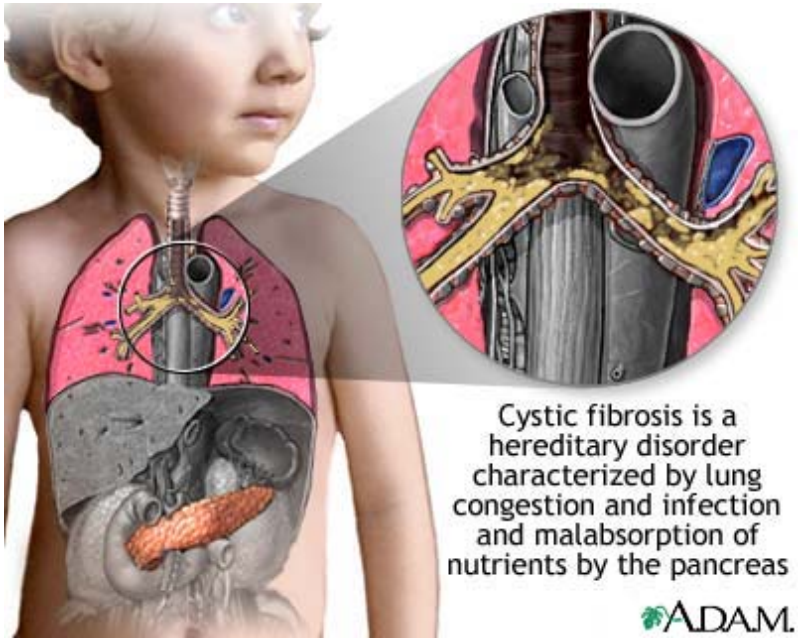
Basic Local Alignment Search Tool

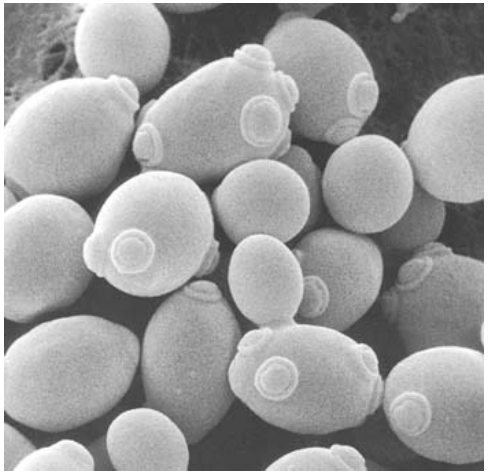
BLAST

Why Use BLAST?

Finding Model Organisms for Study of Disease

Can yeast be used as a model organism to study cystic fibrosis?





Model Organisms

- **Cystic fibrosis is a genetic disorder that affects humans**
  - If **yeast** contain a protein that is related (**homologous**) to the protein involved in cystic fibrosis
  - Then **yeast can be used as a model organism** to study this disease
    - **Study of the protein in yeast will tell us about the function of the protein in humans**

BLAST helps you to find homologous genes and proteins

Homologous Proteins (or genes)

- **Have a common ancestor (they're related)**
- **Have similar structures**
- **Have similar functions**

Criteria for considering two sequences to be homologous

- **Proteins are homologous if**
  - **Their amino acid sequences are at least 25% identical**

- **DNA sequences are homologous if**
  - **they are at least 70% identical**

  - Note that sequences must be over 100 a.a. (or bp) in length

Whenever possible, it is better to compare proteins than to compare genes

What does BLAST do?

BLAST compares sequences

- **BLAST** takes a **query** sequence
- **Compares** it with millions of sequences in the **Genbank databases**
  - By constructing local alignments
- **Lists** those that appear to be similar to the query sequence
  - The **"hit list"**
- **Tells you why** it thinks they are **homologs**
  - **BLAST makes suggestions**
  - **YOU make the conclusions**

How do I input a query into BLAST?

Choose which "flavor" of BLAST to use

- **BLAST comes in many "flavors"**
  - **Protein BLAST (BLASTp)**
    - **Compares a protein query with sequences in GenBank protein database**
  - **Nucleotide BLAST (BLASTn)**
    - **Compare nucleotide query with sequences in GenBank nucleotide database**

Enter your "query" sequence

- **A sequence can be input as a (an)**
  - **FASTA** format sequence
  - **Accession number**

  - Protein blast can only accept amino acid sequences

Choose search set

- Choose which database to search
  - **Default** is non-redundant protein sequences (nr)
    - Searches all databases that contain protein sequences

Choose organism

- **Default is all organisms represented in databases**
- **Use this to limit your search to one organism (eg. Yeast)**

BLAST off!!

- Click on the BLAST button at the bottom of the page!

How do I interpret the results of a BLAST search?

BLAST creates local alignments

- **What is a local alignment?**
  - **BLAST looks for similarities between regions of two sequences**

```
Global  FGFTALILLAVKV
        F--TAL-LLA--V

Local   FGFTALILL-AVKAV
        --FTAL-LLAAV---
```

The BLAST output then describes how these aligned regions are similar

- How long are the aligned segments?
- Did BLAST have to introduce gaps in order to align the segments?
- How similar are the aligned segments?

The BLAST Output

The Graphic Display

1. How good is the match?

- **Red = excellent!**
- Pink = pretty good
- **Green = OK, but look at other factors**
- **Blue = bad**
- **Black = really bad!**

2. How long are the matched segments?

Longer = better

The hit list

- **BLAST lists the best matches (hits)**
  - **For each hit, BLAST provides:**
    - **Accession number – links to Genbank flatfile**
    - **Description**
    - **"G" = genome link**
    - **E-value**
      - **An indicator of how good a match to the query sequence**
    - **Score**
      - **Link to an alignment**

What is an E-value?

- **E-value**
  - **The chance that the match could be random**
  - **The lower the E-value, the more significant the match**
    - **E = 10^{-4}** is considered the **cutoff point**
    - **E = 0** means that the two sequences are statistically **identical**

Most people use the E- value as their first indication of similarity!

The Alignment

- **Look for**:
  - **Long regions of alignment**
  - With few gaps
  - % identity should be >25% for proteins
    - (>70% for DNA)

BLAST makes suggestions, You draw the conclusions!

- Look at E-value
- Look at graphic display
- If necessary, look at alignment

- Make your best guess!