Boris Steipe
Department of Biochemistry
Department of Molecular Genetics
University of Toronto

BCH441 BIOINFORMATICS

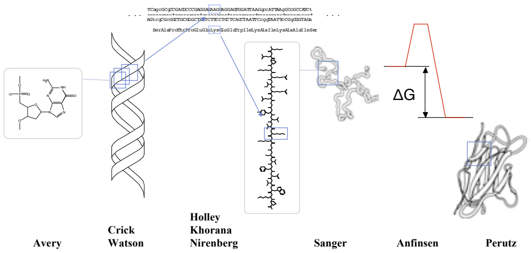
INTRODUCTION - 2012

BCH441

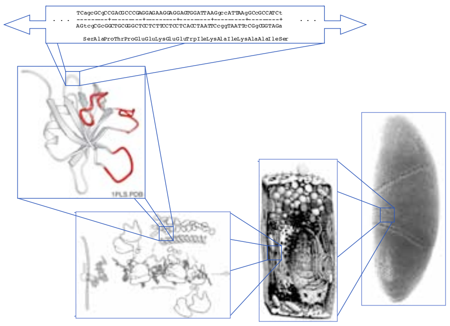
Why is Bioinformatics relevant?

molecular biology

Inheritable information is a substance.


TCagcGCgCCGACGCCCGAGGAGAAGGAGGAGTGGATTAAGgccATTAAgGCcGCCATCt
. . . ---------+---------+---------+---------+---------+---------+ . . .
AGtcgCGcGGCTGCGGGCTCCTCTTCCTCCTCACCTAATTCcggTAATTcCGgCGGTAGa
SerAlaProThrProGluGluLysGluGluTrpIleLysAlaIleLysAlaAlaIleSer
ΔG
Avery
Crick
Watson
Holley
Khorana
Nirenberg
Sanger
Anfinsen
Perutz

genomic biology


TCagcGCgCCGACGCCCGAGGAGAAGGAGGAGTGGATTAAGgccATTAAgGCcGCCATCt
. . . ---------+---------+---------+---------+---------+---------+ . . .
AGtcgCGcGGCTGCGGGCTCCTCTTCCTCCTCACCTAATTCcggTAATTcCGgCGGTAGa
SerAlaProThrProGluGluLysGluGluTrpIleLysAlaIleLysAlaAlaIleSer
1PLS.PDB

(post)genomic biology current practice

1. **Industrial scale** (Data intensive)
2. **Multiple genes** (Cross-sectional)
3. **Model Organisms** (Inference by analogy)
4. **Complete, exhaustive description** (Missing entities are important)
5. **Discovery Science** (Association, not Hypothesis)

Too much data to process by hand,
too many entities to keep in your mind.

(post)genomic biology current practice

In any modern life-science paper you will likely find that conclusions derived from computational inference exceed conclusions derived from direct observation.

That is to be expected, given the importance of "context".

BCH441

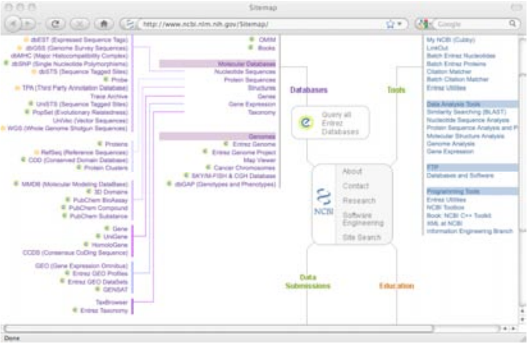
What is Bioinformatics ?

bioinformatics

Data management is the fundamental task of bioinformatics.

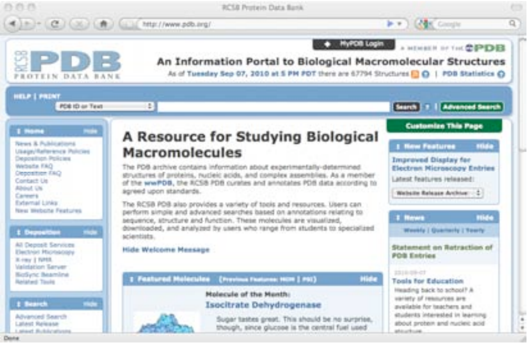
NCBI

http://www.ncbi.nlm.nih.gov



PDB

http://www.pdb.org/



RCSB PDB PROTEIN DATA BANK

An Information Portal to Biological Macromolecular Structures

As of Tuesday Sep 07, 2010 at 5 PM PDT there are 67794 Structures | PDB Statistics

A Resource for Studying Biological Macromolecules

The PDB archive contains information about experimentally-determined structures of proteins, nucleic acids, and complex assemblies. As a member of the wwPDB, the RCSB PDB curates and annotates PDB data according to agreed upon standards.

The RCSB PDB also provides a variety of tools and resources. Users can perform simple and advanced searches based on annotations relating to sequence, structure and function. These molecules are visualized, downloaded, and analyzed by users who range from students to specialized scientists.

Hide Welcome Message

Featured Molecules (Previous Features: MOM | PSI)

Molecule of the Month: Isocitrate Dehydrogenase

Sugar tastes great. This should be no surprise, though, since glucose is the central fuel used

Improved Display for Electron Microscopy Entries

Latest features released:

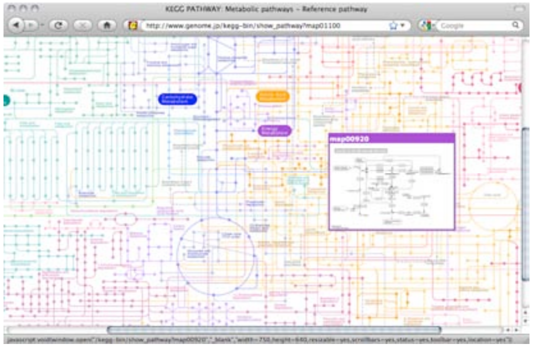
Statement on Retraction of PDB Entries

Tools for Education

Heading back to school? A variety of resources are available for teachers and students interested in learning about protein and nucleic acid structure.

KEGG

http://www.genome.jp/kegg/


KEGG PATHWAY: Metabolic pathways - Reference pathway
http://www.genome.jp/kegg-bin/show_pathway?map01100
map00920

challenges

1. Data overload (2009 NAR: 179 databases, 95 new - 1170 in the Molecular Biology Database Collection)
2. Service overload (2009 NAR: 112 Web services)
3. Poor integration
4. Peer review and expert opinions lacking
5. Cultural gap between life- and computer sciences

How will bioinformatics contribute to our understanding of biology?

The question becomes less: "What can you do?" but: "What should you do?" !

bioinformatics

Modeling is the fundamental task of bioinformatics.

bioinformatics

Problems of modeling:

Models can be right or wrong ...

bioinformatics

Problems of modeling:

Models can be right or wrong ...
... but worse, they can also be irrelevant.

cargo cult science

cargo cult science

[...] In the South Seas there is a cargo cult of people. During the war they saw airplanes land with lots of good materials, and they want the same thing to happen now. So they've arranged to imitate things like runways, to put fires along the sides of the runways, to make a wooden hut for a man to sit in, with two wooden pieces on his head like headphones and bars of bamboo sticking out like antennas--he's the controller--and they wait for the airplanes to land. They're doing everything right. The form is perfect. It looks exactly the way it looked before. But it doesn't work. No airplanes land. So I call [some examples of pseudoscience] cargo cult science, because they follow all the apparent precepts and forms of scientific investigation, but they're missing something essential, because the planes don't land.

Now it behooves me, of course, to tell you what they're missing. But it would be just about as difficult to explain to the South Sea Islanders how they have to arrange things so that they get some wealth in their system. It is not something simple like telling them how to improve the shapes of the earphones. [...].

Richard Feynman

cargo cult science

Avro Lancaster, WikiMedia Commons

BCH441

What we will cover in this course:

learn for change

- **We'll discuss principles and examples of how the facts of biology can be expressed as computable abstractions.** As our knowledge of the facts changes, you should be able to think of novel models.
- **We'll use key databases that store publicly available molecular data.** As the databases grow and change, you should be able to work with new types of data, because you are familiar with the principles.
- **We'll use key procedures that analyse sequence, structure, function and phylogeny.** As new tools become available, you should be able to identify those that are useful to support **your own, changing objectives**.

Bioinformatics

Sources of information

Problem: outdated information has much inertia

web

bioinformatics or "computational biology" - Google Search
http://www.google.com/search?source=ig&hl=en&rlz=&=&q=bioinformat
Google
Web Images Videos Maps News Shopping Gmail more ▾
Web History | Search settings | Sign in
Google
Search
Instant is on ▾
SafeSearch off ▾
Advanced search
Everything
Images
Videos
Books
More
All results
Related searches
More search tools
Done

bioinformatics or "computational biology"

About 852,000 results (0.28 seconds)

Bioinformatics and Computational Biology - Overview
The Bioinformatics and Computational Biology initiative has taken shape in the National Centers for Biomedical Computing. Through this program, the NIH ...
nihroadmap.nih.gov › Initiatives - Cached - Similar

Bioinformatics - Wikipedia, the free encyclopedia
The actual process of analyzing and interpreting data is referred to as computational biology. Important sub-disciplines within bioinformatics and ...
Introduction - Major research areas - Software and tools - See also
en.wikipedia.org/wiki/Bioinformatics - Cached - Similar

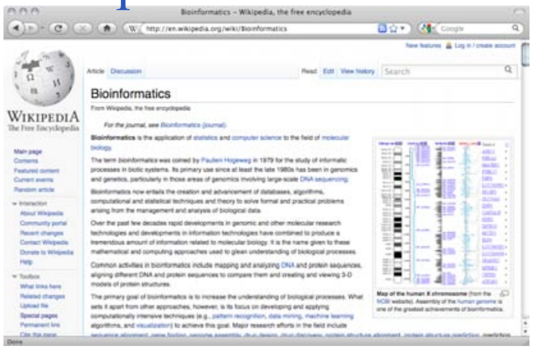
Journal of Bioinformatics and Computational Biology (JBCB)
The Journal of Bioinformatics and Computational Biology aims to publish high quality, original research articles, expository tutorial papers and review ...
www.worldscinet.com › ... › COMPUTER SCIENCE - Similar

Bioinformatics and Computational Biology at ISU
An interdisciplinary Graduate Program offering Ph.D. and MS degrees in Bioinformatics and Computational Biology (BCB) at Iowa State University.
www.bcb.iastate.edu/ - Cached - Similar

UMD Center for Bioinformatics and Computational Biology
The University of Maryland Center for Bioinformatics and Computational Biology is a multidisciplinary center dedicated to research on questions arising from ...
www.cbcb.umd.edu/ - Cached - Similar

George Mason University : Bioinformatics and Computational Biology ...
26 Jul 2010 ... The Department of Bioinformatics and Computational Biology is dedicated to advancing the application of computational methods to ...

Wikipedia http://en.wikipedia.org/wiki/Bioinformatics



Bioinformatics

From Wikipedia, the free encyclopedia

For the journal, see Bioinformatics (journal).

Bioinformatics is the application of statistics and computer science to the field of molecular biology.

The term bioinformatics was coined by Paulien Hogeweg in 1979 for the study of informatic processes in biotic systems. Its primary use since at least the late 1980s has been in genomics and genetics, particularly in those areas of genomics involving large-scale DNA sequencing.

Bioinformatics now entails the creation and advancement of databases, algorithms, computational and statistical techniques and theory to solve formal and practical problems arising from the management and analysis of biological data.

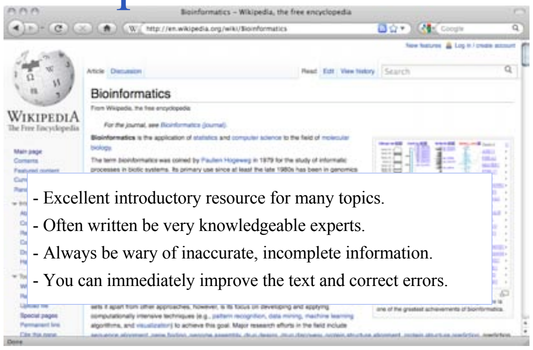
Over the past few decades rapid developments in genomic and other molecular research technologies and developments in information technologies have combined to produce a tremendous amount of information related to molecular biology. It is the name given to these mathematical and computing approaches used to glean understanding of biological processes.

Common activities in bioinformatics include mapping and analyzing DNA and protein sequences, aligning different DNA and protein sequences to compare them and creating and viewing 3-D models of protein structures.

The primary goal of bioinformatics is to increase the understanding of biological processes. What sets it apart from other approaches, however, is its focus on developing and applying computationally intensive techniques (e.g., pattern recognition, data mining, machine learning algorithms, and visualization) to achieve this goal. Major research efforts in the field include sequence alignment, gene finding, genome assembly, drug design, drug discovery, protein structure alignment, protein structure prediction, prediction

Map of the human X chromosome (from the NCBI website). Assembly of the human genome is one of the greatest achievements of bioinformatics.

Wikipedia http://en.wikipedia.org/wiki/Bioinformatics


Bioinformatics

From Wikipedia, the free encyclopedia

For the journal, see Bioinformatics (journal).

Bioinformatics is the application of statistics and computer science to the field of molecular biology.

The term bioinformatics was coined by Paulien Hogeweg in 1979 for the study of informatic processes in biotic systems. Its primary use since at least the late 1980s has been in genomics

sets it apart from other approaches, however, is its focus on developing and applying computationally intensive techniques (e.g., pattern recognition, data mining, machine learning algorithms, and visualization) to achieve this goal. Major research efforts in the field include

one of the greatest achievements of bioinformatics.

- Excellent introductory resource for many topics.
- Often written be very knowledgeable experts.
- Always be wary of inaccurate, incomplete information.
- You can immediately improve the text and correct errors.

portals

http://bioinformatics.ca

Canadian Bioinformatics Workshops | Bioinformatics.ca

http://bioinformatics.ca/

bioinformatics.ca

PRESENTS THE Canadian Bioinformatics Workshops

About · Workshops · Faculty · Testimonials · Resources · Sponsors · Contact

Search

Apply To Workshops
Apply To Hosted Workshops
Jobs Board
Workshop Alumni
Bioinformatics Links Directory
People

Next Workshop

Interpreting Gene Lists from -omics Studies
Location: Toronto, ON

Canadian Bioinformatics Workshops

Welcome to Bioinformatics.ca, the home of the Canadian Bioinformatics Workshops (CBW) series. This series has been operating since 1999 under the direction of Francis Ouellette with on-going support from CIHR Institute of Genetics. Since being established the CBW series has trained over 1000 students from academia, government laboratories and industry.

The CBW began offering a new series of advanced workshops in 2008 and has continuously sold out since then. The CBW is pleased to offer this series again in 2010 and to extend the series across Canada with workshops in both **Vancouver** and **Toronto**.

Workshops for 2010

Exploratory Data Analysis and Essential Statistics using R

May 6-7, 2010 in Vancouver, B.C. - SOLD OUT
September 30 - October 1, 2010 in Toronto, Ontario
Due to a down server, applications will be accepted until Sept 3

Clinical Genomics and Biomarker Discovery

September 23-24, 2010 in Toronto, Ontario

Done

portals

http://bioinformatics.org

Bioinformatics Organization – Bioinformatics.Org

http://www.bioinformatics.org/

Bioinformatics Organization

NEW! Professional Membership
- 20% discount on events
- IMAP4 e-mail account
- And more!

Your donations keep us online!
I pledge to donate (USD)...

Continue

Groups

Not logged in
- Log in

The Organization
- About us
- Membership (29357+)
  - Basic Free!
  - Professional New!
- Hosted groups (461)
- Bioinformatics courses
- Core Facility
- Career Center
- Wiki New!
- Franklin Award
- Sponsorship
- Contribute

News & Commentary

Submit · Archive

Resource: ProtozoaDB 1.0 (beta) released
Submitted by Alberto Davila posted on Saturday, September 11, 2010

We are happy to announce the new release of ProtozoaDB (1.0 beta):

http://protozoadb.biowebdb.org

Since the development of the database and related publication in 2008, a number of modifications and features have been added aiming to improve this resource. For this purpose, some systems needed to be improved and new pipelines needed to be developed. The OrthoSearch [1,2] system was developed for the identification of orthologous genes using HMMER profiles. These HMMER profiles have showed to be useful for the functional (re)annotation of protozoan genomes [3] and are available for download at "ProtozoaDB" horizontal tab / "Orthologs by OrthoSearch" section. The STINGRAY system (http://stingray.biowebdb.org) used in several sequencing/annotation projects (including the analysis and annotation of T. rangeli EST) has been improved too, and a manuscript describing those new features will

Acknowledgments

We wish to thank the following for their support:

DNAnexus

Human Genome Bioinformatics Service from US$4000 in 7 days

[Become a sponsor]

Done

portals

http://gchelpdesk.ualberta.ca

Canadian Bioinformatics Help Desk — At the Nexus of Science and Service

http://gchelpdesk.ualberta.ca/

GenomeCanada
BIOINFORMATICS

Print Page

Home
Montreal Seminar Links
Montreal Workshop Links

About Us
Presentation
Brochure
Bioinformatics Needs Survey Report
Services
Service Policy
Bioinformatics Platform
Platform Services
Platform Team

Personnel
Contact Us

Directory
Access the Directory

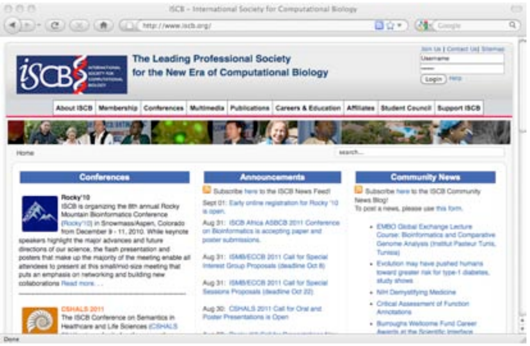
News

HELP DESK
At the nexus of science and service

Done

societies

http://iscb.org



ISCB - International Society for Computational Biology

http://www.iscb.org/

ISCB INTERNATIONAL SOCIETY FOR COMPUTATIONAL BIOLOGY

The Leading Professional Society for the New Era of Computational Biology

Join Us | Contact Us | Sitemap
Username
Login Help

About ISCB · Membership · Conferences · Multimedia · Publications · Careers & Education · Affiliates · Student Council · Support ISCB

Home

Conferences

Rocky'10
ISCB is organizing the 8th annual Rocky Mountain Bioinformatics Conference (Rocky'10) in Snowmass/Aspen, Colorado from December 9 - 11, 2010. While keynote speakers highlight the major advances and future directions of our science, the flash presentation and posters that make up the majority of the meeting enable all attendees to present at this small/mid-size meeting that puts an emphasis on networking and building new collaborations Read more . . .

CSHALS 2011
The ISCB Conference on Semantics in Healthcare and Life Sciences (CSHALS

Announcements

Subscribe here to the ISCB News Feed!

Sept 01: Early online registration for Rocky '10 is open.

Aug 31: ISCB Africa ASBCB 2011 Conference on Bioinformatics is accepting paper and poster submissions.

Aug 31: ISMB/ECCB 2011 Call for Special Interest Group Proposals (deadline Oct 8)

Aug 31: ISMB/ECCB 2011 Call for Special Sessions Proposals (deadline Oct 22)

Aug 30: CSHALS 2011 Call for Oral and Poster Presentations is Open

Community News

Subscribe here to the ISCB Community News Blog!
To post a news, please use this form.

- EMBO Global Exchange Lecture Course: Bioinformatics and Comparative Genome Analysis (Institut Pasteur Tunis, Tunisia)
- Evolution may have pushed humans toward greater risk for type-1 diabetes, study shows
- NIH Demystifying Medicine
- Critical Assessment of Function Annotations
- Burroughs Wellcome Fund Career Awards at the Scientific Interface

Done

journals

Bioinformatics
NAR (esp. Databases and WebServices issues)
BMC Bioinformatics
PLoS Computational Biology
others ...

... all available electronically via U of T Library
... all have e-mail contents alert service or RSS.

textbook

Zvelebil & Baum:
Understanding Bioinformatics

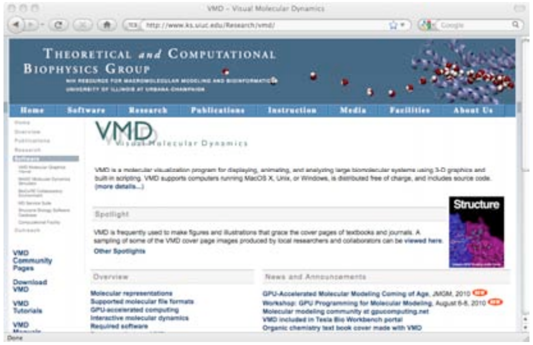
Garland Science, 2008

Tools

- VMD
- R
- Perl
- Jalview
- Phylip
- UCSD genome browser
- [...]

VMD

http://www.ks.uiuc.edu/Research/vmd/


VMD - Visual Molecular Dynamics
http://www.ks.uiuc.edu/Research/vmd/
Theoretical and Computational Biophysics Group
Home
Software
Research
Publications
Instruction
Media
Facilities
About Us
VMD
Visual Molecular Dynamics
Structure
VMD Community Pages
Download VMD
VMD Tutorials
Done

VMD is a molecular visualization program for displaying, animating, and analyzing large biomolecular systems using 3-D graphics and built-in scripting. VMD supports computers running MacOS X, Unix, or Windows, is distributed free of charge, and includes source code. (more details...)

Spotlight

VMD is frequently used to make figures and illustrations that grace the cover pages of textbooks and journals. A sampling of some of the VMD cover page images produced by local researchers and collaborators can be viewed here.

Other Spotlights

Overview

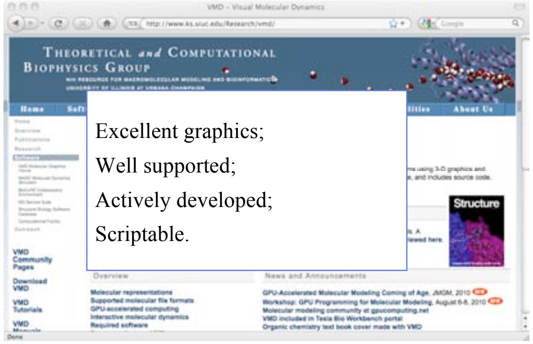
Molecular representations
Supported molecular file formats
GPU-accelerated computing
Interactive molecular dynamics
Required software

News and Announcements

GPU-Accelerated Molecular Modeling Coming of Age, JMGM, 2010
Workshop: GPU Programming for Molecular Modeling, August 6-8, 2010
Molecular modeling community at gpucomputing.net
VMD included in Tesla Bio Workbench portal
Organic chemistry text book cover made with VMD

VMD

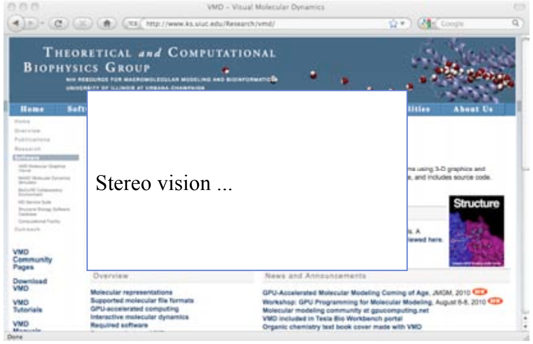
http://www.ks.uiuc.edu/Research/vmd/



Excellent graphics;
Well supported;
Actively developed;
Scriptable.

VMD

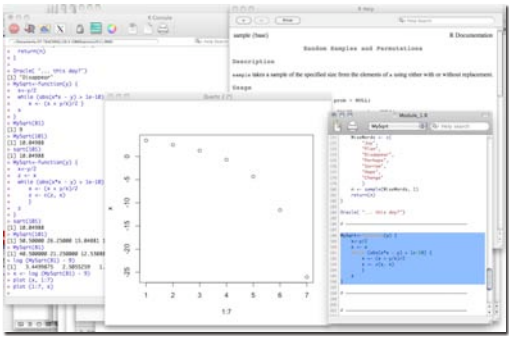
http://www.ks.uiuc.edu/Research/vmd/



Stereo vision ...

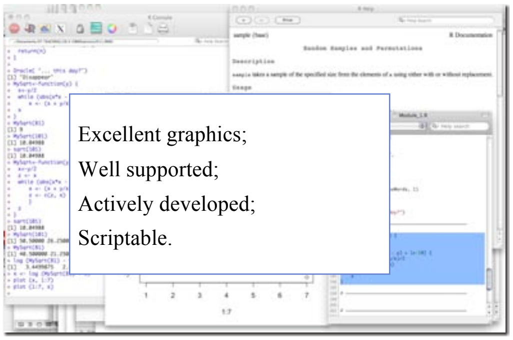
R

http://www.r-project.org

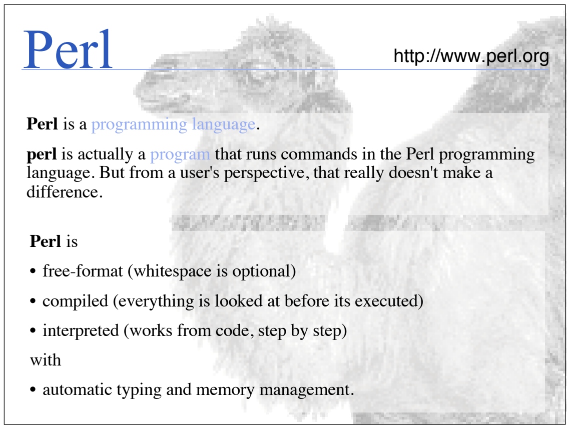


R

http://www.r-project.org



Excellent graphics;
Well supported;
Actively developed;
Scriptable.



Perl

http://www.perl.org

Perl is a programming language.

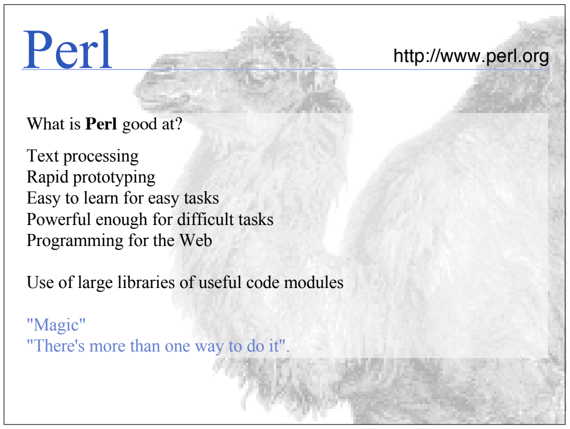
perl is actually a program that runs commands in the Perl programming language. But from a user's perspective, that really doesn't make a difference.

Perl is

- free-format (whitespace is optional)
- compiled (everything is looked at before its executed)
- interpreted (works from code, step by step)

with

- automatic typing and memory management.



Perl

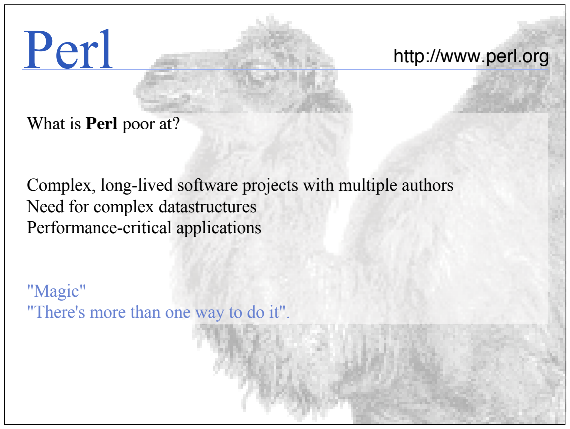
http://www.perl.org

What is **Perl** good at?

Text processing
Rapid prototyping
Easy to learn for easy tasks
Powerful enough for difficult tasks
Programming for the Web

Use of large libraries of useful code modules

"Magic"
"There's more than one way to do it".



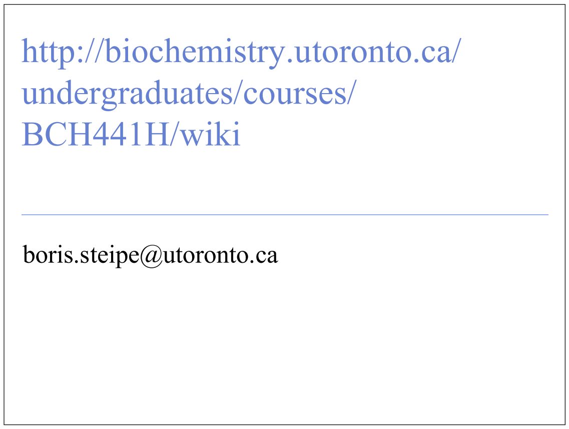
Perl

http://www.perl.org

What is **Perl** poor at?

Complex, long-lived software projects with multiple authors
Need for complex datastructures
Performance-critical applications

"Magic"
"There's more than one way to do it".



http://biochemistry.utoronto.ca/undergraduates/courses/BCH441H/wiki

boris.steipe@utoronto.ca