

Additional file 2. Dataset 2. Coding sequences for selected candidate protective antigens.

Sequence ID: F1

>gi|241896700|emb|FN421429.1| Ctenocephalides felis argk1 gene for arginine kinase 1
ATGGTTGACGCAGCCGTCTTGGACAAATTGGAATCTGGCTACGCCAAATTGGCAGCCTCAGATTCCAAAT
CCCTTTTGAAGAAACATTTAACAAGGAAATCTTTGACAACCTCAAACCAAGAAGACTTCCTTTGGCTC
TACCTTGCTTGACTGCATCCAGTCTGGTAAGATCTAGGAACTAACAATAATTTTCATTGATATTATGAT
ATTATAATAAGCTCATGTCTCATTTTTGTATTATTTAGGTTTGGAAAACCATGATTACAGGTGTAGGTATC
TATGCCCCAGATGCTGAAGCTTACTCCGTTTTTGTGATCTGTTTCGACCCGATCATTGAAGAATATCACT
TCGGTTTTCAAGAAAACCTGACAAGCACCCCCCAAGAAGTGGGGTGATGTAGATACTTTGGGCAACTTGG
CCCCACTGGTGAATTTGTTGTATCCACCCGTGTCCGTTGTGGACGTTCCATGGAAGGCTATCCATTCAAC
CCTTGCTTGACTGAAGCCCAATACAAGGAAATGGAAGAAAAAGTATCCAGCACACTTTCTGGATTGGAAG
GAGAATTGAAAGGAAAATTCACCCATTGACTGGTATGCCAAAAGATGTTTCAGCAACAGCTCATTGATGA
TCACTTCTTATTCAAGGAGGGCGATCGTTTTCTTGCAATCTGCTAATGCTTGCCGTTTTCTGGCCAACCTGGA
CGTGGTATCTACAAATGATAACAAAACCTTTCTCGTCTGGTGAATGAGGAAGATCACTTGAGAATCA
TCTCCATGCAAAATGGGAGGAGATTTAGGCTGTGTGTACCGTCTGTTGAAATGCTGTGAATGATATTGA
GAAGCGTATTCTTTCTCTCACCACGATAGATTAGGTTTCTCACTTTCTGTCCCACTAACTTGGGAACT
ACTGTCCGTGCCTCTGTGCACATCAAGGTCCTAAGTTGGCTGCTAACCGCGCCAAGTTGGAAGAAGTTG
CTGGACGTTACAATCTCCAAGTTCGTGGAACCTCGTGGTGAACACACTGAAGCTGAAGGTGGTGTCTACGA
TATCTCCAACAAGAGGCGCATGGGCCTGACTGAATACCAAGCTGTCAAGGAGATGCACGATGGCATTGCT
GAACTCATTAAGATGGAGAAAGAAATGTAA

Sequence of the primers used for cloning coding regions by RT-PCR:
Forward: 5'-CACCATGGTTGACGCAGCCGTCTT-3'
Reverse: 5'-TTACATTTCTTTCTCCATCTTA-3'

Sequence ID: F2

>gb|S66043.1|:122-3235 sodium pump alpha subunit [Ctenocephalides felis=cat fleas, mRNA, 3369 nt]
ATGGATGATAAGCATGGGCGTTCCGATTTCGTATCGCGTGGCTACAGTACCTACCATAGATGACAATTTGA
CAGCAGACGGTCAATACAAGTCGCGACGTAACGCAACGAAAAAGCAAAGGAAGGAGAGCTTGA
TGACTTGAACAAGAATTAGATATCGATTTTACAAAAGTATCACCCGAAGAATTATCAACGATTTAAT
ACTCACCCCGAAAATGGTCTTAGTACAGCCAAAGCGAAAAGAAAACCTAGAAAGAGATGGACCGAATGCTT
TGACCCCGCCGAAAACAAACACCAGAATGGGTCAAATTTTGAAGAACTTGTTTGGAGGATTGCGCTTGTT
GTTGTGGATCGGTGCCATTTTATGTTTTGTGCGATACTCCATCCAAGCTAGTACTGTGGAAGAACCAGCA
GATGATAATTTGTATCTTGGTATTGTATTGGCGGCTGTAGTTATAGTTACTGGTATATTTTCGTATTACC
AAGAATCCAAGAGTTCCAAAATTATGGAAAGTTTCAAAAACATGGTTCCACAGTTTGCAACAGTATTACG
TGAAGGTGAAAAATTAACCTTACGAGCGGAAGATTTAGTACTCGGAGATGTCGTAGAAGTGAATTCGGC
AGCAGGATACCTGCAGATATCCGTATTATTGAAAGCCGAGGATTCAAGGTAGACAAGTCTTCCTTGACTG
GTGAATCAGAACCTCAATCTCGAGGTCCCAGTTTTACAAATGAAAAGCCTTTAGAAACGAAGAAGTGGC
GTTCTTCTCTATCAACGCCGTGCAAGGTACTGCCAAAGGTGTGCTTATCAGCTGTGGAGACAACACTGTC
ATGGGTGCTATTGCCGGCTTGGCTTCAGGCTTGGACACTGGGGAGACTCCAATCGCTAAAGAAATTCATC
ACTTCATTCATCTCATCACTGGAGTCGCTGATTTTTAGGTGTAACATTCTTTGTTATTGCAATATTTTT
GAACTACCATTGGTTAGACGCTGTATCTTTGATTGGTATCATCGTCGCTAATGTCCCTGAAGGTTTA
TTGGCTACTGTAACCGTATGTCTAACCTTACTGCTAAGCGTATGGCATCCAAGAATTGTCTTGTCAAGA
ATCTTGAAGCTGTAGAACTCTTGGTTCTACCTCAACTATCTGCTCAGACAAAACCTGGCACACTGACCCA
GAACAGAATGACTGTAGCCACATGTGGTTTGAACAACAGATTATTGAAGCCGACACCACTGAAGATCAA
TCTGGAGTCGTATATGACAGAACCAGCCCTGGTTTTCAAAGCTTTGGCGCGCATTGCAACTTTGTGCAACA
GAGCAGAATTAAGGGAGGTGAGGAAGGTGATCCCATCTTGAAGAAAGTCAAGTGGTGTATGCATCTGA
AGCTGCTCTTCTCAAATGTATGGAAGTGGCTTTAGGAGATGTTATGTCTATTGAAAACGAAATAAGAAA
GTCTGTGAAATTCATTTAACTCCACAAAACAAATACCAGGTTTTCCATTACGAAAACCTGAAGATGCGTCA
ACCCCGTCATGTAATGGTTATGAAAGGAGCTCCTGAAAGAATCTTAGAAAATGTTCCACCATCTTCAT

TGGAGGAAAGGAAAACTACTGGACGAAGAGATGAAAGAAGCTTTCAATAATGCATATCTGGAATTGGGC
GGTCTTGGAGAGCGTGTATTGGGCTTTTGTGATCTCATGTTGCCTACAGACAAATTCCCCTTAGGTTTCA
AATTGCAGACGCGATGATCCCAACTTCCCAATTGAAAACCTTAGATTTGTTGGACTCATGTCTATGATTGA
TCCTCCTAGAGCTGCCGTACCTGACGCTGTTGCCAAGTGCCGATCTGCTGGTATCAAGGTTATCATGGTT
ACAGGAGATCATCCAATCACTGCAAAAGCCATTGCTAAATCAGTGGGTATCATCTCAGAGGGTAATGAAA
CTGTAGAAGATATCGCGCAAAGATTGAATATTCCTGTATCAGAGGTAAATCCACGAGAAGCCAAGGCAGC
TGTTGTACATGGAAGCTGAGCTTAGGGAAGCTCAACTCTGATCAGCTCGATGAAATTTCTTAGGTATCACACT
GAAATTGTATTTGCTCGGACATCTCCTCAACAAAAGCTGATTATTGTTGAAGGATGCCAACGTATGGGTG
CTATTGTCGCCGTAAGTGGTGTGATGGTGTGAATGACTCACCTGCTTTGAAAAAGGCTGATATTGGTGTTC
CATGGGTATTGCCGGATCTGATGTATCAAAACAGGCTGCTGACATGATTTTATTAGATGACAACCTTTGCA
TCTATTGTCACTGGTGTGGAAGAGGGTCTGTTGATATTCGACAATCTGAAGAAATCTATTGCTTACACAT
TGACTTCAAATATCCAGAAATTTACCATTCTTGGCATTTCATCTTATGTGATATCCCGCTACCTTTGGG
AACTGTAACAATCTTGTGCATTGACTTGGGAAGTACATGGTGCCTGCCATCTCATTGGCCTACGAACAT
GCTGAAGCTGATATCATGAAGAGGCCGCTAGAGATCCAGTCAATGACAAACTTGTAATTTCCAGACTTA
TCTCTATGGCTTATGGGCAAATCGGAATGATTCAAGCAGCTGCTGGATTCTTTGTATACTTTGTAATCAT
GGCTGAAAATGGATTCTTACCCATGAAATTTGTTTGGAAATTAGAAAACAATGGGACTCGAAAGCTGTCAAT
GATCTAACAGATTCTTATGGACAAGAATGGACATACAGAGATCGCAAGACTCTTGAATATACCTGCCACA
CTGCCTTCTTCGTCTCTATTGTCGTTGTACAATGGGCTGATTTGATTGTCTGCAAAACGCGCCGTAATTC
CTTGTACACCAGGGAATGAGAAATTTGGGCTCTCAACTTTGGTCTAGTTTTTGAAGTGCCTTAGCAGCA
TTCTGTATACACACCAGGAATGGACAAGGGACTGAGGATGTTCCCACTGAAGTTTTGTTTGGTGGCTGC
CTGCTCTGCCATTCATGATTTCCATCTTCATCTATGATGAGACTAGAAGATTTTACCTACGTGCAACCC
TGGTGGTTGGTTAGAACAAGAAACATATTATTA

Sequence of the primers used for cloning coding regions by RT-PCR:

Forward: 5'-CACCATGGATGATAAGCATGGGCGTT-3'

Reverse: 5'-TTAATAATATGTTTCTTGTCTAACC-3'

Sequence ID: F3

>gi|25527510|gb|AY150533.1| Ctenocephalides felis serpin 4 mRNA,
partial cds

ATGGCGGATCCCCAGGAATTGTCTACAAGTATTAACCAGTTTGGCTGGAAGCCTGTACAATACGGTTGCTTC
TGGCAACAAAGACAATCTCATCATGTCCCCATTGTCTGTACAACTGTTCTATCCCTGGTGTCAATGGGAG
CTGGTGGTAATACTGCCACACAAATAGCTGCTGGTTTACGTGAGCCTCAATCAAAAGAAAAAATTCAGAT
GACTACCATGCATTGATGAACACTCTTAATACACAAAAAGGTGTAAGTCTGGAAATTGCCAACAAAGTTTA
CGTTATGGAAGGCTATACATTGAAACCCACCTTCAAAGAAGTTGCCACCAACAAATTCCTTAGCTGGAGCAG
AAAACCTGAACTTTGCCAAAATGCTGAAAGCGCTAAAGTTATCAACACTTGGGTTGAAGAAAAAATCAT
GACAAAATTCATGATTTGATCAAAGCCGGTGTCTAGACCAGGATTCAAGAATGGTTCTTGTCAATGCATT
GTACTTCAAGGGTCTTTGGGAGAAACAATTCAAGAAGGAAAACTCAAGACAAACCTTTCTATGTTACTG
AAACAGAGACAAAGAATGTACGAATGATGCACATTAAGGATAAATTCGGTTATGGAGAATTTGAAGAATTA
GATGCCAAGGCTGTAGAATTGCCCTACAGGAAGTCAAGATTTGGCCATGTTAATCATTTTTGCCAAACAGCAA
AACTGGTCTCCCCGCTCTTGAAGAAAAATTACAAAATGTTGACTTGCAAAACTTGACTCAACGCATGTACT
CTGTTGAAGTTATTTGGATCTGCCTAAATTCAGATTGAATCTGAAATTAATTTGAATGATCCTCTGAAA
AAGTTGGGTATGTCTGATATGTTTGTTCCTGGAAAAGCTGATTTCAAAGGATTGCTTGAAGGATCTGATGA
GATGTTATATATTTCTAAAGTAATTCAAAAAGCTTTCATTGAAGTAAATGAAGAAGGTGCTGAAGCTGCAG
CTGCCACAGGCGTGTATGTTAATGATGCGTTGTATGCCAATGATGCCAATGGCCTTCAATGCTGAGCATCCA
TTCCTGTACTTCTTACACAGCAAAAATTCGTTCTATTCAATGGTTCGTCTTGTAAACCAACAACCTGAATA
A

Sequence of the primers used for cloning coding regions by RT-PCR:

Forward: 5'-CACCATGGCGGATCCCCAGGAATTGT-3'

Reverse: 5'-TTATTCAGTTGTTGGTTTAACT-3'

Sequence ID: F4

>gi|20502971|gb|AF503908.1| Ctenocephalides felis juvenile hormone
epoxide hydrolase mRNA, complete cds

ATGGGTAAATGTTGTCGTATGCTCATTGCTGCAATAGCTGGTATAGCAGTGTGATTACCAAATTAC
GAAAGAAGCTTCCAAAACCAAATATACCTCTCGACACTTGGTGGGGTCCAGGAAAGCCACAAAATGTTGACA
TATCAATAAGACCGTTTTAAATTAATATTAATAACAAAGTGATTGAAAATCTGAAACTAAAACCAATGAT
GTTCAATATACTTTACCTTTAGAGGGCATCAATTTTGAATATGGTTTTCAATACAGATTCCCTGAAAAAGAT
TGTAGATTTTTGGCGAACTCAATATAATTGGCGTGAACGTGAAGCATTATTAATAAATATCCACACTTCA
AAACAAATATTCAAGGCCTGGATATTCATTATGTCCACATAAAAACCACAGGTCTCTAAAAATATTGAAGTT
TTGCCTTTGGTAATGATCCATGGTTGGCCAGGATCTTTTGTGGAATTCTACAAGATCATACTATGTTGAC
AACTCCAAGAGCAGGTTACAATTTTCGTATTTCGAATTGATATTGCCTAGTATTCCCTGGATACGGCTTTTAC
AGGCTGCAGCTAAACCTGGTCTTGGATCGACTCAGGTCGCCGTTATAATGCGCAATTTGATGGAGCGCATT
GGATCAAAAAATATTATGTACAAGGAGGCGACTGGGGTTCTATGATCATAAGCGCGATGTCAACATTATT
TCCAGAAAATGTCCTGGGACAGCACTCAAACATGTGTTTTGTCAATACTCCATCATCAAATATCAAGGCTA
TAATTGGAAGCTTTTTCCCGGAATCGTTTTGCTGGCAGCGGAAATGCGCATAAAAATGTATCCCATGAGTGAA
CACTTTTTTACACTTTTTGGAAGAAATGGGTTATTTGCATCTACAAGCTACCAAACCAGATACAGTGGGCGT
TGCTTTAAGAGATTACCCAGCTGGTTTTAGCAGCTTATATTTTGGAGAAAATTTTCAACATGGACTAACAGAT
CTTGGAGGTCAGTTAAAGATGGAACTTGCTGTTAAAATACAATATTCCTGAACTTTTAGACAATGTCATG
ATATACTACGTTACTGATTCCATTACTACTTCAATGAGATTATATGCAGAATCATTACAAAAGCACACCT
TGCTTTGAACTTAGATAGGGTGCACAATCATGTCCCAGCAGCCTGCGCAAAAATTTCCAAACGAGTTGGCTT
ATGTGACCGATTGCCAATTTGCTGAGAAAATATAAACTTTATTGCAGTCCAATGACATGCCAAGTGGTGGC
CATTTTGCAGCATTTGAGGAACCTGGTCTTTTAGCAGAAGACATTTTCACTGCGGTGAAAAAGTTTAAAGA
ATTTTATTCCAAAAAGCTGAAAGCCAAAAGAAAGCTGATTTGTGA

Sequence of the primers used for cloning coding regions by RT-PCR:

Forward: 5'-CACCATGGGTAAATGTTGTCGTATGC-3'

Reverse: 5'-TCACAAATCAGCTTTCTTTTGG-3'

Sequence ID: F5

>De_novo_sequence_assembly_of_Locus_4025_Transcript_2/7|Locus_18073_Tra
nscript_1/1|Locus_26648_Transcript_1/1|Q7KVA1|Xylosyltransferase_oxt
ATGAAATGTCCTGGAGATCCCAAAGAAATTTGTGGTGGCTTTTTTAAATGAACATATATGAAACAGGAAT
TTCAAATTTCTCACCTCAAACGAGTAACTTTAACAACAAATGATACAGAACCTAAAGTACGCATTGTAT
ATTTGCTGACTTTAAATGGACGAGCAGTACGTCAAGTTTACAGATTAATTAAGCTCTATTTTCATGTAGAT
CATTATTTTTATATTACGTAGACTCTCGCCAAGATTACATGTTTCGTCAACTTTTGCAATTAGAAAATAA
CTTTCCAAATATACGTCTTTCTCGCAATCGTTTTTCTACTATTTGGGGAGGTGCATCTTTACTCCAAATGT
TGCTAGCTTCTATGAATGAACTGTTAAAATCTGATTGGAAGTGGGATTTTATTATAAATTTAAGTGAAGT
GATTTTCCAATTAATCAAATGAAAGACTTGTGCAATTTTAAACAGCAAATAAAAAATAAAACTTTGTGAA
ATCACATGGGCGAGAAGTACAAAGATTTATTACAGAAGCAAGGCTTAGACAAAACCTTTCATTGAATGTGATA
CGCATATGTGGAGAAGTGGTATAGAACCTTGCCATGGGGTATTCAAATTTGATGGCGGTAGTGATTGGATA
ACTCTGACTAGAGACTTTGTGAAATATATCACACAAGAACAAGATGCCCTTGTCAAGGGCCTAATGACTGT
TTTTAGGCATACTTTATTACCTGCTGAATCATTTTTCCATACAGTTTTACGCAACTC

Sequence ID: F6

>De_novo_sequence_assembly_of_Locus_4149_Transcript_2/4_Confidence_0.70
0_Length_2043|Q9VAF0|
ATGAATTCCTCAATTTGGGAGTCAGACTCAGCCTTCAGATTTATTTACACTGATTTAGTTATGGAATACAT
ACCAGAATATTTCAAAGACTGGCAATATCATCCATGGATTTTTTTCGGTCTTAGGTTCTATTTTAAATTGGAT
TAAGTGGTGTGTTGCCTTTACTTATAATTCCATTGGAAGGAGGAGAACATTTTTAAAGTGGAGGAGGGGGC
AAACTTTTTGCGCGTATTGCTCAGTTTTGTCAGTAGGAGGCTGCTCGGAGATGTGTTTTTGCATTTGTTGCC
AGAAGCATGGGAGAAGTACAGTAACAAAGGAGAGCGACTCGTAGTCATGTACGAATTCGTAGATTCCAGGAG
GAGGTAGTCCATCAATGACATGCGGGTGTGGGATTAGCTGGCCTCTTGTATTTCACGATGGTTCGAGAAG
TTATTTGCTGGTTACATGGAAGATGATGACGTTTCCGAGGAAAAGACCGAACGAAAGCGCAAGAAAATGTC
TAAAATTAATTTGTTTATAAAGAACAACAATGTCTTGAGGGCTAGAAAACGAAAATGGACGATTCTCGACCA
GCAAATGCTCAGGCGGACAAAATGGAAAATGTGTGACATTGAAGTTATAAATGGAGCTATCAAAAACGGC
TGTTGTATGGTTCGAGGAGCTAGAGAAAAGAGGCTGTGAGCAACAAGTCTAAGCATGTTTCTGGATACTTAAA
TCTCATGGCAAACCAATTGATAATTTACACACGGTCTTGCTGTTGGTGGTCTTTTCATTGTTTCATTCA
AACTAGGAATTTTAAACAACATTTCGAATTTCTAGTGCATGAGATTCCACACGAAGTGGGCGATTTTGCATTT

TTGTTGCGGTCAGGTTTCTCGCGGTGGGACGCGGCTCGAGCGCAACTCATGACCGCTGGCGCCGGCCTGCT
TGGAGCTATGACCGCAATCGGCGGAAGTCGGGCTACGAGTGCAATGGAAGCTAGAACATGCTGGATAATGC
CCTTCACAGCCGGCGGATTTCTCCACATAAGTTTAGTTACAGTGCTACCTGACTTGCTGCAAGAAAGGGAT
CCAAAAGAGTCGATTAAACAATTTGGAGCATTGATTGGTGGTATAGTTCTAATGGCATCATTGACCACATT
ATTGGAGTGA

Additional file 2. Dataset 3. De novo sequence assembly process for sequences F5 and F6.

De novo sequence assembly for sequence F5: *Drosophila melanogaster* Xylosyltransferase (oxt) Q7KVA1.

1. Protein oriented transcripts for Q7KVA1.

>Locus_167_Transcript_1/1_Confidence_1.000_Length_302|Q7KVA1|Xylosyltransferase_oxt

TTAACAGCAAATAAAAAATAAAAACTTTGTGAAATCACATGGGCGAGAAGTACAAAGATTT
ATTCAGAAGCAAGGCTTAGACAAAACCTTCATTGAATGTGATACGCATATGTGGAGAACA
GGTGATAGAACATTGCCATGGGGTATTCAAATTGATGGCGGTAGTGATTGGATAACTCTG
ACTAGAGAGTTTGTGAAATATATCACACAAGAACAAGATGCCCTTGTCAAGGGCCTAATG
ACTGTTTTTAGGCATACTTTATTACCTGCTGAATCATTTTTCCATACAGTTTTACGCAAC
TC

>Locus_4025_Transcript_1/7_Confidence_0.188_Length_245|Q7KVA1|Xylosyltransferase_oxt

GAACCTAAAGTGCGCATTGTATATTTGCTAACTTTAAATGGGCGAGCAGTACGTCAAGTT
TACAGATTAATTAAGCTCTATTTTCATGTAGATCATTATTTTTATATTCACGTAGACTCT
CGCCAAGATTACATGTTTCGTCAACTTTTGCAATTAGAAAATAACTTTCCAAATATACGT
CTTTCTCGCAATCGTTTTTCTACTATTTGGGGAGGTGCATCTTTACTCCAAATGTTGCTA
GCTTC

>Locus_4025_Transcript_2/7_Confidence_0.312_Length_368|Q7KVA1|Xylosyltransferase_oxt

AATATGAAATGTCCTGGAGATCCCAAAGAAATTTGTGGTGGCTTTTTTAAAATGAACATA
TATGAAACAGGAATTTCAAATTTCTCACCTCAAACCTGCAGTAACTTTAACAACAAATGAT
ACAGAACCTAAAGTACGCATTGTATATTTGCTGACTTTAAATGGACGAGCAGTACGTCAA
GTTTACAGATTAATTAAGCTCTATTTTCATGTAGATCATTATTTTTATATTCACGTAGAC
TCTCGCCAAGATTACATGTTTCGTCAACTTTTGCAATTAGAAAATAACTTTCCAAATATA
CGTCTTTCTCGCAATCGTTTTTCTACTATTTGGGGAGGTGCATCTTTACTCCAAATGTTG
CTAGCTTC

>Locus_8527_Transcript_1/1_Confidence_1.000_Length_299|Q7KVA1|Xylosyltransferase_oxt

TTTCCAAATATACGTCTTTCTCGCAATCGTTTTTCTACTATTTGGGGAGGTGCATCTTTA
CTCCAAATGTTGCTAGCTTCTATGAATGAACTGTTAAAATCTGATTGGAAATGGGATTTT
ATTATAAATTTAAGTGAAAGTGATTTTCCAATTAATCAAATGAAAGACTTGTGCAATTT
TTAACAGCAAATAAAAAATAAAAACTTTGTGAAATCACATGGGCGAGAAGTACAAAGATTT
ATTCAGAAGCAAGGCTTAGACAAAACCTTCATTGAATGTGATACGCATATGTGGAGAAC

>Locus_18073_Transcript_1/1_Confidence_1.000_Length_299|Q7KVA1|Xylosyltransferase_oxt

TTTCCAAATATACGTCTTTCTCGCAATCGTTTTTCTACTATTTGGGGAGGTGCATCTTTA
CTCCAAATGTTGCTAGCTTCTATGAATGAACTGTTAAAATCTGATTGGAAGTGGGATTTT
ATTATAAATTTAAGTGAAAGTGATTTTCCAATTAATCAAATGAAAGACTTGTGCAATTC
TTAACAGCAAATAAAAAATAAAAACTTTGTGAAATCACATGGGCGAGAAGTACAAAGATTT
ATTCAGAAGCAAGGCTTAGACAAAACCTTCATTGAATGTGATACGCATATGTGGAGAAC

>Locus_26648_Transcript_1/1_Confidence_1.000_Length_302|Q7KVA1|Xylosyltransferase_oxt

TTAACAGCAAATAAAAAATAAAAACTTTGTGAAATCACATGGGCGAGAAGTACAAAGATTT
ATTCAGAAGCAAGGCTTAGACAAAACCTTCATTGAATGTGATACGCATATGTGGAGAAC
GGTGATAGAACCTTGCCATGGGGTATTCAAATTGATGGCGGTAGTGATTGGATAACTCTG
ACTAGAGACTTTGTGAAATATATCACACAAGAACAAGATGCCCTTGTCAAGGGCCTAATG

ACTGTTTTTAGGCATACTTTATTACCTGCTGAATCATTTTTCCATACAGTTTTACGCAAC
TC

2. Assembly of the transcripts clustered by Q7KVA1.

- Reference protein Q7KVA1:

```
>sp|Q7KVA1|XYLT_DROME Xylosyltransferase oxt OS=Drosophila melanogaster  
GN=oxt PE=2 SV=1  
MEQSVSARWLKRYRAFFLILLIVAIQLFLAYKSLDIVGGGSGSGFDAAEAPASPPPPHA  
QARVQPPARTKLTAAQLGFQPECDILAREAISALQRAKTKDCREHIAQIACAIQAGRFYA  
PQLRSSCPAGNHTANVSLGCFKDEKDRLLAGYYSSSKTSNSPAKCVLCLQSGYPYAGV  
QYGRECFCGYDPPPKASKLPDSSCNTKCLGNAKEICGGFYAMNIYETGIAKFTAQLAATT  
PSEETKRVRIAFLLTLNLRALRQVHRLKALYAPEHVYIYHVDERQDYLYRKLLELESKF  
PNIRLARKRFSTIWGASLLTMLLQCMEDLLQSNWHWDFVINLSESDFPVKTLDKLVDL  
SANPGRNFVKGHGRETQKFIQKQGLDKTFVECDTHMWRIGDRKLPAGIQVDGGSWDVALS  
RPFVGYVTHPREDELQALLKLFRTLLPAESFFHTVLRNTHKCHTSYVDNNLHVTNWKR  
KQGCKCQYKHVVWDWCGCSPNDFKPEDWPRLQATEQKSLFFARKFEPVINQAVLLQLEEWL  
YGPYTSEYANLHGYSLYHHEDVHSGDDLARSIGDSVMRLSARQAKLYPLELIELTHY  
LHRDQYKGFVRYRARGSTGKPLHLETRVRPTQQGKLARNARFSKRLRNFEVSTDFDQKE  
QIARNFGKLLGPDLLSYTLQANADSGAASHSYNLTLWIDPLGRLQDFNELHVEDST  
SDVINHSKTLKHPITPGIWTAKLIGRNSIYAQLKFLIAPLAYYKGYPLAKSSDAEALNA  
GLTVALPEDFEMPVEWQQHLQTDDEQFTMREESLAKGKMLGQELHSWIDGLVGGFFQLRE  
SCVVEADSEVSLPLCSDAPWSSLAPDPKSDVDALLK
```

- Transcripts to be assembled (protein Q7KVA1):

```
>Locus_4025_Transcript_2/7_Confidence_0.312_Length_368|Q7KVA1|Xylosylt  
ransferase_oxt  
AATATGAAATGTCCTGGAGATCCCAAAGAAATTTGTGGTGGCTTTTTTAAAATGAACATA  
TATGAAACAGGAATTTCAAATTTCTCACCTCAAAGTGCAGTAACCTTTAACAACAAATGAT  
ACAGAACCTAAAGTACGCATTGTATATTTGCTGACTTTAAATGGACGAGCAGTACGTCAA  
GTTTACAGATTAATTAAGCTCTATTTTCATGTAGATCATTATTTTTATATTACGTAGAC  
TCTCGCCAAGATTACATGTTTCGTCAACTTTTGCAATTAGAAAATAACTTTCCAAATATA  
CGTCTTTCTCGCAATCGTTTTTCTACTATTTGGGGAGGTGCATCTTTACTCCAAATGTTG  
CTAGCTTC
```

```
>Locus_18073_Transcript_1/1_Confidence_1.000_Length_299|Q7KVA1|Xylosylt  
ransferase_oxt  
TTTCCAAATATACGTCTTTCTCGCAATCGTTTTTCTACTATTTGGGGAGGTGCATCTTTA  
CTCCAAATGTTGCTAGCTTCTATGAATGAACTGTTAAAATCTGATTGGAAGTGGGATTTT  
ATTATAAATTTAAGTGAAAGTGATTTTCCAATTAATCAAATGAAAGACTTGTGCAATTC  
TTAACAGCAAATAAAAAATAAAAACTTTGTGAAATCACATGGGCGAGAAGTACAAAGATTT  
ATTCAGAAGCAAGGCTTAGACAAAACCTTTCATTGAATGTGATACGCATATGTGGAGAAC
```

```
>Locus_26648_Transcript_1/1_Confidence_1.000_Length_302|Q7KVA1|Xylosylt  
ransferase_oxt  
TTAACAGCAAATAAAAAATAAAAACTTTGTGAAATCACATGGGCGAGAAGTACAAAGATTT  
ATTCAGAAGCAAGGCTTAGACAAAACCTTTCATTGAATGTGATACGCATATGTGGAGA  
GGTGATAGAACCTTGCCATGGGGTATTCAAATTTGATGGCGGTAGTGATTGGATAACTCTG  
ACTAGAGACTTTGTGAAATATATCACACAAGAACAAGATGCCCTTGTCAAGGGCCTAATG  
ACTGTTTTTAGGCATACTTTATTACCTGCTGAATCATTTTTCCATACAGTTTTACGCAAC  
TC
```

- Assembly of the three transcripts:

```
>Locus_4025_Transcript_2/7|Locus_18073_Transcript_1/1|Locus_26648_Transcript_1/1|Q7KVA1|Xylosyltransferase_oxt
AATATGAAATGTCCTGGAGATCCCAAAGAAATTTGTGGTGGCTTTTTTAAAATGAACATA
TATGAAACAGGAATTTCAAATTTCTCACCTCAAAGTGCAGTAACTTTAACAACAAATGAT
ACAGAACCTAAAGTACGCATTGTATATTTGCTGACTTTAAATGGACGAGCAGTACGTCAA
GTTTACAGATTAATTAAGCTCTATTTTCATGTAGATCATTATTTTTTATATTCACGTAGAC
TCTCGCCAAGATTACATGTTTCGTCAACTTTTGCAATTAGAAAATAACTTTCCAAATATA
CGTCTTTTCTCGCAATCGTTTTTCTACTATTTGGGGAGGTGCATCTTTACTCCAAATGTTG
CTAGCTTCTATGAATGAACTGTTAAAATCTGATTGGAAGTGGGATTTTATTATAAATTTA
AGTGAAGTGAATTTTCCAAATTAATCAAATGAAAGACTTGTGCAATTTTAAACAGCAAAT
AAAAATAAAAACTTTGTGAAATCACATGGGCGAGAAGTACAAAGATTTTATTCAGAAGCAA
GGCTTAGACAAAACCTTTCATTGAATGTGATACGCATATGTGGAGAACTGGTGATAGAACC
TTGCCATGGGGTATTCAAATTTGATGGCGGTAGTGATTGGATAACTCTGACTAGAGACTTT
GTGAAATATATCACACAAGAACAAGATGCCCTTGTCAAGGGCCTAATGACTGTTTTTAGG
CATACTTTATTACCTGCTGAATCATTTTTTCCATACAGTTTTTACGCAACTC
```

- Alignment of the protein encoded by assembled transcript to the reference protein Q7KVA1:

```
Query 205 NTKCLGNAKEICGGFYAMNIYETGIAKFTAQLAATTPSEETK-RVRI AFLLLTLNGRALRQ 263
      N KC G+ KEICGGF+ MNIYETGI+KF+ Q A T + +T+ +VRI +LLTLNGRA+RQ
Sbjct 1 NMKCPGDPKEICGGFFKMNIYETGISKFSQPQAVTLTTNDTEPKVRIVYLLTLNGRAVRQ 180

Query 264 VHRLLKALYAPEHVYIYHVDERQDYLYRKLLELESKFPNIRLARKRFSTIWGGASLLTML 323
      V+RL+KAL+ +H +YIHVD RQDY++R+LL+LE+ FPNIRL+R RFSTIWGGASLL ML
Sbjct 181 VYRLIKALFHDHYFYIHVDSRQDYMFRQLLQLENNFPNIRLSRNRFSSTIWGGASLLQML 360

Query 324 LQCMEDLLQSNWHWDFVINLSESDFPVKTLDKLVDFLSANPGRNFVKHGRETQKFIQKQ 383
      L M +LL+S+W WDF+INLSESDFP+K+ ++LV FL+AN +NFVK HGRE Q+FIQKQ
Sbjct 361 LASMNELLKSDWKWDFIINLSESDFPIKSNERLVQFLTANKKNFVKSHGREVQRFIQKQ 540

Query 384 GLDKTFVECDTHMWRIGDRKLPAGIQVDGGSWVALSRPFVGYVTHPREDELQALLKL 443
      GLDKTF+ECDTHMWR GDR LP GIQ+DGGSDW+ L+R FV Y+T +E D L++ L+ +
Sbjct 541 GLDKTFIECDTHMWRIGDRKLPAGIQVDGGSWVALSRPFVGYVTHPREDELQALLKL 714

Query 444 FRHTLLPAESFFHTVLRN 461
      FRHTLLPAESFFHTVLRN
Sbjct 715 FRHTLLPAESFFHTVLRN 768
```

- Translated assembled transcript for protein Q7KVA1:

```
NMKCPGDPKEICGGFFKMNIYETGISKFSQPQAVTLTTNDTEPKVRIVYLLTLNGRAVRQ
VYRLIKALFHDHYFYIHVDSRQDYMFRQLLQLENNFPNIRLSRNRFSSTIWGGASLLQML
```

LASMNELLKSDWKWDFIINLSEDFPIKSNERLVQFLTANKNKNFVKSHGREVQRFIQKQ
GLDKTFIECDTHMWRRTGDRTLPWGIQIDGGSDWITLTRDFVKYITQEQLVKGMLMTVFR
HTLLPAESFFHTVLRN

3. Sequence used for synthesis and expression.

ATGAAATGTCCTGGAGATCCCAAAGAAATTTGTGGTGGCTTTTTTAAAATGAACATATATGAAACAGG
AATTTCAAATTTCTCACCTCAAACCTGCAGTAACTTTAACAACAAATGATACAGAACCTAAAGTACGCA
TTGTATATTTGCTGACTTTAAATGGACGAGCAGTACGTCAGTTTACAGATTAATTAAGCTCTATTT
CATGTAGATCATTATTTTTATATTACGTAGACTCTCGCCAAGATTACATGTTTCGTCAACTTTTGCA
ATTAGAAAATAACTTTCCAAATATACGTCTTTCTCGCAATCGTTTTTCTACTATTTGGGGAGGTGCAT
CTTTACTCCAAATGTTGCTAGCTTCTATGAATGAACTGTTAAAATCTGATTGGAAGTGGGATTTTATT
ATAAATTTAAGTGAAAGTGATTTTCCAATTAATCAAATGAAAGACTTGTGCAATTCTTAACAGCAAA
TAAAAATAAAACTTTGTGAAATCACATGGGCGAGAAGTACAAAGATTTATTTCAGAAGCAAGGCTTAG
ACAAAACCTTTCATTGAATGTGATACGCATATGTGGAGAAGTGGTGATAGAACCTTGCCATGGGGTATT
CAAATTTGATGGCGGTAGTGATTGGATAACTCTGACTAGAGACTTTGTGAAATATATCACACAAGAACA
AGATGCCCTTGTCAAGGGCCTAATGACTGTTTTTAGGCATACTTTATTACCTGCTGAATCATTTTTCC
ATACAGTTTTACGCAACTC

De novo sequence assembly for sequence F5: *Drosophila melanogaster* Metal ion transporter Q9VAF0.

1. Protein oriented transcripts for Q9VAF0.

>Locus_4149_Transcript_1/4_Confidence_0.600_Length_1883|Q9VAF0|
CTTTCCTTCCCGTGTCTATGATAGACTAACTCCGACTTCGTATGCGCAGGCTCGACACGA
CGGACGTATTTCGCATTTCGCACGATCTACGAACCAACTGACATAGTAGTACAGCGGGTGAC
AGTCGTCGTAACCGTAGAATAATGCGTTACAATTTTCGATCTGAAGTGGACAATGACTTC
AATAATTTATAACATCAAAGCTTCTGCACGACCATAAATTGATTTTAAATCTTTTTAAATGA
TTTTCGTTAACGCTAGCTCAATCTCAGTGCAGATTAGCTAGTCGAACGTCGAATCTTCCCA
AGAAAAACGGCTTGTGTTTCTATTGGATTCAATTTGTTTTATGGATTTTGATGCAACATAA
ACGGTGTATTGTAAAAACCTGATGATAAAACAAGTATTTTTAAATAATAATTACGTTACG
TTATTTGTGATAGAGAGTGAAAGTAGTGGCAATGAATTCCTCAATTTGGGAGTCAGACTC
AGCCTTCAGATTTATTTACTGATTTAGTTATGGAATACATACCAGAATATTTCAAAGA
CTGGCAATATCATCCATGGATTTTTTCGGTCTTAGGTTCTATTTTAAATTGGATTAAGTGG
TGTGTTGCCCTTTACTTATAATTCATTGGAAGGAGGAGAACATTTTAAAAGTGGAGGAGG
GGGCAAACCTTTGCGCGTATTGCTCAGTTTTGCAGTAGGAGGTCTGCTCGGAGATGTGTT
TTTGCATTTGTTGCCAGAAGCATGGGAGAAGTACAGTAACAAAGGAGAGCGACTCGTAGT
CATGTACGAATTCGTAGATTTCAGGAGGAGGTAGTCCATCAATGACATGCGGGTGCTGGGT
ATTAGCTGGCCTCTTGTTATTCACGATGGTCGAGAAGTTATTTGCTGGTTACATGGAAGA
TGATGACGTTTCCGAGGAAAAGACCGAAGCAAGAAAGCGCAAGAAAATGTCTAAAATTAATTG
TTTGATAAAGAACAACAATGTCTTGAGGGCTAGAAAACGAAAATGGACGATTCTCGACCAG
CAAATGCTCAGGCGCAAAAATGGAAAATGTGTGACATTGAAGTTATAAATGGAGCTAT
CAAAAACGGCTGTTGTATGGTTCGAGGAGCTAGAGAAAAGAGGCTGTGAGCAACAAGTCTAA
GCATGTTTCTGGATACTTAAATCTCATGGCAAACCTCAATTGATAATTTACACACGGTCT
TGCTGTTGGTGGTTCTTTTCATTGTTTCATTCAAACCTAGGAATTTTAAACAACATTTCGCAAT
TCTAGTGCATGAGATTCCACACGAAGTGGGCGATTTTGCAATTTTGTGCGGTGAGGTTT
CTCGCGGTGGGACGCGGCTCGAGCGCAACTCATGACCGCTGGCGCCGGCCTGCTTGGAGC
TATGACCGCAATCGGCGGAAGTGGGCTACGAGTGCAATGGAAGCTAGAACATGCTGGAT
AATGCCCTTCACAGCCGGCGGATTTCTCCACATAAGTTTAGTTACAGTGCTACCTGACTT
GCTGCAAGAAAGGGATCCAAAAGAGTCGATTAACAATTTGGAGCATTGATTGGTGGTAT
AGTTCTAATGGCATCATTGACCACATTATTGGAGTGAGGCATATATTTTTCTATTTCTAA

TTAGATTTTACATATCGGTGCCATCATCACATTAAGTCACTGTGATTTGACTAATATGAT
ACAGACTTGGATGAAATAAAGTAAAAATAGGGAACCTATTTTACAAAGAAAACCTTATTGACA
TAATTAGATTAATCTATTTTGTCTGTCCGGGAGAATTACTTACTTACTTACATATTATGG
TTTTGAAAAGTAATTTCTCAAATAATGATATTTGTGCATGTAAATAGTTATATTTTTTAATA
CTTGTTATATATATATATATATA

>Locus_4149_Transcript_2/4_Confidence_0.700_Length_2043|Q9VAF0|

CTTTCCTTCCCGTGTCTATGATAGACTAACTCCGACTTCGTATGCGCAGGCTCGACACGA
CGGACGTATTTCGCATTTCGCACGATCTACGAACCAACTGACATAGTAGTACAGCGGGTGC
AGTCGTGCGTAAAACGTAGAATAATGCGTTACAATTTTCGATCTGAAGTGGACAATGACTTC
AATAATTTATAACATCAAAGCTTCTGCACGACCATAAAATGATTTTAAATCTTTTAAATGA
TTTTCGTTAACGCTAGCTCAATCTCAGTGCAGATTAGCTAGTCGAACGTCGAATCTTTCCCA
AGAAAAACGGTGCAGACTATACGACGTTTTGATTTTTTTGTTTACTGCGCGTTCTATTGGTCC
TTAAGTGACATTCTCAGATGTTTTTAATGTGATTTTTCTTGTGATATAAATGCGTGCTCAT
ATCATTTGTGTTAATTTTTGAAATTAACGATTTTTTATTTTTCCGTTCCAGGCTTGTGTTTT
TATTGGATTCAATTTGTTTTATGGATTTTTGATGCAACATAAACGGTGTATTGTA AAAACCT
GATGATAAAAACAAGTATTTTTAAATAATAATTACGTTACGTTATTTGTGATAGAGAGTGA
AAGTAGTGGCAATGAATTCCTCAATTTGGGAGTCAGACTCAGCCTTCAGATTTATTTACA
CTGATTTAGTTATGGAATACATAACCAGAATATTTCAAAGACTGGCAATATCATCCATGGA
TTTTTTCCGTTCTTAGGTTCTATTTTTAATTGGATTAAGTGGTGTGTTGCCTTTACTTATAA
TTCCATTGGAAGGAGGAGAACATTTTTAAAAGTGGAGGAGGGGGCAAACCTTTTGCAGTAT
TGCTCAGTTTTGCAGTAGGAGGTCTGCTCGGAGATGTGTTTTTGCATTTGTTGCCAGAAG
CATGGGAGAACTACAGTAACAAAGGAGAGCGACTCGTAGTCATGTACGAATTCGTAGATT
CAGGAGGAGGTAGTCCATCAATGACATGCGGGTGTGGGTATTAGCTGGCCTCTTGTTAT
TCACGATGGTTCGAGAAGTTATTTGCTGGTTACATGGAAGATGATGACGTTTCCGAGGAAA
AGACCGAACGAAAGCGCAAGAAAATGTCTAAAATTAATTGTTTGATAAAGAACAACAATG
TCTTGAGGGCTAGAAAACGAAAATGGACGATTCTCGACCAGCAAATGCTCAGGCGGACAAA
ATGGAAAAATGTGTGACATTGAAGTTATAAATGGAGCTATCAAAAACGGCTGTTGTATGG
TCGAGGAGCTAGAGAAAAGAGGCTGTGAGCAACAAGTCTAAGCATGTTTCTGGATACTTAA
ATCTCATGGCAAATCAATTTGATAAATTTACACACAGGTCTTGTCTGTTGGTGGTTCTTTCA
TTGTTTTCAATCAAACCTAGGAATTTTTAAACAACATTCGCAATTCATAGTCATGAGATTCCAC
ACGAAGTGGGCGATTTTTGCAATTTTTGTTGCGGTGAGTTTTCTCGCGGTGGGACGCGGCTC
GAGCGCAACTCATGACCGCTGGCGCCGGCCTGCTTGGAGCTATGACCGCAATCGGCGGAA
GTCGGGCTACGAGTGCAATGGAAGCTAGAACATGCTGGATAATGCCCTTCACAGCCGGCG
GATTTCTCCACATAAGTTTAGTTACAGTGCTACCTGACTTGCTGCAAGAAAGGGATCCAA
AAGAGTCGATTAACAATTTGGAGCATTGATTGGTGGTATAGTTCTAATGGCATCATTGA
CCACATTATTGGAGTGAGGCATATATTTTTCTATTTCTAATTAGATTTACATATCGGTG
CCATCATCACATTAAGTCACTGTGATTTGACTAATATGATACAGACTTGTGAAATAAAG
TAAAAATAGGGAACCTATTTTACAAAGAAAACCTTATTGACATAATTAGATTAATCTATTTG
ATCTGTCCGGGAGAATTACTTACTTACTTACATATTATGGTTTTGAAAAGTAATTTCTCAA
ATAATGATATTTGTGCATGTAAATAGTTATATTTTTTAATACTTGTATATATATATATAT
ATA

>Locus_4149_Transcript_3/4_Confidence_0.600_Length_1932|Q9VAF0|

CTTTCCTTCCCGTGTCTATGATAGACTAACTCCGACTTCGTATGCGCAGGCTCGACACGA
CGGACGTATTTCGCATTTCGCACGATCTACGAACCAACTGACATAGTAGTACAGCGGGTGC
AGTCGTGCGTAAAACGTAGAATAATGCGTTACAATTTTCGATCTGAAGTGGACAATGACTTC
AATAATTTATAACATCAAAGCTTCTGCACGACCATAAAATGATTTTAAATCTTTTAAATGA
TTTTCGTTAACGCTAGCTCAATCTCAGTGCAGATTAGCTAGTCGAACGTCGAATCTTTCCCA
AGAAAAACGGTGCAGACTATACGACGTTTTGATTTTTTTGTTTACTGCGCGTTCTATTGGTCC
TTAAGTGACATTCTCAGATGTTTTTAATGCTTGTGTTTCTATTGGATTCATTTGTTTTAT
GGATTTTTGATGCAACATAAACGGTGTATTGTA AAAACCTGATGATAAAAACAAGTATTTTT
AAATAATAATTACGTTACGTTATTTGTGATAGAGAGTGAAGTAGTGGCAATGAATTCCT
CAATTTGGGAGTCAGACTCAGCCTTCAGATTTATTTACACTGATTTAGTTATGGAATACA
TACCAGAATATTTCAAAGACTGGCAATATCATCCATGGATTTTTTCCGTTCTTAGGTTCTA
TTTTAATTGGATTAAGTGGTGTGTTGCCTTTACTTATAATTCCATTGGAAGGAGGAGAAC
ATTTTAAAAGTGGAGGAGGGGGCAAACCTTTTGCAGTATTGCTCAGTTTTGCAGTAGGAG
GTCTGCTCGGAGATGTGTTTTTGCATTTGTTGCCAGAAGCATGGGAGAACTACAGTAACA
AAGGAGATTCAGGAGGAGGTAGTCCATCAATGACATGCGGGTGTGGGTATTAGCTGGCC

TCTTGTATTACGATGGTCGAGAAGTTATTTGCTGGTTACATGGAAGATGATGACGTTT
CCGAGGAAAAGACCGAACGAAAGCGCAAGAAAATGTCTAAAATTAATTGTTTGATAAAGA
ACAACAATGTCTTGAGGGCTAGAAACGAAAATGGACGATTCTCGACCAGCAAATGCTCAG
GCGGACAAAATGGAAAATGTGTGACATTGAAGTTATAAATGGAGCTATCAAAAACGGCT
GTTGTATGGTCGAGGAGCTAGAGAAAAGAGGCTGTGAGCAACAAGTCTAAGCATGTTTCTG
GATACTTAAATCTCATGGCAAACCTCAATTGATAATTTACACACCGGTCTTGCTGTTGGTG
GTTCTTTTCATTGTTTCATTCAAACCTAGGAATTTTAAACAACATTCGCAATTCTAGTGCATG
AGATTCCACACGAAGTGGGCGATTTTGCAATTTTGTGCGGTGAGGTTTCTCGCGGTGGG
ACGCGGCTCGAGCGCAACTCATGACCGCTGGCGCCGGCCTGCTTGGAGCTATGACCGCAA
TCGGCGGAGTCGGGCTACGAGTGAATGGAAGCTAGAACATGCTGGATAATGCCCTTCA
CAGCCGGCGGATTTCTCCACATAAGTTTAGTTACAGTGCTACCTGACTTGCTGCAAGAAA
GGGATCCAAAAGAGTCGATTAACAATTTGGAGCATTGATTGGTGGTATAGTTCTAATGG
CATCATTGACCACATTATTGGAGTGAGGCATATATTTTTCTATTTCTAATTAGATTTAC
ATATCGGTGCCATCATCACATTAAGTCACTGTGATTTGACTAATATGATACAGACTTGAT
GAAATAAAGTAAAAATAGGGAACATTTTTACAAAGAAAACCTTATTGACATAATTAGATTA
ATCTATTTGATCTGTCCGGGAGAATTACTTACTTACTTACATATTATGGTTTTGAAAAGT
AATTCTCAAATAATGATATTTGTGCATGTAAATAGTTATATTTTTAATACTTGTTATATA
TATATATATATA

>Locus_4149_Transcript_4/4_Confidence_0.600_Length_1962|Q9VAF0|
CTTTCCTTCCCCTGTCTATGATAGACTAACTCCGACTTCGTATGCGCAGGCTCGACACGA
CGGACGTATTCGCATTTCGCACGATCTACGAACCAACTGACATAGTAGTACAGCGGGTGAC
AGTCGTCGTA AACGTTAGAATAATGCGTTACAATTTTCGATCTGAAGTGGACAATGACTTC
AATAATTTATAACATCAAAGCTTCTGCACGACCATAAATTTGATTTTTAATCTTTTTAAATGA
TTTTCGTTAACGCTAGCTCAATCTCAGTGCAGATTAGCTAGTTCGAACGTCGAATCTTCCCA
AGAAAACGGTGCAGACTATACGACGTTTGTATTTTTGTTTACTGCGCGTTCTATTGGTCC
TTAAGTGACATTCTCAGATGTTTTAATGCTTGTGTTTCTATTGGATTCAATTTGTTTTAT
GGATTTTGATGCAACATAAACGGTGTATTGTA AAAACCTGATGATAAAAACAAGTATTTTT
AAATAATAATTACGTTACGTTATTTGTGATAGAGAGTAAAAGTAGTGGCAATGAATTCCT
CAATTTGGGAGTCAGACTCAGCCTTCAGATTTTATTACACTGATTTAGTTATGGAATACA
TACCAGAATATTTCAAAGACTGGCAATATCATCCATGGATTTTTTCGGTCTTAGGTTCTA
TTTTAATTGGATTAAGTGGTGTGTTGCCTTTACTTATAAATCCATTGGAAGGAGGAGAAC
ATTTTTAAAAGTGGAGGAGGGGGCAAACTTTTGCGCGTATTGCTCAGTTTTGCAGTAGGAG
GTCTGCTCGGAGATGTGTTTTTGCATTTGTTGCCAGAAGCATGGGAGA ACTACAGTAACA
AAGGAGAGCGACTCGTAGTCATGTACGAATTCGTAGATTTCAGGAGGAGGTAGTCCATCAA
TGACATGCGGGTGTGGGTATTAGCTGGCCTCTTGTATTACGATGGTCGAGAAGTTAT
TTGCTGGTTACATGGAAGATGATGACGTTTCCGAGGAAAAGACCGAACGAAAGCGCAAGA
AAATGTCTAAAATTAATTGTTTGATAAAGAACAACAATGTCTTGAGGGCTAGAAACGAAA
ATGGACGATTCTCGACCAGCAAATGCTCAGGCGGACAAAATGGAAAATGTGTGACATTG
AAGTTATAAATGGAGCTATCAAAAACGGCTGTTGTATGGTCGAGGAGCTAGAGAAAAGAGG
CTGTGAGCAACAAGTCTAAGCATGTTTCTGGATACTTAAATCTCATGGCAAACCTCAATTG
ATAATTTACACACCGGTCTTGCTGTTGGTGGTTCTTTTCATTGTTTCATTCAAACCTAGGAA
TTTTAACAACATTCGCAATTCTAGTGCATGAGATTCCACACGAAGTGGGCGATTTTGCAA
TTTTGTTGCGGTGAGGTTTCTCGCGGTGGGACGCGGCTCGAGCGCAACTCATGACCGCTG
GCGCCGGCCTGCTTGGAGCTATGACCGCAATCGGCGGAAGTCCGGGCTACGAGTGAATGG
AAGCTAGAACATGCTGGATAATGCCCTTCACAGCCGGCGGATTTCTCCACATAAGTTTAG
TTACAGTGCTACCTGACTTGCTGCAAGAAAGGGATCCAAAAGAGTCGATTAACAATTTG
GACATTGATTGGTGGTATAGTTCTAATGGCATCATTGACCACATTATTGGAGTGAGGCA
TATATTTTTCTATTTCTAATTAGATTTTACATATCGGTGCCATCATCACATTAAGTCACT
GTGATTTGACTAATATGATACAGACTTGATGAAATAAAGTAAAAATAGGGAACATTTTTA
CAAAGAAAACCTTATTGACATAATTAGATTAATCTATTTGATCTGTCCGGGAGAATTACTT
ACTTACTTACATATTATGGTTTTGAAAAGTAATTCTCAAATAATGATATTTGTGCATGTA
AATAGTTATATTTTTAATACTTGTTATATATATATATATATA

2. Assembly of the transcripts clustered by Q9VAF0.

- Reference protein Q9VAF0:

```
>sp|Q9VAF0|Y816_DROME Uncharacterized protein CG7816 OS=Drosophila
melanogaster GN=CG7816 PE=2 SV=1
MTTNSFFDEHIAMIYNSLMDQYMPEYFKSFEYTPWVFSLLGSVIGLSGIFPLIIIPTE
EKMAKEGYKDPADSKLLRVLLSFAVGGLLDVFLHLLPEAWEGDNQDPSSHPSLRSLGWV
LSGILIFTIVEKIFSGYASADEENPQPKCVEIANCLLRRHGGQLPEGETSESCGGACDIE
DVGKVCFLREQEQKSKERKEQPKKVAGYLNLLANSIDNFTHGLAVAGSFLVSFRHGILAT
FAILLHEIPHEVGDFAILLRSGFSRWDAARAQLLTAGAGLLGALVAIGGSGVTSAMEART
SWIMPFTAGGFLHIALVTVLPDLLKEEERKESIKQLLLALVFGIALMAVMTMLFEH
```

- Transcripts to be assembled (protein Q9VAF0):

The four transcripts are very similar with some indels that could correspond to different isoforms. This is the alignment of all the transcripts clustered by the protein Q9VAF0:

CLUSTAL 2.1 multiple sequence alignment

```
Locus_4149_Transcript_3/4_Conf      CTTTCCTTCCCGTGTCTATGATAGACTAACTCCGACTTCGTATGCGCAGG
Locus_4149_Transcript_4/4_Conf      CTTTCCTTCCCGTGTCTATGATAGACTAACTCCGACTTCGTATGCGCAGG
Locus_4149_Transcript_1/4_Conf      CTTTCCTTCCCGTGTCTATGATAGACTAACTCCGACTTCGTATGCGCAGG
Locus_4149_Transcript_2/4_Conf      CTTTCCTTCCCGTGTCTATGATAGACTAACTCCGACTTCGTATGCGCAGG
*****
```

```
Locus_4149_Transcript_3/4_Conf      CTCGACACGACGGACGTATTCGCATTTCGCACGATCTACGAACCAACTGAC
Locus_4149_Transcript_4/4_Conf      CTCGACACGACGGACGTATTCGCATTTCGCACGATCTACGAACCAACTGAC
Locus_4149_Transcript_1/4_Conf      CTCGACACGACGGACGTATTCGCATTTCGCACGATCTACGAACCAACTGAC
Locus_4149_Transcript_2/4_Conf      CTCGACACGACGGACGTATTCGCATTTCGCACGATCTACGAACCAACTGAC
*****
```

```
Locus_4149_Transcript_3/4_Conf      ATAGTAGTACAGCGGGTGACAGTCGTCGTAAAACGTAGAATAATGCGTTA
Locus_4149_Transcript_4/4_Conf      ATAGTAGTACAGCGGGTGACAGTCGTCGTAAAACGTAGAATAATGCGTTA
Locus_4149_Transcript_1/4_Conf      ATAGTAGTACAGCGGGTGACAGTCGTCGTAAAACGTAGAATAATGCGTTA
Locus_4149_Transcript_2/4_Conf      ATAGTAGTACAGCGGGTGACAGTCGTCGTAAAACGTAGAATAATGCGTTA
*****
```

```
Locus_4149_Transcript_3/4_Conf      CAATTTTCGATCTGAAGTGGACAATGACTTCAATAATTTATAACATCAAAG
Locus_4149_Transcript_4/4_Conf      CAATTTTCGATCTGAAGTGGACAATGACTTCAATAATTTATAACATCAAAG
Locus_4149_Transcript_1/4_Conf      CAATTTTCGATCTGAAGTGGACAATGACTTCAATAATTTATAACATCAAAG
Locus_4149_Transcript_2/4_Conf      CAATTTTCGATCTGAAGTGGACAATGACTTCAATAATTTATAACATCAAAG
*****
```

```
Locus_4149_Transcript_3/4_Conf      CTTCTGCACGACCATAAATTGATTTTAAATCTTTTAAATGATTCGTTAAC
Locus_4149_Transcript_4/4_Conf      CTTCTGCACGACCATAAATTGATTTTAAATCTTTTAAATGATTCGTTAAC
Locus_4149_Transcript_1/4_Conf      CTTCTGCACGACCATAAATTGATTTTAAATCTTTTAAATGATTCGTTAAC
Locus_4149_Transcript_2/4_Conf      CTTCTGCACGACCATAAATTGATTTTAAATCTTTTAAATGATTCGTTAAC
*****
```

```
Locus_4149_Transcript_3/4_Conf      GCTAGCTCAATCTCAGTGCAGATTAGCTAGTCGAACGTCGAATCTTCCCA
Locus_4149_Transcript_4/4_Conf      GCTAGCTCAATCTCAGTGCAGATTAGCTAGTCGAACGTCGAATCTTCCCA
Locus_4149_Transcript_1/4_Conf      GCTAGCTCAATCTCAGTGCAGATTAGCTAGTCGAACGTCGAATCTTCCCA
Locus_4149_Transcript_2/4_Conf      GCTAGCTCAATCTCAGTGCAGATTAGCTAGTCGAACGTCGAATCTTCCCA
*****
```

```
Locus_4149_Transcript_3/4_Conf      AGAAAAACGGTGCAGCTATACGACGTTTGTATTTTTGTTTACTGCGGTT
Locus_4149_Transcript_4/4_Conf      AGAAAAACGGTGCAGCTATACGACGTTTGTATTTTTGTTTACTGCGGTT
Locus_4149_Transcript_1/4_Conf      AGAAAAACG-----
Locus_4149_Transcript_2/4_Conf      AGAAAAACGGTGCAGCTATACGACGTTTGTATTTTTGTTTACTGCGGTT
*****
```

```
Locus_4149_Transcript_3/4_Conf      CTATTGGTCCTTAAGTGCATTTCTCAGATGTTTTTAATG-----
Locus_4149_Transcript_4/4_Conf      CTATTGGTCCTTAAGTGCATTTCTCAGATGTTTTTAATG-----
```

```

Locus_4149_Transcript_1/4_Conf -----
Locus_4149_Transcript_2/4_Conf CTATTGGTCCTAAGTGACATTCTCAGATGTTTTTAATGTGATTTTCTTG

Locus_4149_Transcript_3/4_Conf -----
Locus_4149_Transcript_4/4_Conf -----
Locus_4149_Transcript_1/4_Conf -----
Locus_4149_Transcript_2/4_Conf TGATATAAATGCGTGCATATCATTGTGTAAATTTGAAATTAACGAT

Locus_4149_Transcript_3/4_Conf -----CTGTGTTTCTATTGGATTCAATTTGTTTTA
Locus_4149_Transcript_4/4_Conf -----CTGTGTTTCTATTGGATTCAATTTGTTTTA
Locus_4149_Transcript_1/4_Conf -----GCTGTGTTTCTATTGGATTCAATTTGTTTTA
Locus_4149_Transcript_2/4_Conf TTTTTATTTTCCGTTCCAGGCTTGTGTTTCTATTGGATTCAATTTGTTTTA
*****

Locus_4149_Transcript_3/4_Conf TGGATTTTGATGCAACATAAACGGTGTATTGTAAAAACCTGATGATAAAA
Locus_4149_Transcript_4/4_Conf TGGATTTTGATGCAACATAAACGGTGTATTGTAAAAACCTGATGATAAAA
Locus_4149_Transcript_1/4_Conf TGGATTTTGATGCAACATAAACGGTGTATTGTAAAAACCTGATGATAAAA
Locus_4149_Transcript_2/4_Conf TGGATTTTGATGCAACATAAACGGTGTATTGTAAAAACCTGATGATAAAA
*****

Locus_4149_Transcript_3/4_Conf CAAGTATTTTTAAATAATAATTACGTTACGTTATTTGTGATAGAGAGTGA
Locus_4149_Transcript_4/4_Conf CAAGTATTTTTAAATAATAATTACGTTACGTTATTTGTGATAGAGAGTGA
Locus_4149_Transcript_1/4_Conf CAAGTATTTTTAAATAATAATTACGTTACGTTATTTGTGATAGAGAGTGA
Locus_4149_Transcript_2/4_Conf CAAGTATTTTTAAATAATAATTACGTTACGTTATTTGTGATAGAGAGTGA
*****

Locus_4149_Transcript_3/4_Conf AAGTAGTGGCAATGAATTCCTCAATTTGGGAGTCAGACTCAGCCTTCAGA
Locus_4149_Transcript_4/4_Conf AAGTAGTGGCAATGAATTCCTCAATTTGGGAGTCAGACTCAGCCTTCAGA
Locus_4149_Transcript_1/4_Conf AAGTAGTGGCAATGAATTCCTCAATTTGGGAGTCAGACTCAGCCTTCAGA
Locus_4149_Transcript_2/4_Conf AAGTAGTGGCAATGAATTCCTCAATTTGGGAGTCAGACTCAGCCTTCAGA
*****

Locus_4149_Transcript_3/4_Conf TTTATTTACACIGATTTAGTTATGGAATACATACCAGAATATTTCAAAGA
Locus_4149_Transcript_4/4_Conf TTTATTTACACIGATTTAGTTATGGAATACATACCAGAATATTTCAAAGA
Locus_4149_Transcript_1/4_Conf TTTATTTACACIGATTTAGTTATGGAATACATACCAGAATATTTCAAAGA
Locus_4149_Transcript_2/4_Conf TTTATTTACACIGATTTAGTTATGGAATACATACCAGAATATTTCAAAGA
*****

Locus_4149_Transcript_3/4_Conf CTGGCAATATCATCCATGGATTTTTTCGGTCTTAGGTTCTATTTTAAATG
Locus_4149_Transcript_4/4_Conf CTGGCAATATCATCCATGGATTTTTTCGGTCTTAGGTTCTATTTTAAATG
Locus_4149_Transcript_1/4_Conf CTGGCAATATCATCCATGGATTTTTTCGGTCTTAGGTTCTATTTTAAATG
Locus_4149_Transcript_2/4_Conf CTGGCAATATCATCCATGGATTTTTTCGGTCTTAGGTTCTATTTTAAATG
*****

Locus_4149_Transcript_3/4_Conf GATTAAGTGGTGTGTTGCCTTTACTTATAATTCATTGGAAGGAGGAGAA
Locus_4149_Transcript_4/4_Conf GATTAAGTGGTGTGTTGCCTTTACTTATAATTCATTGGAAGGAGGAGAA
Locus_4149_Transcript_1/4_Conf GATTAAGTGGTGTGTTGCCTTTACTTATAATTCATTGGAAGGAGGAGAA
Locus_4149_Transcript_2/4_Conf GATTAAGTGGTGTGTTGCCTTTACTTATAATTCATTGGAAGGAGGAGAA
*****

Locus_4149_Transcript_3/4_Conf CATTTTAAAAGTGGAGGAGGGGGCAAACCTTTGCGCGTATTGCTCAGTTT
Locus_4149_Transcript_4/4_Conf CATTTTAAAAGTGGAGGAGGGGGCAAACCTTTGCGCGTATTGCTCAGTTT
Locus_4149_Transcript_1/4_Conf CATTTTAAAAGTGGAGGAGGGGGCAAACCTTTGCGCGTATTGCTCAGTTT
Locus_4149_Transcript_2/4_Conf CATTTTAAAAGTGGAGGAGGGGGCAAACCTTTGCGCGTATTGCTCAGTTT
*****

Locus_4149_Transcript_3/4_Conf TGCAGTAGGAGGTCTGCTCGGAGATGTGTTTTTGCATTTGTTGCCAGAAG
Locus_4149_Transcript_4/4_Conf TGCAGTAGGAGGTCTGCTCGGAGATGTGTTTTTGCATTTGTTGCCAGAAG
Locus_4149_Transcript_1/4_Conf TGCAGTAGGAGGTCTGCTCGGAGATGTGTTTTTGCATTTGTTGCCAGAAG
Locus_4149_Transcript_2/4_Conf TGCAGTAGGAGGTCTGCTCGGAGATGTGTTTTTGCATTTGTTGCCAGAAG
*****

Locus_4149_Transcript_3/4_Conf CATGGGAGAACACAGTAACAAAG-----
Locus_4149_Transcript_4/4_Conf CATGGGAGAACACAGTAACAAAGGAGAGCGACTCGTAGTCATGTACGAA
Locus_4149_Transcript_1/4_Conf CATGGGAGAACACAGTAACAAAGGAGAGCGACTCGTAGTCATGTACGAA
Locus_4149_Transcript_2/4_Conf CATGGGAGAACACAGTAACAAAGGAGAGCGACTCGTAGTCATGTACGAA
*****

Locus_4149_Transcript_3/4_Conf ----GAGATTCAGGAGGAGGTAGTCCATCAATGACATGCGGGTGTGGGT

```

Locus_4149_Transcript_4/4_Conf	TTCGTAGATTCAGGAGGAGGTAGTCCATCAATGACATGCGGGTGCTGGGT
Locus_4149_Transcript_1/4_Conf	TTCGTAGATTCAGGAGGAGGTAGTCCATCAATGACATGCGGGTGCTGGGT
Locus_4149_Transcript_2/4_Conf	TTCGTAGATTCAGGAGGAGGTAGTCCATCAATGACATGCGGGTGCTGGGT *****
Locus_4149_Transcript_3/4_Conf	ATTAGCTGGCCTCTTGTATTACGATGGTCGAGAAGTTATTTGCTGGTT
Locus_4149_Transcript_4/4_Conf	ATTAGCTGGCCTCTTGTATTACGATGGTCGAGAAGTTATTTGCTGGTT
Locus_4149_Transcript_1/4_Conf	ATTAGCTGGCCTCTTGTATTACGATGGTCGAGAAGTTATTTGCTGGTT
Locus_4149_Transcript_2/4_Conf	ATTAGCTGGCCTCTTGTATTACGATGGTCGAGAAGTTATTTGCTGGTT *****
Locus_4149_Transcript_3/4_Conf	ACATGGAAGATGATGACGTTTCCGAGGAAAAGACCGAACGAAAGCGCAAG
Locus_4149_Transcript_4/4_Conf	ACATGGAAGATGATGACGTTTCCGAGGAAAAGACCGAACGAAAGCGCAAG
Locus_4149_Transcript_1/4_Conf	ACATGGAAGATGATGACGTTTCCGAGGAAAAGACCGAACGAAAGCGCAAG
Locus_4149_Transcript_2/4_Conf	ACATGGAAGATGATGACGTTTCCGAGGAAAAGACCGAACGAAAGCGCAAG *****
Locus_4149_Transcript_3/4_Conf	AAAATGCTAAAAATTAATTGTTTGATAAAGAACAACAATGTCTTGAGGGC
Locus_4149_Transcript_4/4_Conf	AAAATGCTAAAAATTAATTGTTTGATAAAGAACAACAATGTCTTGAGGGC
Locus_4149_Transcript_1/4_Conf	AAAATGCTAAAAATTAATTGTTTGATAAAGAACAACAATGTCTTGAGGGC
Locus_4149_Transcript_2/4_Conf	AAAATGCTAAAAATTAATTGTTTGATAAAGAACAACAATGTCTTGAGGGC *****
Locus_4149_Transcript_3/4_Conf	TAGAAACGAAAATGGACGATTCTCGACCAGCAAATGCTCAGCGGACAAA
Locus_4149_Transcript_4/4_Conf	TAGAAACGAAAATGGACGATTCTCGACCAGCAAATGCTCAGCGGACAAA
Locus_4149_Transcript_1/4_Conf	TAGAAACGAAAATGGACGATTCTCGACCAGCAAATGCTCAGCGGACAAA
Locus_4149_Transcript_2/4_Conf	TAGAAACGAAAATGGACGATTCTCGACCAGCAAATGCTCAGCGGACAAA *****
Locus_4149_Transcript_3/4_Conf	ATGGAAAAATGTTGACATTGAAGTTATAAATGGAGCTATCAAAAACGGC
Locus_4149_Transcript_4/4_Conf	ATGGAAAAATGTTGACATTGAAGTTATAAATGGAGCTATCAAAAACGGC
Locus_4149_Transcript_1/4_Conf	ATGGAAAAATGTTGACATTGAAGTTATAAATGGAGCTATCAAAAACGGC
Locus_4149_Transcript_2/4_Conf	ATGGAAAAATGTTGACATTGAAGTTATAAATGGAGCTATCAAAAACGGC *****
Locus_4149_Transcript_3/4_Conf	TGTTGTATGGTCGAGGAGCTAGAGAAAGAGGCTGTGAGCAACAAGTCTAA
Locus_4149_Transcript_4/4_Conf	TGTTGTATGGTCGAGGAGCTAGAGAAAGAGGCTGTGAGCAACAAGTCTAA
Locus_4149_Transcript_1/4_Conf	TGTTGTATGGTCGAGGAGCTAGAGAAAGAGGCTGTGAGCAACAAGTCTAA
Locus_4149_Transcript_2/4_Conf	TGTTGTATGGTCGAGGAGCTAGAGAAAGAGGCTGTGAGCAACAAGTCTAA *****
Locus_4149_Transcript_3/4_Conf	GCATGTTTCTGGATACTTAAATCTCATGGCAAACCTCAATTGATAATTTCA
Locus_4149_Transcript_4/4_Conf	GCATGTTTCTGGATACTTAAATCTCATGGCAAACCTCAATTGATAATTTCA
Locus_4149_Transcript_1/4_Conf	GCATGTTTCTGGATACTTAAATCTCATGGCAAACCTCAATTGATAATTTCA
Locus_4149_Transcript_2/4_Conf	GCATGTTTCTGGATACTTAAATCTCATGGCAAACCTCAATTGATAATTTCA *****
Locus_4149_Transcript_3/4_Conf	CACACGGTCTTGCTGTTGGTGGTTCTTTTCATTGTTTCATTCAAACCTAGGA
Locus_4149_Transcript_4/4_Conf	CACACGGTCTTGCTGTTGGTGGTTCTTTTCATTGTTTCATTCAAACCTAGGA
Locus_4149_Transcript_1/4_Conf	CACACGGTCTTGCTGTTGGTGGTTCTTTTCATTGTTTCATTCAAACCTAGGA
Locus_4149_Transcript_2/4_Conf	CACACGGTCTTGCTGTTGGTGGTTCTTTTCATTGTTTCATTCAAACCTAGGA *****
Locus_4149_Transcript_3/4_Conf	ATTTTAACAACATTCGCAATTCAGTGCATGAGATTCACACGAAAGTGGG
Locus_4149_Transcript_4/4_Conf	ATTTTAACAACATTCGCAATTCAGTGCATGAGATTCACACGAAAGTGGG
Locus_4149_Transcript_1/4_Conf	ATTTTAACAACATTCGCAATTCAGTGCATGAGATTCACACGAAAGTGGG
Locus_4149_Transcript_2/4_Conf	ATTTTAACAACATTCGCAATTCAGTGCATGAGATTCACACGAAAGTGGG *****
Locus_4149_Transcript_3/4_Conf	CGATTTTGCAATTTTGTGCGGTCAGGTTTCTCGCGGTGGGACGCGGCTC
Locus_4149_Transcript_4/4_Conf	CGATTTTGCAATTTTGTGCGGTCAGGTTTCTCGCGGTGGGACGCGGCTC
Locus_4149_Transcript_1/4_Conf	CGATTTTGCAATTTTGTGCGGTCAGGTTTCTCGCGGTGGGACGCGGCTC
Locus_4149_Transcript_2/4_Conf	CGATTTTGCAATTTTGTGCGGTCAGGTTTCTCGCGGTGGGACGCGGCTC *****
Locus_4149_Transcript_3/4_Conf	GAGCGCAACTCATGACCGCTGGCGCCGGCCTGCTTGAGACTATGACCGCA
Locus_4149_Transcript_4/4_Conf	GAGCGCAACTCATGACCGCTGGCGCCGGCCTGCTTGAGACTATGACCGCA
Locus_4149_Transcript_1/4_Conf	GAGCGCAACTCATGACCGCTGGCGCCGGCCTGCTTGAGACTATGACCGCA
Locus_4149_Transcript_2/4_Conf	GAGCGCAACTCATGACCGCTGGCGCCGGCCTGCTTGAGACTATGACCGCA *****

```

Locus_4149_Transcript_3/4_Conf      ATCGGGCGAAGTCGGGCTACGAGTGC AATGGAAGCTAGAACATGCTGGAT
Locus_4149_Transcript_4/4_Conf      ATCGGGCGAAGTCGGGCTACGAGTGC AATGGAAGCTAGAACATGCTGGAT
Locus_4149_Transcript_1/4_Conf      ATCGGGCGAAGTCGGGCTACGAGTGC AATGGAAGCTAGAACATGCTGGAT
Locus_4149_Transcript_2/4_Conf      ATCGGGCGAAGTCGGGCTACGAGTGC AATGGAAGCTAGAACATGCTGGAT
*****

Locus_4149_Transcript_3/4_Conf      AATGCCCTTACAGCCGGCGGATTTCTCCACATAAGTTTAGTTACAGTGC
Locus_4149_Transcript_4/4_Conf      AATGCCCTTACAGCCGGCGGATTTCTCCACATAAGTTTAGTTACAGTGC
Locus_4149_Transcript_1/4_Conf      AATGCCCTTACAGCCGGCGGATTTCTCCACATAAGTTTAGTTACAGTGC
Locus_4149_Transcript_2/4_Conf      AATGCCCTTACAGCCGGCGGATTTCTCCACATAAGTTTAGTTACAGTGC
*****

Locus_4149_Transcript_3/4_Conf      TACCTGACTTGCTGCAAGAAAGGGATCCAAAAGAGTCGATTAACAATTT
Locus_4149_Transcript_4/4_Conf      TACCTGACTTGCTGCAAGAAAGGGATCCAAAAGAGTCGATTAACAATTT
Locus_4149_Transcript_1/4_Conf      TACCTGACTTGCTGCAAGAAAGGGATCCAAAAGAGTCGATTAACAATTT
Locus_4149_Transcript_2/4_Conf      TACCTGACTTGCTGCAAGAAAGGGATCCAAAAGAGTCGATTAACAATTT
*****

Locus_4149_Transcript_3/4_Conf      GGAGCATTGATTGGTGGTATAGTTCTAATGGCATCATTGACCACATTATT
Locus_4149_Transcript_4/4_Conf      GGAGCATTGATTGGTGGTATAGTTCTAATGGCATCATTGACCACATTATT
Locus_4149_Transcript_1/4_Conf      GGAGCATTGATTGGTGGTATAGTTCTAATGGCATCATTGACCACATTATT
Locus_4149_Transcript_2/4_Conf      GGAGCATTGATTGGTGGTATAGTTCTAATGGCATCATTGACCACATTATT
*****

Locus_4149_Transcript_3/4_Conf      GGAGTGAGGCATATATTTTTCTATTTCTAATTAGATTTACATATCGGTG
Locus_4149_Transcript_4/4_Conf      GGAGTGAGGCATATATTTTTCTATTTCTAATTAGATTTACATATCGGTG
Locus_4149_Transcript_1/4_Conf      GGAGTGAGGCATATATTTTTCTATTTCTAATTAGATTTACATATCGGTG
Locus_4149_Transcript_2/4_Conf      GGAGTGAGGCATATATTTTTCTATTTCTAATTAGATTTACATATCGGTG
*****

Locus_4149_Transcript_3/4_Conf      CCATCATCACATTAAGTCACTGTGATTTGACTAATATGATACAGACTTGA
Locus_4149_Transcript_4/4_Conf      CCATCATCACATTAAGTCACTGTGATTTGACTAATATGATACAGACTTGA
Locus_4149_Transcript_1/4_Conf      CCATCATCACATTAAGTCACTGTGATTTGACTAATATGATACAGACTTGA
Locus_4149_Transcript_2/4_Conf      CCATCATCACATTAAGTCACTGTGATTTGACTAATATGATACAGACTTGA
*****

Locus_4149_Transcript_3/4_Conf      TGAAATAAAGTAAAAATAGGGAAC TATTTTACAAGAAAAC TATTGACA
Locus_4149_Transcript_4/4_Conf      TGAAATAAAGTAAAAATAGGGAAC TATTTTACAAGAAAAC TATTGACA
Locus_4149_Transcript_1/4_Conf      TGAAATAAAGTAAAAATAGGGAAC TATTTTACAAGAAAAC TATTGACA
Locus_4149_Transcript_2/4_Conf      TGAAATAAAGTAAAAATAGGGAAC TATTTTACAAGAAAAC TATTGACA
*****

Locus_4149_Transcript_3/4_Conf      TAATTAGATTAATCTATTTGATCTGTCCGGGAGAATTACTTACTTACTTA
Locus_4149_Transcript_4/4_Conf      TAATTAGATTAATCTATTTGATCTGTCCGGGAGAATTACTTACTTACTTA
Locus_4149_Transcript_1/4_Conf      TAATTAGATTAATCTATTTGATCTGTCCGGGAGAATTACTTACTTACTTA
Locus_4149_Transcript_2/4_Conf      TAATTAGATTAATCTATTTGATCTGTCCGGGAGAATTACTTACTTACTTA
*****

Locus_4149_Transcript_3/4_Conf      CATATTATGGTTTTGAAAAGTAATTC TCAAATAATGATATTTGTGCATGT
Locus_4149_Transcript_4/4_Conf      CATATTATGGTTTTGAAAAGTAATTC TCAAATAATGATATTTGTGCATGT
Locus_4149_Transcript_1/4_Conf      CATATTATGGTTTTGAAAAGTAATTC TCAAATAATGATATTTGTGCATGT
Locus_4149_Transcript_2/4_Conf      CATATTATGGTTTTGAAAAGTAATTC TCAAATAATGATATTTGTGCATGT
*****

Locus_4149_Transcript_3/4_Conf      AAATAGTTATATTTTTAATACTTGTTATATATATATATATATA
Locus_4149_Transcript_4/4_Conf      AAATAGTTATATTTTTAATACTTGTTATATATATATATATATA
Locus_4149_Transcript_1/4_Conf      AAATAGTTATATTTTTAATACTTGTTATATATATATATATATA
Locus_4149_Transcript_2/4_Conf      AAATAGTTATATTTTTAATACTTGTTATATATATATATATATA
*****

```

-Translation of the largest contig,
Locus_4149_Transcript_2/4_Confidence_0.700_Length_2043|Q9VA
F0:

```

atgaattcctcaatttgggagtcagactcagccttcagattttatttacact
M N S S I W E S D S A F R F I Y T
gatttagttatggaatacataccagaatatttcaaagactggcaatatcatccatggat

```

t
D L V M E Y I P E Y F K D W Q Y H P W I
ttttcgggtcttaggttctatttttaattggattaagtgggtgtggtgcctttacttataat
t
F S V L G S I L I G L S G V L P L L I I
ccattggaaggaggagaacatttttaaaagtggaggagggggcaaacttttgcgcggtatt
g
P L E G G E H F K S G G G G K L L R V L
ctcagttttgacagtaggaggtctgctcggagatgtggtttttgcatttggtgccagaagc
a
L S F A V G G L L G D V F L H L L P E A
tgggagaactacagtaacaaaggagagcgactcgtagtcacgtacgaattcgtagattc
a
W E N Y S N K G E R L V V M Y E F V D S
ggaggaggtagtcacatcaatgacatgcgggtgctgggtattagctggcctcttggttatt
c
G G G S P S M T C G C W V L A G L L L F
acgatgggtcgagaagttatttgctgggtacatggaagatgatgacgtttccgaggaaaa
g
T M V E K L F A G Y M E D D D V S E E K
accgaacgaaagcgcaagaaaatgtctaaaattaattgtttgataaagaacaacaatgt
c
T E R K R K K M S K I N C L I K N N N V
ttgagggctagaaacgaaaatggacgattctcgcaccagcaaattgctcaggcggacaaaa
t
L R A R N E N G R F S T S K C S G G Q N
ggaaaaatgtgtgacattgaagttataaatggagctatcaaaaacggctggtgtatggt
c
G K M C D I E V I N G A I K N G C C M V
gaggagctagagaaagaggctgtgagcaacaagtctaagcatgtttctggataacttaaa
t
E E L E K E A V S N K S K H V S G Y L N
ctcatggcaactcaattgataatttcacacacgggtcttgctggttggtggttctttcat
t
L M A N S I D N F T H G L A V G G S F I
gtttcattcaaactaggaatttttaacaacattcgcaattctagtgcatgagattccaca
c
V S F K L G I L T T F A I L V H E I P H
gaagtgggagatttttgcaattttggtgcggtcagggtttctcgcgggtgggacgcggctcg
a
E V G D F A I L L R S G F S R W D A A R
gcgcaactcatgaccgctggcgccggcctgcttgagctatgaccgcaatcggcgggaag
t
A Q L M T A G A G L L G A M T A I G G S
cgggctacgagtgcaatggaagctagaacatgctggataatgcccttcacagccggcgg
a
R A T S A M E A R T C W I M P F T A G G

tttctccacataagtttagttacagtgctacctgacttgctgcaagaaagggatccaaa
a

F L H I S L V T V L P D L L Q E R D P K
gagtcgattaaacaatttggagcattgattgggtggtatagttctaattggcatcattgac
c

E S I K Q F G A L I G G I V L M A S L T
acattattggagtga
T L L E -

- DNA sequence and protein for the largest open reading frame
inLocus_4149_Transcript_2/4_Confidence_0.700_Length_2043|Q9VAF0:

ATGAATTCCTCAATTTGGGAGTCAGACTCAGCCTTCAGATTTATTTACA
CTGATTTAGTTATGGAATACATACCAGAATATTTCAAAGACTGGCAATATCATCCATGGA
TTTTTTTCGGTCTTAGGTTCTATTTTAAATTGGATTAAGTGGTGTGTTGCCTTTACTTATAA
TTCCATTGGAAGGAGGAGAACATTTTAAAAGTGGAGGAGGGGGCAAACCTTTGCGCGTAT
TGCTCAGTTTTGCAGTAGGAGGTCTGCTCGGAGATGTGTTTTTGCATTTGTTGCCAGAAG
CATGGGAGAACTACAGTAACAAAGGAGAGCGACTCGTAGTCATGTACGAATTCGTAGATT
CAGGAGGAGGTAGTCCATCAATGACATGCGGGTCTGGGTATTAGCTGGCCTCTTGTTAT
TCACGATGGTCGAGAAGTTATTTGCTGGTTACATGGAAGATGATGACGTTTCCGAGGAAA
AGACCGAACGAAAGCGCAAGAAAATGTCTAAAATTAATTGTTTTGATAAAGAACAACAATG
TCTTGAGGGCTAGAAACGAAAATGGACGATTCTCGACCAGCAAATGCTCAGGCGGACAAA
ATGGAAAAATGTGTGACATTGAAGTTATAAATGGAGCTATCAAAAACGGCTGTTGTATGG
TCGAGGAGCTAGAGAAAGAGGCTGTGAGCAACAAGTCTAAGCATGTTTCTGGATACTTAA
ATCTCATGGCAAACCTCAATTGATAATTTACACACCGGTCTTGCTGTTGGTGGTTCTTTCA
TTGTTTCATTCAAACCTAGGAATTTTAAACAACATTTCGCAATTCTAGTGCATGAGATTCCAC
ACGAAGTGGGCGATTTTGCATTTTGTGCGGTGAGTTTCTCGCGGTGGGACGCGGCTC
GAGCGCAACTCATGACCGCTGGCGCCGGCCTGCTTGGAGCTATGACCGCAATCGGCGGAA
GTCGGGCTACGAGTGAATGGAAGCTAGAACATGCTGGATAATGCCCTTACAGCCGGCG
GATTTCTCCACATAAAGTTTAGTTACAGTGTACCTGACTTGTGCAAGAAAGGGATCCAA
AAGAGTCGATTAACAATTTGGAGCATTGATTGGTGGTATAGTTCTAATGGCATCATTGA
CCACATTATTGGAGTGA

MNSSIWESDSAFRFIYTDLMVEYIPEYFKDWQYHPWIFSVLGSILIGLSGVLPLLIIPLE
GGEHFKSGGGKLLRVLLSFAVGGLLDVFLHLLPEAWENYSNKGERLVVMYEFVDSGGG
SPSMTCGCWVLAGLLLFTMVEKLFAGYMEDDDVSEEKTERKRKKMSKINCLIKNNNVLRA
RNENGRFSTSKCSGGQNGKMCIEVINGAIKNGCCMVEELEKEAVSNKSKHVSGLNLMA
NSIDNFTHGLAVGGSFIVSFKLGILTTFAILVHEIPHEVGDFAILLRSGFSRWDAARAQL
MTAGAGLLGAMTAIGSRATSAMEARTCWIMPFTAGGFLHISLVTVLPDLLQERDPKESI
KQFGALIGGIVLMASLTTLLE-

- Alignment to the reference protein Q9VAF0:

Score	Expect	Method	Identities	Positives	Gaps
402 bits (1033)	8e-143	Compositional matrix adjust.	221/385 (57%)	269/385 (69%)	39/385 (10%)
Query 4	NSSFF--DEHIAMIYSNLMDQYMPYFKSFEYTPWVFSLLGSVVIGLSGIFPLLIIPTEE				61
Sbjct 2	NSS + D IY++L+ +Y+PEYFK ++Y PW+FS+LGS++IGLSG+ PL+IIP E				60
Query 62	KMAKEGYKDPADSKLLRVLLSFAVGGLLDVFLHLLPEAWEGDNQ-----DP				108
Sbjct 61	E +K KLLRVLLSFAVGGLLDVFLHLLPEAWE +				118
Query 109	SSHPSLRSLWVLSGILIFTIVEKIFSGYASAD---EENPQPKCVEIA--NCLLRRHGGQ				163
Sbjct 119	PS+ G WV L+G+L+FT+VEK+F+GY D EE + K +++ NCL++ +				178
Query 164	LPEGE-----TSESCGGA-----CDIEDVGK-----VCFLREQEQKSKERKEQPKKVAGYL				209
Sbjct 179	E TS+ GG CDIE + C + E E+++ K K V+GYL				236
Query 210	NLLANSIDNFTHGLAVAGSFLVSRFHGILATFAILLHEIPHEVGDFAILLRSGFSRWDA				269
Sbjct 237	NL+ANSIDNFTHGLAV GSF+VSF+ GIL TFAIL+HEIPHEVGDFAILLRSGFSRWDA				296

Query 270 RAQLLTAGAGLLGALVAIGGSGVTSAMEARTSWIMPFTAGGFLHIALVTVLPDLLKEEER 329
 RAQL+TAGAGLLGA+ AIGGS TSAMEART WIMPFTAGGFLHI+LVTVLPDLL+E +
 Sbjct 297 RAQLMTAGAGLLGAMTAIGGSRATSAMEARTCWIMPFTAGGFLHISLVTVLPDLLQERDP 356

 Query 330 KESIKQLLALVFGIALMAVMTMLFE 354
 KESIKQ AL+ GI LMA +T L E
 Sbjct 357 KESIKQFGALIGGIVLMAASLTTLLE 381

3. Sequence used for synthesis and expression.

ATGAATTCCTCAATTTGGGAGTCAGACTCAGCCTTCAGATTTATTTACACTGATTTAGTTATGGAATA
 CATAACCAGAATATTTCAAAGACTGGCAATATCATCCATGGATTTTTTCGGTCTTAGGTTCTATTTTAA
 TTGGATTAAGTGGTGTGTTGCCTTTACTTATAATTCCATTGGAAGGAGGAGAACATTTTAAAAGTGGA
 GGAGGGGGCAAACTTTTGCGCGTATTGCTCAGTTTTGCAGTAGGAGGTCTGCTCGGAGATGTGTTTTT
 GCATTTGTTGCCAGAAGCATGGGAGAACTACAGTAACAAAGGAGAGCGACTCGTAGTCATGTACGAAT
 TCGTAGATTCAGGAGGAGGTAGTCCATCAATGACATGCGGGTGCTGGGTATTAGCTGGCCTCTTGTTA
 TTCACGATGGTCGAGAAGTTATTTGCTGGTTACATGGAAGATGATGACGTTTTCCGAGGAAAAGACCGA
 ACGAAAATGGACGATTCTCGACCAGCAAATGCTCAGGCGGACAAAATGGAAAATGTGTGACATTGAA
 GTTATAAATGGAGCTATCAAAAACGGCTGTTGTATGGTCGAGGAGCTAGAGAAAAGAGGCTGTGAGCAA
 CAAGTCTAAGCATGTTTCTGGATACTTAAATCTCATGGCAAACCAATTGATAATTTACACACCGGTC
 TTGCTGTTGGTGGTTCTTTTCATTGTTTCATTCAAAC TAGGAATTTTAAACAACATTGCAATTCTAGTG
 CATGAGATTCCACACGAAGTGGGCGATTTTGAATTTTGTGCGGTGAGGTTTCTCGCGGTGGGACGC
 GGCTCGAGCGCAACTCATGACCGCTGGCGCCGGCCTGCTTGGAGCTATGACCGCAATCGGCGGAAGTC
 GGGCTACGAGTGCAATGGAAGCTAGAACATGCTGGATAATGCCCTTCACAGCCGGCGGATTTCTCCAC
 ATAAGTTTTAGTTACAGTGCTACCTGACTTGCTGCAAGAAAGGGATCCAAAAGAGTCGATTAACAATT
 TGGAGCATTGATTGGTGGTATAGTTCTAATGGCATCATTGACCACATTATTGGAGTGA