

Bi 188

- Course Syllabus
- Textbook
- Mussa
- Genome Browsers
- Extra Credit Lecture

Syllabus

- Webpage: <http://woldlab.caltech.edu/bi188/>
- Midterm (April 30)
- Project (May 18)
- Final (May 30 - June 1)
- (Problem Sets associated with Readings)

Office Hours

- Barbara Wold
 - Tuesday, 3-4
- Ali Mortazavi
 - Friday, 1-2
- Steven Kuntz
 - Thursday, 3-4

Textbook

*Recombinant DNA: Genes and Genomes –
A Short Course, 3rd edition 2007*

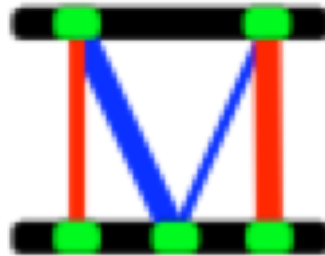
J. Watson, A. Caudy, R. Myers, and J.
Witkowski

ISBN: 0-7167-2866-4



Mussa GL

An Introduction



Mussa GL

- Mussa - *Multiple Species Sequence Analysis*
- Runs on any major platform (Linux, MacOSX, WindowsXP)
- Graphics require OpenGL

What it is for

- Identify conserved regions of DNA
- Conserved regions must be under evolutionary pressure
- Conserved regions possess some functionality
- Some conserved regions are associated with protein-binding sites
- Thus conserved regions may be associated with cis-regulatory elements

What it does

- Uses a sliding window for comparison
- Identifies all windows that match within a certain threshold
- Higher thresholds:
 - Less noise
 - More likely to miss a signal
- Window-size is clade-dependent (for mammals use 30)

What it does

TGGTATCATGGGGATATATGTAGGCCTAATACAGG

TCGTATCATGCAGATATATGCAGGCCTATTACAAG

What it does

TGGTATCATGGGGATATATGTAGGCCTAATACAGG

TCGTATCATGCAGATATATGCAGGCCTATTACAAG

What it does

TGGTATCATGGGGATATATGTAGGCCTAATACAGG

TCGTATCATGCAGATATATGCAGGCCTATTACAAG

Why is this useful?

- Identify ungapped-matches
- Ungapped matches are a sign of evolutionary conservation
- Ungapped matches can be used on larger elements even if part of the element is frame-shifted or flipped
- Can split palindromes

C. elegans

TCCTTATGCTTCATGTCAAAATGATGTGGAGAGAGGGGAAAAGAAGGGG

C. briggsae

TCCTTATGCTTCATGTCAAAATGATTGGAGTCTTCTAGTGGGGTACTG

CB5161

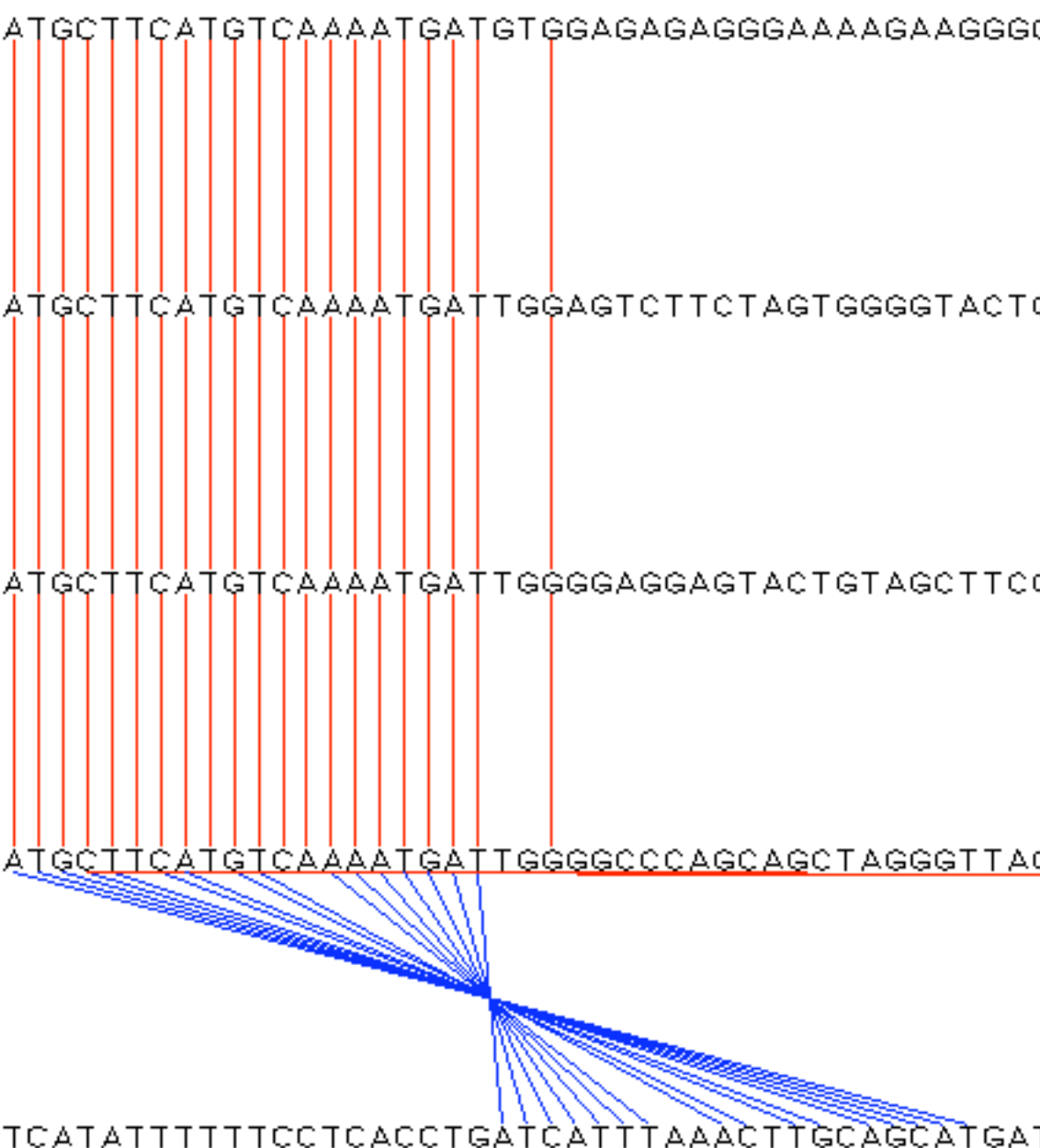
CCCGTATGCTTCATGTCAAAATGATTGGGGAGGAGTACTGTAGCTTCC

C. remanei

GACGTATGCTTCATGTCAAAATGATTGGGGCCCAGCAGCTAGGGTTAC

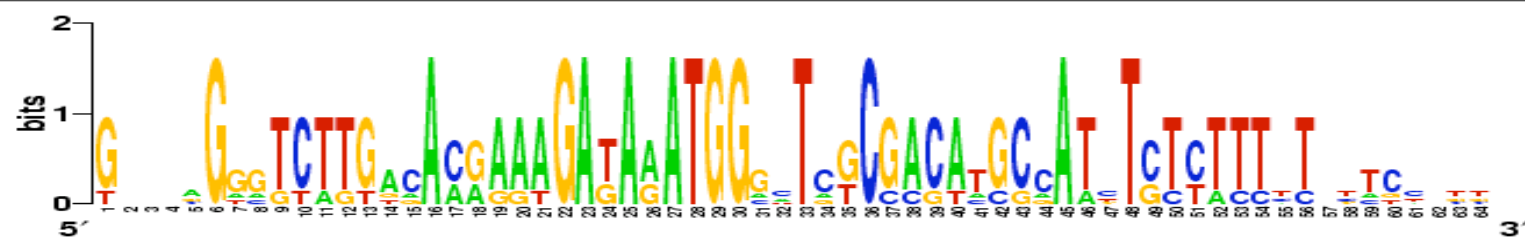
PSI010

CATTTTCATATTTTTTCCTCACCTGATCATTTAAACTTGCAGCATGAT



Why is this useful?

- Identify ungapped-matches
- Ungapped matches are a sign of evolutionary conservation
- Ungapped matches can be used on larger elements even if part of the element is frame-shifted or flipped
- Can split palindromes



C. elegans

GGGGTCTTGACACGAAAGATAAAATGGGCTCGCGACATGCCATCTCTCTTTTGGCTCCGCCAC

C. briggsae

CGGGTCTTGACACGAAAGATAAAATGGGTTTCGCGACATGCCATTCTTTTCTCCTTATTTTT

CB5161

AGGGTCTTGACACGAAAGATAAAATGGGCTCGCGACATGCCATCTCTCTTTCCATTCTTGGCT

C. remanei

GGGGTCTTGACACGAAAGATAAAATGGGCTCGCGACATGCCATCTCTCTCTTTTCCCTTTCT

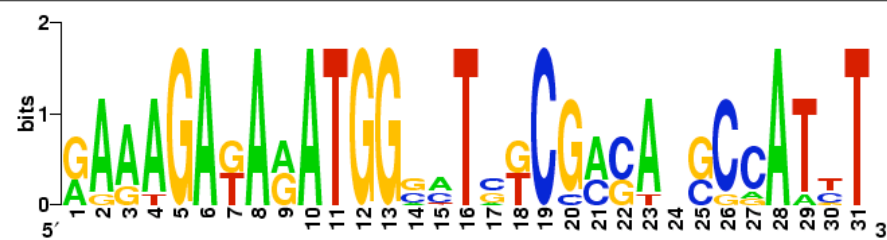
PS1010

CGAGTCTTGACACGACTCTAAAAAGACAAATCGGTTTCGCGACATGCCATCTCTCTTTCTTTCC

D. rerio hoxb1a a2

AGACTCTTGTA AAAAGTGAGAGATGGAATATCCACTCGGAAAGTGCTATCATATTCCATGGC

L2-I
Sub7
11/20



C. elegans

CTCGCGACATGCCATCTCTCTTTTGTCTCCGCCACTGACTAT

C. briggsae

GAAAGATAAATGGGTTTCGCGACAT

CB5161

GAAAGATAAATGGGCTCGCGACAT

C. remanei

GAAAGATAAATGGGCTCGCGACAT

CB5161 f2

CTCGCGACATGCCATCTCTCTTTCC

C. remanei f2

CTCGCGACATGCCATCTCTCTTTT

PS1010

TTCGCGACATGCCATCTCTCTTCTT

D. rerio hoxb1a a2 f1

AAGTGAGAGATGGAATATCCACTC

L2-I
Sub7Corr2
12/20

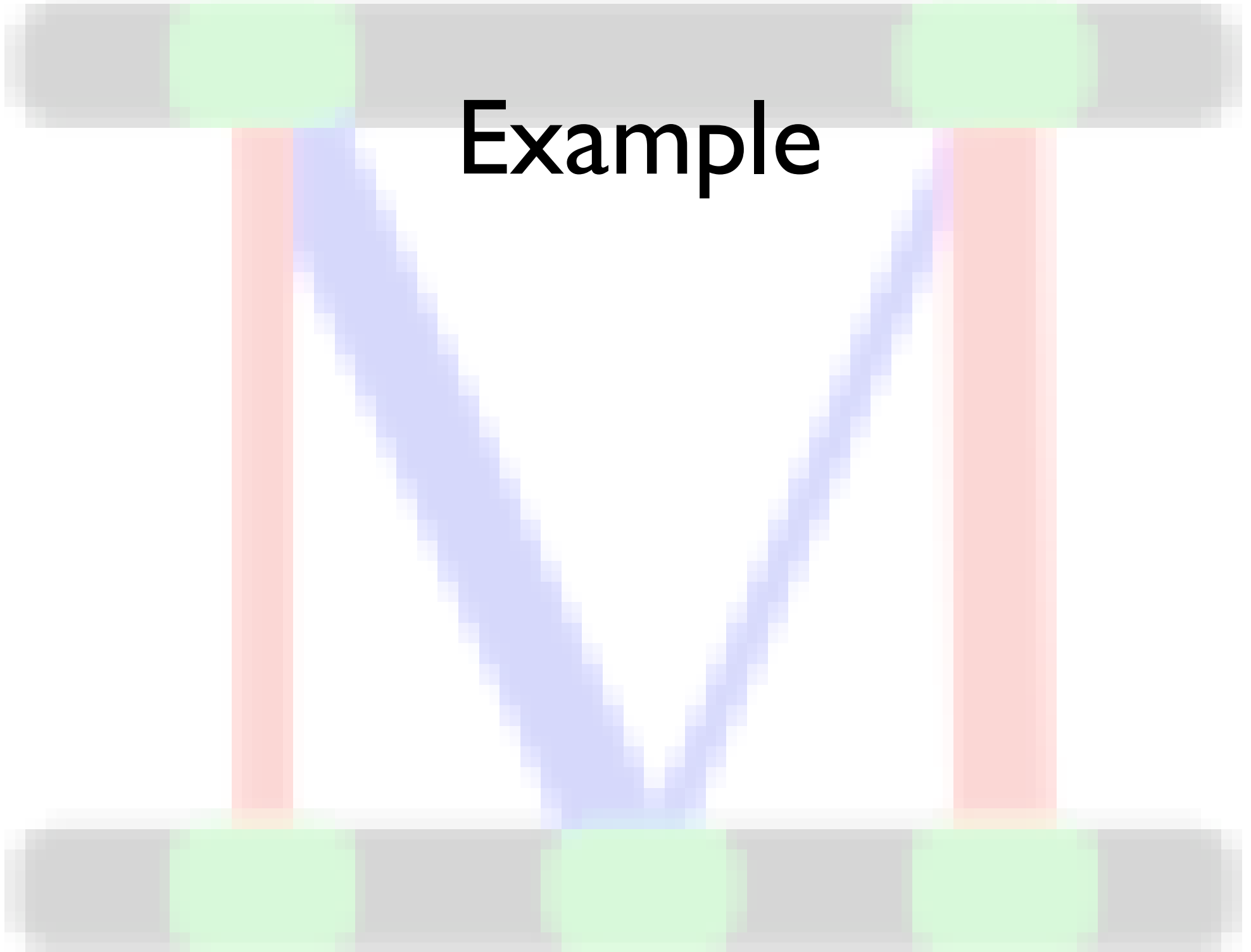
What it will miss

- Any significant insertion or deletion can 'hide' a conserved region
- Luckily, most regulatory regions have rare insertions or deletions (in-dels)
- Conservation is very species-dependent
- Repeat-masking can mask real conserved regions

What you need to do to make it run

- Identify relevant sequence
- Download sequence
- Repeat-mask the sequence
- 2 options (results will not vary):
 - GUI set-up
 - Analysis from .mupa file
- Pick a window-size and threshold

Example



Lecture Today

- James Noonan
- “Neanderthal Genomics and the Evolution of Modern Humans”
- 4 pm
- 119 Kerckhoff
- EC Problem Set to be posted online