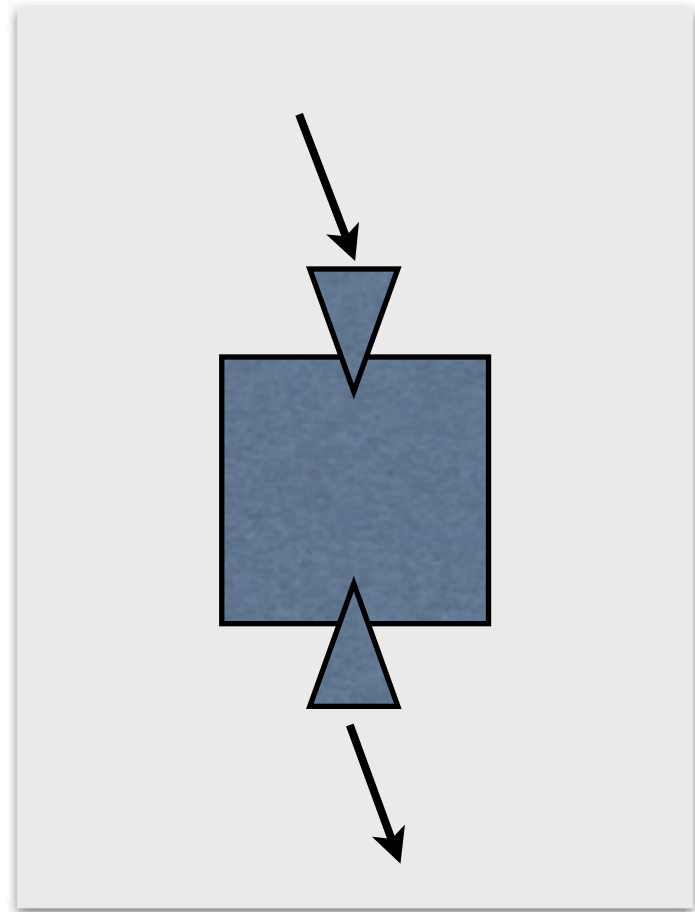


# BLAST Configuration

Changing the parameters of the algorithm

# Black Box

- Input goes in
- Something (unknown) happens
- Output comes out



# Understanding BLAST

- Better understanding means better usage
- Basic algorithm
  - Find a word, expand outwards
- Variations in the algorithm
  - Word size
  - Gapping
  - Scoring matrices

ATACACAG  
GACTTACTGACA

ATACACAG  
GACTTACTGACA

# Scoring by Points

- Gain points for correct nucleotide matches
- Lose points for
  - Gaps
  - Nucleotide mismatches
- Threshold defines cutoff

# Scoring Matrices

- PAM Matrices
  - Measures evolutionary distance
  - Based on global alignment of closely related sequences
- BLOSUM Matrices
  - Measures percent identity
  - Based on local alignments of distantly related sequences

# General Variables

- Expectation value (E-value)
  - Statistical significance of a find
  - The number of matches expected to be found by chance
- Word size
  - Size of the initial word that must be found
  - As word size grows, false positives (and sensitivity) decrease

# Search Type

- Protein sequences
  - Can be better at finding conserved function
  - Some matrices designed for finding protein structure similarity
- Nucleotide sequences
  - More specific; finds very closely related sequence
  - More readily available

# Pitfalls

- Low stringency / high sensitivity
  - Large E-value
  - Small word size
- Sequence similarity does not *necessarily* imply evolutionary closeness
- Homologous genes with divergent functionality can lead to confusing results
- BLAST does *not* really understand biology

# Resources

- <http://www.arabidopsis.org/Blast/BLASToptions.jsp>
- [http://bioinformatics.ubc.ca/resources/tutorials/commandblast\\_tutorial/](http://bioinformatics.ubc.ca/resources/tutorials/commandblast_tutorial/)

```
> blastall -p blastp -d ecoli -i takifugu_cftr.faa -o takifugu_cftr_blast_ecoli.out  
BLASTP 2.1.2 [Nov-13-2000]
```

Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997),

.....