

Global Alignment with Affine Gap Penalties

Jocelyn Hansson

Constant vs. Affine Gap Penalties

- Mutations:

GATCTAAAAAAAAAAGATCTA

GATCT-----GATCTA

GATCCAG

GATCCAG

GA-C-AG

GA--CAG

- 2 penalties:

- σ : gap opening penalty
- ϵ : gap extension penalty

Constant vs Affine Gap Penalties

Rhesus monkey DNA vs mRNA:

DNA:

```
ATGGAGGGCGAGCTCGCAGCCAACCTGGAGCACCGAGGCGAGTCAACTCCAG
CGCCCGCGCCGCCGGGGGCGAGAGGGCAACTGCACCGCCGAGCCCCCGCGG
CGCAACGAGGCCCTGGCGCGCTGGAGGTGGCGGTGCTCTGTCTCATCTCT
GTTTCTGGCGCTGAGCGGGAATGCGTGTGTGCTGGCGCTGCGCACCA
CGCGCCACAAAGCACTCGCGCCTCTTCTTTCATGAAGCACTGAGCATCG
CCGACCTGGTGGTGGCAGTCTCCAGGTGCTGCGCGAGTTGCTGTGGGAC
ATCACCTTCCGCTTCTACGGGCCCGACCTGCTGTGCCGCTGGTCAAGTAC
TTGCAGGTGGTGGGCATGTTGCGCTCCACCTACCTGCTGCTACTCATGTCC
CTGGACCGCTGCTGGCCATCTGCGAGCCGCTGCGCTGCGCTGCGCCGCG
CACGGACCCGCTGGCAGTGTGCGCCAGTGGCTGCGCTGCGTGGTGGCCA
GCGCGCCGCGAGGTGCACATCTTCTCTGCGCGAGGTGGCTGACGGCGCT
TTGCACTGCTGGCCGCTTTCATCCAGCCCTGGGGACCCAAAGGCTACATC
ACGTGGATCACGCTAGCTGTCTACATCGTGGCGGTATCGTGTGCTGCGTGC
TGCTATGGCCTTATCAGCTTCAAGATCTGCGAGAATTTGCGGCTCAAGACCG
CTGCAGCTGCGCGCGCAGAGGCGCCAGAGGCGCGCGGCTGGCGATGG
GGGGCGCATGGCCCTGGCAGCTGTACGACGCGTCAAGCTCATCTCCAAGG
CCAAGATCCGACGGTCAAGATGACTTTCATCATCGTGTGGCCTTTCATCGT
GTGCTGGACGCTTCTTCTTCTGTCAGATGTGGAGCGTCTGGGATGCCAA
CGCGCCCAAGGAAGGTAGCCAGGGCTGGGAGACCCAGGAGGAGGGAGCC
TGGTGGCTGGGGGACGCCCTTATCTTGTGCTCGCAGAAATGCCAGGGGCT
TTGACTTCTCTGGGGGATAAGAGGGTTTGAATCCCAAGAGTCAAGTCTT
CCATCATCCCTTGCCAAAGTACCTAGGCAATTAAGCTCCCTGAGGCTCCA
CTTTCTCATCTGTGAGGTGGCAATAAGGATAAAAGTACCAACTGTCACTGGG
CATAGGGGCGAGCCATGAGAAAATGCAAGTTAAAGCTTTAGCACAGTCCCTTG
GGCTGCATATGGGCTGCATGGTTTACTGCGGTGGTGGAAACAGGTTCAAGGG
ACTCCATCTGCTTTCCACGTGGTTAGGAGGAGTGTAGGAGGAGATGG
CAGAGGTAAGTTTGAACCCCTGGCCAGGCTCAAACTCTTCAAGTTTAAAGT
TCACATTAAGCTGAACCTTCACTTTAAGTTCAAGAAATCCATTTGAAGCCAA
AAGTCTGGTTTGGACAAGGACAGCCTTCAGGGGGTGGGCGATTGTCCAGC
CAAGCCCTAGTAGTTGAAGGGGGTGTGGGGGCGAGGATTCAAGGG
AGAGGTGAAGACAAATCCCGGAACCCCTCATCGAGTGAAGGTGAGTCT
CCATGCAAGAGGTGAGTACTGCAAAAAGCGGTGGGCAGGCTTAGGGAGG
AGAGTGCCCGAGACTGAGACTTCACTTCTACTACTTGGGGA.....
```

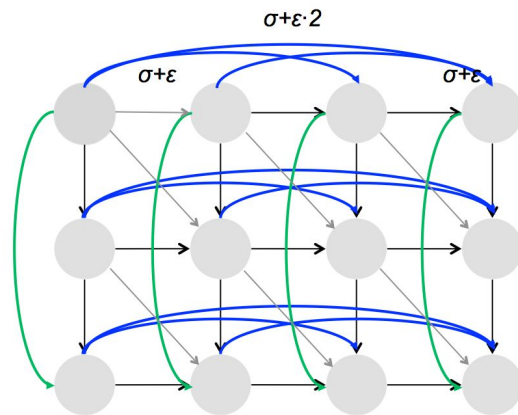
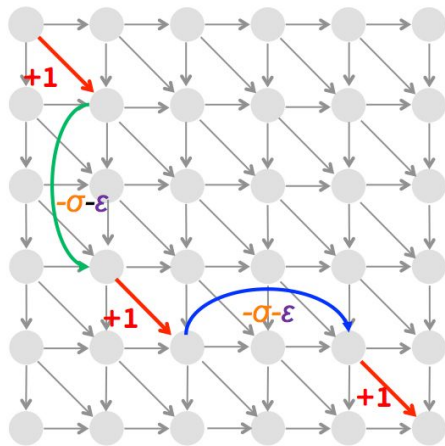
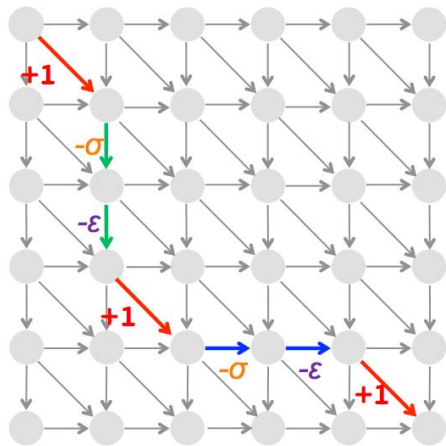
mRNA with constant gap penalties:

```
ATGGAGGGCGAGCTCGCAGCCAACCTGGAGCACCGAGGCGAGTCAACTCCA
GCGCCGCGCCGCCGGGGGCG-----C-G-A-G-GG-C-----AAC-----TG-
C-----A-C-C-G-C-C-G-----G-A-C-C-----C-C-CG-C-G-G-
--GC--A-A-C-G-----A-G-GC-CGTG-GC--GC-G-C-G-TGG--AG-
--G-TG--GC-G-GT-GCT--C-T-----G-T-CT-C-----A-T-C-C-TG-T-
T-T-C-TGG--C-G-C-T-----G-A-G-C--GG-G-----A-A-
--TGCG-T-G-TG-----TG-C-TGCT-G--G-C-G-CTGC-G--C-A-C-
C-----A-C-----GCGC-----C-AC-----A-----A-GC--AC-
--T-C-G--G--C-GC-CT-CT--TC-T--TC-T-T-C-----ATG--A-AG
C--A-C-C--T-G--A-G-C--A-T-CG-C-C-GA-C-C-----TGG-
-TGG--TGGC--A-GT-C-----T--TC-C-A-GG--T-GC--T-G-
-C--G--C--A-GT-TGCTG--T--G-G-GA--C-----AT-C-A-
C-C-----T-----T--CC-G--C-T-T-CT--ACG--G-GCC-C
-GA--CC-----T-G-CT--G--T-G-----CC--G-TC-TG-----G-
--T-C-A-A--A--G-T-ACIT--G--CA-GG-T-G--GT-
-GGGAGT-G--T--T-C-G--C--C--TCC--A--C-T--A
C-C-TG-----C--T--G-CT-A-CT-C--A-TG-T-----C-----
--CC-TGG-A-C-C-----G--C--T-GC-CT-----G--G-CC
AT--C--T-G--C--CAGCG-GC--TG-GC--C-----T-----
--C-G--C--T--G--C--G--CC--G-----C-C-GC
A-----C-G--G--A-C-----C-G-C-CT-----GG-----
--C-AG--T-G--CT--C-G--C-----C-A-----C
-G-TG-G-C-T-----G--G-C-TG-CC--T-GG--T-G-G-
C-C-AG--C-G--C-G-C--G-C-----A-GGTGC-----A-
-CA-----T-C-T-TC-T-C--T-C-----T-G-C-G--C-G-AGG-T
G--GCTG--A-----C--G-G-C-G--T-C--TTC-G--
-A-C--T-GCT--G--G-GC-G--GT--C-TTCA--T-
C-C-A--G--C--C-C-TG--GGGA--C-CCA-----A-G-G-C-
CT-----A-CA--TC-A-C--G--T--G-G-A-T--C-A--C--G-C
-T-A-----G-C-TGT--T--ACA-T-C-G--T--G-CC-G-G-
--T-CA--TC--G-T-G--CT--C--G--C--T--G-C-C-TG-
-----C--TA-T-G-G-C-CT--TA--T--C-A--G--C-T--T--C-AA
GA-----TC--T--G-G--C--AG-A-A-T-T-----TG--C-G-----
```

mRNA with affine gap penalties:

```
ATGGAGGGCGAGCTCGCAGCCAACCTGGAGCACCGAGGCGAGTCAAC
CTCCAGCGCCGCGCCGCCGGGGGCGAGGGAACCTGCACCGCC
GGACCCCCCGCGCGCAACGAGGCCCTGGCGCGCGTGGAGGTGG
CGGTGCTCTGTCTCATCCTGTTTCTGGCGCTGAGCGGGAATGCGT
GTGTGCTGCTGGCGCTGCGCACCACGCGCCACAAGCACTGCGCG
CTCTTCTTCTCATGAAGCACTGAGCATCGCCGACCTGGTGGTG
GCAGTCTTCCAGGTGCTGCCGCAAGTTGCTGTGGGACATCACCTTC
CGCTTCTACGGGCCGACCTGCTGTGCCGTCTGGTCAAGTACTT
GCAGGTGGTGGGCATGTTGCGCTCCACCTACCTGCTACTCAT
GTCCCTGGACCGCTGCTGCGCCATCTGCCAGCCGCTGCGCTCGC
TGCGCCCGCGCAGGACCGCTGGCAGTGTCTGCCACGCTGGCT
CGGTGCTGGTGGCCAGCGCGCCGAGGTGCACATCTTCTCTC
TGCAGGAGGTGGCTGACGGCGTCTTCACTGCTGGCGGTCTTC
ATCCAGCCCTGGGGACCCAAAGGCTACATCACGTGGATCACGCT
AGCTGTCTACATCGTGGCCGTCATCGTCTGCTGCTGCTATGG
CCTTATCAGCTTCAAGATCTGCGAGAATTTGCGGCTCAAGACCGC
TGCAGCTGCGCGCGCAGAGGCCCAAGGCGCGCGCGCTGAG
CGATGGGGGGCGCATGGCCCTGGCAGTGTGACGACGCTGAG
CTCATCTCCAAGGCCAAGATCCGACGGTCAAGATGACTTTCATC
ATCGTGTGCGCCTTTCATCGTGTGCTGGACACCTTCTTCTCTG
CAGATGTGGAGCGTCTGGGATGCCAATGCGCCCAAGGA-----
-----
-----AGCCTCGGCCCTTCATCATCGTATGCTCCT
GGCCAGCCTCAACAGCTGCTGCAACCCCTGGATCTACATGCTGT
CACGGGTCACTCTTCCACGAACCTCGTGACGCGCTTCTGTGCTG
CTCCGCCAGCTACCTAAAGGGCAACCGCTAGGAGAGCAAGTA
CCAGCAAAAAGAGCAACTCGTCTTCTTTGCTGAGCCATCGCA
GCTCCAGCCAGAGGAGTGTCTCCCAACCATCCAGCGCATGACCT
ACCAGCCAGGGCGAGGGCTGCTCCTGAGGTCTGGGCTGTGCTG
GCATAAGTGGTCTGCCCTAGGTGATGGCGTATGTT
```

Long edges in graph



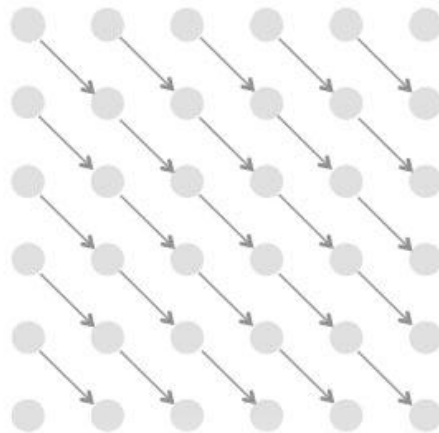
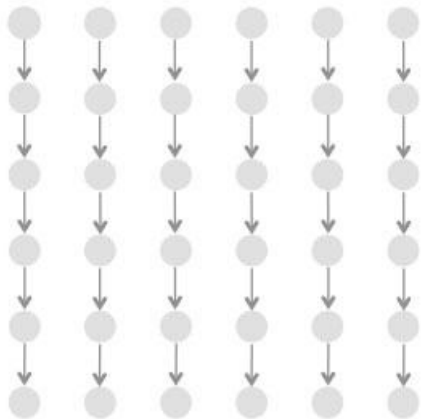
Alignment graph as 3 levels

$$\text{lower}_{ij} = \max \begin{cases} \text{gap extension} \\ \text{Initialize gap} \end{cases}$$

$$\begin{aligned} &\text{lower}_{i-1,j} - \epsilon \\ &\text{middle}_{i-1,j} - \sigma \end{aligned}$$

lower

(Insertions on string 2)

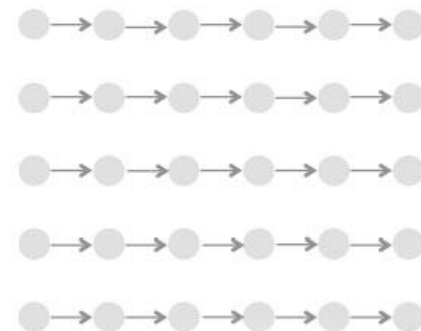


middle

(matches/mismatches)

$$\text{middle}_{ij} = \max \begin{cases} \text{ending gap on string 1} \\ \text{match/mismatch} \\ \text{ending gap on string 2} \end{cases}$$

$$\begin{aligned} &\text{upper}_{ij} \\ &\text{middle}_{i-1,j-1} + \text{score} \\ &\text{lower}_{ij} \end{aligned}$$



upper

(Insertions on string 1)

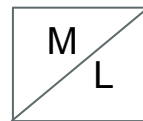
$$\text{upper}_{ij} = \max \begin{cases} \text{gap extension} \\ \text{Initialize gap} \end{cases} \quad \begin{aligned} &\text{upper}_{i,j-1} - \epsilon \\ &\text{middle}_{i,j-1} - \sigma \end{aligned}$$

Backtracking

Why you need 3 backtracking tables: tied scores

- Where you came from matters!
- This ended up in gaps being opened that were not penalized

$$\text{middle}_{ij} = \max \left\{ \begin{array}{l} \text{ending gap on string 1} \\ \text{match/mismatch} \\ \text{ending gap on string 2} \end{array} \begin{array}{l} \text{upper}_{ij} \\ \text{middle}_{i-1,j-1} \\ \text{lower}_{ij} \end{array} + \text{score} \right.$$



Making 3 backtrack tables

- Keep track of which move had highest score
- Tables: 'L' 'M' 'U'
- Set backtrack value to table you came from

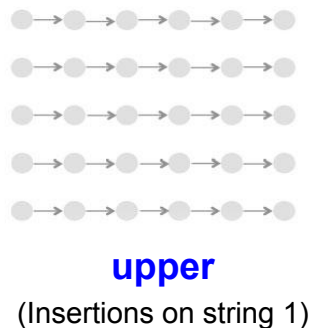
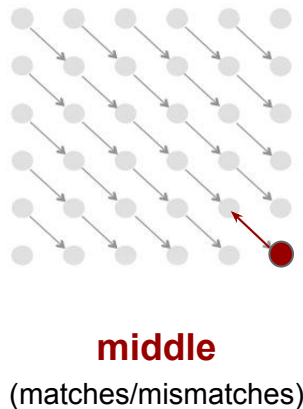
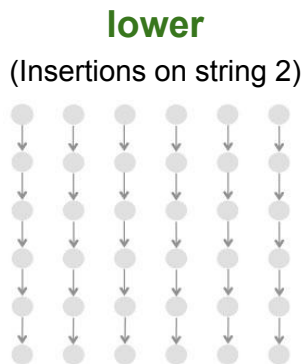
$$\text{upper}_{ij} = \max \left\{ \begin{array}{ll} \text{gap extension} & \text{upper} \rightarrow \text{upper} \\ \text{Initialize gap} & \text{middle} \rightarrow \text{upper} \end{array} \right.$$

$$\text{middle}_{ij} = \max \left\{ \begin{array}{ll} \text{ending gap on string 1} & \text{upper} \rightarrow \text{middle} \\ \text{match/mismatch} & \text{middle} \rightarrow \text{middle} \\ \text{ending gap on string 2} & \text{lower} \rightarrow \text{middle} \end{array} \right.$$

$$\text{lower}_{ij} = \max \left\{ \begin{array}{ll} \text{gap extension} & \text{lower} \rightarrow \text{lower} \\ \text{Initialize gap} & \text{middle} \rightarrow \text{lower} \end{array} \right.$$

Backtracking

- Start at end of middle table
- Keep track of what table you're in with: `table='M'`
 - Change this value when you move to different tables



- If `table = 'U'` and backtrack value = `'M'`
 - Add letter to alignment string 1
 - Add `'-'` to alignment string 2
 - Change table value to `'M'`

Data

- OXTR gene
 - Oxytocin receptor
 - Relevant for several social functions such as bonding and maternal behavior
 - Gene that may be relevant for Autism Spectrum Disorder (ASD)
 - Mouse models of ASD are frequently used
- Comparing OXTR gene between humans, mice and other animals
- All data was found on [NCBI.nlm.nih.gov](https://www.ncbi.nlm.nih.gov)

Results

- OXTR mRNA: Human vs Mouse:

- Global alignment with affine gap score: 1,113

Human mRNA:

GTGTTAAGGCTCTGGGACCAACGCTGGGCCGCCAGCTCCGCTCCGGAGGGGTCTGCGCGGCTGGCCCT--CGCCCCGCCCTACGCGAACCGTGCGATAGTGCAGCCTCAGCCCCA
CGCGCAACGCGCGCATCCAGACGCTCTCCCGCGCGGACGCTGGGAGGGGCTCTCCCTCGCCCTCTGATCCCATCCAGCGACCGACGAGCTCGCGCGAGGGGATCCCAACCG
AGGCTCCAGTAGAGACACCTCACTTAGTACATAGTTCAGCGCGGACCGGACCTCCAACTCGGGCG--GAGCGACGCGCTCAGTGGGCGGTGTCAGTCGCCGT
GCAACTTCCCCGGGGGAGTCAACTTTAGGTTCCGCTGCGGACTCGGTGCAATGGAAGCCGCTGAACATCCGAGGAAGTGGCACGCTGGGGGCT-----CTGGGCTTG
TGGCCGGTAGAGGATTTCCCGCTATTTGCAGTGGCTCAGAGGGGTGGACCCAGCACTCCGCTCGTGGAGTCTCCAGGATGGAGCGCCCGGCTACACCCCTCCGACAC
CGCGGATCCGGCCGACCGCGCCAAAGCGCTAAAGGGCTGAAGGCGGGGCGACCGCTCGCGCCAGGCTCATGGAAGGGCGCGCTCGACGCAACTGGAAGCGCGGAGGACG
CAACGCCACAGCGCCGCGCCGCGGGGGCGAGGGCAACCGCACCGCCGACCCCGCGGCGCAACGAGGCCCTGGCCGCGCGTGAGGTGGCGGTGCTGTGTCTATCTGCTGCT
CTGGCGCTGAGCGGGAACCGGTGCTGCTCGGCTCGGCCACACAGCCAGCAAGCACTCGCCCTTCTTCTCATGAAGCACTAAGCATCGCGACCTGCGGAGCTGGTGGGACG
TGTTTCAGGTGCTCGCGAGTTGCTGTGGGACATACCTTCGCTTACGGGCGGACACTGCTGTGGCGCTGGTCAAGTACTTCAGAGTGTGGGCATGTTCCGCTCCACCTAC
CTGCTGCTGCTCATGTCC--CTGACCGCTGCTGCTGGCCATCTGCGACGCCGTCGCGCTGCTGCGCGCGCCACCGACCGCTGGCAGTCTGCGCCACGTGGCTCGGCTGCCGTGT
GGCAGCGCGCGCGAGGTGCACATCTTCTCTGCGAGGTGCTGACGCGCTTTCAGTCTGGGCGTCTTATCCACGCCCTGGGGACCCCAAGCGCTACATACATGGATC
ACGCTAGCTGTCTACATCTGCGCGGTATCGTGTCTGCTGCTGCTGCTACGGCTTATACGTTCAAGATCTGGCAGAACTTCGCGCTCAAGACCGCTGACGCGCGCGCGCGGAGGC
GCCAGAGGCGCGCGGCTGCGCATGCGGCGCGGCTGGCCCTGGCGGCTGACGACGCTCAAGCTCATCTCAAAGGCCAAGATCCGACCGGTCAAGATGACTTTATCATCGT
GCTGGCCTTATCGTGTGCTGACGCGCTTCTTCTGTGTCAGATGTGGAGCGTCTGGGATGCCAACGCGCCAAAGGACCTCGGCCT-----TCATC
ATCGTCTATGCTCTTGGCCAGCCTCAACAGCTGCTGCAACCCCTGGATCTACATGCTGTTACGCGGCCACCTCTTCCACGAACCTGTGCAGCGCTTCTGTGCTGCTCCGCCAGCTAC
CTGAAGGCGCAGACGCTGGGAGAGACGAGTGCCAGCAAAAAGGCAACTGCTCTCTTCTTCTGCTGAGCCATCGAGCTCGACGACGAGAGGCTGCTCCGAC--CCATCCACGGCG
TGACCCACGACGCGACGGCGAGGCTGACGCTGAGCTGAGCTGCTGTGCGTAAAGTCTGCTCTAGGTGATGGCGTATGTTTGTATGAAGTACCTATCAGTTTGTATCC
TCCCTCTCTTGGGGTGGCTTCACTGGGGTGGAGAGTGGCTCCATGATGGAAGATGATAGGGGACTCAGCCATCAGACAACACCCCTGGCCTCTACACGTACTTCTACCCCTGA
ACCCATCGCTGCCCTGGGAGTGTAGTGCTGTTTTCTCTCGTGACTTGAATTTACCTCAGTATATTTTACTTCTTCTATCTGGGATTGTGAAAGCCGGAATAATAGGATTGG
TGACCAATTGGGTGAGGAAGTCCAGTGTCTTGAGCTGGGGTAAGCAGTGGGGTGGGACCTCAGATGGG--AAGGGTGGTGCATAGATCTCTCTGAC
CTCAAAGTGATTTGCTTTAAGCGAACAAATGCTGGGTCTTGGGACCAAGCTTGTGAGAGGTAGCCCTAAGAGAAGGGGATTACCTTGAAGACCATCTGGCGCAGTGGACCT
ATTAGAAGCTGGGTAAAAATGTTTAAAGAGCTAATGTTTAAAGAGCATTTGGGAAAGAAAGAAATAAATGATCAGACATGGAAGAAGAAGT-----AAAC
TATTTGCAGATGACACAGTTTGTATTAAGAAATCTCAAGAAATCACACACACACACACGACGACACAGCTATTAGAAGTAATAAGCAAGTTCGCGAAGGTTTCAAGAT
ACAAGATCAATATACAAAAATGAATTTGATTTCTTATACTAGCAACAACAATATGAAACGAAGTTAAATAATTCCATTTATAATACCACAGAAGAATAAAATAGGAATCAACTTAAC
AAAAACAAGTGCAGACGTGAAACATCAAAATGGAAAGAAATTAAGAAAGCTTAAATAAATGGAAGACATCTGTGTTCAATGGATCAGACTAGTATTGTAAAGATTGGAATCTATC
CTAAGTACATCGAGATTCACTGCAACTCTTATGAAAACTATAGCT-----GGCTTCTTACAGAAATGATAGAGTAGTCCCAAATTCATAAGAAATGAAGGACCGCAAGAT....

Results

- OXTR mRNA: Mouse vs Rat
 - Global alignment with affine gap penalties score: 2,451

Mouse mRNA:

```
-----CGCGGTTGGTCATGGAG
GGCACGCCCCGACGCCAACTGGAGTATCGAGTTGGACCTCGGGAGTGGAGTGCCACCAAGGGGCGGA
GGGTAACCTCACGGCCGGGCCGCCACGACGCAACGAGGCCCTGGCGCGCGTGGAGGTGGCGGTC
CTGTGTCATACTGTTCTGGCTCTGAGTGGCAACGCGTGCGTGCTGCTGGCGCTGCGTACGACG
CGCCACAAGCACTCGCGCCTCTTCTTTTCATGAAGCACTGAGCATCGCCGACCTGGTGGTGGCCG
TGTTCCAGGTTCTCCCGCAGCTGCTGTGGGACATCACCTTCCGCTTCTACGGGCCCCACCTGCTGTG
TCGTCTGGTCAAATCTTGCAGGTGGTGGGCATGTTGCGCTCCACCTACCTGCTGTTGCTGATGCG
CTCGACCGCTGCCTGGCCATCTGCCAGCCGCTGCGCTCACTGCGCCGCCGAACCGACCGCCTGGC
GGTGCTGGCGACGTGGCTCGGCTGCTGTTGGCCAGCGTGCCGCGAGGTGCACATTTTCTCGCTGC
GCGAAGTGGCGGACGCGCTTTCGATTGCTGGGCGGTCTTCATCCAGCCCTGGGGACCCAAGGCCCT
ACGTCACGTGGATCACGCTCGCCGCTTACATTGTACCGGTCACTGCTGCTGGCCGCTGCTATGTTCT
CATCAGCTTCAAGATCTGGCAGAATCTGCGACTCAAGACGGCAGCCGCGGCGGACAGCTGCCGAGGG
GAGTGCAGCGACGCCGGTGGAGCTGGCCGTGCGGCGTTGGCACGGGTCACTAGTGTCGAAGCTTATCTC
CAAGGCCAAAATCCGACAGTGAAGATGACCTTTCATCATTGTTCTGGCCTTCATCGTGTGCTGGACG
CTTTTCTTCTCGTGCAGATGTGGAGCTCTGGGACGTCAATGCGCCAAAGCTTCTGCCCTCA
TCATTGCCATGCTCTTGGGAGCCTCAACAGCTGCTGCAACCCATGAGTCTACATGCTCTTCAAGGG
CCATCTCTTCCAGAACTCGTGACGCGCTTCTCTGCTGCTGCTGCTGCGTACTGAAGGGCAGCCGG
CCTGGAGACGACGACGACCTTAGCAAGAAAAGCAACTCTCCACCTTCGCTCGTGAAGTCTCGCAGCTCGA
GTCAGAGGAGCTGTTCTCAACCATCTCGGCATGAGCCATAGGCTGGCCACCAGGCCAGCCGCG--
--TGGGTGGTGTAGCCGCCGATCTTCCCTCTGGTGGCTGTGCATGGAGCTGTATATGGTGCCTATTG
ATTGGTTTGCATCCCTCTGTGCTTGGGCTGGTTAGATTCTGTGTCATATTCTGACTTGGGGTAGGGA
AATGTCTCCATGGGAGATGACAGGGTCACTCAGCCATCAGAGGATCACCTGGACTCCACCC--TATT
CTACTACCTGACTCAAGTGTTCTTGGGCACTGGGTAGATTCTCTGACCTCACTTCTGCTCCCTCC
ACCCCTTGTATTCTGGTATCTTGTGAAAGACGGCACACCTAGGATTGGTACAGGATTAAGCTGAGGA
AGTCCAGTGGGATGAGACAAGGAGTGGGAAATGAGGCTGCAGACGCAGAGTGGTGCTAATTTCTCT
CTGACCTGAGTTGTGTGGACAGATCTTAGCATCTTAGGACATCCCAAGCAAGCATTTGGAGCTGAA
ACTGAGCCGTGAAGCCACCTGGCTCTGGAAGGAATGAGAACGAGGGCTGAAAGGTTGAAGATGTTG
TTGCTTGACAACATTTGGGAAGGAAAAAGAAATGAGAGCCACCCAGGACAGACGAGGGAGGAGAG
CTGTCTGTACACACACTGATCCAT-----GCATATGTGATCTAGGAGCGCTCACAC
```

Rat mRNA:

```
GGAGAGGACAGGCAGGTAGACACCTGGACCGAGCACACGAGCCCCGCCACAGTGCCACGCTTTAAAGA
ATCGTAAGACTTAGTGCTCCCTGGAGGTGCGGGTGCGGGATGAGAGGGCAGCCGACGACGCC
AACTGGAGTGTGAGTTGGACCTCGGAGTGGAGTGCCACCAGGGGAGGAGGGCAACCGCAGGCCG
GGCCACCACAACGCAACGAGGCCCTGGCGCGCGTGGAGGTGGCGGTCTGTGCCTCATTTCTGTCCTG
GCTCTGAGTGGCAACGCGTGTGTCTGCTGGCGCTGCGCACCACGCGCCACAAGCACTCGCGCCTCTT
CTTTTTCATGAAGCACCTGAGCATCGTGACCTGGTGGTGGCTGTGTTTCAGGTGCTTCCGACAGTGTG
GTGGGACATCACTTCCGCTTCTATGGGCCGACCTGCTGTGCTGCTGGTCAAATACTTGACAGTGGT
GGGCATGTTGCGCTCCACCTACCTGCTGCTGCTGATGTCGCTCGACCGCTGTCTGGCCATCTGCCAGCC
GCTGCGCTCCCTGCGCCGCCGAACCGACCGCTGGCGGTGCTGGGACAGTGGCTTGGCTGCTGGTG
GCCAGTGCTCCGAGGTGCACATTTTCTGCTGCGCAAGTGGCGGACGGAGTCTTCAATTGCTGGGC
GGTCTTCATCAAACCTTGGGACCCAAGGCCCTACGTCAATGATCAGCTCGCCGCTCACATTGTACC
GGTCATCGTGTGGCCGCTGCTACGGCTCATCAGCTTCAAGATCTGGCAGAATCTGAGACTCAAGAC
GGCAGCGCGCGCGCGCTGCCGAGGGAATGACGCGCGGGTGGAGCTGGCGCTGCGGCGTTGGC
ACGGGCTCAGTAGTGTCAAGCTTATCTCAAAGGCCAAAATCCGACGAGTGAAGATGACCTTTCATCATCGTA
CTGGCTTCACTGTGTGCTGGACACCTTCTTCTCTGTCGATGTGGAGCTTTGGGACGTCATAGCG
CCCAAGGAAGCTTCTGCGCTTTCATCATTGGCCATGCTCTTGGCCAGCTCAACAGCTGCTGCAACCCGTGG
ATCTACATGCTCTTCAAGGTCACTCTTCCACGAACCTGTGTCAGCGCTTCTTCTGCTGCTGCTGCTGTT
ACCTGAAGGGCAGCCGCGCGCGGAGACGAGTGTGACGAAGAAGAGCAACTCATCCACCTTTGTCTG
AGCCGTCGACGTCAAGCCAGAGGAGCTGCTCTCAGCCATCTTCAAGATGAGCCATAGGCTGGCCCA
CCGGGCCAGCCGCGAGGTTGGTGGTGGAGTCCAGTCTTCCCCGCTGGTGGCTGTGATGAAGCTGCG
ATTAGGTGCTCT----ATTGGTTTGCATCCCTTCTCTCCTTGAAGTGGCTAGAGCGGTGTGTCATATTCTGACGT
GGGTGGGGAAATGTCTATGGGGATGATCGGGTCACTCAGCCATCAGAGGCTCACCTGAACCTCCC
ACCTGTGTTTCTACTACCTGACTCAAGTGCTCTGGTGGTGGTGGAGCTTGGACACCTGGCTTTTG
TCCCTTTC-CTCTTGTGTTATCCTGGTATCTTGTGAAAGACAGCAGCGCTAGGATAGGTGACTTGAAGCC--
-----CAGTGGGCTGACAAGGATGGGAAATGAGGCTGCGGACGCAAGTGGTGCTAATTTCCCTGTG
ACCTGAGTGTGTGGACAGATCTTAGCATCTTAGGCGATCCTGGAAGCACTGGGAACCGGAAACTGA
GCTGGGAGGCTGCTGGCTGTGAAAGGTGATGAGAACGAGGCCAAAATGTTGAAGATGCTGCTGCTTG
GCAACATTTAGGAAGGAAAGGAGAGTATCCAGGACAGGTGAGGGAGGGGAGCTGTTGTATCATGAC
ACTGACCCGTGCATATGTGATCTAGGAGCACTCACATACATATGCTACGTGCACACAC
```

Results

Main Results:

- Human and mouse mRNA were very similar, but rat and mouse mRNAs were even more similar

Comparison	Affine gap score
Mouse vs rat	2,451
Mouse vs human	1,113
Human vs rat	418

Other comparisons:

Human vs rhesus monkey	-1,919
Dog vs monkey	794

- Some unexpected differences in mRNA length

organism	mRNA length
Rat	5,335
Mouse	4,568
Human	4,361
Dog (boxer)	1,275
Rhesus monkey	1,253