In Memoriam – July 2006
Elisa Santry, 16, Boston, MA
Dave Bushow, 29, River Vale, NJ
www.sltrib.com 7-20-06

A Genetic Element Common to Tumor Viruses and Human

Angelo Kontgas – BBSI Intern
Dr. Bino John – Mentor

Background

- Research Risks
- Virology & Evolution
- Mutation Rates
- Evolutionary Perspective
- Exogenous/Endogenous viruses

Research Risks

- ~ NIH R21
  - – New, exploratory, developmental research
  - – High risk/payoff
- ~ Wildcatters
  - – Drill in low(?) possibility geology
    - Gusher
    - *Dry Hole*
- Great learning experience

Lakeview 18k bp/day – 18 mn
©www.sjgs.com

7/27/2006

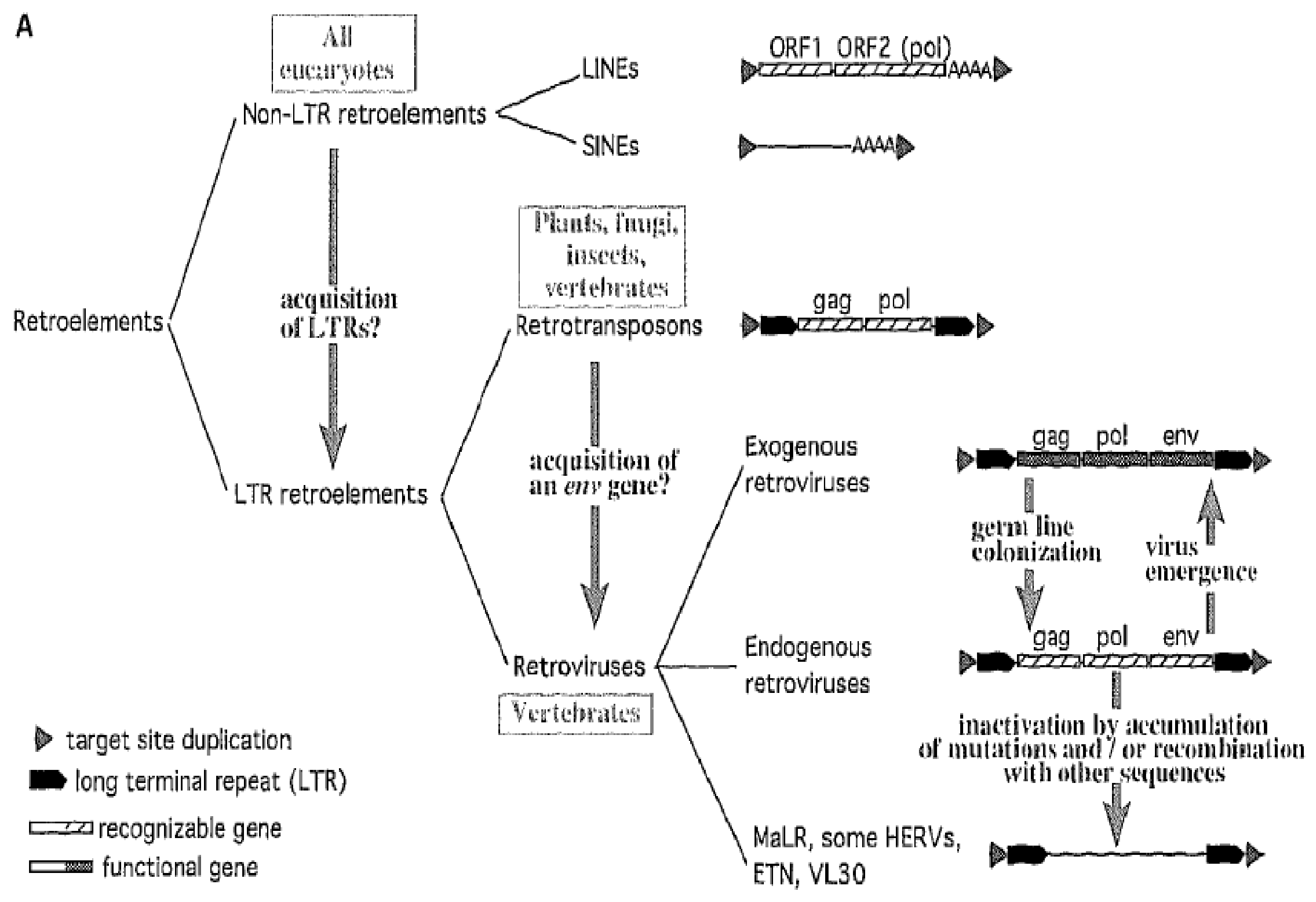
Virology & Evolution

- Virus – infectious, non living agent
  - Oncogenic – uncontrolled cell growth
- Antiviral agents difficult
  - Virus variants, mutation (evolution)
  - Distinguish viral vs cellular processes
- Viral Evolution Theories
  - Intracellular parasites – essential replication, xmission
  - Normal cellular components – replicate autonomously
  - Self-replicating RNA molecules

Virology & Evolution

- Viral processes ~ Cellular Processes
  - Viral mechanisms → treat viral illnesses
- Directed (Virus) Evolution
  - Designed, avoided human immune system
  - Virus as genetic delivery vehicle
  - Adeno-associated virus: infects 90% humans
    - **Maheshri, N., et. al. *Nat Biotechnol.* 24, 198-204 (2006).**
- RNA Viruses model system
- Quasispecies model
  - Depends on own replication rate
  - Mutant sequence replication rate
  - Predecessor sequence replication rate
  - Predicts mutational clouds: *in vitro* RNA viruses
  - System of linear equations

Evolutionary Perspective



Parseval, Heidmann, *Cytogenet Genome Res*. 2005;110(1-4):318-32

Virus Mutation Rates

Endogenous

- 80K proviruses human
- 6-8% of Genome
- 2-4 bp muta/1000 bp /million years

Exogenous

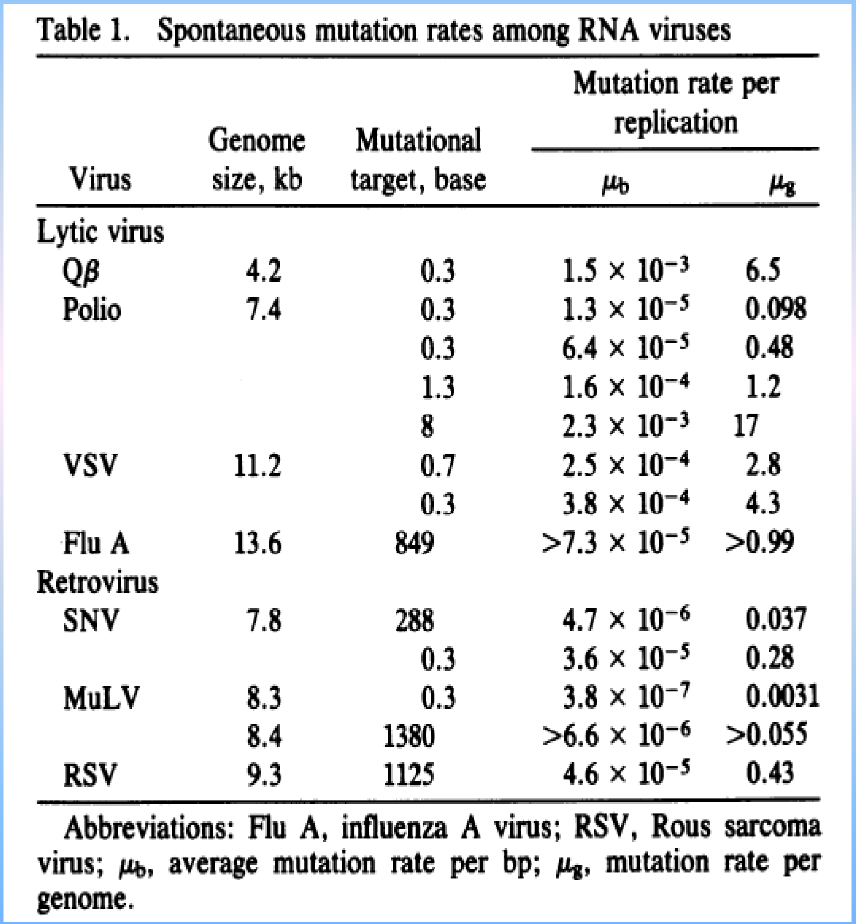
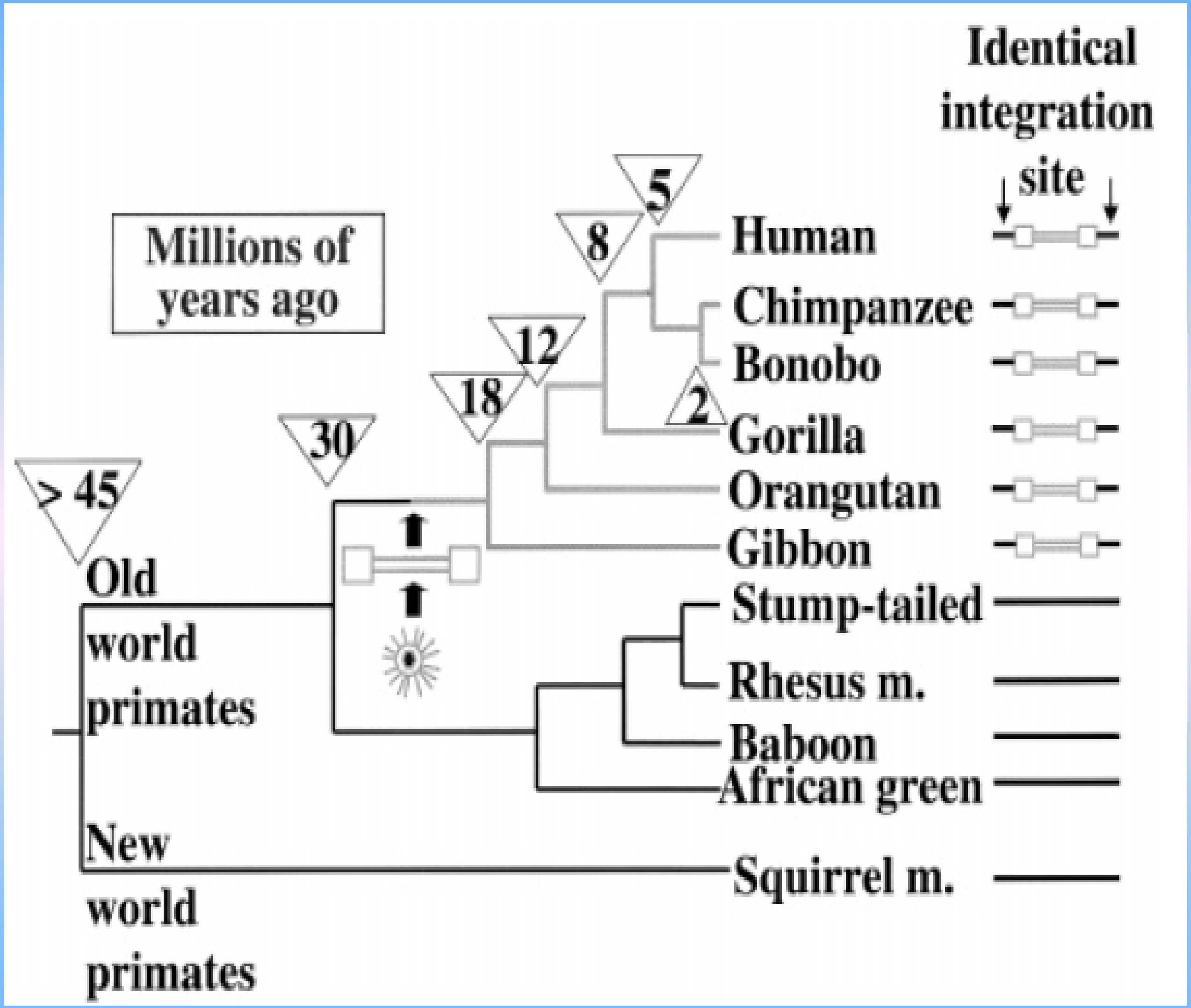


Table 1. Spontaneous mutation rates among RNA viruses

Virus	Genome size, kb	Mutational target, base	Mutation rate per replication μ_b	Mutation rate per replication μ_g
Lytic virus				
Qβ	4.2	0.3	1.5×10^{-3}	6.5
Polio	7.4	0.3	1.3×10^{-5}	0.098
		0.3	6.4×10^{-5}	0.48
		1.3	1.6×10^{-4}	1.2
		8	2.3×10^{-3}	17
VSV	11.2	0.7	2.5×10^{-4}	2.8
		0.3	3.8×10^{-4}	4.3
Flu A	13.6	849	$>7.3 \times 10^{-5}$	>0.99
Retrovirus				
SNV	7.8	288	4.7×10^{-6}	0.037
		0.3	3.6×10^{-5}	0.28
MuLV	8.3	0.3	3.8×10^{-7}	0.0031
	8.4	1380	$>6.6 \times 10^{-6}$	>0.055
RSV	9.3	1125	4.6×10^{-5}	0.43

Abbreviations: Flu A, influenza A virus; RSV, Rous sarcoma virus; μ_b, average mutation rate per bp; μ_g, mutation rate per genome.

Drake JW, *Proc Natl Acad Sci* U S A. 1993 May 1; 90(9): 4171–4175



Coffin JM, *Proc Am Philos Soc.* 2004 Sep;148(3):264-80.

HIV Coevolution

… No doubt, the end point of the coevolution of HIV and humans will be a more-host-friendly interaction, but *no one* should want to participate in the process of evolution that will take us there. (J. M. Coffin – 2004)

Project Overview – The Hunt Is On

- Curated Genome List
  - Viruses, bacteria: associated with cancer
- BLAST of Genomes → Interesting Set(IS)
- BLAST of IS

  → nr, REF_SEQ genome, EST, Human genome
- *Mfold* Secondary Structures →

  miRNA encoded, ala KSHV ?
- CLUSTALW Alignments →

  Phylogeny ?

Oncogenic Pathogen Genomes

Search:

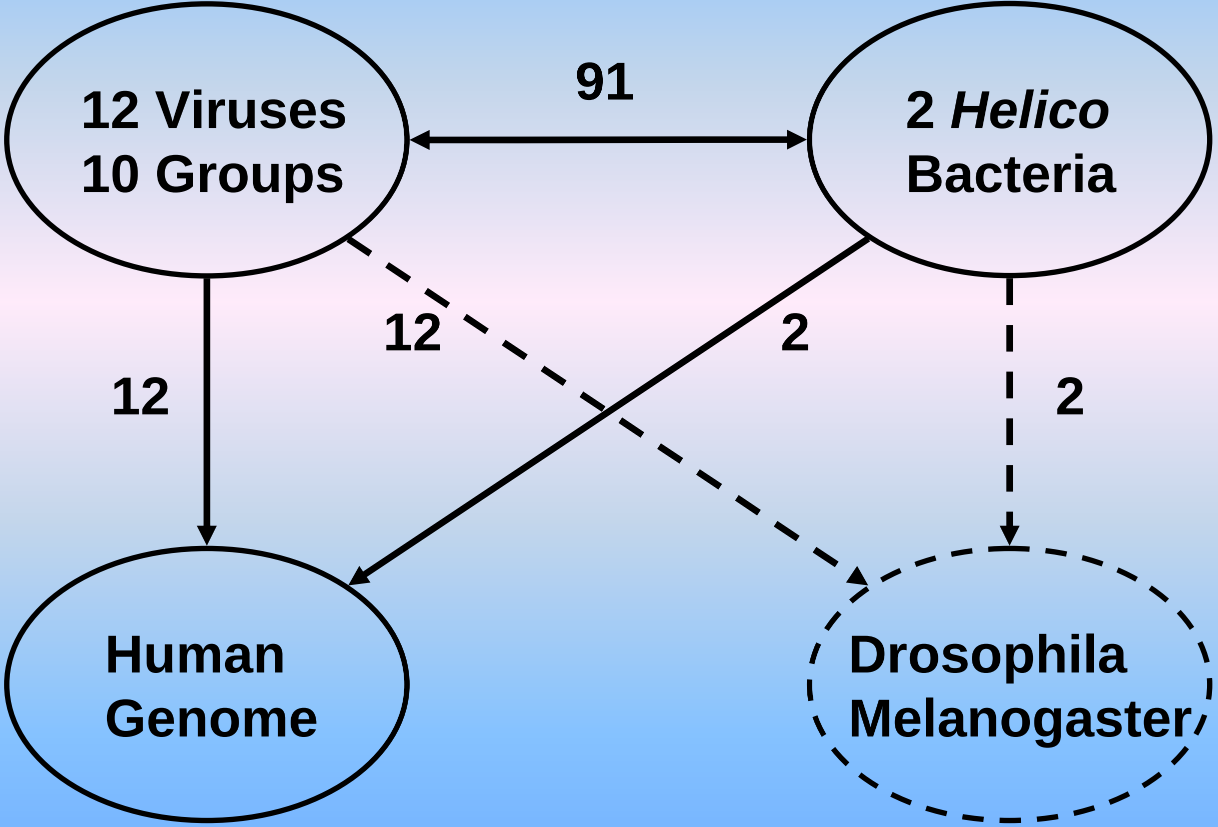
- Books – Creative Titles
  - Online book sites
  - U Pitt Library
  - Serendipity
- Peer Review Literature
  - PubMed Entrez

Results:

- Several Genomes
  - 29 viruses
  - 2 bacteria

Initial BLAST Experiments

Command Line: Perl Scripts/Linux



12 Viruses
10 Groups

2 *Helico*
Bacteria

Human
Genome

Drosophila
Melanogaster

BLAST Hits Between Genomes

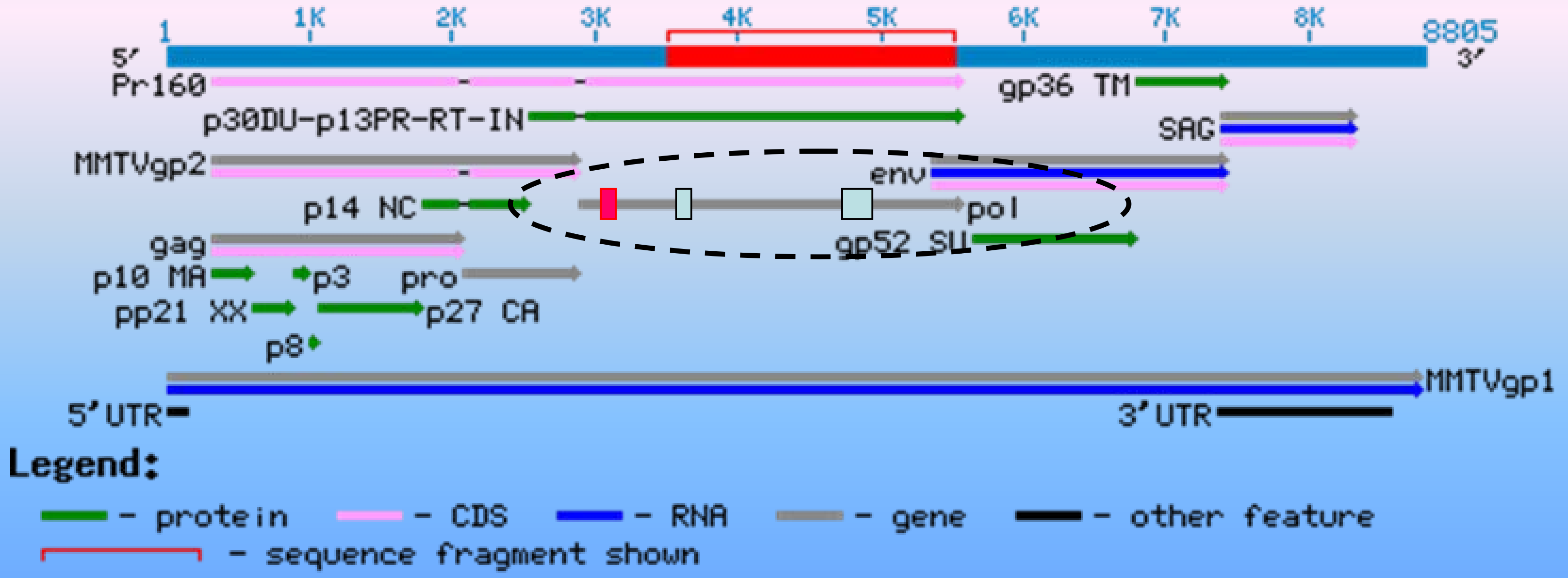
◯ = Intersection of interest …

		DROS_MELA	HBV	HCV	HHV4_EBV	HHV8_KSHV	HPV16	HPV5	HPV6	HTLV_1	HUM_GENO	JCPV	MMTV
Drosophila melanogaster	*DROS_MELA*												
Hepatitis B virus	HBV												
Hepatitis C virus	HCV												
Human herpesvirus 4 (Epstein Bar Virus)	HHV4_EBV												
Human herpesvirus 8 (Kaposi Sarcoma)	HHV8_KSHV				9								
Human herpesvirus 16	HPV16				1								
Human papillomavirus 5	HPV5				1		4						
Human papillomavirus 6	HPV6	1											
Primate T-lymphotropic virus 1	HTLV_1												
Homo sapien	HUM_GENO				1	1							
JC polyomavirus	JCPV												
Mouse mammary tumor virus	MMTV									1	28	1	

MMTV Genome

Retroviruses: 1 RNA genome strand

- gag = capsid, matrix, nucleocapsid proteins
- *pol = reverse transcriptase, protease, RNaseH, integrase*
- env = viral envelope proteins
- LTR = promoters, enhancers
- Package viral DNA, RNA splice cites



BLAST Objective:

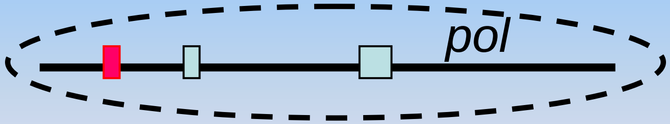
Seek Annotation Absence

		Target	
		Annotated	$\overline{\text{Annotated}}$
Query	Annotated	n/a	**X**
	$\overline{\text{Annotated}}$	**X**	**X**

Follow-up Blast Experiments

NCBI Server

- 3 MMTV *pol* sequences



‘nr’	264
REF_SEQ Genome	15

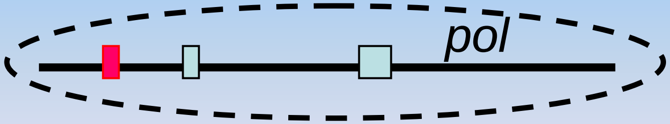
🟦 = Nothing significant apparent

- 1st MMTV Sequence(🟥) → Human EST => 4 Hits !
  - neuronal precursor
  - DKFZ (synonym: hlcc3) mRNA
  - NCI_CGAP_Phe1 mRNA
  - NCI_CGAP_GCB1 mRNA

Follow-up Blast Experiments

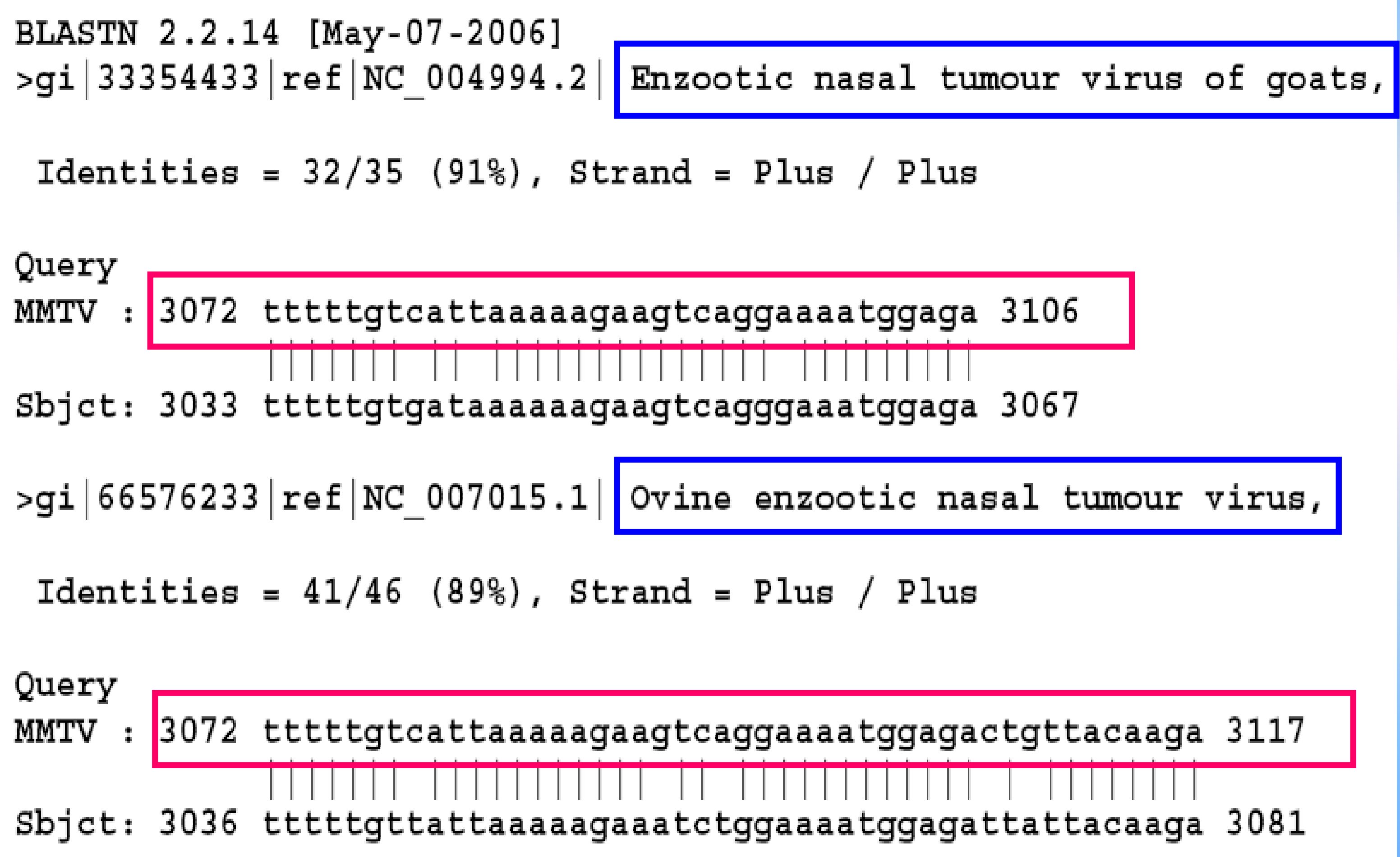
NCBI Server

- MMTV Genome → All Virus Genomes



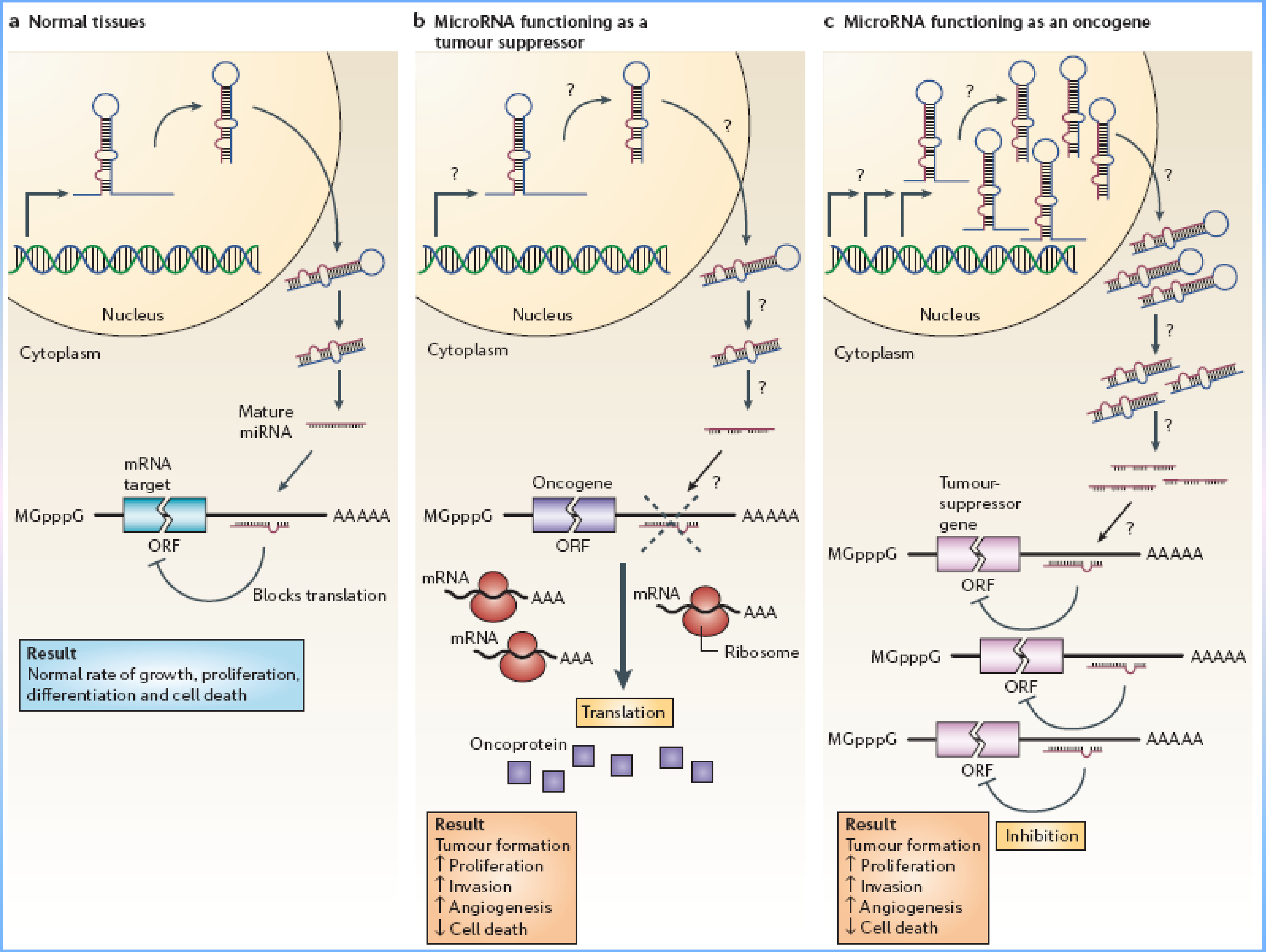
  - Goat Enzootic Nasal Tumor Virus (RV)
    - Hit = 🟥
  - Sheep Enzootic Nasal Tumor Virus (RV)
    - Hit = 🟥
  - Sheep Pulmonary Adenocarcinoma Virus (RV)
    - Hit = ▯

Representative Blast Data



```
BLASTN 2.2.14 [May-07-2006]
>gi|33354433|ref|NC_004994.2| Enzootic nasal tumour virus of goats,

 Identities = 32/35 (91%), Strand = Plus / Plus

Query
MMTV : 3072 tttttgtcattaaaaagaagtcaggaaaatggaga 3106
            ||||||| || |||||||||||||| |||||||||
Sbjct: 3033 tttttgtgataaaaaagaagtcagggaaatggaga 3067

>gi|66576233|ref|NC_007015.1| Ovine enzootic nasal tumour virus,

 Identities = 41/46 (89%), Strand = Plus / Plus

Query
MMTV : 3072 tttttgtcattaaaaagaagtcaggaaaatggagactgttacaaga 3117
            ||||||| ||||||||||| || |||||||||||| | ||||||||
Sbjct: 3036 tttttgttattaaaaagaaatctggaaaatggagattattacaaga 3081
```

Recall: µRNA → KSHV



Esquela-Kerscher *et al. Nature Reviews Cancer* 6, 259–269 (April 2006) | doi:10.1038/nrc1840

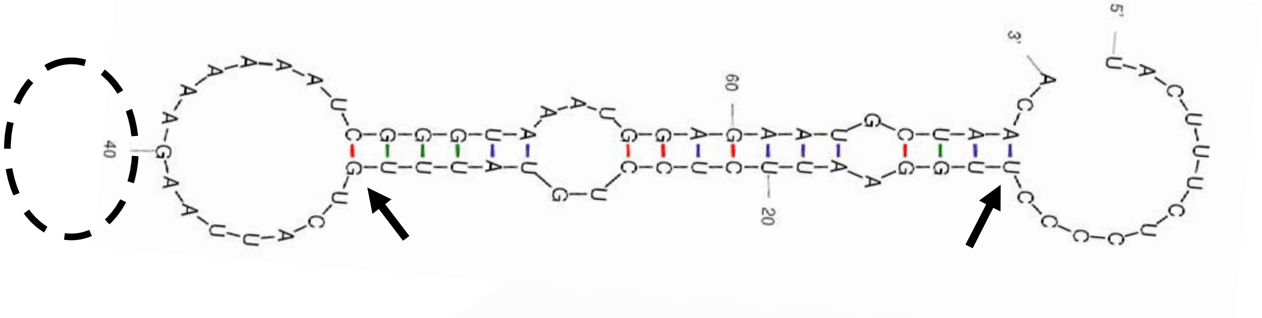
Stem – Loop Structures

Mfold:

- **2º Structure**
- **Energy minimum**

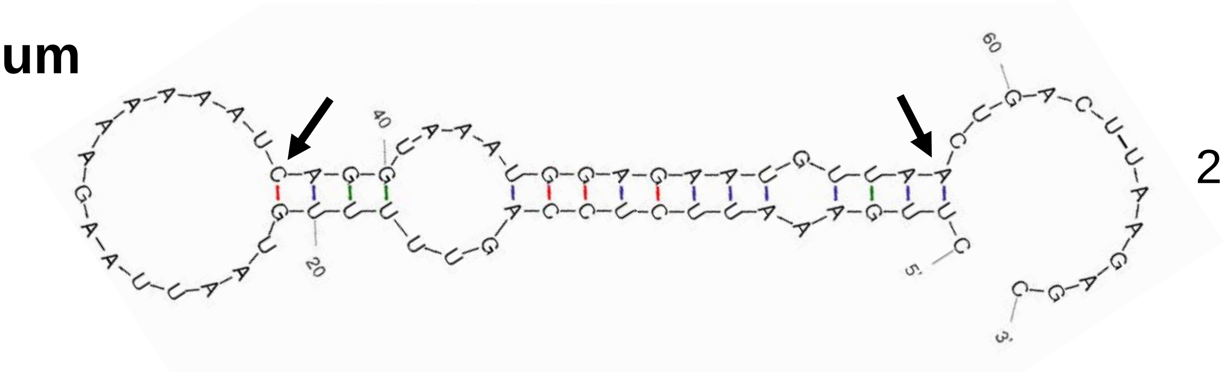
Issues:

- **dG <= - 25**
- **double loop**
- **19 – 26 nt's**



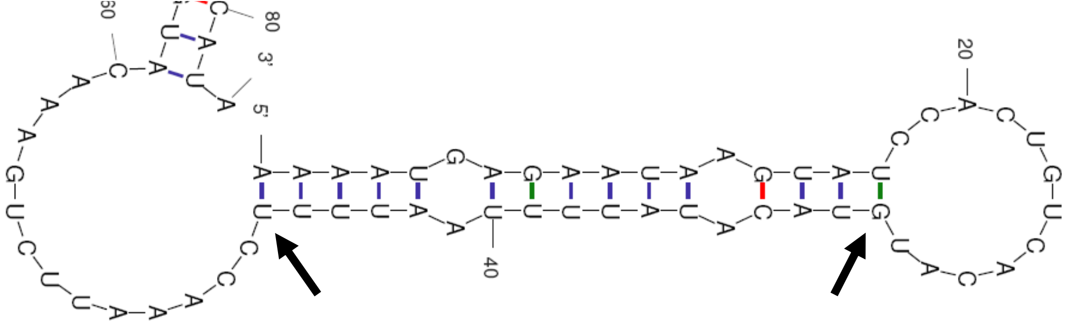
dG = -13.7 [initially -13.70] EST_STRAT_N2_NEURO

20 nt's



dG = -11.4 [initially -11.40] EST_DFKZ_hlcc3_mRNA

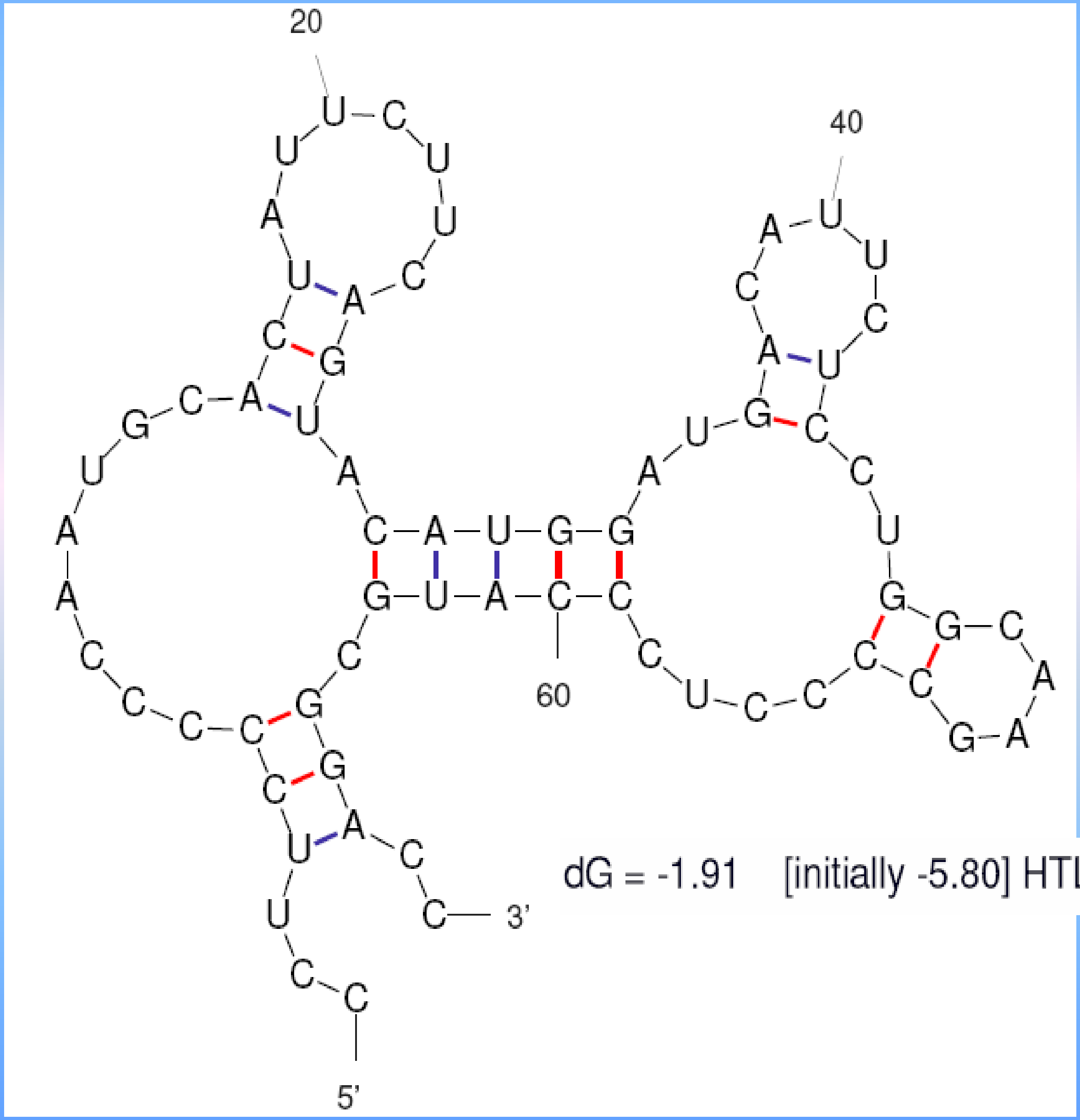
20 nt's



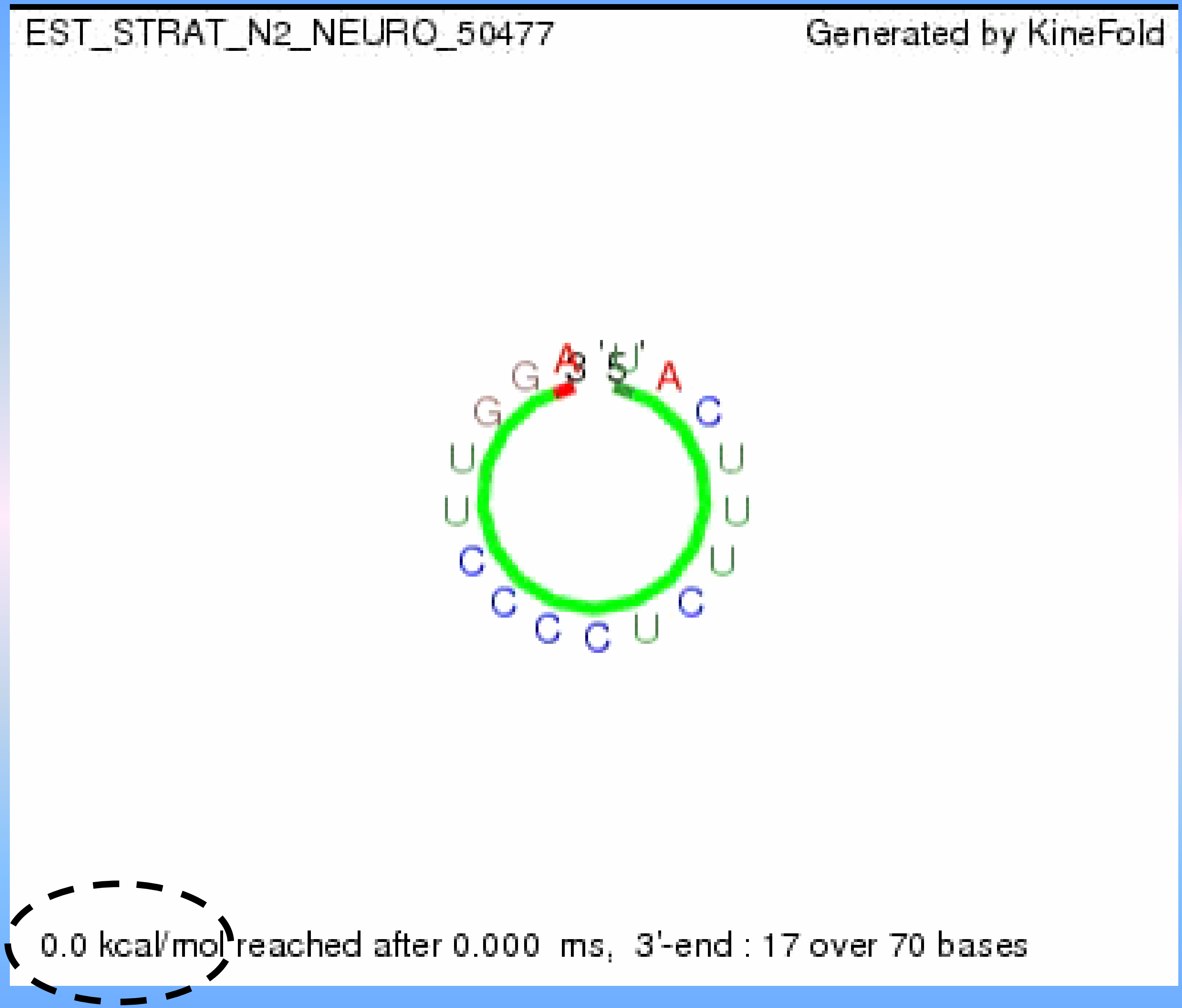
dG = -5.9 [initially -5.90] MMVT_4800_4882_28_Homo_s_hits

17 nt's (short)

Funky symmetry



dG = -1.91 [initially -5.80] HTLV_1_2768_2746

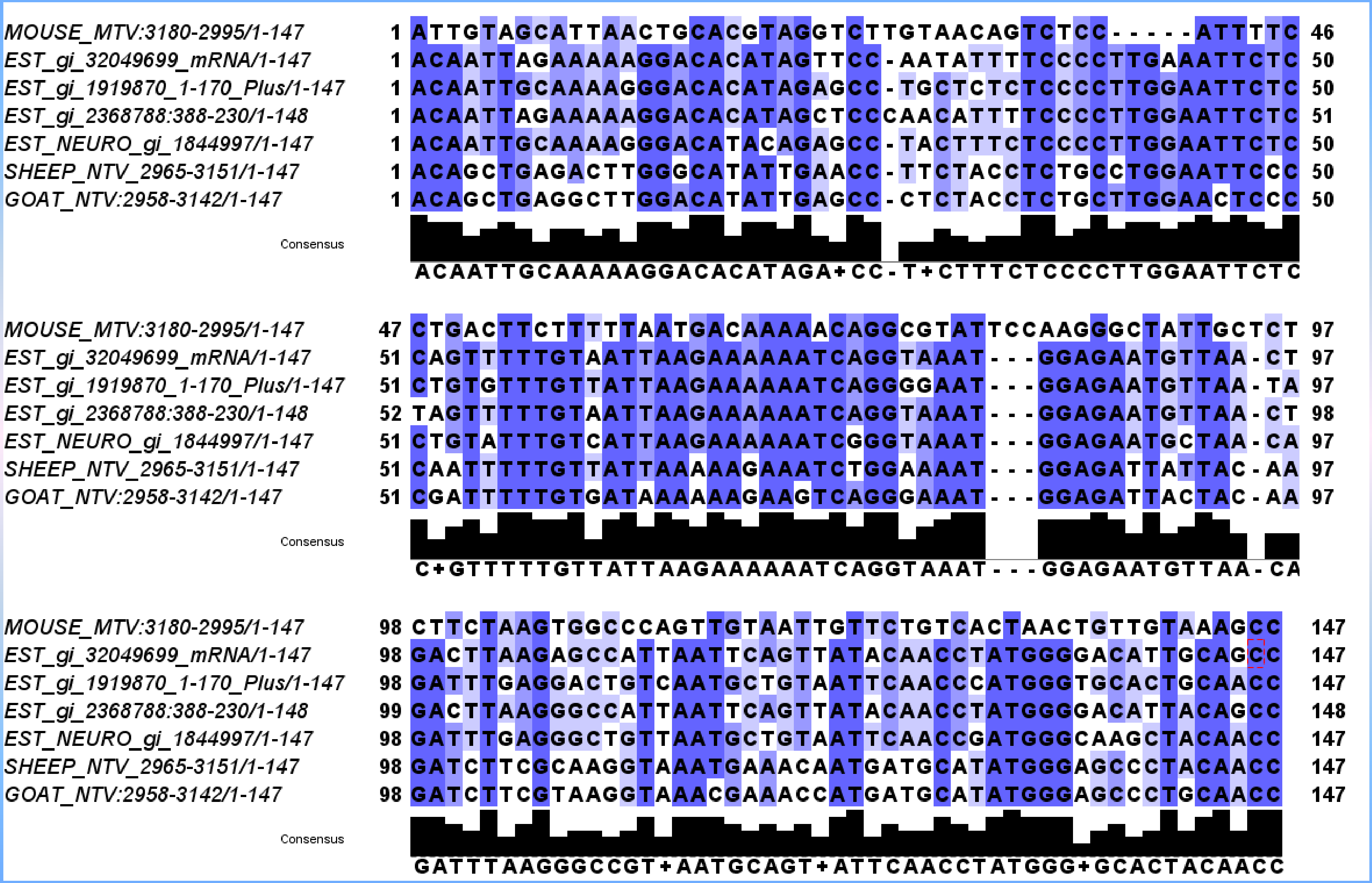


EST_STRAT_N2_NEURO_50477

Generated by KineFold

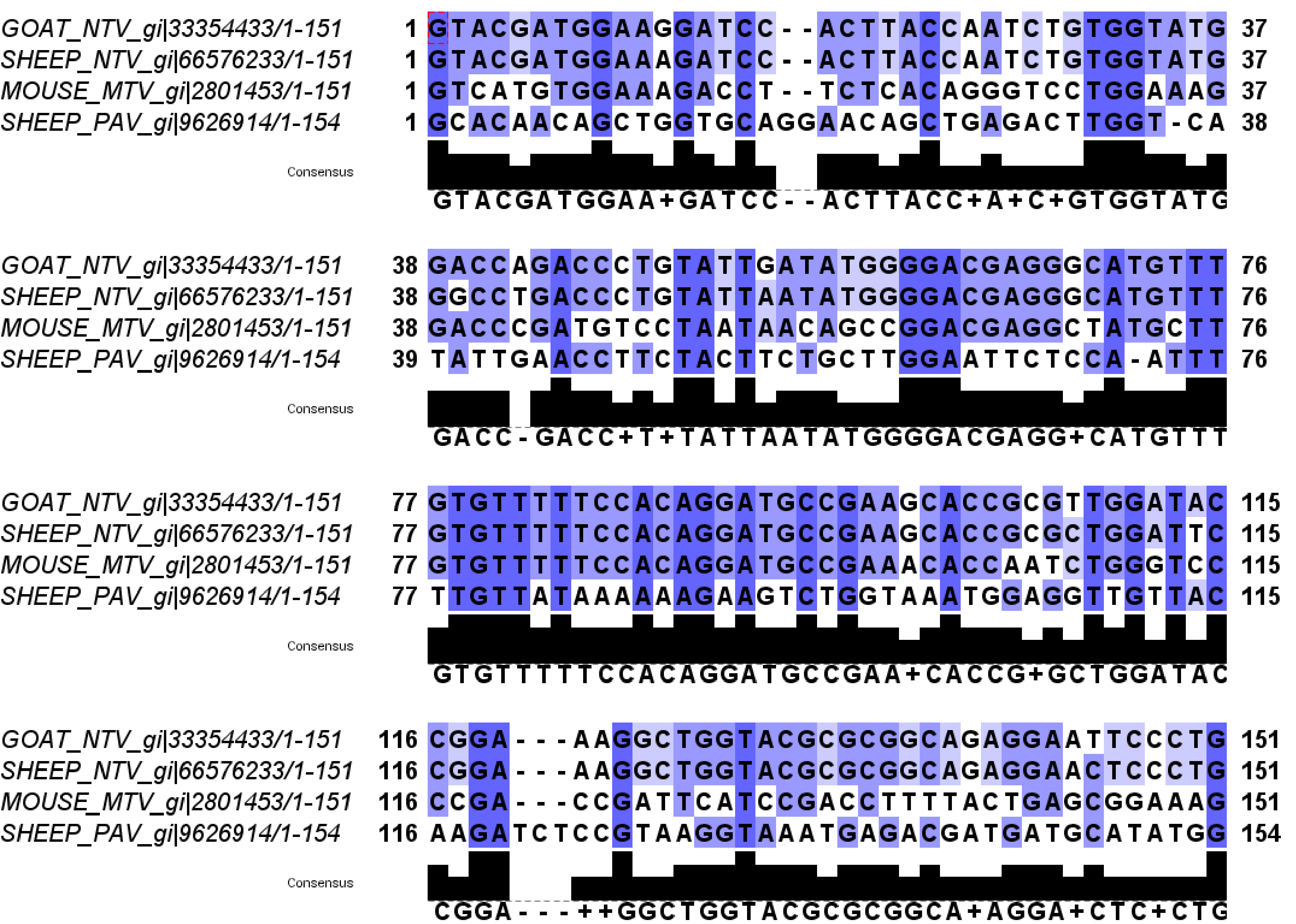
0.0 kcal/mol reached after 0.000 ms, 3'-end : 17 over 70 bases

CLUSTALW Analysis: Human Closer to Sheep & Goat



```
MOUSE_MTV:3180-2995/1-147          1 ATTGTAGCATTAACTGCACGTAGGTCTTGTAACAGTCTCC-----ATTTTC 46
EST_gi_32049699_mRNA/1-147         1 ACAATTAGAAAAAGGACACATAGTTCC-AATATTTTCCCCTTGAAATTCTC 50
EST_gi_1919870_1-170_Plus/1-147    1 ACAATTGCAAAAGGGACACATAGAGCC-TGCTCTCTCCCCTTGGAATTCTC 50
EST_gi_2368788:388-230/1-148       1 ACAATTAGAAAAAGGACACATAGCTCCCAACATTTTCCCCTTGGAATTCTC 51
EST_NEURO_gi_1844997/1-147         1 ACAATTGCAAAAGGGACATACAGAGCC-TACTTTCTCCCCTTGGAATTCTC 50
SHEEP_NTV_2965-3151/1-147          1 ACAGCTGAGACTTGGGCATATTGAACC-TTCTACCTCTGCCTGGAATTCCC 50
GOAT_NTV:2958-3142/1-147           1 ACAGCTGAGGCTTGGACATATTGAGCC-CTCTACCTCTGCTTGGAACTCCC 50
                       Consensus
                                     ACAATTGCAAAAAGGACACATAGA+CC-T+CTTTCTCCCCTTGGAATTCTC

MOUSE_MTV:3180-2995/1-147         47 CTGACTTCTTTTTAATGACAAAAACAGGCGTATTCCAAGGGCTATTGCTCT 97
EST_gi_32049699_mRNA/1-147        51 CAGTTTTTGTAATTAAGAAAAAATCAGGTAAAT---GGAGAATGTTAA-CT 97
EST_gi_1919870_1-170_Plus/1-147   51 CTGTGTTTGTTATTAAGAAAAAATCAGGGGAAT---GGAGAATGTTAA-TA 97
EST_gi_2368788:388-230/1-148      52 TAGTTTTTGTAATTAAGAAAAAATCAGGTAAAT---GGAGAATGTTAA-CT 98
EST_NEURO_gi_1844997/1-147        51 CTGTATTTGTCATTAAGAAAAAATCGGGTAAAT---GGAGAATGCTAA-CA 97
SHEEP_NTV_2965-3151/1-147         51 CAATTTTTGTTATTAAAAAGAAATCTGGAAAAT---GGAGATTATTAC-AA 97
GOAT_NTV:2958-3142/1-147          51 CGATTTTTGTGATAAAAAAGAAGTCAGGGAAAT---GGAGATTACTAC-AA 97
                       Consensus
                                     C+GTTTTTGTTATTAAGAAAAAATCAGGTAAAT---GGAGAATGTTAA-CA

MOUSE_MTV:3180-2995/1-147         98 CTTCTAAGTGGCCCAGTTGTAATTGTTCTGTCACTAACTGTTGTAAAGCC 147
EST_gi_32049699_mRNA/1-147        98 GACTTAAGAGCCATTAATTCAGTTATACAACCTATGGGGACATTGCAGCC 147
EST_gi_1919870_1-170_Plus/1-147   98 GATTTGAGGACTGTCAATGCTGTAATTCAACCCATGGGTGCACTGCAACC 147
EST_gi_2368788:388-230/1-148      99 GACTTAAGGGCCATTAATTCAGTTATACAACCTATGGGGACATTACAGCC 148
EST_NEURO_gi_1844997/1-147        98 GATTTGAGGGCTGTTAATGCTGTAATTCAACCGATGGGCAAGCTACAACC 147
SHEEP_NTV_2965-3151/1-147         98 GATCTTCGCAAGGTAAATGAAACAATGATGCATATGGGAGCCCTACAACC 147
GOAT_NTV:2958-3142/1-147          98 GATCTTCGTAAGGTAAACGAAACCATGATGCATATGGGAGCCCTGCAACC 147
                       Consensus
                                     GATTTAAGGGCCGT+AATGCAGT+ATTCAACCTATGGG+GCACTACAACC
```

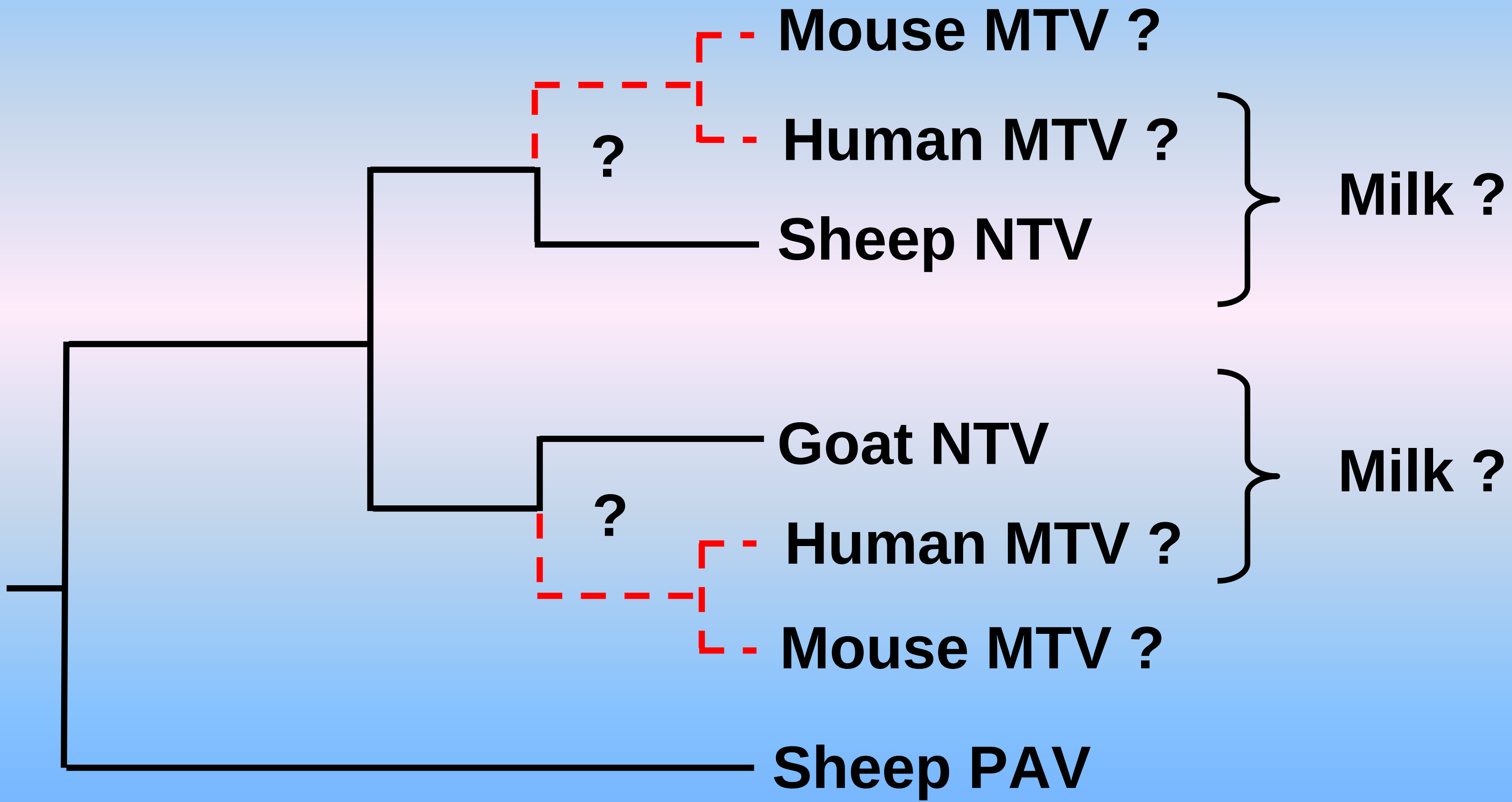
CLUSTALW Analysis: Goat & Sheep NTV Similarity



```
GOAT_NTV_gi|33354433/1-151     1 GTACGATGGAAGGATCC--ACTTACCAATCTGTGGTATG 37
SHEEP_NTV_gi|66576233/1-151    1 GTACGATGGAAAGATCC--ACTTACCAATCTGTGGTATG 37
MOUSE_MTV_gi|2801453/1-151     1 GTCATGTGGAAAGACCT--TCTCACAGGGTCCTGGAAAG 37
SHEEP_PAV_gi|9626914/1-154     1 GCACAACAGCTGGTGCAGGAACAGCTGAGACTTGGT-CA 38
                   Consensus     GTACGATGGAA+GATCC--ACTTACC+A+C+GTGGTATG

GOAT_NTV_gi|33354433/1-151    38 GACCAGACCCTGTATTGATATGGGGACGAGGGCATGTTT 76
SHEEP_NTV_gi|66576233/1-151   38 GGCCTGACCCTGTATTAATATGGGGACGAGGGCATGTTT 76
MOUSE_MTV_gi|2801453/1-151    38 GACCCGATGTCCTAATAACAGCCGGACGAGGCTATGCTT 76
SHEEP_PAV_gi|9626914/1-154    39 TATTGAACCTTCTACTTCTGCTTGGAATTCTCCA-ATTT 76
                   Consensus     GACC-GACC+T+TATTAATATGGGGACGAGG+CATGTTT

GOAT_NTV_gi|33354433/1-151    77 GTGTTTTTCCACAGGATGCCGAAGCACCGCGTTGGATAC 115
SHEEP_NTV_gi|66576233/1-151   77 GTGTTTTTCCACAGGATGCCGAAGCACCGCGCTGGATTC 115
MOUSE_MTV_gi|2801453/1-151    77 GTGTTTTTCCACAGGATGCCGAAACACCAATCTGGGTCC 115
SHEEP_PAV_gi|9626914/1-154    77 TTGTTATAAAAAAGAAGTCTGGTAAATGGAGGTTGTTAC 115
                   Consensus     GTGTTTTTCCACAGGATGCCGAA+CACCG+GCTGGATAC

GOAT_NTV_gi|33354433/1-151   116 CGGA---AAGGCTGGTACGCGCGGCAGAGGAATTCCCTG 151
SHEEP_NTV_gi|66576233/1-151  116 CGGA---AAGGCTGGTACGCGCGGCAGAGGAACTCCCTG 151
MOUSE_MTV_gi|2801453/1-151   116 CCGA---CCGATTCATCCGACCTTTTACTGAGCGGAAAG 151
SHEEP_PAV_gi|9626914/1-154   116 AAGATCTCCGTAAGGTAAATGAGACGATGATGCATATGG 154
                   Consensus     CGGA---++GGCTGGTACGCGCGGCA+AGGA+CTC+CTG
```

HMTV Evolution ?

Phylogram: length ~ change
Cladogram: ancestry *



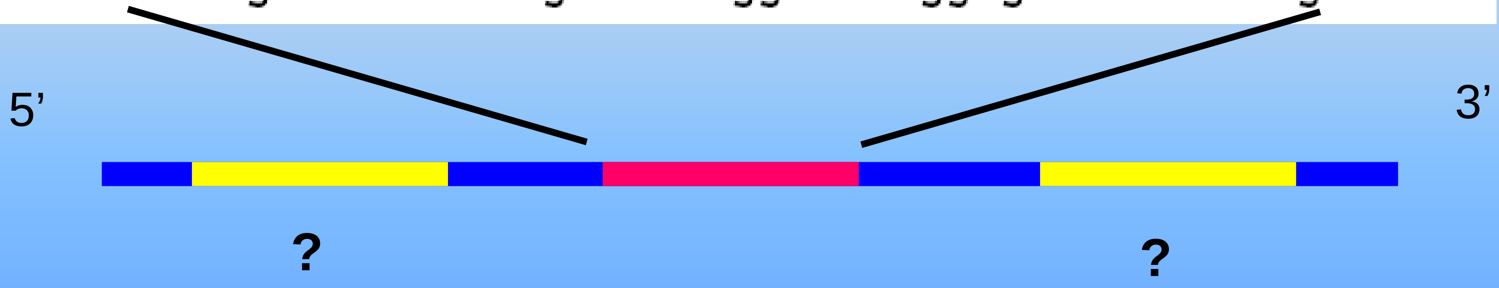
? Adjacent Sequences of Interest . . .

```
>gi|66576233|ref|NC_007015.1| Ovine enzootic nasal tumor virus,
           complete genome
         Length = 7434
```

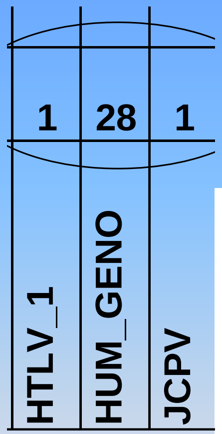
```
 Score = 52.0 bits (26), Expect = 8e-05
 Identities = 41/46 (89%)
 Strand = Plus / Plus
```

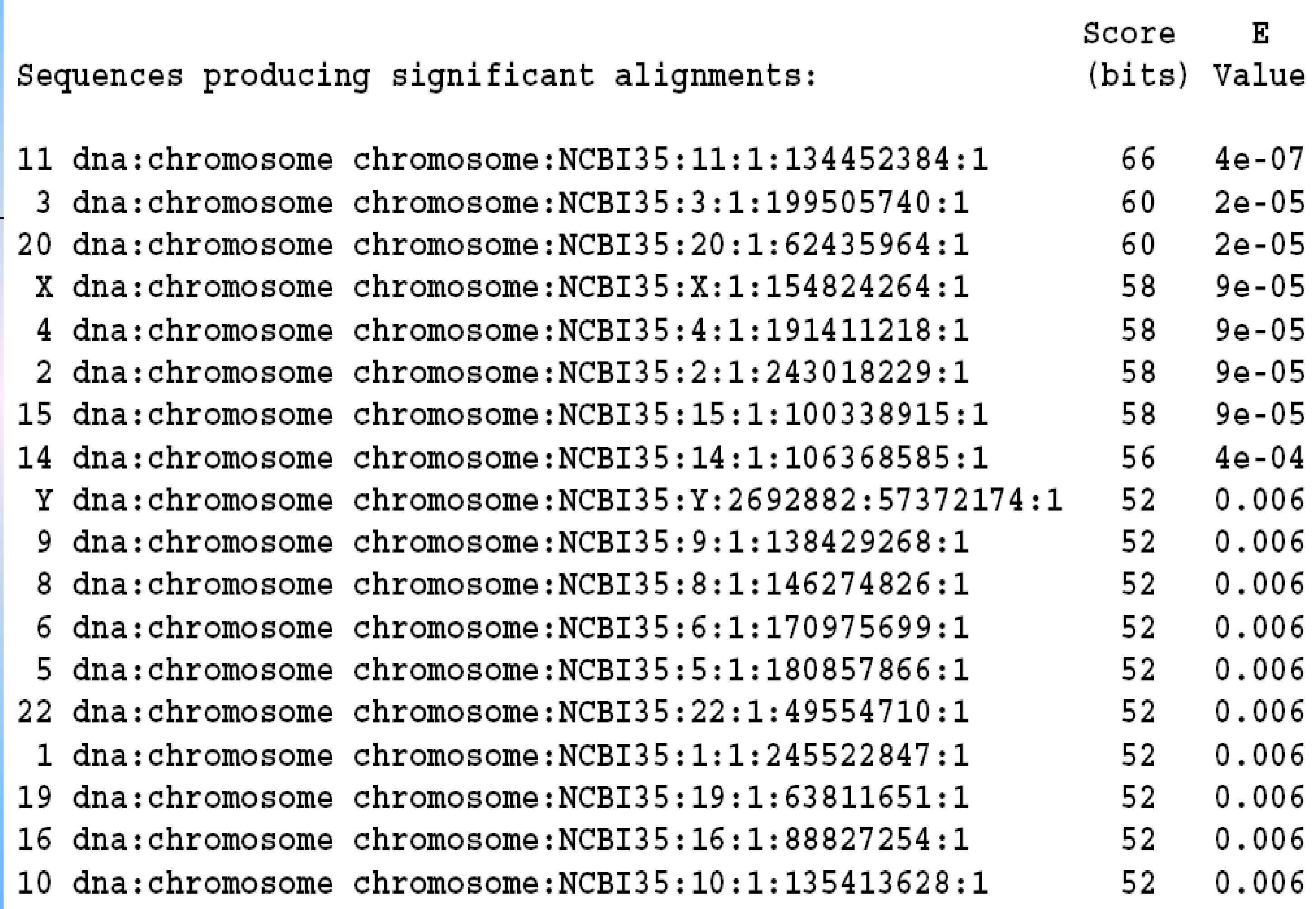
MMTV

```
Query: 3072 tttttgtcattaaaaagaagtcaggaaaatggagactgttacaaga 3117
            ||||||| ||||||||||| || |||||||||||| | ||||||||
Sbjct: 3036 tttttgttattaaaaagaaatctggaaaatggagattattacaaga 3081
```



? Significance of 28 ‘Hits,’
18 Human Chromosomes





```
                                                                    Score    E
Sequences producing significant alignments:                        (bits) Value

11 dna:chromosome chromosome:NCBI35:11:1:134452384:1                 66   4e-07
 3 dna:chromosome chromosome:NCBI35:3:1:199505740:1                  60   2e-05
20 dna:chromosome chromosome:NCBI35:20:1:62435964:1                  60   2e-05
 X dna:chromosome chromosome:NCBI35:X:1:154824264:1                  58   9e-05
 4 dna:chromosome chromosome:NCBI35:4:1:191411218:1                  58   9e-05
 2 dna:chromosome chromosome:NCBI35:2:1:243018229:1                  58   9e-05
15 dna:chromosome chromosome:NCBI35:15:1:100338915:1                 58   9e-05
14 dna:chromosome chromosome:NCBI35:14:1:106368585:1                 56   4e-04
 Y dna:chromosome chromosome:NCBI35:Y:2692882:57372174:1             52   0.006
 9 dna:chromosome chromosome:NCBI35:9:1:138429268:1                  52   0.006
 8 dna:chromosome chromosome:NCBI35:8:1:146274826:1                  52   0.006
 6 dna:chromosome chromosome:NCBI35:6:1:170975699:1                  52   0.006
 5 dna:chromosome chromosome:NCBI35:5:1:180857866:1                  52   0.006
22 dna:chromosome chromosome:NCBI35:22:1:49554710:1                  52   0.006
 1 dna:chromosome chromosome:NCBI35:1:1:245522847:1                  52   0.006
19 dna:chromosome chromosome:NCBI35:19:1:63811651:1                  52   0.006
16 dna:chromosome chromosome:NCBI35:16:1:88827254:1                  52   0.006
10 dna:chromosome chromosome:NCBI35:10:1:135413628:1                 52   0.006
```

? Significance of evolutionarily preserved sequences between: Human Genome, MMTV, SNTV, GNTV

Tools

blastall
formatdb

NCBI BLAST

LINUX POWERED

mfold server

Supported by a grant from the
National Institutes of Health

Algorithms, thermodynamics & statistics for RNA folding.

GM54250

National Institute of General Medical Sciences

National Institutes of Health

NIGMS

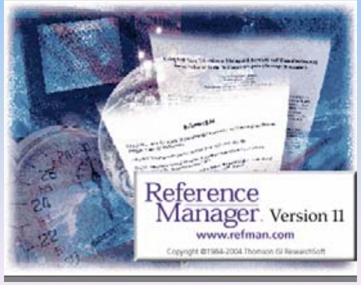
PubMed www.pubmed.gov

A service of the National Library of Medicine and the National Institutes of Health

Entrez Genome

SuSE

Entrez Nucleotide



NCBI nucleotide–nucleotide BLAST

Google

2can Bioinformatics Support Portal

EBI product

Entrez search and retrieval system

Perl

Jalview

EMBL-EBI European Bioinformatics Institute

ClustalW

mfold

DELL

OpenOffice.org

Sequences miRBase

HS LS Health Sciences Library System

University of Pittsburgh

Microsoft Office

Microsoft Windows

The MAVID/AMAP multiple alignment server

Thank you ☺


Dr. Ivet Bahar


Dr. Rajan Munshi


Dr. Bino John

Dr. Judy Wieber Mr. Jason Boles Ms. Sandy Yates

Many program instructors, guest lecturers

BBSI Interns

Mr. Robert Baird
Dr. Timothy Beagley

Dr. Tamara Goetz
Dr. Deborah Neklason

SLCC Study Buddies – Lots!

Common Thread ?

“Walking to the Sky”
Jonathan Borofsky
@Carnegie Mellon

Cathedral of Learning
U of Pittsburgh

John Climacus
Ladder of Divine Ascent

LED·ZEPPELIN

Stairway To Heaven

Your questions…