

RNA Secondary Structures: A Case Study on Viruses

Bioinformatics Senior Project

John Acampado

Under the guidance of Dr. Jason Wang



Table of Contents

- Overview
- RSpredict JAVA
- RSpredict WebServer
- RNAstructure
- Cis-Regulatory Element
- Virus Data
- Alignment
- Phylogenetic Tree
- RSpredict WebServer Results
- Analysis / Conclusion
- Resources
- Contact Information



Overview

- Secondary structure analysis of RNA in Bioinformatics
- Take various virus sequences that are cis-reg elements – see how viruses are related
- Use both RSpredict and RNAstructure programs
- Phylogenetic tree shows distance and relationships between sequences

RSpredict JAVA

- Used to effectively predict the secondary structure
- Takes into account sequence variation
- Uses FASTA file format for input, outputs CT and Vienna format
- Machine Settings:
 - Microsoft Windows XP Service Pack 2
 - Intel Pentium M 1.59GHz, 512MB RAM
- Link for [RSpredict JAVA](#)

RSpredict WebServer

- RSpredict program also available via a WebServer
- Accepts the more universal FASTA format
- Output still in CT and Vienna format
- Link for [RSpredict WebServer](#)



[Home](#) [Download](#) [Documentation](#) [Help](#)

RSpredict is an RNA secondary structure prediction tool that works on multiple sequence alignments. It takes into account sequence covariation and employs effective heuristics for accuracy improvement. RSpredict accepts, as input data, a multiple sequence alignment in either the Fasta or ClustalW format and outputs the consensus secondary structure of the input sequences in both the Vienna style Dot Bracket format and the Connectivity Table (CT) format.

Please enter the multiple sequence alignment in FASTA format below ([example](#)):

```
>AF047977.1/876-1015
AACUUYSCAUAUCGUAACAAAGCCGCGUACCGUAGCCGCUUUGAAGAGAGGUU
UACACGACUUCUUGUGGUGUAGCACGUAACUACCCUUGCCGACAGAUAAAGCCUU
UGUUUUCACAGUUUUUUUU
>S26017.1/6226-6865
AACUUYSCAUAUCGUAACAAAGCCGCGUACCGUAGCCGCUUUGAAGAGAGGUU
UACACGACUUCUUGUGGUGUAGCACGUAACUACCCUUGCCGACAGAUAAAGCCUU
UGUUUUCACAGUUUUUUUU
>AF077962.1/600-789
AACUUYSCAUAUCGUAACAAAGCCGCGUACCGUAGCCGCUUUGAAGAGAGGUU
UACACGACUUCUUGUGGUGUAGCACGUAACUACCCUUGCCGACAGAUAAAGCCUU
UGUUUUCACAGUUUUUUUU
>S77962.1/6225-6864
AACUUYSCAUAUCGUAACAAAGCCGCGUACCGUAGCCGCUUUGAAGAGAGGUU
UACACGACUUCUUGUGGUGUAGCACGUAACUACCCUUGCCGACAGAUAAAGCCUU
UGUUUUCACAGUUUUUUUU
```

RNAstructure

- Uses CT (Connectivity Table) from RSPredict to draw structure of sequence
- Developed at the University of Rochester Medical Center
- Used for prediction and analysis of RNA secondary structure
- Link to [RNAstructure](#)



Cis-Regulatory Elements

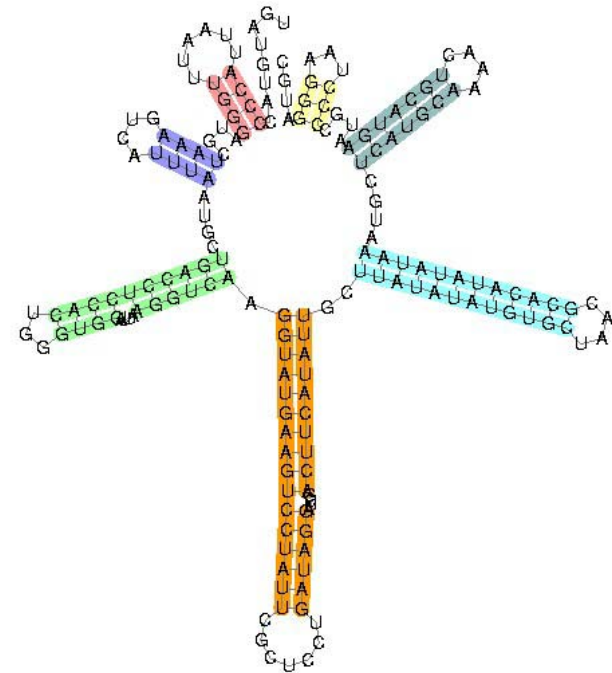
- Region of RNA that is able to regulate the expression of genes
- Often on binding sites of one or more trans-acting factors
- May be located in the promoter 5' region, or the 3' untranslated region
- Eleven viruses were used and analyzed for this project

Virus Data

- Gathered from RNA Families Database of Alignments and CMs ([Rfam](#))
- Sequences were chosen and entered manually
- Sequences of type “cis-reg”
- Sequences listed as virus within description
- All sequences chosen to have the ability to regulate gene expression
- Brief description and Rfam structure provided

Virus Data – Alfamo_CPB

- RNA element found in 3' UTR of genome
- Stimulates translation of AMV RNA up to 100 times more
- Contains at least two binding sites thought to be essential for efficient RNA translation

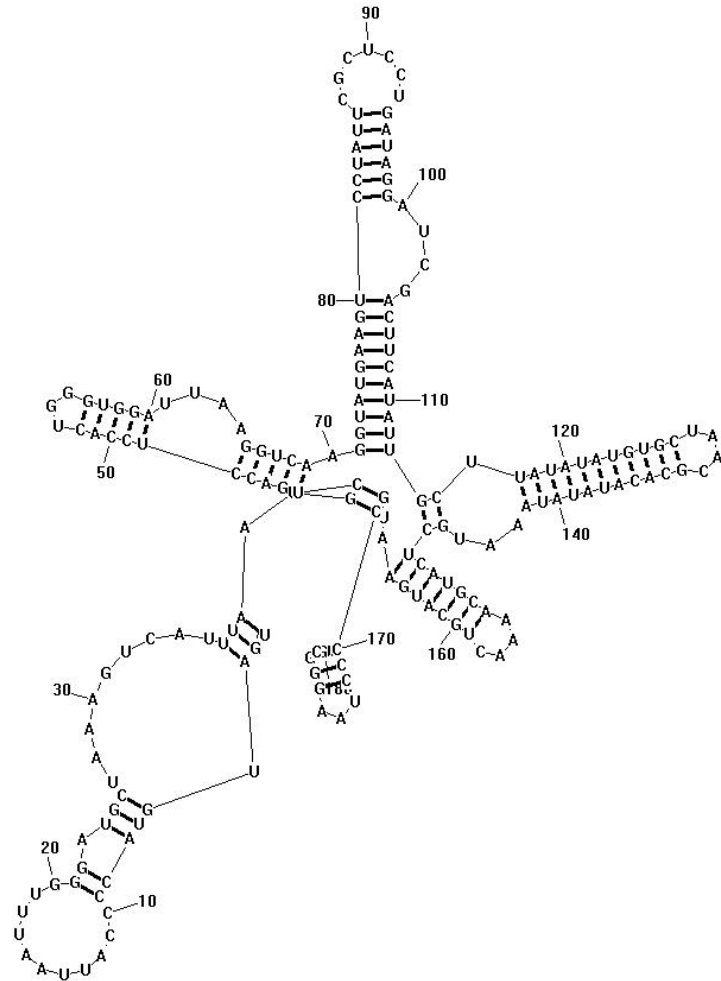


Virus Data – Alfamo_CPB

```
Command Prompt
C:\Documents and Settings\John Acampado>cd RSpredict
C:\Documents and Settings\John Acampado\RSpredict>java -Xmx512m RSpredict Alfamo_CPB.aln
Average Pairwise Sequenc Identity:      90%
Refining sequence alignment...
Computing the pairing threshold...
The pairing threshold is:      -0.15
Sequence alignment has 18 sequences.
Alignment length is: 182
Folding sequence alignment...
The final score is:      -17.38
Runing miliseconds:2133

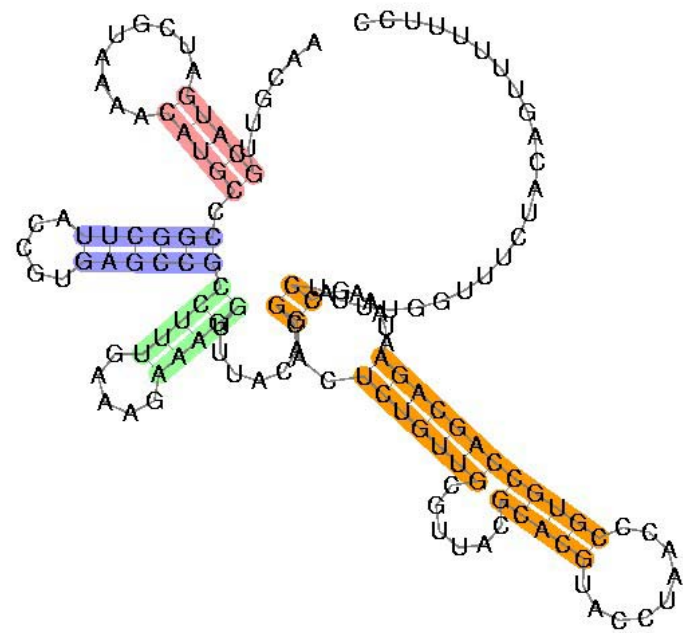
Predicted Structure:
UGAUGUACCCCAUUAAUUUGGGGaugcUAAAGUCAUUUAAUgcuGACCCUCCACUGGGUGGAUUAAAGGUCAAGGUaUGAAGU
CCUAUUcgcuCCUGAUAGGAUGGACUUCAUAUUgcuUAUAUAUGUGGCUAAcGCACAUAUAUAUAUAUGCUCAUGCAAACUG
CAUGAAUGCCCCUAAGGGGaugc
<<<<(((((.....))..)).....))>>>>..<<<<<<<<<<.....>>>>.....>>>>..<<<<<<<<<
<<<<<<.....>>>>.....>>>>)>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
>>>>..>>>><<<<.....>>>>.....
```

Virus Data – Alfamo_CPB

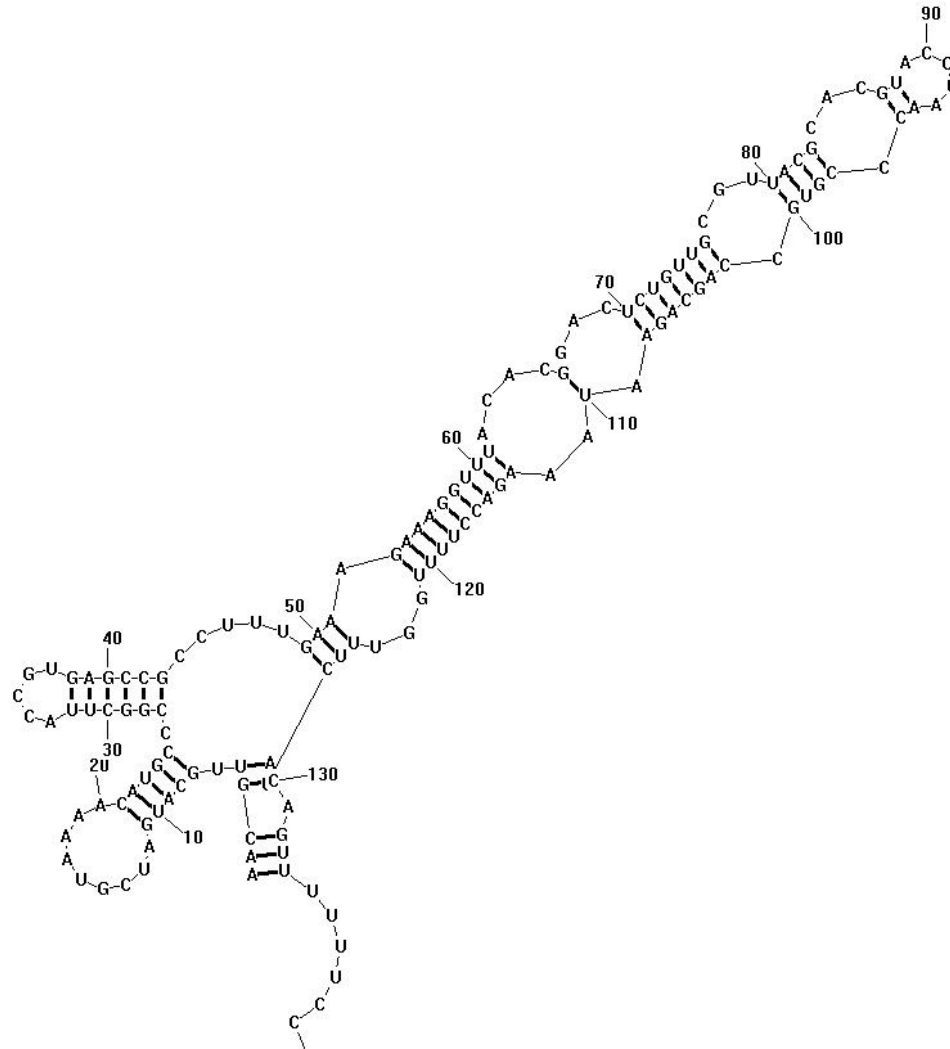


Virus Data – BaMV_CRE

- Family represents complex cloverleaf structure found in 3'UTR of genome
- Thought to play important role in initiation of minus strand RNA synthesis
- May also be involved with regulation of viral replication

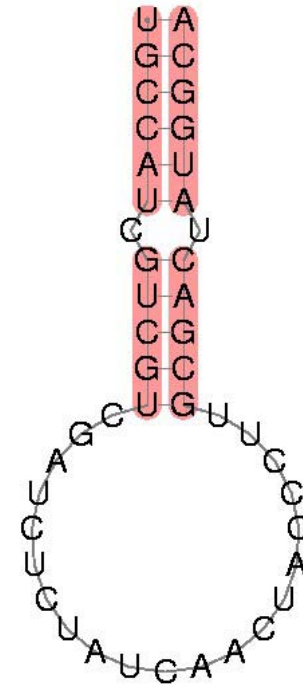


Virus Data – BaMV_CRE



Virus Data – EAV_LTH

- RNA element thought to be key structural element in subgenomic RNA synthesis
- Critical for leader transcription-regulating sequences
- Similar structures have been predicted in related arteriviruses and coronaviruses

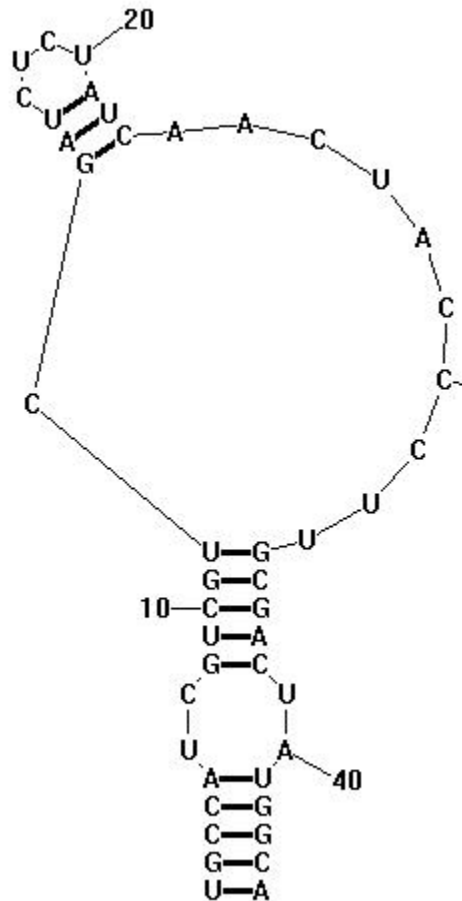


Virus Data – EAV_LTH

```
Command Prompt
C:\Documents and Settings\John Acampado\RSpredict>java -Xmx512m RSpredict EAV_LTH
H.aln
Average Pairwise Sequenc Identity:      98%
Refining sequence alignment...
Computing the pairing threshold...
The pairing threshold is:      -0.22
Sequence alignment has 4 sequences.
Alignment length is: 45
Folding sequence alignment...
The final score is:      -4.85
Runing miliseconds:60

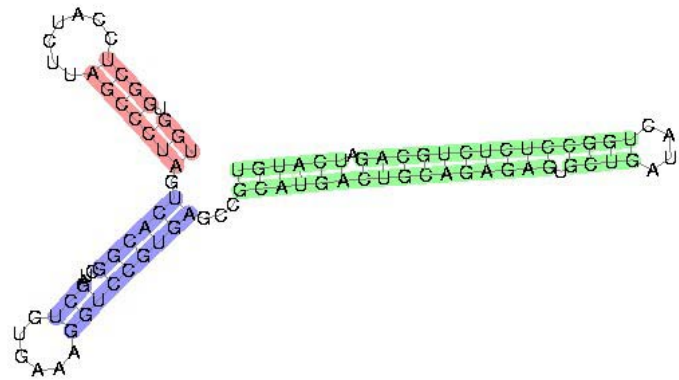
Predicted Structure:
UGCCAUCGUCGUCGAUCUCUAUGCAACUACCCUUGCGACUAUGGCA
(<<<<..(<<<<.((<<.....))>>.....>>>>..>>>>)
```


Virus Data – EAV_LTH



Virus Data – HCV_X3

- Thought to contain three stem-loop structure
- Structure of sequence essential for replication of the viral strand

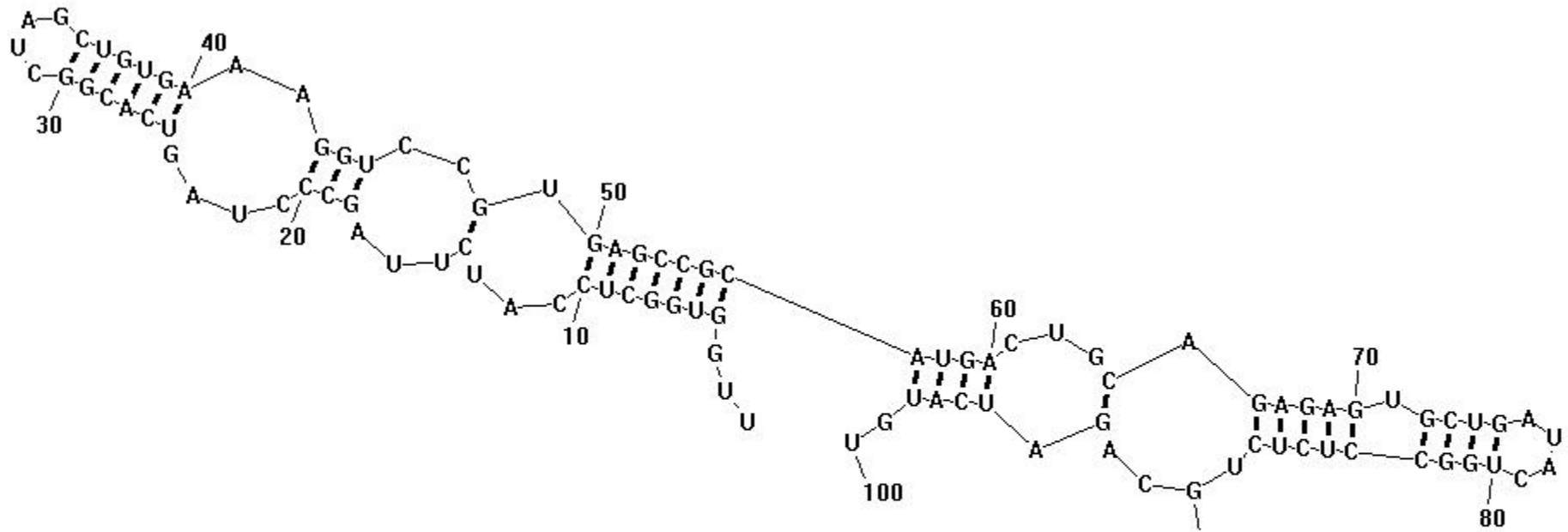


Virus Data – HCV_X3

```
Command Prompt
C:\Documents and Settings\John Acampado\RSpredict>java -Xmx512m RSpredict HCV_X3
.aln
Average Pairwise Sequenc Identity:    96%
Refining sequence alignment...
Computing the pairing threshold...
The pairing threshold is:    -0.09
Sequence alignment has 22 sequences.
Alignment length is: 100
Folding sequence alignment...
The final score is:    -12.54
Runing miliseconds:391

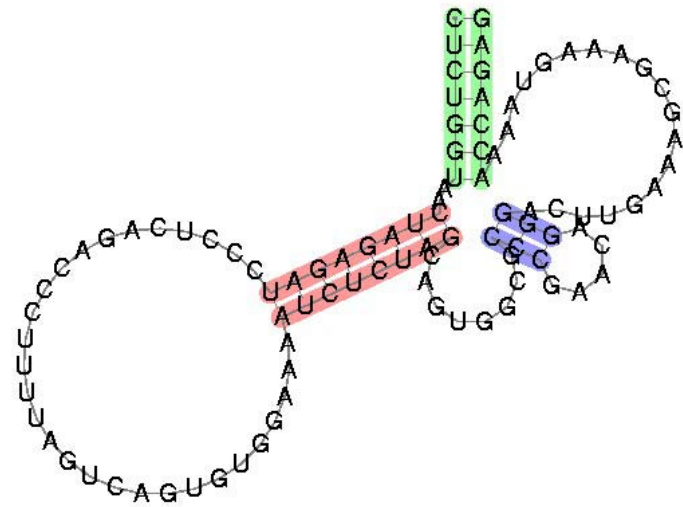
Predicted Structure:
UUGGUGGGCUCCAUCUUAGCCCUAGUCACGGCUAGCUGUGAAAGGUCCGUGAGCCGCAUGACUGCAGAGAGUGCUGAUACU
GGCCUCUCUGCAGAUCAUGU
...<<<<<<...<...<<...<<<<<...>>>>>>...>>>...>>>>>>>><<<<...<.<<<<<.<<<<...>
>>>>>>>...>>>>>>>..
```

Virus Data – HCV_X3

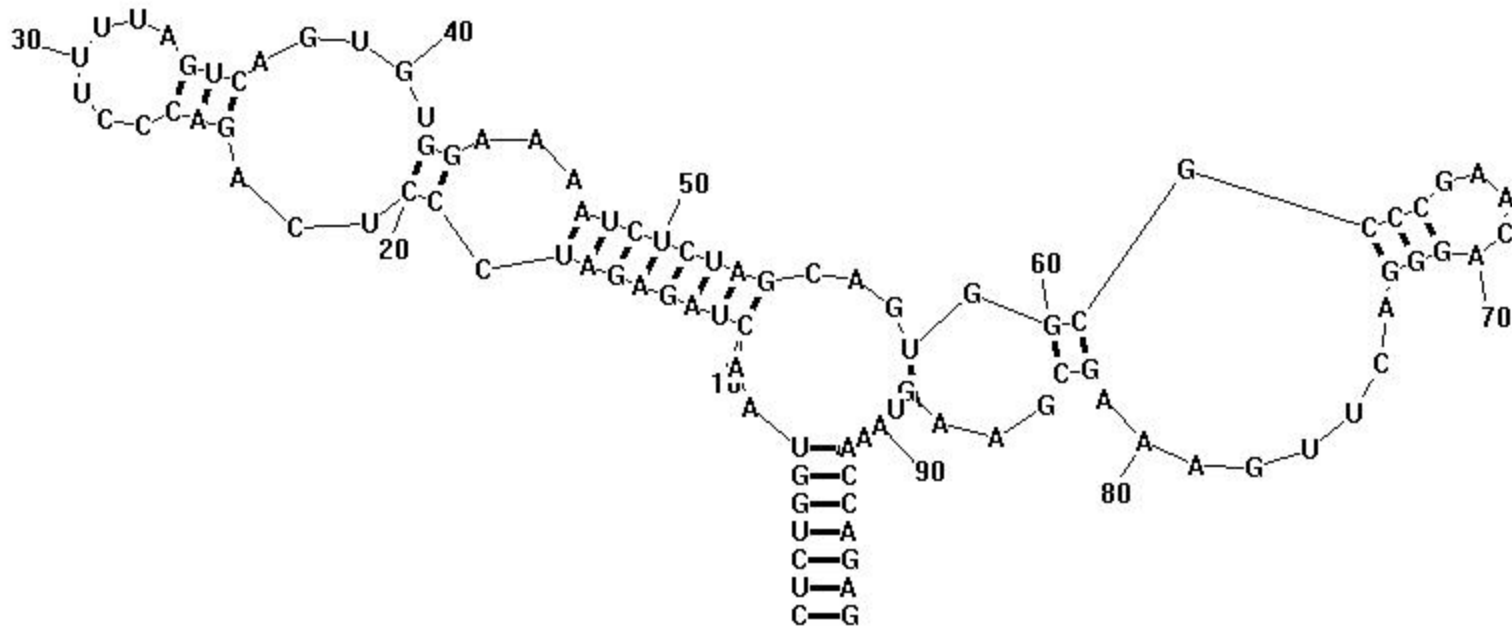


Virus Data – HIV_PBS

- Primer binding site is structured RNA element in genomes of retroviruses
- tRNA binds to site to initiate reverse transcription

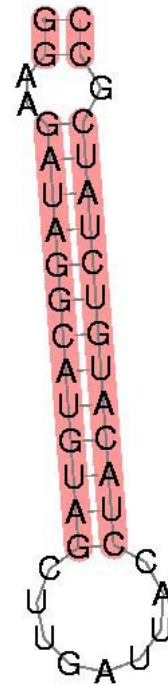


Virus Data – HIV_PBS

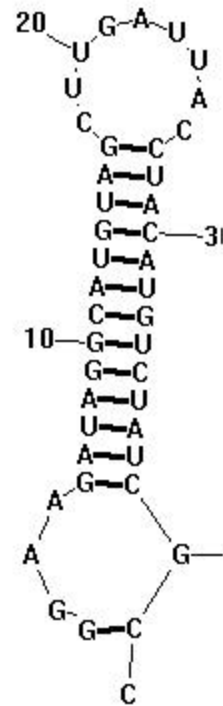


Virus Data – IBV_D-RNA

- RNA element known as defective or D-RNA
- Essential for viral replication and efficient packaging

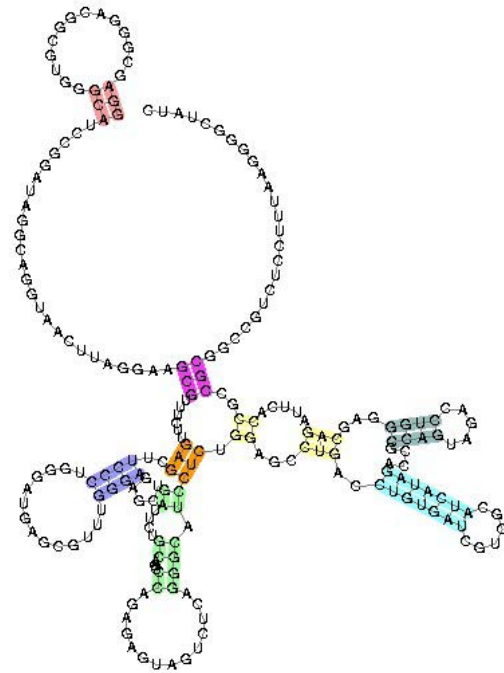


Virus Data – IBV_D-RNA

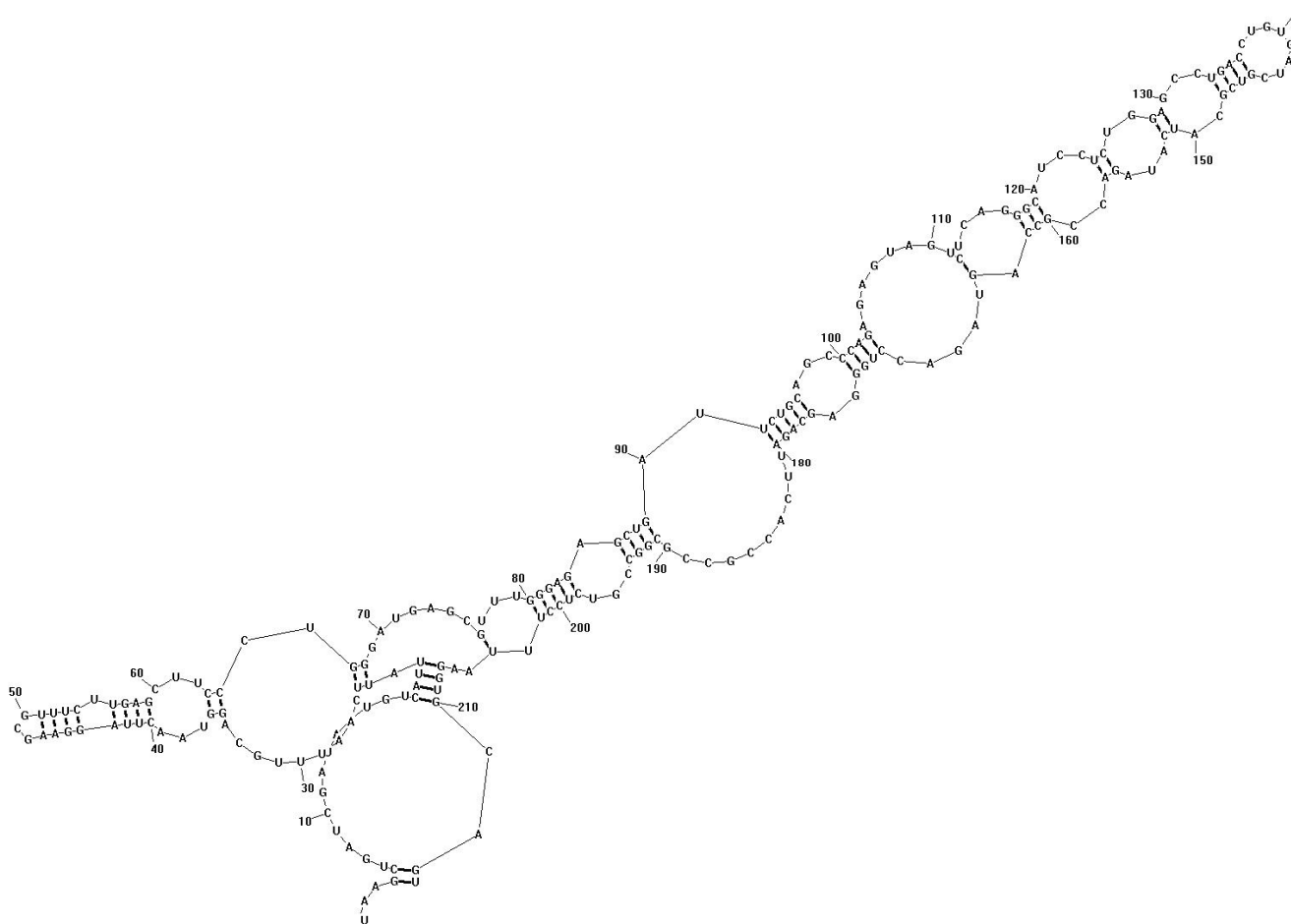


Virus Data – IRES_EBNA

- Found on U leader exon of 5' UTR
- Allows translation to occur when initiation is reduced
- Thought to be necessary for latent gene expression

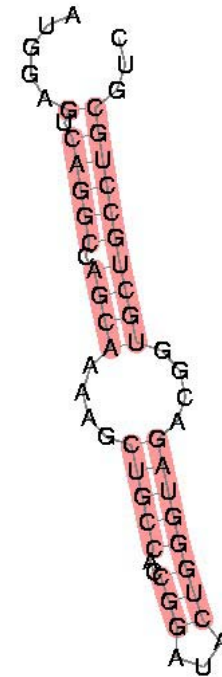


Virus Data – IRES_EBNA

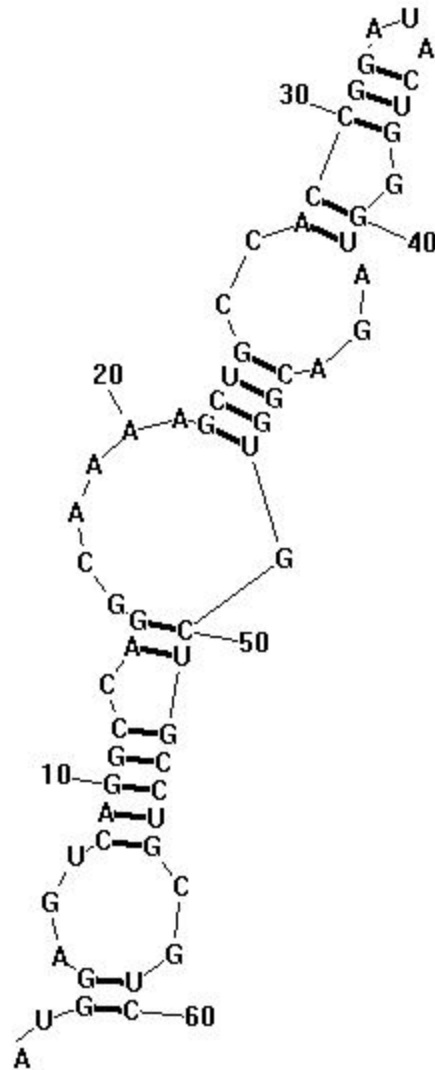


Virus Data – JEV_hairpin

- Small hairpin structure found in Japanese encephalitis virus
- May play a role in RNA synthesis

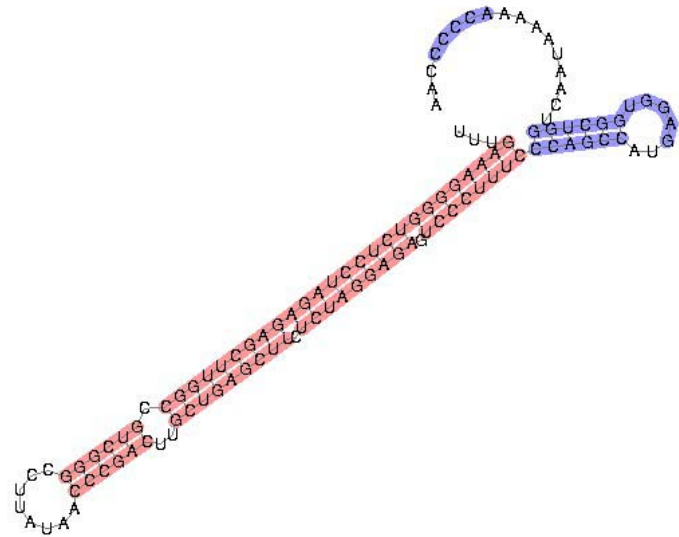


Virus Data – JEV_hairpin



Virus Data – Parecho_CRE

- Located in the 5' terminal of genome
- Consists of two stem-loop structures
- Disruption impairs both viral replication and growth



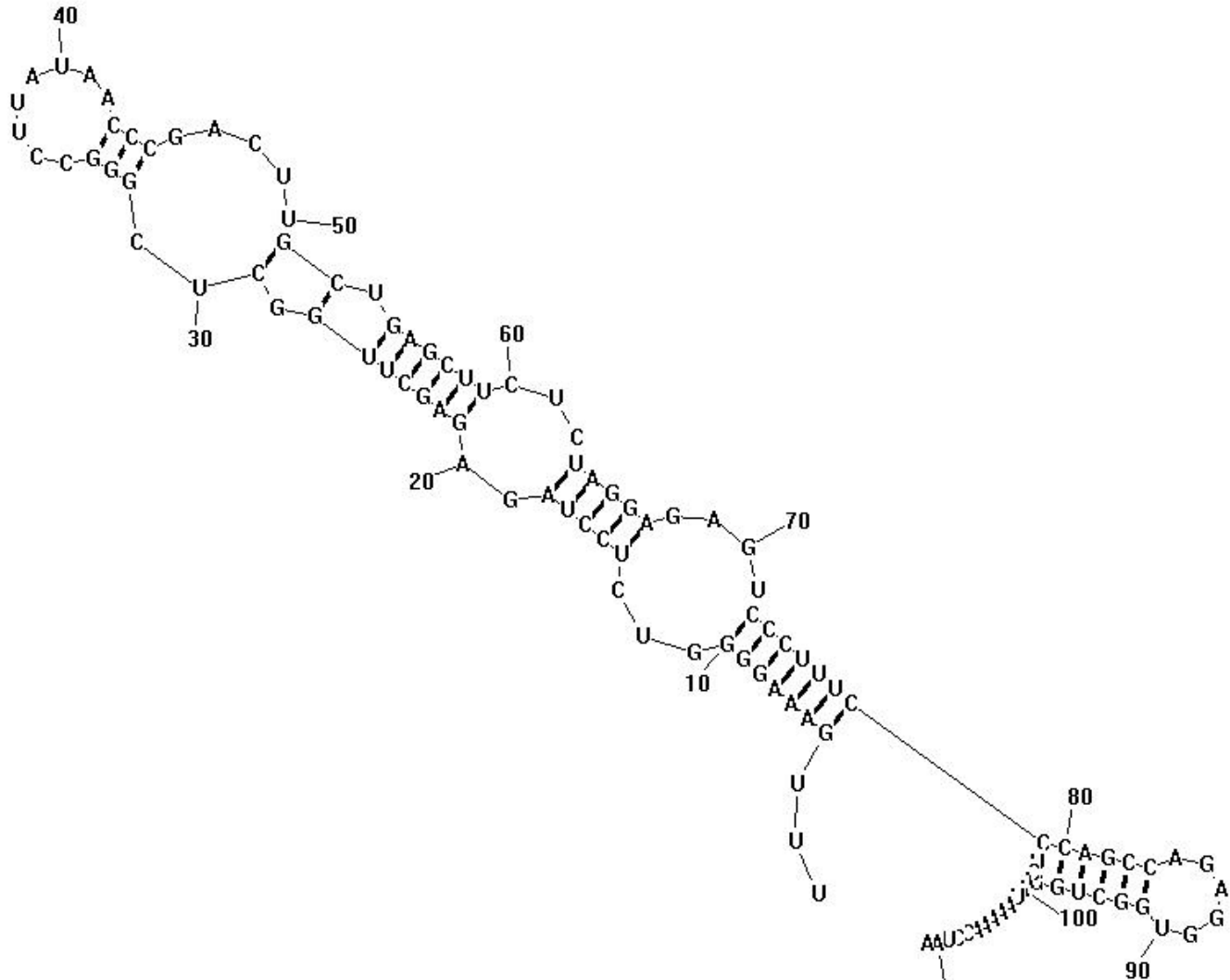
Virus Data – Parecho_CRE

```
C:\> Command Prompt

C:\Documents and Settings\John Acampado\RSpredict>java -Xmx512m RSpredict Parecho_CRE.aln
Average Pairwise Sequenc Identity:      87%
Refining sequence alignment...
Computing the pairing threshold...
The pairing threshold is:      -0.20
Sequence alignment has 5 sequences.
Alignment length is: 111
Folding sequence alignment...
The final score is:      -10.69
Runing miliseconds:341

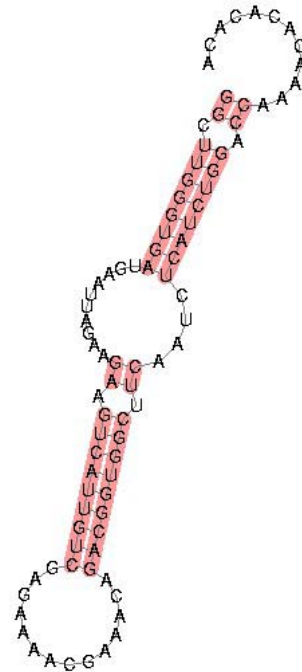
Predicted Structure:
UUUGAAAGGGGUCUCCUAGAGAGCUUGGCUGGGGCCUUAUAACCCGACUUGCUGAGCUUCUCUAGGAGAGUCCCUUCCC
AGCCAGAGGGUGGCUGGUCAAUAAAAACCUAA
...<<<<<<...<<<<...<<<<<<...<<<...>>>>>>...>>>>>>...>>>>>>>><<
<<<<...>>>>>>>>.....
```

Virus Data – Parecho_CRE

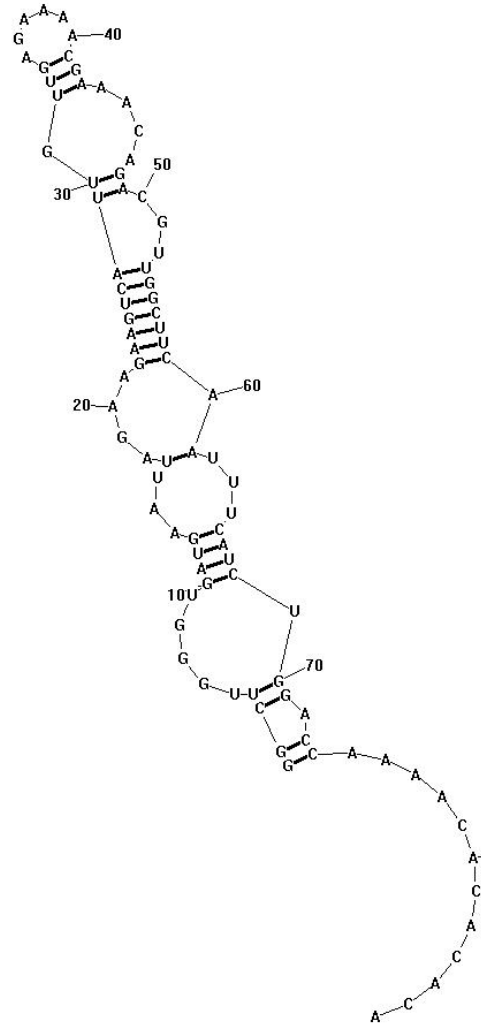


Virus Data – Rhino_CRE

- Cis-acting regulatory element for family of rhinoviruses (common cold)
- Located in protein coding region
- Essential for efficient viral replication

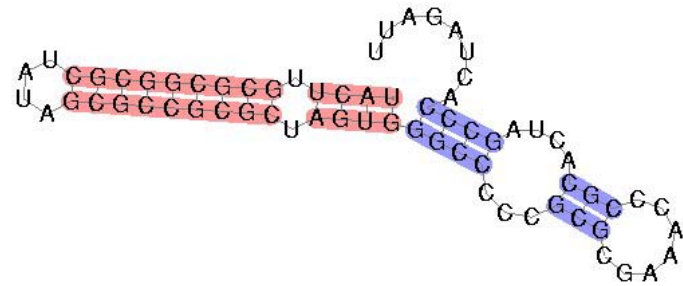


Virus Data – Rhino_CRE



Virus Data – Rubella_3

- Found in 3' UTR of rubella virus
- All loop structures thought to be vital for efficient viral replication
- Deletion of stem loop three is known to be lethal



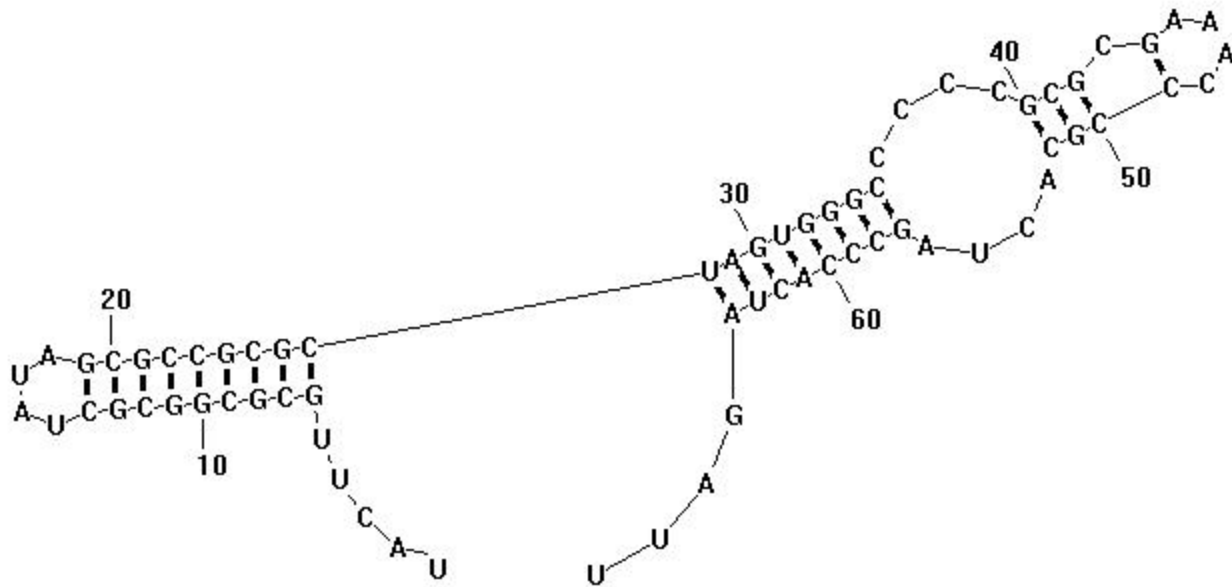
Virus Data – Rubella_3

```
C:\ Command Prompt
C:\Documents and Settings\John Acampado\RSpredict>java -Xmx512m RSpredict Rubella_3.aln
Average Pairwise Sequenc Identity:    95%
Refinining sequence alignment...
Computing the pairing threshold...
The pairing threshold is:    -0.17
Sequence alignment has 25 sequences.
Alignment length is: 68
Folding sequence alignment...
The final score is:    -9.45
Runing miliseconds:180

Predicted Structure:
UACUUGCGCGGGCGCUAUAAGCGCGCGCUAGUGGGCCCCCGCGCGAAACCCGGCACUAGCCCACUAGAUAU
.....<<<<<<<<<.....>>>>>>>>>>>><<<<<<<<.....<<<.....>>>>>>>>>>>>.....

C:\Documents and Settings\John Acampado\RSpredict>
```


Virus Data – Rubella_3



Alignment

- Alignment generated from Vienna sequences from output of RSPredict
- ClustalW2 alignment tool used to align sequences
- ClustalW2 aligned all eleven sequences

Alignment

```

Alfamo_CPB -----UGAUGUACCCCAUUAAUUUUGGGAUGCUAAAGUCAUUUAAUGCUGACCUCCACU 53
IBV_D-RNA -----
IRES_EBNA UAAGCUGAUCGAUAUUGUCAUUUAUCAAUUUGCAGGUAAUUAGGAAGCGUUUCUUGAGC 60
JEV_hairpin -----
EAV_LTH -----
Parecho_CRE -----
Rhino_CRE -----
Rubella_3 -----
BaMV_CRE -----AACGUUGCAUGAUCGUAAAACAUGCCCGGCU 31
HIV_PBS -----
HCV_X3 -----

```

```

Alfamo_CPB CUAAGGGAUGC----- 182
IBV_D-RNA -----
IRES_EBNA UUCACCGCCCGCGCGUCUCCUUUAAGGUGCAGU 214
JEV_hairpin -----
EAV_LTH -----
Parecho_CRE -----
Rhino_CRE -----
Rubella_3 -----
BaMV_CRE -----
HIV_PBS -----
HCV_X3 -----

```

```

Alfamo_CPB GGGUGGAUUAAAGGUCAGGUUAUGAAGUCCUAUUCGCUCUUGAUGGAUCGACUU-CAUUAU 112
IBV_D-RNA -----GGAAGAUAGGCAUGUAGCU 19
IRES_EBNA UUCUCCUGGGGAUGAGCGUUUGGAGAGCUGAUUCUGCAGCCCAGAGAGUAGUCUCAGGGCA 120
JEV_hairpin -----AUG-GAGUCAAGCCAGGCCAAAAGCUGCCACC-GGAUACUGGGUAGACGGUG 49
EAV_LTH -----UGCCAUCGUCGUGCGAUCU 18
Parecho_CRE ---UUUGAAAAGGGUUCUCCUAGAGAGCUUGGCUCGGGCCUUUAACCCGACUUGCUGAGC 57
Rhino_CRE -GGCUUGGGUGAUGAAUUAGAAGAGUCAUUGUUAGAGAAA-ACGAAAACAGACGUUGGCCUU 58
Rubella_3 -----UACUUGCGCGCGCUAUAAGCGCCGCGCUAGUGGGCCCCCGCGGAAA 47
BaMV_CRE UACCGUGAGCCGCCUUUUGAAAAGAAAGUUUACACGGACUCUGUUUGGCUUA-CGCACGUAC 90
HIV_PBS CUCUGGUAAUCUAGAGAUCCUCAGACCCUUUUAGUCAGUGUGGAAAAUCUCUAGCAGUGG 60
HCV_X3 --UUGGUGGCUCUUCUAGCCCUAGUCACGGCUA-GCUGUGAAAGGUCGGUGAGCCGCA 57

```

```

Alfamo_CPB UGCUUUAUAUUGUG-CUAAACGCACAUAUAUAAAUGCUCUAGCAAAAUCGCAUGAAUGCCC 171
IBV_D-RNA UGAUUACCUACAUGUCUAUCGCC----- 42
IRES_EBNA UCCUCUGGAGCCUGACCCUGUGAUCGUCGCAUCAUAGACCGCCAGUAGACCUUGGGAGCAGA 180
JEV_hairpin CUGCCUGCGUC----- 60
EAV_LTH CUAUCAACUAACCCUUGCGCAUAUGGCA----- 45
Parecho_CRE UUCUCUAGGAGAGUCCUUUCCAGCCAGAGGUGGCUGGUCAAUAAAAACCUAA----- 111
Rhino_CRE CAAUUUCAUCUGGACCAAAACACACACA----- 86
Rubella_3 CCGCACUAAGCCCAUAGAUA----- 68
BaMV_CRE CUAAACCCGUGCCAGCAGAAUAAAGACCUUUUGGUUUUCUACAGUUUUUCC----- 140
HIV_PBS CGCCCCGAACAGGGACUUUGAAAGCGAAAUAUAAACCAAG----- 99
HCV_X3 UGACUGCAGAGAGUGCUAUAUCUGGCCUCUCUGCAGAUCAUGU----- 100


```

Phylogenetic Tree

- Phylogenetic tree generated from Vienna output of RSpredict
- Shows the distances of the sequences from each other
- ClustalW2 tool from [EMBL-EBI](#) website
- Generated phylogenetic tree, with gaps turned off, and neighbor-joining clustering

Phylogenetic Tree





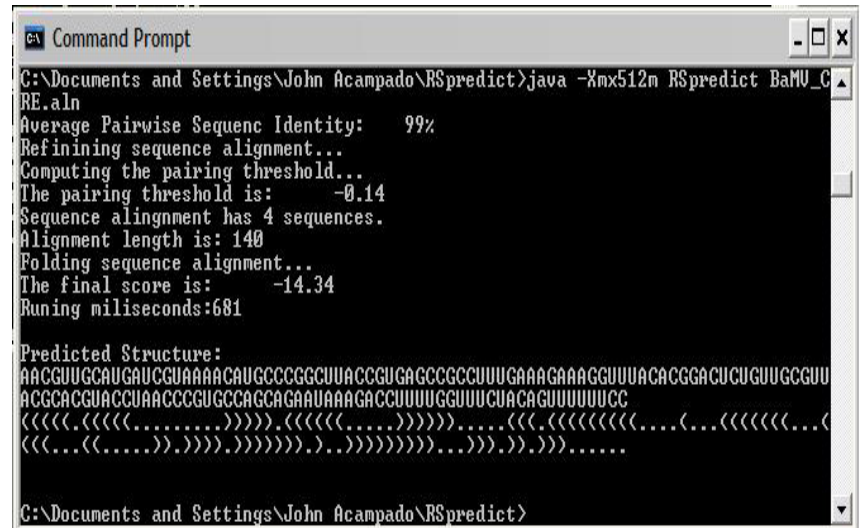
RSpredict WebServer Results

- Ran RSpredict via WebServer on the same eleven sequences as with the RSpredict JAVA
- Identical results to JAVA, but with a friendlier interface
- No need to use command line interface, everything on website
- CT and Vienna files available for download, to then be input into RNAstructure
- Side-by-side comparison of results on following slide

RSpredict WebServer Results



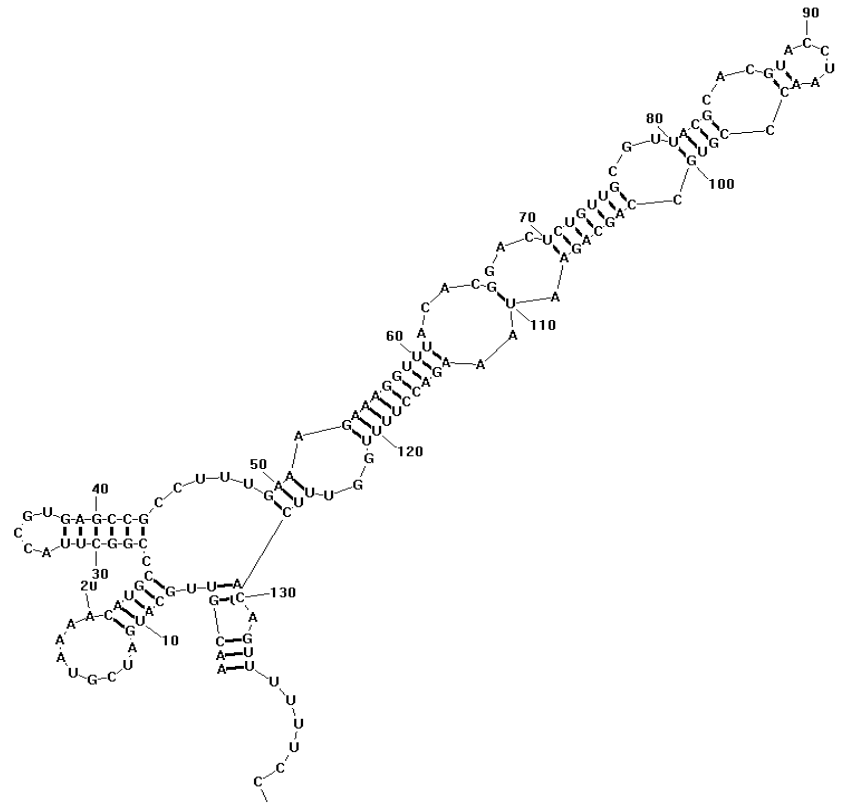
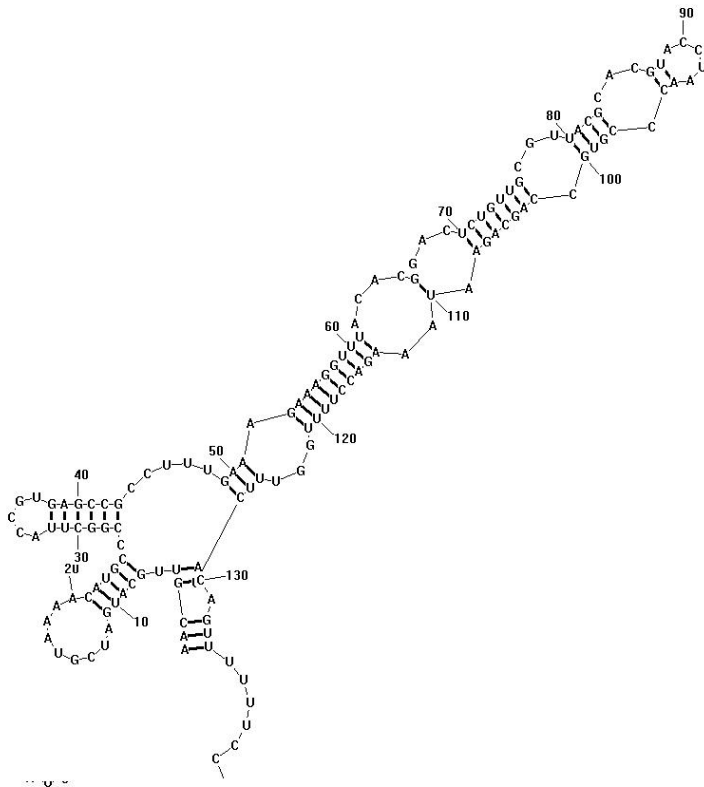
```
RSpredict  
Average Pairwise Sequenc Identity: 99%  
The pairing threshold is: -0.14  
Sequence alignment has 4 sequences.  
Alignment length is: 140  
The final score is: -14.34  
Predicted Structure:  
AACGUUGCAGUGAUGCAAAAACAUGCCCGCUUACCGUGAGCCGCUUUGAAAGAAAGUUACACGGACUUGUGCGUUAACGACGUACCUAACCCGUGCCAGCAGAAUAAAGACUUUUGGUUUUCACAGUUUUUCC  
(((((((((.....))))))(((((((.....)))))).....(((.....(((.....(((.....(((.....)))..))))))..))..))..)).....  
To get the predicted structure in .ct format please click here  
To get the predicted structure in .vienna format please click here
```



```
Command Prompt  
C:\Documents and Settings\John Acampado\RSpredict>java -Xmx512m RSpredict BaMV_C  
RE.alm  
Average Pairwise Sequenc Identity: 99%  
Refining sequence alignment...  
Computing the pairing threshold...  
The pairing threshold is: -0.14  
Sequence alignment has 4 sequences.  
Alignment length is: 140  
Folding sequence alignment...  
The final score is: -14.34  
Runing miliseconds:681  
Predicted Structure:  
AACGUUGCAGUGAUGCAAAAACAUGCCCGCUUACCGUGAGCCCGCUUUGAAAGAAAGUUUACACGGACUCUGUUGCGUU  
ACGCACGUACCUAACCCGUGCCAGCAGAAUAAAGACCUUUUGGUUUUCACAGUUUUUCC  
(((((((((.....))))))(((((((.....)))))).....(((.....(((.....(((.....(((.....)))..))))))..))..))..)).....  
C:\Documents and Settings\John Acampado\RSpredict>
```

Side-by-side comparison of Webserver and JAVA RSpredict with identical results.

RSpredict WebServer Results



Identical results after CT file was input into RNAstructure to get sequence structure.

Analysis / Conclusion

- Average length and sequence identity correct when compared to Rfam
- Structure from RNAstructure does not match that of Rfam exactly
- RSpredict takes FASTA files as input and outputs CT and Vienna files that effectively predict structure
- There are many similarities between Rfam and RSpredict/RNAstructure pictures
- Phylogenetic tree shows relationships between the different viruses

Resources

- EMBL-EBI: <http://www.ebi.ac.uk/>
- Rfam: <http://www.sanger.ac.uk/Software/Rfam/browse/index.shtml>
- RNAstructure: <http://rna.urmc.rochester.edu/rnastructure.html>
- RSpredict: <http://datalab.njit.edu/biology/RSpredict/index.html>
- Senior Project: <http://web.njit.edu/~jsa4/SeniorProject/>

Contact Information

- John Acampado
Bioinformatics Major, Senior Year
New Jersey Institute of Technology
e-mail: jsa4@njit.edu