

# Use of DNA in Bioinformatics

Let's say some one gave you a DNA sequence like this:

```
TTAACACATGCAAGTCGAACGGAAAGGCCCTTCGGGGTGCTCGAGTGGCGAACGGGTGAGTAACACGTGG
GTGATCTGCCCTGCACTTTGGGATAAGCCTGGGAAACTGGGTCTAATACCGAATAGGACTCCGGCCTTCA
TGGGTGGGGTGGAAAGCTTTTGCGGTGTGGGATGGGCCCCGGCCTATCAGCTTGTTGGTGGGGTAATG
GCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGTGACCGGCCACACTGGGACTGAGATACGGCCC
AGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCGACGCCGCGT
GAGGGATGACGGCCTTCGGGTGTAAACCTCTTTCAGCACAGACGAAGCGCAAGTGACGGTATGTGCAGA
AGAAGGACCGGCCAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGTCCGAGCGTTGTCCGGAATTACT
GGGCGTAAAGAGCTCGTAGGTGGTTTGTGCGGTTGTTTCGTGAAAACCTCACAGCTTAACTGTGGGCGTGCG
GGCGATACGGGCAGACTGGAGTACTGCAGGGGAGACTGGAATTCCTGGTGTAGCGGTGGAATGCGCAGAT
ATCAGGAGGAACACCGGTGGCGAAGGCGGGTCTCTGGGCAGTAACTGACGCTGAGGAGCGAAAGCGTGGG
GAGCGAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGGTGGGTACTAGGTGTGGGTTTCCTTCC
TTGGGATCCGTGCCGTAGCTAACGCATTAAGTACCCCCGCTGGGGAGTACGGCCGCAAGGCTAAAACCTCA
AAGGAATTGACGGGGGCCCGCACAAAGCGGCGGAGCATGTGGATTAATTCGATGCAACGCGAAGAACCTTA
CCTGGGTTTGACATGCACAGGACGCTGGTAGAGATATCAGTTCCTTGTGGCCTGTGTGCAGGTGGTGCA
TGGCTGTGCTCAGCTCGTGTGCTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTCTATGT
TGCCAGCGGGTTATGCCGGGGACTCGTAGGAGACTGCCGGGGTCAACTCGGAGGAAGGTGGGGATGACGT
CAAGTCATCATGCCCTTATGTCCAGGGCTTCACACATGCTACAATGGCCGGTACAAAGGGCTGCGATGC
CGTGAGGTGGAGCGAATCCTTGTAAAGCCGGTCTCAGTTCGGATCGGGGTCTGCAACTCGACCCCGTGAA
GTCGGAGTCGCTAGTAATCGCAGATCAGCAACGCTGCGGTGAATACGTTCCCAGGCTTGTACACACCCG
CCGTCACGTCATGAAAGTCGGTAAACCCGAAGCCGGTGGCCTAACCCCTTGTGGGAGGGAGCCGTGCAA
GGTGGGATCGGCGATTGGGACGAAGTCGAACAA
```

What is the gene (function)?

Which organism?

Sense or anti-sense?

Contamination? Is it new to science?

# BLAST in Biology

- **Tuberous sclerosis complex (TSC)** 1 in 6000 newborns. Named 1880 for the firm, potato-like nodules that form in the cerebral cortex; also sprout in the kidneys, lungs, heart, eyes, and skin.

- **BLAST found *TSC1* and *TSC2* homologs in *Drosophila*:**  
Fundamental to cell division control.

- **Mutant flies lacking tuberin** possess enlarged cells that contain 10X normal amount of DNA. How?

Tuberous sclerosis



Adenoma sebaceum

Shagreen patch



Ungual fibromas

# A Nobel Prize for BLAST?

2009 OCTOBER 2

by abhishektiwari

Although the [Nobel prizes](#) will be announced in coming weeks there is an increasing interest about possible winners. The winners of the prize for Physiology will be announced on Monday 5th. Like every year Thomson Reuters has predicted a list of likely winners in medicine, chemistry, physics, and economics [for year 2009](#). According to Thomson Reuters predictions scientific works leading to discovery of cell aging, solar cells and functional magnetic resonance imaging or fMRI, are all top runners for the Nobel prizes for this year. As Thomson Reuters website suggests about methodology used for prediction

*Since 1989, the Scientific business of Thomson Reuters has correctly predicted at least one Nobel Laureate each year with the exception of the years 1993 and 1996. David Pendlebury's predictions, combined with those of Dr Eugene Garfield – Founder and Chairman Emeritus of the Institute for Scientific Information (ISI), now part of Thomson Reuters – have shown that citations in the journal literature reveal influential researchers who have contributed significantly to scientific advances, such as those recognized by the Nobel Prize committees*

If higher citations counts and number of high impact papers are effective predictor of Nobel Prize winner then what are the odds that the creators of Basic Local Alignment Search Tool, or BLAST will win a Nobel Prize in Physiology. I know this is an extreme scenario but if you look on few facts then may be you will be equally surprise that what is coming in their way to win very first Nobel Prize for the Bioinformatics. For example the original paper about BLAST published by Altschul, et al. is cited more than 26,587 times. Similarly paper describing the Gapped BLAST and PSI-BLAST is cited more than 22,827 times. Apart from that early papers by David J. Lipman and W. R. Pearson describing rapid similarity searches of nucleic acid and protein data banks are highly cited too. Following is a list of 5 highly influential papers (in top 0.1% of their field) by [Stephen Altschul](#), [David J. Lipman](#) and W. R. Pearson



## **BASIC LOCAL ALIGNMENT SEARCH TOOL**

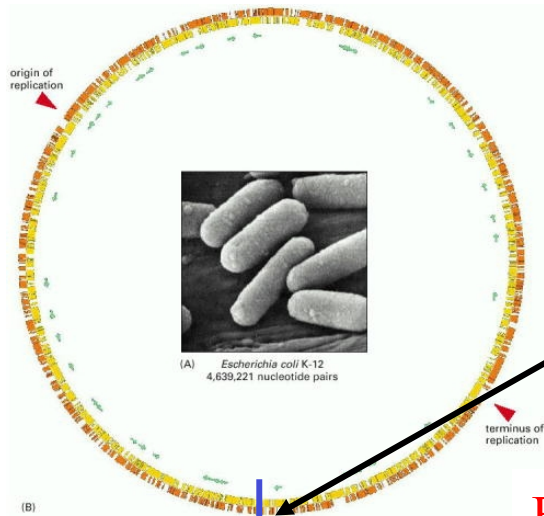
**Author(s)** ALTSCHUL, SF; GISH,  
W; MILLER, W; MYERS, EW;  
LIPMAN, DJ

**Source** JOURNAL OF  
MOLECULAR BIOLOGY 215 (3):  
403-410 1990

# MOST CITED PAPER 1990s

# BLAST

- Easy argument: Majority of modern gene function discovery comes via BLAST
- Everything we know about newly sequenced genomes
- Genes involved in disease, new bacteria/viruses, saved millions of hours of lab work (billions of dollars)
- Developed by government scientists! NCBI, OMG!
- Sequence analysis: Ebola, SARS, Flu, HIV



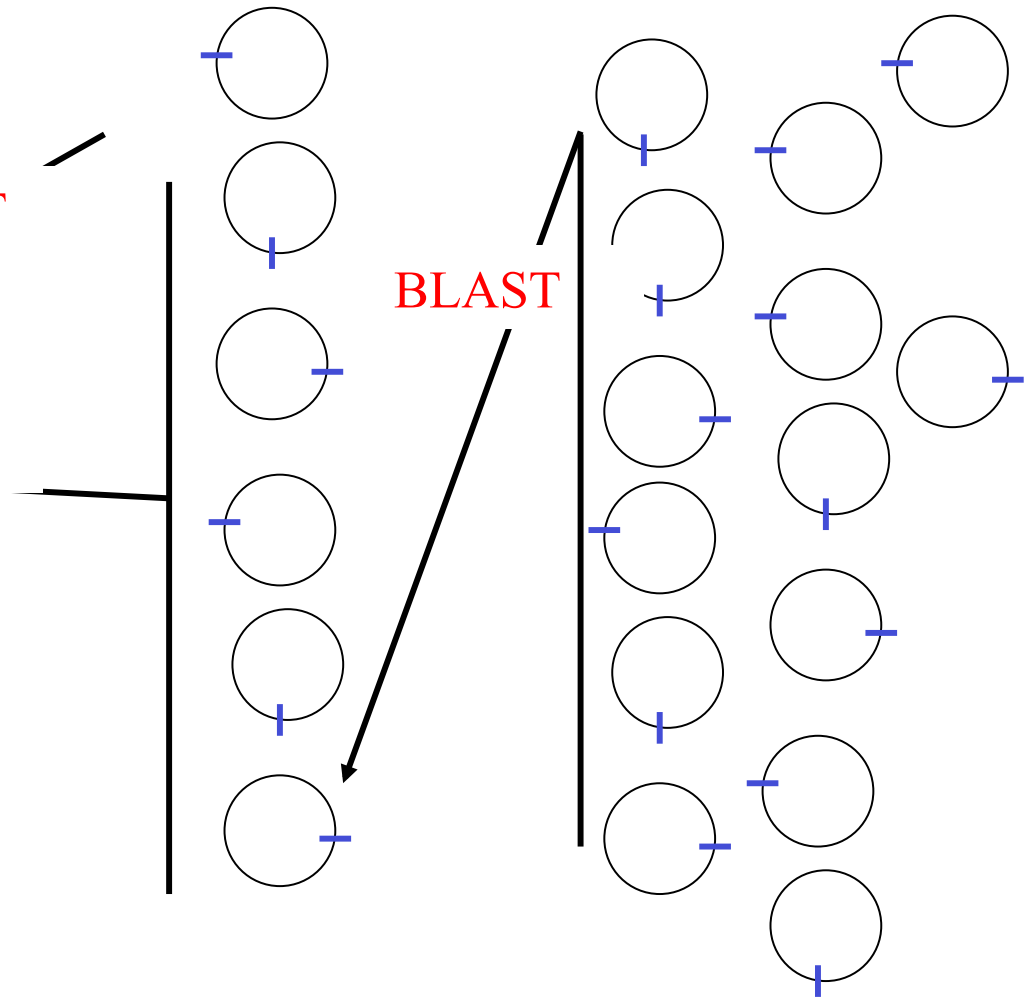
**BLAST**

**BLAST**

2 months  
Knockout Gene X

10 months  
Biochemically Characterize

1 Year &  
\$30,000 Later  
Basic Function  
(e.g., Porin)



**BLAST**

**SAVINGS = 21 Years  
and \$600,000!**