

BioPython

Fragaria x ananassa

Vs

Fragaria vesca

Comparative Genomic Analysis

Link to video:

<https://youtu.be/wSam4itsuF4>

Python and numpy basic lectures referenced in the video:

<https://github.com/jrjohansson/scientific-python-lectures>

BioPython's official manual:

<http://biopython.org/DIST/docs/tutorial/Tutorial.pdf>

—

By Eliza Doering



<http://strawberry-garden.kazusa.or.jp/images/strawberry1.jpg>

INTRODUCTION/BACKGROUND

The strawberry (genus *Fragaria*) is a diverse plant with over 22 described species. Two of these are *Fragaria x ananassa* and *Fragaria vesca*. The diversity of the *Fragaria* genus can be shown via the comparison of these two organisms using the BioPython library.

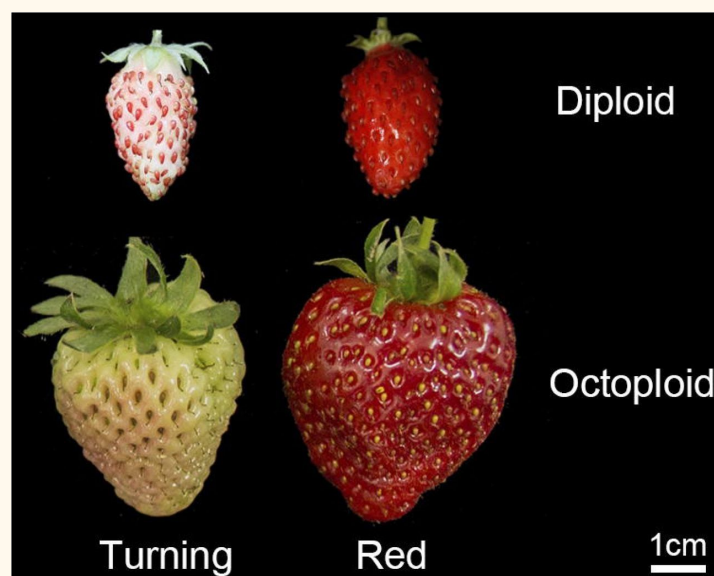
BioPython

Biopython is a great tool for bioinformaticians. It includes functionalities for the reading and parsing of multiple types of files, analysis of sequences (such as GC content and length), plotting, visualizations, and more! It is an open source technology, meaning that anyone can look at its source code. It also includes a wiki with user uploaded functions and tools currently in development. This project includes all three of these types of functions.

Datasets: Cultivated vs Wild

As the image displayed below demonstrates these two organisms' respective sizes differ greatly. For each organism: fasta files containing the annotated proteins and transcripts. These particular files are useful as they are the same format and they are all from two resources that interact with each other (GDR sources Strawberry Garden and vice versa). The larger

counterpart is more commonly grown than its wild counterpart and they can both grow under similar conditions. Anecdotally I have found that the wild counterpart is more difficult to plant in a garden, but it has a tendency of maintaining its sweetness through extreme weather than its cultivated relative. I wanted to see if they GC content and transcript length could account for any of this.



<https://www.biorxiv.org/content/biorxiv/early/2020/01/03/2020.01.02.893453/F6.large.jpg>

1. *Fragaria x ananassa* (cultivated)

As stated above, *Fragaria x ananassa* is a much larger species than its wild counterpart. Fasta files downloaded from GDR (www.rosaceae.org) and Strawberry Garden (<http://strawberry-garden.kazusa.or.jp/>).

<i>File Name + Link</i>	<i>Description</i>	<i>Size</i>
Fxa_v1.2_makerStandard_proteins_woTposases.fasta	Annotated proteins from GDR	52.6 MB
FAN_r2.3_transcripts.fa	Transcripts from Strawberry Garden	222 MB

2. *Fragaria vesca* (wild/alpine)

Files were also accessed from GDR and Strawberry Garden for the smaller species.

<i>File Name + Link</i>	<i>Description</i>	<i>Size</i>
Fragaria_vesca_v4.0.a1_makerStandard_proteins.fasta	Annotated proteins from GDR	13.5 MB
Fragaria_vesca_v4.0.a2.transcripts.fa	Transcripts from GDR	134 MB

The file size differences echo the different organism sizes.

BioPython installation

```
pip install biopython
```

Installing BioPython is easy! It can be done in a single command if you have the python library installer ‘pip’ installed. That option is shown in the image above. Python3 and JupyterLab should also be installed. The directions for how to install these are in the link on the first page of this manual.

```

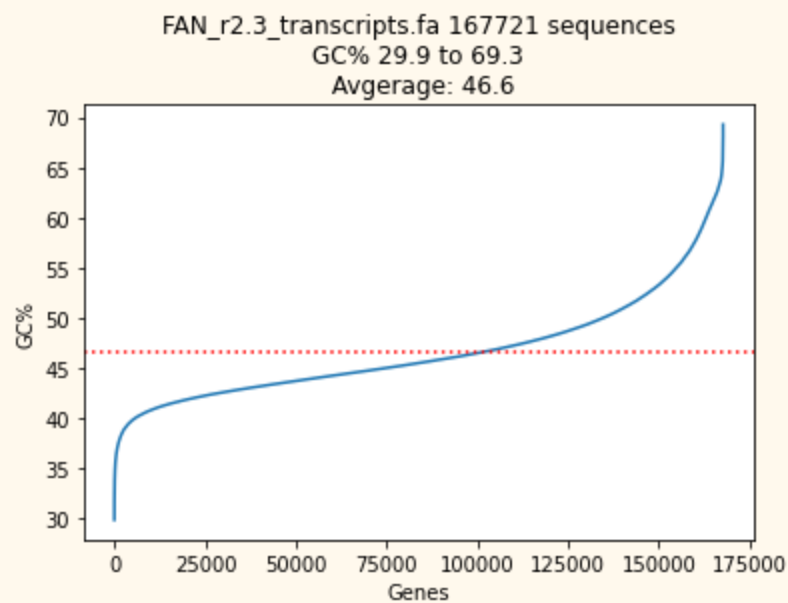
: !pip install biopython
Collecting biopython
  Downloading https://files.pythonhosted.org/packages/76/02/8b606c4aa92ff61b5eda71d23b499ab1de57d5e818be33f77b01a6f435a8/biopython-1.78-cp36-cp36m-manylinux1_x86_64.whl (2.3MB)
    | 2.3MB 5.8MB/s
Requirement already satisfied: numpy in /usr/local/lib/python3.6/dist-packages (from biopython) (1.18.5)
Installing collected packages: biopython
Successfully installed biopython-1.78

```

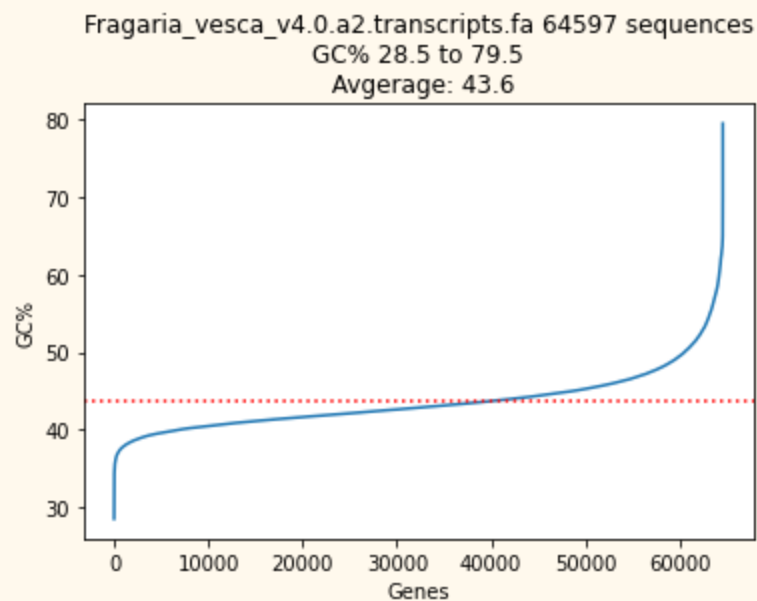
Alternatively biopython can be installed directly into a google colab notebook. This option is shown above. Once installation is complete it is time to begin! Before any programming can take place the correct libraries must be imported. All imports needed to complete this tutorial are shown in the image above.

The rest of the tutorial will be in the last pages of the manual.

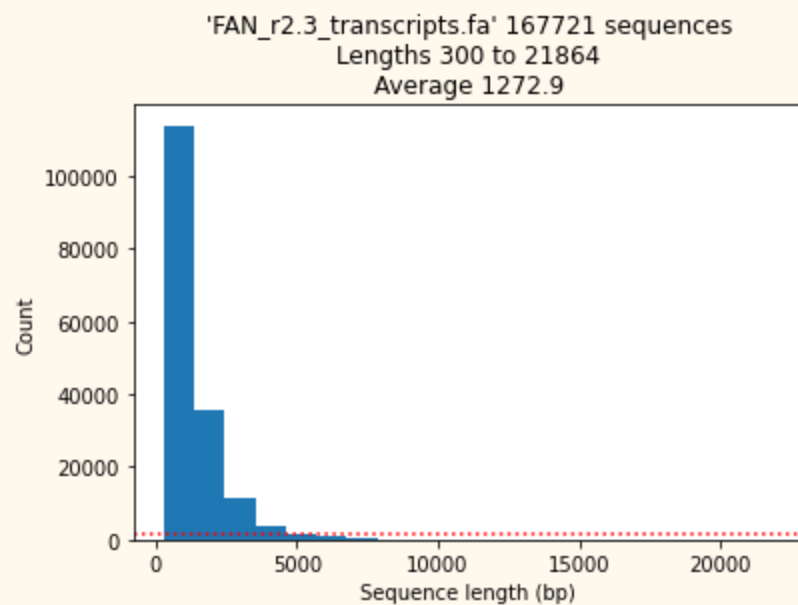
BioPython Use with the Organisms Output



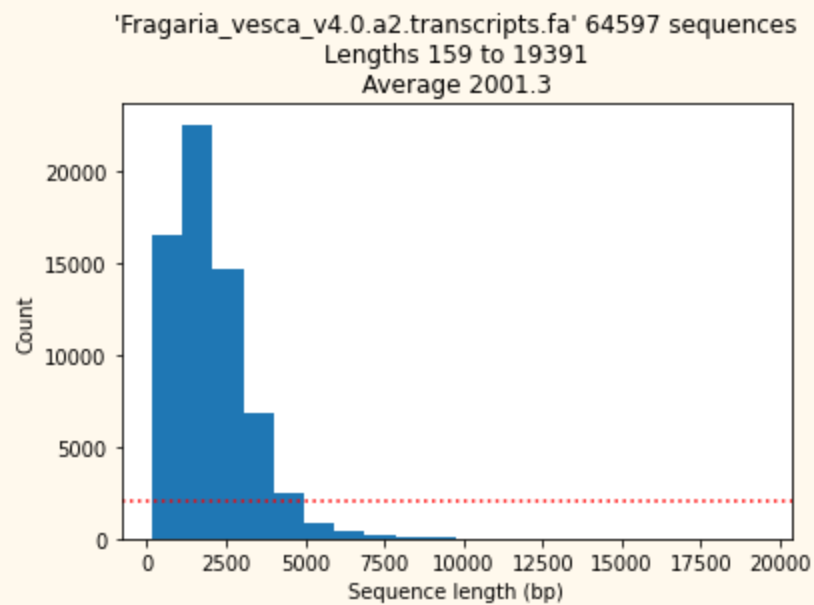
Here is the GC content of *Fragaria x ananassa* transcripts.



Here is the GC content of *Fragaria vesca* transcripts.



Here is the average sequence length of *Fragaria x ananassa* transcripts.



Here is the average sequence length of *Fragaria vesca* transcripts.

Before any programming can take place the correct libraries must be imported. These are all of the imports necessary for this tutorial. Importing `SeqIO`, `GC`, and `StringIO` in this way allows for direct calling. This will be more clear in a later cell.

```
In [1]: from Bio.Seq import Seq
from Bio import SeqIO
from Bio.SeqUtils import GC
from io import StringIO
import sys
import pylab
import os
import numpy as np

Now the imports are done it is time to load in and parse some fasta files. What 'direct calling' means will become more obvious. Calling
SeqIO's parse function is simple now. Since this is a fasta file the second argument (filetype) written is fasta. This can be done for
different types of files, but fasta will be the only one used in this tutorial.
```

```
In [2]: FAN_r2 = 'FAN_r2.3.transcripts.fa'
anan_Trs = SeqIO.parse(FAN_r2,'fasta') #ananas transcripts
```

To view it's contents I'll make a function that loops through the contents of the parsed file. I'll stop at 10 lines to reduce computation time. Since I'll only be passing fasta objects through it I can grab Python's fasta parameters. In this case I'm grabbing 'id' and 'seq'.

```
In [3]: def printFastaContents(fasta):
    i=0
    for record in fasta:
        i+=1
        print("ID:", record.id)
        print((repr(record.seq)))
        print("Length:",len(record))
        if i>10: break

Now I'll run it through with the ananassa transcript variable I made in cell 4.
```

```
In [4]: printFastaContents(anan_Trs)

ID: FAN_r2.3.chiIaVa_g000010.1
Seq('ATGCTCTGTCACAGATGCTTATCAACCTCGACACTCTCTGACTCCACTGACAG...GAG')
Length: 666
ID: FAN_r2.3.chiIaVa_g000020.1
Seq('CCCCCTTGCACAGATGACCCCTTAGACGAATCTTTATCTCTAGAGGGGATCTAC...CTC')
Length: 342
ID: FAN_r2.3.chiIaVa_g000030.1
Seq('ATGAGGAGAGAGAGAGCTTATGCTCAAGATGCTGCTCAATCTCGAGAGAGGCT...TAG')
Length: 2178
ID: FAN_r2.3.chiIaVa_g000040.1
Seq('ATGTTTACTGCTATGACAGAGTATATCAAAAGGCTTTCTATTGGGATCAAGCT...TAG')
Length: 873
ID: FAN_r2.3.chiIaVa_g000050.1
Seq('CCCCCTCTGAAATCTTTGATCTCATAACCAAGGTAGCTGTGTGATCATCTCAA...TGA')
Length: 609
ID: FAN_r2.3.chiIaVa_g000060.1
Seq('ATGCTGTTTACATGAAAGAGAGGATTTACTTACCAAAACCAAGAGAGCTGTGCAAGGCG...GTC')
Length: 675
ID: FAN_r2.3.chiIaVa_g000070.1
Seq('ATGTTGGGCAATCATCATCATCAGCAGCAGCTCTCTGCTACTCTGCTCTCTCCAGCA...TAA')
Length: 885
ID: FAN_r2.3.chiIaVa_g000080.1
Seq('ATGAGATGCGCCCTCTTTGTTTGCATATAAAGATCTCTCTCGGCCCTCAAGAT...TGA')
Length: 588
ID: FAN_r2.3.chiIaVa_g000090.1
Seq('ATGATTTTGGAAGAACCAATTTAGTCAGCAGGTAATATACATCAATGAATAAGCGG...TAA')
Length: 360
ID: FAN_r2.3.chiIaVa_g000100.1
Seq('ATGCTGCGAGAGAGCTGCGCGAGGCGAGACAGGAGATGCCCGAGAGAGCTGGCGA...TAG')
Length: 462
ID: FAN_r2.3.chiIaVa_g000110.1
Seq('ATGTTGGTGTTCGCGCGCGCTGGCGAGCTCATCTCCGCGTCTGGTTGGGAGAG...TAA')
Length: 318
```

You'll notice it prints out a shortened version of the transcript. The Python function 'repr' shortens outputs and makes them 'printable'. I've decided to use this because Python files can become too large and 'break' if the outputs are too long. I'll now load in ananassa's proteins and fragaria vesca's transcripts and proteins in the same way I did ananassa's transcripts.

```
In [5]: type(anan_Trs)

Out[5]: Bio.SeqIO.FastaIO.FastaIterator
```

```
In [7]: FXAv = 'Fxa_v1.2_makerStandard_proteins_wPotposases.fasta'
anan_Prot = SeqIO.parse(FXAv,'fasta')
```

```
In [8]: vesca_v4 = 'Fragaria vesca_v4.0.a2.transcripts.fa'
vesca_Trs = SeqIO.parse(vesca_v4,'fasta') #vesca transcript
```

```
In [9]: vesca_v4_prot = 'Fragaria vesca_v4.0.a1_makerStandard_proteins.fasta'
vesca_Prot = SeqIO.parse(vesca_v4_prot,'fasta') #vesca proteins
```

Now lets look through the contents of our newly loaded files!

```
In [10]: print('Ananassa Transcripts')
printFastaContents(vesca_Trs)

Vesca Transcripts
ID: FvH4_lg00010.t1
Seq('TCTCTTTGCTCTACACTTTATTCTTCTGTCAGATATATATTATTTCAGTTTATA...AAG')
Length: 1436
ID: FvH4_lg00020.t2
Seq('CCTTACTTAATCAATCAATCAACGAGCAAGCTGACACAGATTACAGATCGTGA...ATG')
Length: 1625
ID: FvH4_lg00020.t11
Seq('CTGATGTAATCAATCAATCAACGAGCAAGCTGACACAGATTACAGATCGTGA...GCA')
Length: 2898
ID: FvH4_lg00020.t12
Seq('CCTTACTTAATCAATCAATCAACGAGCAAGCTGACACAGATTACAGATCGTGA...GCA')
Length: 1721
ID: FvH4_lg00020.t14
Seq('CCTTACTTAATCAATCAATCAACGAGCAAGCTGACACAGATTACAGATCGTGA...GCA')
Length: 1721
ID: FvH4_lg00020.t3
Seq('CCTTACTTAATCAATCAATCAACGAGCAAGCTGACACAGATTACAGATCGTGA...GCA')
Length: 1633
ID: FvH4_lg00020.t7
Seq('CCTTACTTAATCAATCAATCAACGAGCAAGCTGACACAGATTACAGATCGTGA...GCA')
Length: 1611
ID: FvH4_lg00020.t9
Seq('ACGAGCAAGCAAGCAAGATTACAGATCGATGATTAACCAATTAACATGTCAGATG...GCA')
Length: 1742
ID: FvH4_lg00020.t6
Seq('ACGAGCAAGCAAGCAAGATTACAGATCGATGATTAACCAATTAACATGTCAGATG...GCA')
Length: 1701
ID: FvH4_lg00020.t10
Seq('CGAATCTGAACACAGATTACAGATCGATGATTAACCAATTAACATGTCAGATG...AGG')
Length: 931
```

```
In [11]: print('Ananassa Proteins')
printFastaContents(anan_Prot)

Ananassa Proteins
ID: maker-FvH4-2-snap-gene-0.67-mRNA-1
Seq('MSKVNDEGERTEMENLKLGLMDGLFTNNHQRLKLSLPGSSSTESQSDSPSSA...KKS')
Length: 743
ID: maker-FvH4-2-augustus-gene-0.46-mRNA-1
Seq('MDLYGATISSSLLIARFENFVSKLIVKSPFRILSPDALNDGLCHVFRYPDT...EYL')
Length: 181
ID: augustus_masked-FvH4-2-processed-gene-0.73-mRNA-1
Seq('MGNYISCLATPLIKSTMAARV1FPTGEVQREPIHARLIMLESPPNFIANSK...FSR')
Length: 161
ID: maker-FvH4-2-snap-gene-0.70-mRNA-1
Seq('MGVVNDGDTISGPLRPQVGVHVHPVPSMSRA1DITLGGTQAIHKHASSNNIN...HNV')
Length: 1274
ID: maker-FvH4-2-snap-gene-0.60-mRNA-1
Seq('MVQSSSIVDSNIGFGRLEPQVNLKQSRGLGADAPNKRPLVUSNBSGSM...DNV')
Length: 187
ID: maker-FvH4-2-augustus-gene-0.51-mRNA-1
Seq('MQGSGVSDFTFVPIVILVSVARAKSGCENRIMINADIBGLIKLNDGRIPAT...LVR')
Length: 140
ID: maker-FvH4-2-augustus-gene-0.56-mRNA-1
Seq('MQGSGVSDFTFVPIVILVSVARAKSGCENRIMINADIBGLIKLNDGRIPAT...LVR')
Length: 140
ID: maker-FvH4-2-augustus-gene-0.55-mRNA-1
Seq('MHNHSGGADAVTKRMIRLITVLIMWLAGATASPVKRLMBAIDQQRRLNLANG...PIS')
Length: 345
ID: maker-FvH4-2-snap-gene-0.75-mRNA-1
Seq('MISRACGCSFIVKPAAGVRCBAAGVFTFGRDPNPKVKAQWLRKLAQGIYEEV...AIR')
Length: 333
ID: maker-FvH4-2-snap-gene-0.77-mRNA-1
Seq('MASALATACCSQLLPFRCLGFLPLFTVSPAFVILPVCVKRGRLPILAAAKSNIS...DSD')
Length: 319
```

```
In [12]: print('Vesca Proteins')
printFastaContents(vesca_Prot)

Vesca Proteins
ID: FvH4_c1g00030.1
Seq('MTVVHVLATRSRSDVEGANTSPFPVVPFRQQAATCFKSPQAFKARKTRMGEG...PLQ')
Length: 80
ID: FvH4_c1g00020.1
Seq('MTVVHVLATRSRSDVEGANTSPFPVVPFRQQAATCFKSPQAFKARKTRMGEG...PLQ')
Length: 80
ID: FvH4_c1g00010.1
Seq('MTVVHVLATRSRSDVEGANTSPFPVVPFRQQAATCFKSPQAFKARKTRMGEG...PLQ')
Length: 80
ID: FvH4_c1g00020.1
Seq('MELAYRKRHRKLRSDFPHCLGAPASLSLITDADAPIRPAIGKRRQE...LDT')
Length: 80
ID: FvH4_c1g00010.1
Seq('MCTGTFSLIAMELARPGDILGSHNLHVLNLTIAHAFILPFMMWPMNIGGSSN...YVK')
Length: 476
ID: FvH4_c1g00060.1
Seq('MAFSGCTFKDPKAPVLLRLYLRYKRVKVMRDSIATDILDRLLPFLGEGTILIF...LEP')
Length: 68
ID: FvH4_c1g00130.1
Seq('MHSATIRSGVHAGDPLGLCDIDISFTQGFASRIDELEMLLTGNRIWQRILVDI...VDR')
Length: 242
ID: FvH4_c1g00080.1
Seq('MQGSPASIVLPTVAABASLGLAIFVITFVRVGTIVAVSINSIGCGERTFESLP...QKS')
Length: 196
ID: FvH4_c1g00090.1
Seq('MQGTEGLIGHSKFDYVQVSKPPEFDALIPFLSNDYDPVPPVTGPTGIG...LIA')
Length: 273
ID: FvH4_c1g00250.1
Seq('MAVFPVRKGFDPVAGFLQIPLADGLKLILKEPISPSANFSLFPMAPVTFML...SSN')
Length: 71
ID: FvH4_c1g00460.1
Seq('MHNHSGGADAVTKRMIRLITVLIMWLAGATASPVKRLMBAIDQQRRLNLANG...PIS')
Length: 345
ID: FvH4_c1g00010.1
Seq('MCTGTFSLIAMELARPGDILGSHNLHVLNLTIAHAFILPFMMWPMNIGGSSN...YVK')
Length: 476
ID: FvH4_c1g00060.1
Seq('MAFSGCTFKDPKAPVLLRLYLRYKRVKVMRDSIATDILDRLLPFLGEGTILIF...LEP')
Length: 68
ID: FvH4_c1g00130.1
Seq('MHSATIRSGVHAGDPLGLCDIDISFTQGFASRIDELEMLLTGNRIWQRILVDI...VDR')
Length: 242
ID: FvH4_c1g00080.1
Seq('MQGSPASIVLPTVAABASLGLAIFVITFVRVGTIVAVSINSIGCGERTFESLP...QKS')
Length: 196
ID: FvH4_c1g00090.1
Seq('MQGTEGLIGHSKFDYVQVSKPPEFDALIPFLSNDYDPVPPVTGPTGIG...LIA')
Length: 273
ID: FvH4_c1g00250.1
Seq('MAVFPVRKGFDPVAGFLQIPLADGLKLILKEPISPSANFSLFPMAPVTFML...SSN')
Length: 71
ID: FvH4_c1g00460.1
Seq('MHNHSGGADAVTKRMIRLITVLIMWLAGATASPVKRLMBAIDQQRRLNLANG...PIS')
Length: 345
ID: FvH4_c1g00010.1
Seq('MCTGTFSLIAMELARPGDILGSHNLHVLNLTIAHAFILPFMMWPMNIGGSSN...YVK')
Length: 476
ID: FvH4_c1g00060.1
Seq('MAFSGCTFKDPKAPVLLRLYLRYKRVKVMRDSIATDILDRLLPFLGEGTILIF...LEP')
Length: 68
ID: FvH4_c1g00130.1
Seq('MHSATIRSGVHAGDPLGLCDIDISFTQGFASRIDELEMLLTGNRIWQRILVDI...VDR')
Length: 242
ID: FvH4_c1g00080.1
Seq('MQGSPASIVLPTVAABASLGLAIFVITFVRVGTIVAVSINSIGCGERTFESLP...QKS')
Length: 196
ID: FvH4_c1g00090.1
Seq('MQGTEGLIGHSKFDYVQVSKPPEFDALIPFLSNDYDPVPPVTGPTGIG...LIA')
Length: 273
ID: FvH4_c1g00250.1
Seq('MAVFPVRKGFDPVAGFLQIPLADGLKLILKEPISPSANFSLFPMAPVTFML...SSN')
Length: 71
ID: FvH4_c1g00460.1
Seq('MHNHSGGADAVTKRMIRLITVLIMWLAGATASPVKRLMBAIDQQRRLNLANG...PIS')
Length: 345
ID: FvH4_c1g00010.1
Seq('MCTGTFSLIAMELARPGDILGSHNLHVLNLTIAHAFILPFMMWPMNIGGSSN...YVK')
Length: 476
ID: FvH4_c1g00060.1
Seq('MAFSGCTFKDPKAPVLLRLYLRYKRVKVMRDSIATDILDRLLPFLGEGTILIF...LEP')
Length: 68
ID: FvH4_c1g00130.1
Seq('MHSATIRSGVHAGDPLGLCDIDISFTQGFASRIDELEMLLTGNRIWQRILVDI...VDR')
Length: 242
ID: FvH4_c1g00080.1
Seq('MQGSPASIVLPTVAABASLGLAIFVITFVRVGTIVAVSINSIGCGERTFESLP...QKS')
Length: 196
ID: FvH4_c1g00090.1
Seq('MQGTEGLIGHSKFDYVQVSKPPEFDALIPFLSNDYDPVPPVTGPTGIG...LIA')
Length: 273
ID: FvH4_c1g00250.1
Seq('MAVFPVRKGFDPVAGFLQIPLADGLKLILKEPISPSANFSLFPMAPVTFML...SSN')
Length: 71
ID: FvH4_c1g00460.1
Seq('MHNHSGGADAVTKRMIRLITVLIMWLAGATASPVKRLMBAIDQQRRLNLANG...PIS')
Length: 345
ID: FvH4_c1g00010.1
Seq('MCTGTFSLIAMELARPGDILGSHNLHVLNLTIAHAFILPFMMWPMNIGGSSN...YVK')
Length: 476
ID: FvH4_c1g00060.1
Seq('MAFSGCTFKDPKAPVLLRLYLRYKRVKVMRDSIATDILDRLLPFLGEGTILIF...LEP')
Length: 68
ID: FvH4_c1g00130.1
Seq('MHSATIRSGVHAGDPLGLCDIDISFTQGFASRIDELEMLLTGNRIWQRILVDI...VDR')
Length: 242
ID: FvH4_c1g00080.1
Seq('MQGSPASIVLPTVAABASLGLAIFVITFVRVGTIVAVSINSIGCGERTFESLP...QKS')
Length: 196
ID: FvH4_c1g00090.1
Seq('MQGTEGLIGHSKFDYVQVSKPPEFDALIPFLSNDYDPVPPVTGPTGIG...LIA')
Length: 273
ID: FvH4_c1g00250.1
Seq('MAVFPVRKGFDPVAGFLQIPLADGLKLILKEPISPSANFSLFPMAPVTFML...SSN')
Length: 71
ID: FvH4_c1g00460.1
Seq('MHNHSGGADAVTKRMIRLITVLIMWLAGATASPVKRLMBAIDQQRRLNLANG...PIS')
Length: 345
ID: FvH4_c1g00010.1
Seq('MCTGTFSLIAMELARPGDILGSHNLHVLNLTIAHAFILPFMMWPMNIGGSSN...YVK')
Length: 476
ID: FvH4_c1g00060.1
Seq('MAFSGCTFKDPKAPVLLRLYLRYKRVKVMRDSIATDILDRLLPFLGEGTILIF...LEP')
Length: 68
ID: FvH4_c1g00130.1
Seq('MHSATIRSGVHAGDPLGLCDIDISFTQGFASRIDELEMLLTGNRIWQRILVDI...VDR')
Length: 242
ID: FvH4_c1g00080.1
Seq('MQGSPASIVLPTVAABASLGLAIFVITFVRVGTIVAVSINSIGCGERTFESLP...QKS')
Length: 196
ID: FvH4_c1g00090.1
Seq('MQGTEGLIGHSKFDYVQVSKPPEFDALIPFLSNDYDPVPPVTGPTGIG...LIA')
Length: 273
ID: FvH4_c1g00250.1
Seq('MAVFPVRKGFDPVAGFLQIPLADGLKLILKEPISPSANFSLFPMAPVTFML...SSN')
Length: 71
ID: FvH4_c1g00460.1
Seq('MHNHSGGADAVTKRMIRLITVLIMWLAGATASPVKRLMBAIDQQRRLNLANG...PIS')
Length: 345
ID: FvH4_c1g00010.1
Seq('MCTGTFSLIAMELARPGDILGSHNLHVLNLTIAHAFILPFMMWPMNIGGSSN...YVK')
Length: 476
ID: FvH4_c1g00060.1
Seq('MAFSGCTFKDPKAPVLLRLYLRYKRVKVMRDSIATDILDRLLPFLGEGTILIF...LEP')
Length: 68
ID: FvH4_c1g00130.1
Seq('MHSATIRSGVHAGDPLGLCDIDISFTQGFASRIDELEMLLTGNRIWQRILVDI...VDR')
Length: 242
ID: FvH4_c1g00080.1
Seq('MQGSPASIVLPTVAABASLGLAIFVITFVRVGTIVAVSINSIGCGERTFESLP...QKS')
Length: 196
ID: FvH4_c1g00090.1
Seq('MQGTEGLIGHSKFDYVQVSKPPEFDALIPFLSNDYDPVPPVTGPTGIG...LIA')
Length: 273
ID: FvH4_c1g00250.1
Seq('MAVFPVRKGFDPVAGFLQIPLADGLKLILKEPISPSANFSLFPMAPVTFML...SSN')
Length: 71
ID: FvH4_c1g00460.1
Seq('MHNHSGGADAVTKRMIRLITVLIMWLAGATASPVKRLMBAIDQQRRLNLANG...PIS')
Length: 345
ID: FvH4_c1g00010.1
Seq('MCTGTFSLIAMELARPGDILGSHNLHVLNLTIAHAFILPFMMWPMNIGGSSN...YVK')
Length: 476
ID: FvH4_c1g00060.1
Seq('MAFSGCTFKDPKAPVLLRLYLRYKRVKVMRDSIATDILDRLLPFLGEGTILIF...LEP')
Length: 68
ID: FvH4_c1g00130.1
Seq('MHSATIRSGVHAGDPLGLCDIDISFTQGFASRIDELEMLLTGNRIWQRILVDI...VDR')
Length: 242
ID: FvH4_c1g00080.1
Seq('MQGSPASIVLPTVAABASLGLAIFVITFVRVGTIVAVSINSIGCGERTFESLP...QKS')
Length: 196
ID: FvH4_c1g00090.1
Seq('MQGTEGLIGHSKFDYVQVSKPPEFDALIPFLSNDYDPVPPVTGPTGIG...LIA')
Length: 273
ID: FvH4_c1g00250.1
Seq('MAVFPVRKGFDPVAGFLQIPLADGLKLILKEPISPSANFSLFPMAPVTFML...SSN')
Length: 71
ID: FvH4_c1g00460.1
Seq('MHNHSGGADAVTKRMIRLITVLIMWLAGATASPVKRLMBAIDQQRRLNLANG...PIS')
Length: 345
ID: FvH4_c1g00010.1
Seq('MCTGTFSLIAMELARPGDILGSHNLHVLNLTIAHAFILPFMMWPMNIGGSSN...YVK')
Length: 476
ID: FvH4_c1g00060.1
Seq('MAFSGCTFKDPKAPVLLRLYLRYKRVKVMRDSIATDILDRLLPFLGEGTILIF...LEP')
Length: 68
ID: FvH4_c1g00130.1
Seq('MHSATIRSGVHAGDPLGLCDIDISFTQGFASRIDELEMLLTGNRIWQRILVDI...VDR')
Length: 242
ID: FvH4_c1g00080.1
Seq('MQGSPASIVLPTVAABASLGLAIFVITFVRVGTIVAVSINSIGCGERTFESLP...QKS')
Length: 196
ID: FvH4_c1g00090.1
Seq('MQGTEGLIGHSKFDYVQVSKPPEFDALIPFLSNDYDPVPPVTGPTGIG...LIA')
Length: 273
ID: FvH4_c1g00250.1
Seq('MAVFPVRKGFDPVAGFLQIPLADGLKLILKEPISPSANFSLFPMAPVTFML...SSN')
Length: 71
ID: FvH4_c1g00460.1
Seq('MHNHSGGADAVTKRMIRLITVLIMWLAGATASPVKRLMBAIDQQRRLNLANG...PIS')
Length: 345
ID: FvH4_c1g00010.1
Seq('MCTGTFSLIAMELARPGDILGSHNLHVLNLTIAHAFILPFMMWPMNIGGSSN...YVK')
Length: 476
ID: FvH4_c1g00060.1
Seq('MAFSGCTFKDPKAPVLLRLYLRYKRVKVMRDSIATDILDRLLPFLGEGTILIF...LEP')
Length: 68
ID: FvH4_c1g00130.1
Seq('MHSATIRSGVHAGDPLGLCDIDISFTQGFASRIDELEMLLTGNRIWQRILVDI...VDR')
Length: 242
ID: FvH4_c1g00080.1
Seq('MQGSPASIVLPTVAABASLGLAIFVITFVRVGTIVAVSINSIGCGERTFESLP...QKS')
Length: 196
ID: FvH4_c1g00090.1
Seq('MQGTEGLIGHSKFDYVQVSKPPEFDALIPFLSNDYDPVPPVTGPTGIG...LIA')
Length: 273
ID: FvH4_c1g00250.1
Seq('MAVFPVRKGFDPVAGFLQIPLADGLKLILKEPISPSANFSLFPMAPVTFML...SSN')
Length: 71
ID: FvH4_c1g00460.1
Seq('MHNHSGGADAVTKRMIRLITVLIMWLAGATASPVKRLMBAIDQQRRLNLANG...PIS')
Length: 345
ID: FvH4_c1g00010.1
Seq('MCTGTFSLIAMELARPGDILGSHNLHVLNLTIAHAFILPFMMWPMNIGGSSN...YVK')
Length: 476
ID: FvH4_c1g00060.1
Seq('MAFSGCTFKDPKAPVLLRLYLRYKRVKVMRDSIATDILDRLLPFLGEGTILIF...LEP')
Length: 68
ID: FvH4_c1g00130.1
Seq('MHSATIRSGVHAGDPLGLCDIDISFTQGFASRIDELEMLLTGNRIWQRILVDI...VDR')
Length: 242
ID: FvH4_c1g00080.1
Seq('MQGSPASIVLPTVAABASLGLAIFVITFVRVGTIVAVSINSIGCGERTFESLP...QKS')
Length: 196
ID: FvH4_c1g00090.1
Seq('MQGTEGLIGHSKFDYVQVSKPPEFDALIPFLSNDYDPVPPVTGPTGIG...LIA')
Length: 273
ID: FvH4_c1g00250.1
Seq('MAVFPVRKGFDPVAGFLQIPLADGLKLILKEPISPSANFSLFPMAPVTFML...SSN')
Length: 71
ID: FvH4_c1g00460.1
Seq('MHNHSGGADAVTKRMIRLITVLIMWLAGATASPVKRLMBAIDQQRRLNLANG...PIS')
Length: 345
ID: FvH4_c1g00010.1
Seq('MCTGTFSLIAMELARPGDILGSHNLHVLNLTIAHAFILPFMMWPMNIGGSSN...YVK')
Length: 476
ID: FvH4_c1g00060.1
Seq('MAFSGCTFKDPKAPVLLRLYLRYKRVKVMRDSIATDILDRLLPFLGEGTILIF...LEP')
Length: 68
ID: FvH4_c1g00130.1
Seq('MHSATIRSGVHAGDPLGLCDIDISFTQGFASRIDELEMLLTGNRIWQRILVDI...VDR')
Length: 242
ID: FvH4_c1g00080.1
Seq('MQGSPASIVLPTVAABASLGLAIFVITFVRVGTIVAVSINSIGCGERTFESLP...QKS')
Length: 196
ID: FvH4_c1g00090.1
Seq('MQGTEGLIGHSKFDYVQVSKPPEFDALIPFLSNDYDPVPPVTGPTGIG...LIA')
Length: 273
ID: FvH4_c1g00250.1
Seq('MAVFPVRKGFDPVAGFLQIPLADGLKLILKEPISPSANFSLFPMAPVTFML...SSN')
Length: 71
ID: FvH4_c1g00460.1
Seq('MHNHSGGADAVTKRMIRLITVLIMWLAGATASPVKRLMBAIDQQRRLNLANG...PIS')
Length: 345
ID: FvH4_c1g00010.1
Seq('MCTGTFSLIAMELARPGDILGSHNLHVLNLTIAHAFILPFMMWPMNIGGSSN...YVK')
Length: 476
ID: FvH4_c1g00060.1
Seq('MAFSGCTFKDPKAPVLLRLYLRYKRVKVMRDSIATDILDRLLPFLGEGTILIF...LEP')
Length: 68
ID: FvH4_c1g00130.1
Seq('MHSATIRSGVHAGDPLGLCDIDISFTQGFASRIDELEMLLTGNRIWQRILVDI...VDR')
Length: 242
ID: FvH4_c1g00080.1
Seq('MQGSPASIVLPTVAABASLGLAIFVITFVRVGTIVAVSINSIGCGERTFESLP...QKS')
Length: 196
ID: FvH4_c1g00090.1
Seq('MQGTEGLIGHSKFDYVQVSKPPEFDALIPFLSNDYDPVPPVTGPTGIG...LIA')
Length: 273
ID: FvH4_c1g00250.1
Seq('MAVFPVRKGFDPVAGFLQIPLADGLKLILKEPISPSANFSLFPMAPVTFML...SSN')
Length: 71
ID: FvH4_c1g00460.1
Seq('MHNHSGGADAVTKRMIRLITVLIMWLAGATASPVKRLMBAIDQQRRLNLANG...PIS')
Length: 345
ID: FvH4_c1g00010.1
Seq('MCTGTFSLIAMELARPGDILGSHNLHVLNLTIAHAFILPFMMWPMNIGGSSN...YVK')
Length: 476
ID: FvH4_c1g00060.1
Seq('MAFSGCTFKDPKAPVLLRLYLRYKRVKVMRDSIATDILDRLLPFLGEGTILIF...LEP')
Length: 68
ID: FvH4_c1g00130.1
Seq('MHSATIRSGVHAGDPLGLCDIDISFTQGFASRIDELEMLLTGNRIWQRILVDI...VDR')
Length: 242
ID: FvH4_c1g00080.1
Seq('MQGSPASIVLPTVAABASLGLAIFVITFVRVGTIVAVSINSIGCGERTFESLP...QKS')
Length: 196
ID: FvH4_c1g00090.1
Seq('MQGTEGLIGHSKFDYVQVSKPPEFDALIPFLSNDYDPVPPVTGPTGIG...LIA')
Length: 273
ID: FvH4_c1g00250.1
Seq('MAVFPVRKGFDPVAGFLQIPLADGLKLILKEPISPSANFSLFPMAPVTFML...SSN')
Length: 71
ID: FvH4_c1g00460.1
Seq('MHNHSGGADAVTKRMIRLITVLIMWLAGATASPVKRLMBAIDQQRRLNLANG...PIS')
Length: 345
ID: FvH4_c1g00010.1
Seq('MCTGTFSLIAMELARPGDILGSHNLHVLNLTIAHAFILPFMMWPMNIGGSSN...YVK')
Length: 476
ID: FvH4_c1g00060.1
Seq('MAFSGCTFKDPKAPVLLRLYLRYKRVKVMRDSIATDILDRLLPFLGEGTILIF...LEP')
Length: 68
ID: FvH4_c1g00130.1
Seq('MHSATIRSGVHAGDPLGLCDIDISFTQGFASRIDELEMLLTGNRIWQRILVDI...VDR')
Length: 242
ID: FvH4_c1g00080.1
Seq('MQGSPASIVLPTVAABASLGLAIFVITFVRVGTIVAVSINSIGCGERTFESLP...QKS')
Length: 196
ID: FvH4_c1g00090.1
Seq('MQGTEGLIGHSKFDYVQVSKPPEFDALIPFLSNDYDPVPPVTGPTGIG...LIA')
Length: 273
ID: FvH4_c1g00250.1
Seq('MAVFPVRKGFDPVAGFLQIPLADGLKLILKEPISPSANFSLFPMAPVTFML...SSN')
Length: 71
ID: FvH4_c1g00460.1
Seq('MHNHSGGADAVTKRMIRLITVLIMWLAGATASPVKRLMBAIDQQRRLNLANG...PIS')
Length: 345
ID: FvH4_c1g00010.1
Seq('MCTGTFSLIAMELARPGDILGSHNLHVLNLTIAHAFILPFMMWPMNIGGSSN...YVK')
Length: 476
ID: FvH4_c1g00060.1
Seq('MAFSGCTFKDPKAPVLLRLYLRYKRVKVMRDSIATDILDRLLPFLGEGTILIF...LEP')
Length: 68
ID: FvH4_c1g00130.1
Seq('MHSATIRSGVHAGDPLGLCDIDISFTQGFASRIDELEMLLTGNRIWQRILVDI...VDR')
Length: 242
ID: FvH4_c1g00080.1
Seq('MQGSPASIVLPTVAABASLGLAIFVITFVRVGTIVAVSINSIGCGERTFESLP...QKS')
Length: 196
ID: FvH4_c1g00090.1
Seq('MQGTEGLIGHSKFDYVQVSKPPEFDALIPFLSNDYDPVPPVTGPTGIG...LIA')
Length: 273
ID: FvH4_c1g00250.1
Seq('MAVFPVRKGFDPVAGFLQIPLADGLKLILKEPISPSANFSLFPMAPVTFML...SSN')
Length: 71
ID: FvH4_c1g00460.1
Seq('MHNHSGGADAVTKRMIRLITVLIMWLAGATASPVKRLMBAIDQQRRLNLANG...PIS')
Length: 345
ID: FvH4_c1g00010.1
Seq('MCTGTFSLIAMELARPGDILGSHNLHVLNLTIAHAFILPFMMWPMNIGGSSN...YVK')
Length: 476
ID: FvH4_c1g00060.1
Seq('MAFSGCTFKDPKAPVLLRLYLRYKRVKVMRDSIATDILDRLLPFLGEGTILIF...LEP')
Length: 68
ID: FvH4_c1g00130.1
Seq('MHSATIRSGVHAGDPLGLCDIDISFTQGFASRIDELEMLLTGNRIWQRILVDI...VDR')
Length: 242
ID: FvH4_c1g00080.1
Seq('MQGSPASIVLPTVAABASLGLAIFVITFVRVGTIVAVSINSIGCGERTFESLP...QKS')
Length: 196
ID: FvH4_c1g00090.1
Seq('MQGTEGLIGHSKFDYVQVSKPPEFDALIPFLSNDYDPVPPVTGPTGIG...LIA')
Length: 273
ID: FvH4_c1g00250.1
Seq('MAVFPVRKGFDPVAGFLQIPLADGLKLILKEPISPSANFSLFPMAPVTFML...SSN')
Length: 71
ID: FvH4_c1g00460.1
Seq('MHNHSGGADAVTKRMIRLITVLIMWLAGATASPVKRLMBAIDQQRRLNLANG...PIS')
Length: 345
ID: FvH4_c1g00010.1
Seq('MCTGTFSLIAMELARPGDILGSHNLHVLNLTIAHAFILPFMMWPMNIGGSSN...YVK')
Length: 476
ID: FvH4_c1g00060.1
Seq('MAFSGCTFKDPKAPVLLRLYLRYKRVKVMRDSIATDILDRLLPFLGEGTILIF...LEP')
Length: 68
ID: FvH4_c1g00130.1
Seq('MHSATIRSGVHAGDPLGLCDIDISFTQGFASRIDELEMLLTGNRIWQRILVDI...VDR')
Length: 242
ID: FvH4_c1g00080.1
Seq('MQGSPASIVLPTVAABASLGLAIFVITFVRVGTIVAVSINSIGCGERTFESLP...QKS')
Length: 196
ID: FvH4_c1g00090.1
Seq('MQGTEGLIGHSKFDYVQVSKPPEFDALIPFLSNDYDPVPPVTGPTGIG...LIA')
Length: 273
ID: FvH4_c1g00250.1
Seq('MAVFPVRKGFDPVAGFLQIPLADGLKLILKEPISPSANFSLFPMAPVTFML...SSN')
Length: 71
ID: FvH4_c1g00460.1
Seq('MHNHSGGADAVTKRMIRLITVLIMWLAGATASPVKRLMBAIDQQRRLNLANG...PIS')
Length: 345
ID: FvH4_c1g00010.1
Seq('MCTGTFSLIAMELARPGDILGSHNLHVLNLTIAHAFILPFMMWPMNIGGSSN...YVK')
Length: 476
ID: FvH4_c1g00060.1
Seq('MAFSGCTFKDPKAPVLLRLYLRYKRVKVMRDSIATDILDRLLPFLGEGTILIF...LEP')
Length: 68
ID: FvH4_c1g00130.1
Seq('MHSATIRSGVHAGDPLGLCDIDISFTQGFASRIDELEMLLTGNRIWQRILVDI...VDR')
Length: 242
ID: FvH4_c1g00080.1
Seq('MQGSPASIVLPTVAABASLGLAIFVITFVRVGTIVAVSINSIGCGERTFESLP...QKS')
Length: 196
ID: FvH4_c1g00090.1
Seq('MQGTEGLIGHSKFDYVQVSKPPEFDALIPFLSNDYDPVPPVTGPTGIG...LIA')
Length: 273
ID: FvH4_c1g00250.1
Seq('MAVFPVRKGFDPVAGFLQIPLADGLKLILKEPISPSANFSLFPMAPVTFML...SSN')
Length: 71
ID: FvH4_c1g00460.1
Seq('MHNHSGGADAVTKRMIRLITVLIMWLAGATASPVKRLMBAIDQQRRLNLANG...PIS')
Length: 345
ID: FvH4_c1g00010.1
Seq('MCTGTFSLIAMELARPGDILGSHNLHVLNLTIAHAFILPFMMWPMNIGGSSN...YVK')
Length: 476
ID: FvH4_c1g00060.1
Seq('MAFSGCTFKDPKAPVLLRLYLRYKRVKVMRDSIATDILDRLLPFLGEGTILIF...LEP')
Length: 68
ID: FvH4_c1g00130.1
Seq('MHSATIRSGVHAGDPLGLCDIDISFTQGFASRIDELEMLLTGNRIWQRILVDI...VDR')
Length: 242
ID: FvH4_c1g00080.1
Seq('MQGSPASIVLPTVAABASLGLAIFVITFVRVGTIVAVSINSIGCGERTFESLP...QKS')
Length: 196
ID: FvH4_c1g00090.1
Seq('MQGTEGLIGHSKFDYVQVSKPPEFDALIPFLSNDYDPVPPVTGPTGIG...LIA')
Length: 273
ID: FvH4_c1g00250.1
Seq('MAVFPVRKGFDPVAGFLQIPLADGLKLILKEPISPSANFSLFPMAPVTFML...SSN')
Length: 71
ID: FvH4_c1g00460.1
Seq('MHNHSGGADAVTKRMIRLITVLIMWLAGATASPVKRLMBAIDQQRRLNLANG...PIS')
Length: 345
ID: FvH4_c1g00010.1
Seq('MCTGTFSLIAMELARPGDILGSHNLHVLNLTIAHAFILPFMMWPMNIGGSSN...YVK')
Length: 476
ID: FvH4_c1g00060.1
Seq('MAFSGCTFKDPKAPVLLRLYLRYKRVKVMRDSIATDILDRLLPFLGEGTILIF...LEP')
Length: 68
ID: FvH4_c1g00130.1
Seq('MHSATIRSGVHAGDPLGLCDIDISFTQGFASRIDELEMLLTGNRIWQRILVDI...VDR')
Length: 242
ID: FvH4_c1g00080.1
Seq('MQGSPASIVLPTVAABASLGLAIFVITFVRVGTIVAVSINSIGCGERTFESLP...QKS')
Length: 196
ID: FvH4_c1g00090.1
Seq('MQGTEGLIGHSKFDYVQVSKPPEFDALIPFLSNDYDPVPPVTGPTGIG...LIA')
Length: 273
ID: FvH4_c1g00250.1
Seq('MAVFPVRKGFDPVAGFLQIPLADGLKLILKEPISPSANFSLFPMAPVTFML...SSN')
Length: 71
ID: FvH4_c1g00460.1
Seq('MHNHSGGADAVTKRMIRLITVLIMWLAGATASPVKRLMBAIDQQR
```