRLVYDGVREIRRAVLMNRADEDLDPE-DVELDEN
                                                ** .** **.**  *:****************************:.**** : :.*  *:

LDEC_009905F2F303802AED9B6891B7FFF088           YTLETRSKSSAHTGEHGVDEYPDISGITTAREAMGKMPEEDKQKILQQVEFFRSEKLKFD
gi|17737747|ref|NP_524219.1|                   LTLETRSRSSAHTGDQTVDEYPDISGICTAREAMRKMTEEDKQKIAQQVELFRREKLTFD
Cufflinks_92BA0B7F01612DBCD26CE6DAEA73897C      YTLETRSKSSAHTGEHGVDEYPDISGITTAREAMGKMPEEDKQKILQQVEFFRSEKLKFD
                                                 ******:******:: ********** ****** ** ******* ****:** ***.**

LDEC_009905F2F303802AED9B6891B7FFF088           REVAKWDDTGNDIIVLAKHMCMIMMEMTDFTRGRGPLKTTMDVINAAKKISEAGTKLDKL
gi|17737747|ref|NP_524219.1|                   SEVAKWDDTGNDIIFLAKHMCMIMMEMTDFTRGRGPLKTTMDVINAAKKISEAGTKLDKL
Cufflinks_92BA0B7F01612DBCD26CE6DAEA73897C      REVAKWDDTGNDIIVLAKHMCMIMMEMTDFTRGRGPLKTTMDVINAAKKISEAGTKLDKL
                                                 *************.*********************************************

LDEC_009905F2F303802AED9B6891B7FFF088           TRQIAEQCPESSTKQDLLAYLQRIALYCHQMNITSKVKADVQNISGELIVSGLDSATSLI
gi|17737747|ref|NP_524219.1|                   TREIAEQCPESSTKKDLLAYLQRIALYCHQIQITSKVKADVQNISGELIVSGLDSATSLI
Cufflinks_92BA0B7F01612DBCD26CE6DAEA73897C      TRQIAEQCPESSTKQDLLAYLQRIALYCHQMNITSKVKADVQNISGELIVSGLDSATSLI
                                                **:***********:***************:::***************************

LDEC_009905F2F303802AED9B6891B7FFF088           QAAKNLMNAVVLTVKSSYVASTKYPRQGTIAVSDFSILKQIMFEHICQRFECMKICYEFG
gi|17737747|ref|NP_524219.1|                   QAAKNLMNAVVLTVKYSYVASTKYTRQGTVSSPIVVWKMKAPE----------------K
Cufflinks_92BA0B7F01612DBCD26CE6DAEA73897C      QAAKNLMNAVVLTVKSSYVASTKYPRQGTIAV----------------------------
                                                *************** ******** ****::

LDEC_009905F2F303802AED9B6891B7FFF088           VPLS-KQVPTDLK----------------------------
gi|17737747|ref|NP_524219.1|                   KPLVRPEKPEEVRAKVRKGSQKKVQNPIHALSEFQSPADAV
Cufflinks_92BA0B7F01612DBCD26CE6DAEA73897C      ----------------------------------------
```

LDEC_clustal_2

```
CLUSTAL O(1.2.1) multiple sequence alignment


gi|17737747|ref|NP_524219.1|                MLKPDKMGTLTDFGQIALKWDPKNLEIRTMSVEKTLEPLVLQVTTLVNTKGPSKKKKGKS
Cufflinks_92BA0B7F01612DBCD26CE6DAEA73897C  ----------MTDRFGLKWDPKNLEIRTMSVEKTLEPLVLQVTTLVNTKGPSKKKKGKS
                                                      ::.******************************************

gi|17737747|ref|NP_524219.1|                KRASALVAAVEKATENFIQKGEQIAYENPDITQEMLTAVDEVKKTGDAMSIAAREFSEDP
Cufflinks_92BA0B7F01612DBCD26CE6DAEA73897C  KRANALVSTVEKATENFIEKGEQIAYENPDITDEMLAAVEEVKKTGTAMSVAAREFSEDP
                                            ***.***::*********:*************:***:**:****** ***:*********

gi|17737747|ref|NP_524219.1|                CSSLKRGNMVRAARNLLSAVTRLLILADMVDVHLLLKSLHIVEDDLNKLKNASSQDELMD
Cufflinks_92BA0B7F01612DBCD26CE6DAEA73897C  CSSLKRGNMVRAARNLLSAVTRLLILADMVDVHLLLKSLHVVEDDLEKLKNASSHGELLD
                                            ***************************************:*****:*******: **:*

gi|17737747|ref|NP_524219.1|                NMRQFGRNAGELIKQAAKRQQELKDPQLRDDLAAARAMLKKHSTMLLTASKVYVRHPELD
Cufflinks_92BA0B7F01612DBCD26CE6DAEA73897C  NIKAFGQNANELMNQAAKRQQELKDPQLRDDLAAARAVLKKHSTMLLTASKVYVRHPELA
                                            *:: **:** **::***********************:********************

gi|17737747|ref|NP_524219.1|                LAKVNRDFILKQVCDAVNTISDVAQGKSSQPT-DIYSGAGELAAALDDFDEGIVMDPMTY
Cufflinks_92BA0B7F01612DBCD26CE6DAEA73897C  AAKANRDYVLKQVCEAVHTINDVAQGRTPQPACGPYDGPGELAAALDDFDDHMVMEPLAY
                                             **.***::*****:**.**.*****:: **:   *.* ***********: :**:*::*

gi|17737747|ref|NP_524219.1|                SEKRSRQLLEERLESIISAAALMADADCTRDERRERIVAECNAVRQALQDLLSEYMSNMS
Cufflinks_92BA0B7F01612DBCD26CE6DAEA73897C  NEVHTRPSLEERLESIISGAALMADSSCTRDERRERIVAECNAVRQALQDLLSEYMTNMG
                                            .* ::*  **********.******:.*****************************:**.

gi|17737747|ref|NP_524219.1|                QKDNSPGLSRAIDQMCRKTRDLRRQLRKAVVDHVSDSFLETTTPLLDLIEAAKSGNEKKV
Cufflinks_92BA0B7F01612DBCD26CE6DAEA73897C  NKDKSESLNRAIDNMGRKTRDLRRQLRKAVVDHVSDSFLETNVPLLVLIKAAQNGNEKEV
                                            :**:* .*.****:* ************************..*** **:**:.****:*

gi|17737747|ref|NP_524219.1|                REKSEIFTKHAEKLVEVANLVCSMSNNEDGVKMVRYAAAQIESLCPQVINAASILTVRPN
Cufflinks_92BA0B7F01612DBCD26CE6DAEA73897C  EEYAVVFTEHSNKLVEVANLVCSMSNNEDGVKMVRYAAAQIENLCPEVINAARILAARPR
                                            .* : :**:*::*******************************.***:***** **:.**.

gi|17737747|ref|NP_524219.1|                SKVAQENMTTYRQAWEVQVRILTEAVDDITTIDDFLAVSENHILEDVNKCVMALQVGDAR
Cufflinks_92BA0B7F01612DBCD26CE6DAEA73897C  SKVAQENMAAFKQSWENHVRILTEAVDDITTIDDFLAVSENHILEDVNKCVLALQEGDAD
                                            ********::::*:** :*********************************:*** ***

gi|17737747|ref|NP_524219.1|                DLRATAGAIQGRSSRVCNVVEAEMDNYEPCIYTKRVLEAVKVLRDQVMMKFDQRVGAAVG
Cufflinks_92BA0B7F01612DBCD26CE6DAEA73897C  TLDRTASGIRGRSNRVCNVVSAEMDNYEPCIYTKRVLEAVKVLNDQVMPKFTQRVQVAVQ
                                             *  **..*:***.******.**********************.**** ** *** .**

gi|17737747|ref|NP_524219.1|                ALSNNSNKDVDENDFIDASRLVYDGVREIRRAVLMNRSSEDLDTDTEFEPVEDLTLETRS
Cufflinks_92BA0B7F01612DBCD26CE6DAEA73897C  ALGNNPPKEVDENDFIDASRLVYDGVREIRRAVLMNRADEDLDPE-DVELDENYTLETRS
                                            **.**  *:****************************:.**** : :.*  *: ******

gi|17737747|ref|NP_524219.1|                RSSAHTGDQTVDEYPDISGICTAREAMRKMTEEDKQKIAQQVELFRREKLTFDSEVAKWD
Cufflinks_92BA0B7F01612DBCD26CE6DAEA73897C  KSSAHTGEHGVDEYPDISGITTAREAMGKMPEEDKQKILQQVEFFRSEKLKFDREVAKWD
                                            :******:: ********** ****** ** ******* ****:** ***.** ******

gi|17737747|ref|NP_524219.1|                DTGNDIIFLAKHMCMIMMEMTDFTRGRGPLKTTMDVINAAKKISEAGTKLDKLTREIAEQ
Cufflinks_92BA0B7F01612DBCD26CE6DAEA73897C  DTGNDIIVLAKHMCMIMMEMTDFTRGRGPLKTTMDVINAAKKISEAGTKLDKLTRQIAEQ
                                            *******.**************************************************:****

gi|17737747|ref|NP_524219.1|                CPESSTKKDLLAYLQRIALYCHQIQITSKVKADVQNISGELIVSGLDSATSLIQAAKNLM
Cufflinks_92BA0B7F01612DBCD26CE6DAEA73897C  CPESSTKQDLLAYLQRIALYCHQMNITSKVKADVQNISGELIVSGLDSATSLIQAAKNLM
                                            *******:***************::**********************************

gi|17737747|ref|NP_524219.1|                NAVVLTVKYSYVASTKYTRQGTVSSPIVVWKMKAPEKKPLVRPEKPEEVRAKVRKGSQKK
Cufflinks_92BA0B7F01612DBCD26CE6DAEA73897C  NAVVLTVKSSYVASTKYPRQGTIAV-----------------------------------
                                            ******** ******** ****::

gi|17737747|ref|NP_524219.1|                VQNPIHALSEFQSPADAV
Cufflinks_92BA0B7F01612DBCD26CE6DAEA73897C  ------------------
```

LDEC_clustal_3

```
CLUSTAL O(1.2.1) multiple sequence alignment


B_imp-gi|350397044|ref|XP_003484750.1|          MSDHFGPITLKWDPKNLEIRTMSVEKTLEPLVLQVTTLVNTKGPSKKKKGKSKRASALVG
T_cas-gi|91076138|ref|XP_970159.1|              MNDHFG---LKWDPKNLEIRTMSVEKTLEPLVLQVTTLVNTKGPSKKKKGKSKRANALVS
cufflinks-362CFAFB1F42C3D289AF73E188844BA8      MTDRFG---LKWDPKNLEIRTMSVEKTLEPLVLQVTTLVNTKGPSKKKKGKSKRANALVS
                                                *.*:**   **********************************************.***.

B_imp-gi|350397044|ref|XP_003484750.1|          TVEKATSNFIEKGEQIAYENPDITAEMLSAVEEVKKTGAAMSIAAREFSEDPCSSLKRGN
T_cas-gi|91076138|ref|XP_970159.1|              TVEKATENFIEKGEQIAYENPDITQEMLSAVEEVRKTGNAMSIAAREFSEDPCSSLKRGN
cufflinks-362CFAFB1F42C3D289AF73E188844BA8      TVEKATENFIEKGEQIAYENPDITDEMLAAVEEVKKTGTAMSVAAREFSEDPCSSLKRGN
                                                ******.***************** ***:*****:*** ***:*****************

B_imp-gi|350397044|ref|XP_003484750.1|          MVRAARNLLSAVTRLLILADMVDVHLLLKSLHVVEDDLKKLKNASSQGELLENIKQFGRN
T_cas-gi|91076138|ref|XP_970159.1|              MVRAARNLLSAVTRLLILADMVDVHLLLKSLHVVENDIEKLKNASSQGELLDNIKAFGQN
cufflinks-362CFAFB1F42C3D289AF73E188844BA8      MVRAARNLLSAVTRLLILADMVDVHLLLKSLHVVEDDLEKLKNASSHGELLDNIKAFGQN
                                                ***********************************:*::*******:****:*** **:*

B_imp-gi|350397044|ref|XP_003484750.1|          ASELMNQAAKRQQELKDPQLRDDLAAARAVLKKHSTMLLTASKVYVRHPELAAAKANRDY
T_cas-gi|91076138|ref|XP_970159.1|              ANELMNQAAKRQQELKDPQLRDDLAAARAVLKKHSTMLLTASKVYVRHPELAAAKANRDY
cufflinks-362CFAFB1F42C3D289AF73E188844BA8      ANELMNQAAKRQQELKDPQLRDDLAAARAVLKKHSTMLLTASKVYVRHPELAAAKANRDY
                                                *.************************************************************

B_imp-gi|350397044|ref|XP_003484750.1|          VLKQVCEAVNTINDVAQGKTPVDSQHPYEGPGELAAALDDFDERMVMSPLAYNEVRTRPS
T_cas-gi|91076138|ref|XP_970159.1|              VLKQVCEAVNTINDVAQGRTPQPATGPYDGPGELAAALDDFDDHMVMEPLAYNEVHTRPS
cufflinks-362CFAFB1F42C3D289AF73E188844BA8      VLKQVCEAVHTINDVAQGRTPQPACGPYDGPGELAAALDDFDDHMVMEPLAYNEVHTRPS
                                                *********.********:**  :  **:*************::***.*******:****

B_imp-gi|350397044|ref|XP_003484750.1|          LEERLESIISGAALMADSSCTRDERRERIVAECNAVRQALQDLLSEYMNNMGVKEQSEGL
T_cas-gi|91076138|ref|XP_970159.1|              LEERLESIISGAALMADSSCTRDERRERIVAECNAVRQALQDLLSEYMSNIGNKEKSESL
cufflinks-362CFAFB1F42C3D289AF73E188844BA8      LEERLESIISGAALMADSSCTRDERRERIVAECNAVRQALQDLLSEYMTNMGNKDKSESL
                                                ************************************************.*:* *::**.*

B_imp-gi|350397044|ref|XP_003484750.1|          ERAIDHMCRKTRDLRRQLRKAVVDHVSDSFLETSVPLLVLIEAARNGRDKEVEEYALVFT
T_cas-gi|91076138|ref|XP_970159.1|              NRAIDNMGRKTRDLRRQLRKAVVDHVSDSFLETNVPLLVLIKAAQNGNEKEVEEFAVMFT
cufflinks-362CFAFB1F42C3D289AF73E188844BA8      NRAIDNMGRKTRDLRRQLRKAVVDHVSDSFLETNVPLLVLIKAAQNGNEKEVEEYAVVFT
                                                :****.* *************************.*******:**:**.:*****:*::**

B_imp-gi|350397044|ref|XP_003484750.1|          EHANKLVEVANLVCSMSGNEDGVKMVRYAAAQIGNLCPQVINAARVLAARNRSKVALDNM
T_cas-gi|91076138|ref|XP_970159.1|              EHSNKLVEVANLVCSMSNNEDGVKMVRYAAAQIDNLCPEVINAARILAARPRSKVAQENM
cufflinks-362CFAFB1F42C3D289AF73E188844BA8      EHSNKLVEVANLVCSMSNNEDGVKMVRYAAAQIENLCPEVINAARILAARPRSKVAQENM
                                                **:************** *************** ****:******:**** ***** :**

B_imp-gi|350397044|ref|XP_003484750.1|          EVFRQAWENQVRVLTEAVDDITTIDDFLAVSENHILEDVNKCVLALQEGDADTLDRTAGA
T_cas-gi|91076138|ref|XP_970159.1|              DAFKQSWENQVRILTEAVDDITTIDDFLAVSENHILEDVNKCVLALQEGDADTLDRTAGG
cufflinks-362CFAFB1F42C3D289AF73E188844BA8      AAFKQSWENHVRILTEAVDDITTIDDFLAVSENHILEDVNKCVLALQEGDADTLDRTASG
                                                 .*:*:***:**:*********************************************..

B_imp-gi|350397044|ref|XP_003484750.1|          IRGRSARVCNVVQAEMDNYEPCIYTKRVLEAVKVLREQVMPKFAQRVEVAVDALGSNPAK
T_cas-gi|91076138|ref|XP_970159.1|              IRGRSNRVCNVVTAEMDNYEPCIYTKRVLEAVKVLNDQVMPKFSQRVQVAVQALSSVPTK
cufflinks-362CFAFB1F42C3D289AF73E188844BA8      IRGRSNRVCNVVSAEMDNYEPCIYTKRVLEAVKVLNDQVMPKFTQRVQVAVQALGNNPPK
                                                ***** ****** **********************.******:***:***:**.. * *

B_imp-gi|350397044|ref|XP_003484750.1|          DVDENDFIDASRLVYDGVREIRRAVLMNRADEDLDPEDVELDEHYTLETRSKSSAQTGEH
T_cas-gi|91076138|ref|XP_970159.1|              EVDENDFIDASRLVYDGVREIRRAVLMNRADEDLDPEDVELDENYTLETRSKSSAHTGEH
cufflinks-362CFAFB1F42C3D289AF73E188844BA8      EVDENDFIDASRLVYDGVREIRRAVLMNRADEDLDPEDVELDENYTLETRSKSSAHTGEH
                                                :****************************************.***********:****

B_imp-gi|350397044|ref|XP_003484750.1|          GVDEYPEISGITTAREAMRKMPEEDKQKILQQVEYFKSEKLKFDKEVAKWDDAGNDIIVL
T_cas-gi|91076138|ref|XP_970159.1|              GVDEYPDISGITTAREAMGKMPEEDKQKILQQVEYFRSEKLKFDREVAKWDDTGNDIIVL
cufflinks-362CFAFB1F42C3D289AF73E188844BA8      GVDEYPDISGITTAREAMGKMPEEDKQKILQQVEFFRSEKLKFDREVAKWDDTGNDIIVL
                                                ******:*********** ***************:*:*******:*******:*******

B_imp-gi|350397044|ref|XP_003484750.1|          AKHMCMIMMEMTDFTRGRGPLKTTMDVINAAKKISEAGTKLDKLTRQIADQCPESSTKKD
T_cas-gi|91076138|ref|XP_970159.1|              AKHMCMIMMEMTDFTRGRGPLKTTMDVINAAKKISEAGTKLDKLTRQIAEQCPESSTKKD
cufflinks-362CFAFB1F42C3D289AF73E188844BA8      AKHMCMIMMEMTDFTRGRGPLKTTMDVINAAKKISEAGTKLDKLTRQIAEQCPESSTKQD
                                                ************************************************:********:*

B_imp-gi|350397044|ref|XP_003484750.1|          LLAYLQRIALYCHQMNITSKVKADVQNISGELIVSGLDSATSLIQAAKNLMNAVVLTVKA
T_cas-gi|91076138|ref|XP_970159.1|              LLAYLQRIALYCHQMNITSKVKADVQNISGELIVSGLDSATSLIQAAKNLMNAVVLTVKA
cufflinks-362CFAFB1F42C3D289AF73E188844BA8      LLAYLQRIALYCHQMNITSKVKADVQNISGELIVSGLDSATSLIQAAKNLMNAVVLTVKS
                                                ************************************************************:

B_imp-gi|350397044|ref|XP_003484750.1|          SYVASTKYPRQSTVTSPIVVWKMKAP----------------EKKPLVRPERPEEVRAKV
T_cas-gi|91076138|ref|XP_970159.1|              SYVASTKYPRQGTISSPIVVWKMKAP----------------EKKPLVRPEKPEEVRAKV
cufflinks-362CFAFB1F42C3D289AF73E188844BA8      SYVASTKYPRQGTIAVSDFSILKQIMFEHICQRFECMKICYEFGVPLS-KQVPTDLK---
                                                ***********.*::   .    :                    **   : * :::

B_imp-gi|350397044|ref|XP_003484750.1|          RKGSQKKVQNPIHALSEFQSPTESV
T_cas-gi|91076138|ref|XP_970159.1|              RKGSQKKVQNPIHALSEFQSPTESI
cufflinks-362CFAFB1F42C3D289AF73E188844BA8      ------------------------
```