

Apr 6th, 12:00 PM

Bioinformatics: Alive and Kicking

David L. Osterbur
Harvard Medical School

Follow this and additional works at: http://escholarship.umassmed.edu/escience_symposium



Part of the [Bioinformatics Commons](#), and the [Library and Information Science Commons](#)

Osterbur, David L., "Bioinformatics: Alive and Kicking" (2009). *University of Massachusetts and New England Area Librarian e-Science Symposium*. 5.

http://escholarship.umassmed.edu/escience_symposium/2009/program/5

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in University of Massachusetts and New England Area Librarian e-Science Symposium by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.

Bioinformatics: alive and kicking

David L. Osterbur

Too Central

- “Bioinformatics has become too central to biology to be left to specialist bioinformaticians. Biologists are all bioinformaticians now.”

Taken from Stein, L.D. (2008). Bioinformatics: alive and kicking. *Genome Biol* 9, 114.

Google Generation

BLAST *Basic Local Alignment Search Tool*

[Home](#) [Recent Results](#) [Saved Strategies](#) [Help](#)

▶ [NCBI/BLAST/blastp suite: BLASTP programs search protein databases using a protein query.](#) [more...](#)

Enter Query Sequence

Enter accession number, gi, or FASTA sequence [?](#) [Clear](#) **Query subrange** [?](#)

From

To

Or, upload file [Browse...](#) [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

[I'm Feeling Lucky](#) Search database **nr** using **Blastp (protein-protein BLAST)**

Show results in a new window

▶ [Algorithm parameters](#)

History



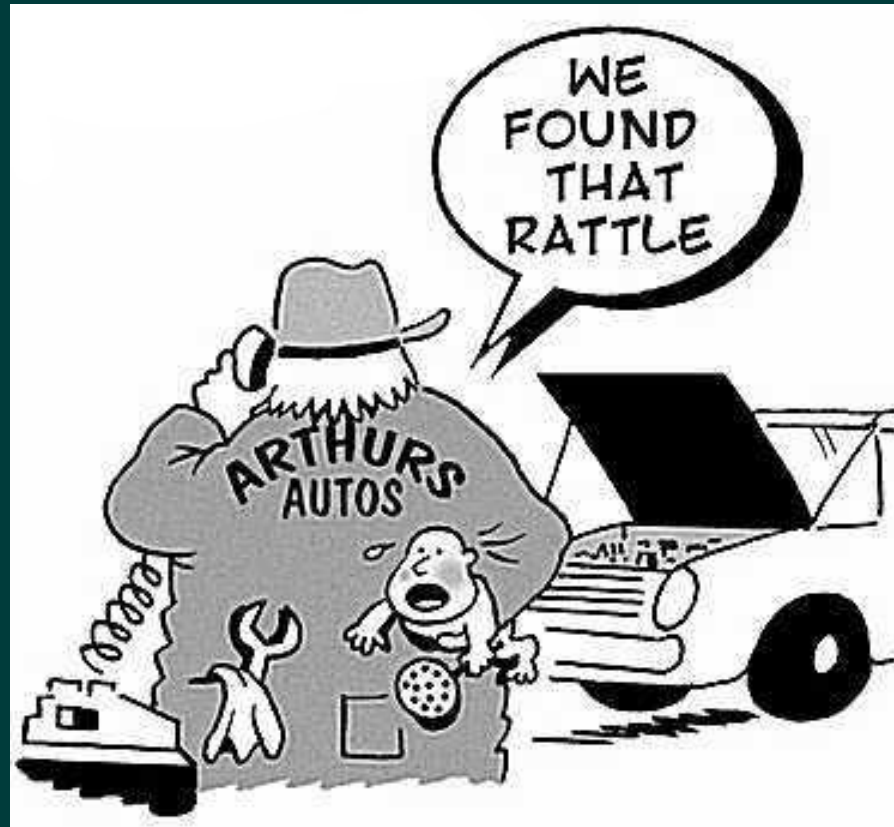
August 21,
2011

Harvard Medical School

Auto Designers



Auto Mechanics



Driver's Ed



August 21,
2011

Harvard Medical School

Keeping up to speed...



Not just how to drive but choosing the right tools.

...to reach your goal on time.



The Library's Role

Why Libraries?

- We are a service organization.
- We are already good at organizing, distributing and teaching access to many different types of information.
- We are a shared organization, not “owned” by any one department or unit.
- “Librarians like to search...
...everyone else likes to find”

Why Your Library?

(15) Translational Science

- 15-LM-101*** **Presenting genome information in electronic health records.** Develop approaches for presenting relevant genomic information in an understandable way, in the context of a patient's electronic health record. As genomic data becomes available for more individuals, these data must be integrated into electronic health records in ways that: help clinicians and patients to understand the significance of the data; provide an avenue for alerting clinicians and patients when new knowledge from GWAS, etc. rises to the level of potential clinical impact; and enable linking to effective decision support. Contact: Dr. Jane Ye, 301-594-4882, yej@mail.nih.gov.
- 15-LM-102** **Computational hypothesis generation for biology and medicine.** Employing two or more sources, use advanced computational approaches to generate a new and meaningful hypothesis in biomedical science, capable of being tested by bench or clinical research. One source must be full-text published biomedical literature; the other source should be either (1) a database storing primary data from basic biomedical research or (2) data drawn from the electronic health records used for routine clinical care or from the data accumulated for a clinical research project. The user interface of an integrated hypothesis generation system should support easy use by the intended users (i.e., by biomedical researchers or clinicians). Mining techniques should involve minimal human intervention. Contact: Dr. Valerie Florance, 301-594-4882, florancev@mail.nih.gov.

Google Generation

BLAST *Basic Local Alignment Search Tool*

Home Recent Results Saved Strategies Help

▶ [NCBI/BLAST/blastp suite: BLASTP programs search protein databases using a protein query.](#) [more...](#)

Enter Query Sequence

Enter accession number, gi, or FASTA sequence [?](#) [Clear](#) Query subrange [?](#)

From

To

Or, upload file [Browse...](#) [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

[I'm Feeling Lucky](#) Search database **nr** using **Blastp (protein-protein BLAST)**

Show results in a new window

▶ [Algorithm parameters](#)

BLAST Results

```
> ref|NP\_002070.1 UG aspartate aminotransferase 1 [Homo sapiens]
  sp|P17174.3|AATC\_HUMAN G RecName: Full=Aspartate aminotransferase, cytoplasmic; AltName:
  Full=Transaminase A; AltName: Full=Glutamate oxaloacetate
  transaminase 1
  gb|AAA35563.1 G aspartate aminotransferase
  ▶ 7 more sequence titles
  Length=413

  GENE ID: 2805 GOT1 | glutamic-oxaloacetic transaminase 1, soluble (aspartate
  aminotransferase 1) [Homo sapiens] (Over 10 PubMed links)

  Score = 860 bits (2223), Expect = 0.0, Method: Compositional matrix adjust.
  Identities = 413/413 (100%), Positives = 413/413 (100%), Gaps = 0/413 (0%)

  Query 1  MAPPSVFAEVPQAQPVLVFKLTADFREDPDPKVNLVGVGAYRTDDCHPWVLPVVKKVEQK 60
           MAPPSVFAEVPQAQPVLVFKLTADFREDPDPKVNLVGVGAYRTDDCHPWVLPVVKKVEQK
  Sbjct 1  MAPPSVFAEVPQAQPVLVFKLTADFREDPDPKVNLVGVGAYRTDDCHPWVLPVVKKVEQK 60

  Query 61  IANDNSLNHEYLPILGLAEFRSCASRLALGDDSPALKEKRVGGVQSLGGTGALRIGADFL 120
           IANDNSLNHEYLPILGLAEFRSCASRLALGDDSPALKEKRVGGVQSLGGTGALRIGADFL
  Sbjct 61  IANDNSLNHEYLPILGLAEFRSCASRLALGDDSPALKEKRVGGVQSLGGTGALRIGADFL 120

  Query 121 ARWYNGTNNKNTPVYVSSPTWENHNAVFSAGFKDIRSYRYWDAEKRGDLQGFNLNDLEN 180
           ARWYNGTNNKNTPVYVSSPTWENHNAVFSAGFKDIRSYRYWDAEKRGDLQGFNLNDLEN
  Sbjct 121 ARWYNGTNNKNTPVYVSSPTWENHNAVFSAGFKDIRSYRYWDAEKRGDLQGFNLNDLEN 180

  Query 181 APEFSIVVLHACAHNPTGIDPTPEQWKQIASVMKHRFLFPFFDSAYQGFASGNLERDAWA 240
           APEFSIVVLHACAHNPTGIDPTPEQWKQIASVMKHRFLFPFFDSAYQGFASGNLERDAWA
  Sbjct 181 APEFSIVVLHACAHNPTGIDPTPEQWKQIASVMKHRFLFPFFDSAYQGFASGNLERDAWA 240

  Query 241 IRYFVSEGFEFFCAQSFSKNFGLYNERVGNLTVVGKEPESILQVLSQMEKIVRITWSNPP 300
           IRYFVSEGFEFFCAQSFSKNFGLYNERVGNLTVVGKEPESILQVLSQMEKIVRITWSNPP
  Sbjct 241 IRYFVSEGFEFFCAQSFSKNFGLYNERVGNLTVVGKEPESILQVLSQMEKIVRITWSNPP 300

  Query 301 AQQGARIVASTLSNPELFEEWIGNVKTMADRILTMRSELRARLEALKTPGTWNHITDQIGM 360
           AQQGARIVASTLSNPELFEEWIGNVKTMADRILTMRSELRARLEALKTPGTWNHITDQIGM
  Sbjct 301 AQQGARIVASTLSNPELFEEWIGNVKTMADRILTMRSELRARLEALKTPGTWNHITDQIGM 360

  Query 361 FSFTGLNPKQVEYLVNEKHIYLLPSGRINVSGLTTKNLDYVATSIHEAVTKIQ 413
           FSFTGLNPKQVEYLVNEKHIYLLPSGRINVSGLTTKNLDYVATSIHEAVTKIQ
  Sbjct 361 FSFTGLNPKQVEYLVNEKHIYLLPSGRINVSGLTTKNLDYVATSIHEAVTKIQ 413
```

NCBI has already done it...

Pre-computed BLAST results for: [gij4504067|ref|NP_002070.1](#) aspartate an

Matching gis: [12653453;3360464;3452380;5902703;55663625;119570](#):

Total (score > 100) : 3731 hits in 3727 proteins in 872 species

Selected: 3731 hits in 3727 proteins in 872 species Filter: Min Sco

Other views (Reports): Reset all filters

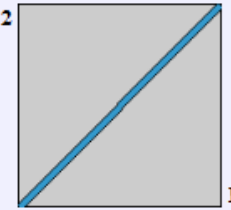
Archaea Bacteria Metazoa Fungi Pl

Score = 844 bits (2181), Expect = 0.0
Identities = 413/413 (100%), Positives = 413/413 (100%), Gaps = 0/413 (0%)

Query 1 MAPPSVFAEVPQAQPVLVFKLTADFREDPPDRKVNLVGAYRTDDCHPWVLPVVKKVEQK 60
Sbjct 1 MAPPSVFAEVPQAQPVLVFKLTADFREDPPDRKVNLVGAYRTDDCHPWVLPVVKKVEQK 60

Query 61 IANDNSLNHEYLPILGLAEFRSCASRLALGDDSPALKEKRVGGVQSLGGTGALRIGADFL 120
Sbjct 61 IANDNSLNHEYLPILGLAEFRSCASRLALGDDSPALKEKRVGGVQSLGGTGALRIGADFL 120

kecname: full=Aspartate aminotransferase, cytoplasmic; AltName: full=Transaminase
glutamic-oxaloacetic transaminase 1, soluble (aspartate aminotransferase 1) [Homo
glutamic-oxaloacetic transaminase 1, soluble (aspartate aminotransferase 1), isofo
aspartate aminotransferase



NOTE: Bitscore and expect value are calculated based on the size of the nr database.

SCORE	ACCESSION	Length
2181	AAH00498	413
2181	AAC28622	413
2181	AAC32951	413
2181	P17074	413
2181	CH73859	413
2181	EAW49869	413
2181	AAA35563	413

Conserved Domain Database h

413 aa

reset selection

blink

8

...and more.

Pairwise Alignment Scores							
Gene		Identity (%)		Substitution Rates ¹			
Species	Symbol	Protein	DNA	d	d _N /d _S	d _{NR} /d _{NC}	
Homo sapiens		GOT1					
vs. Pan troglodytes	GOT1	100.0	99.8	0.002	0.000	undef	Blast
vs. Canis lupus familiaris	GOT1	92.5	89.4	0.114	0.087	0.464	Blast
vs. Bos taurus	GOT1	91.5	89.5	0.113	0.100	0.490	Blast
vs. Mus musculus	Got1	91.0	86.9	0.144	0.069	0.541	Blast
vs. Rattus norvegicus	Got1	89.8	87.2	0.140	0.090	0.731	Blast
vs. Gallus gallus	GOT1	80.4	76.1	0.287	0.076	0.538	Blast
vs. Danio rerio	got1	77.0	72.5	0.343	0.052	0.685	Blast
vs. Drosophila melanogaster	Got1	56.9	59.1	0.591	0.182	0.912	Blast
vs. Anopheles gambiae	AgaP_AGAP004142	61.1	58.6	0.603	undef	0.867	Blast
vs. Caenorhabditis elegans	aminotransferase	54.5	58.5	0.604	0.110	0.891	Blast
vs. Schizosaccharomyces pombe	SPAC10F6.13c	47.4	51.0	0.794	undef	0.828	Blast
vs. Saccharomyces cerevisiae	AAT2	48.1	51.6	0.777	undef	0.872	Blast
vs. Kluyveromyces lactis	KLLA0F17754g	50.0	50.8	0.801	undef	0.792	Blast
vs. Eremothecium gossypii	AGOS_AFR211C	48.1	51.9	0.770	0.157	0.818	Blast
vs. Magnaporthe grisea	MGG_04156	53.4	56.1	0.660	undef	0.731	Blast
vs. Neurospora crassa	NCU07941.1	53.1	54.6	0.698	undef	0.795	Blast
vs. Arabidopsis thaliana	ASP3	50.6	55.6	0.672	undef	0.764	Blast
vs. Oryza sativa	Os01g0760600	51.4	55.9	0.664	undef	0.814	Blast
Pan troglodytes		GOT1					
vs. Homo sapiens	GOT1	100.0	99.8	0.002	0.000	undef	Blast
vs. Canis lupus familiaris	GOT1	92.5	89.6	0.112	0.089	0.464	Blast
vs. Bos taurus	GOT1	91.5	89.7	0.111	0.103	0.490	Blast

Why not Libraries?

LIBRARIAN

Librarian Action Figure with Movable Arms



Criteria for Success

- Bioinformaticist interested in service
- Money to support software licensing
- No micromanaging

Bioinformatics Support at Countway



Countway Bioinformatics Education Program

- **R/Bioconductor**
- **GeneGO - Metacore**
- **Biobase – ExPlain, TransFac and HGMD**
- **Ingenuity Pathway Analysis**
- **SNP Data**
- **Sequence Alignment – BLAST & Clustal**
- **Genome Browsing**
- **Beginning Unix**
- **ENSEMBL**
- **Matlab**
- **Rosetta Resolver**

Community

- BITS – Bioinformatics Tutorial Series – In collaboration with Courtney Crummett at MIT.
 - <https://www.countway.harvard.edu/lenya/countway/live/menuNavigation/libraryServices/classes/videoTutorials.html>
 - <http://libguides.mit.edu/content.php?pid=14149&sid=145112>

Harvard's Favorites

- Survey to find out applications that are used by various labs around campus
 - This will inform us of what software we need to support if we are not already
 - It will help researchers to see what their colleagues are using
 - A way for others to see what Harvard is doing

Impact For the Library



August 21,
2011

Harvard Medical School

Impact

For the Library

- First time ever library invited to participate in both the graduate and medical curriculums.
- Collaborative opportunities
 - MIT
 - MLA – other libraries
 - Across Harvard

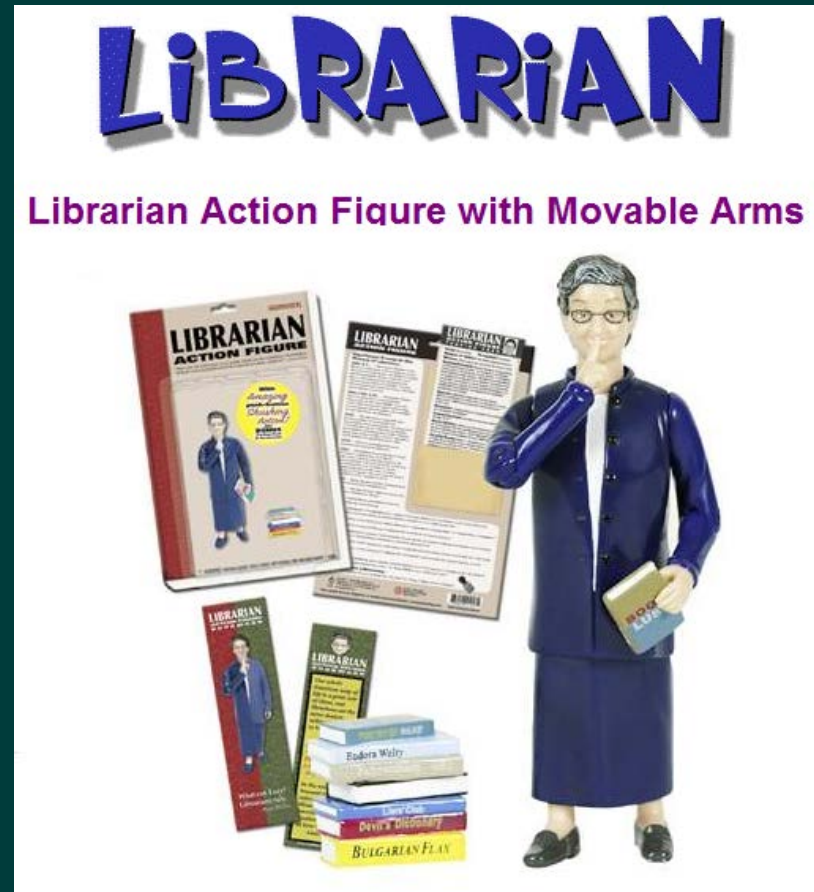
Impact

For HMS and the CTSA

- More productive researchers
- Better educated students and postdocs
- ... (and faculty)

“The greatest obstacle to discovery is not ignorance - it is the illusion of knowledge.” Daniel J. Boorstin (1914–2004) Historian and Librarian of Congress

When you think of libraries



Don't think of Marian the Librarian

Think of Conan the Librarian



August 21,
2011

Harvard Medical School

Bioinformatics: alive and kicking... in the library.