
B-TEAM Progress Report

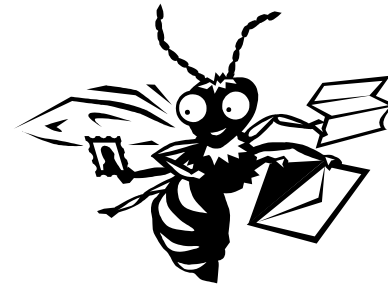
30 June, 2005

Tracy Gabridge

Erja Kajosalo

Louisa Worthington Rogers

Howard Silver

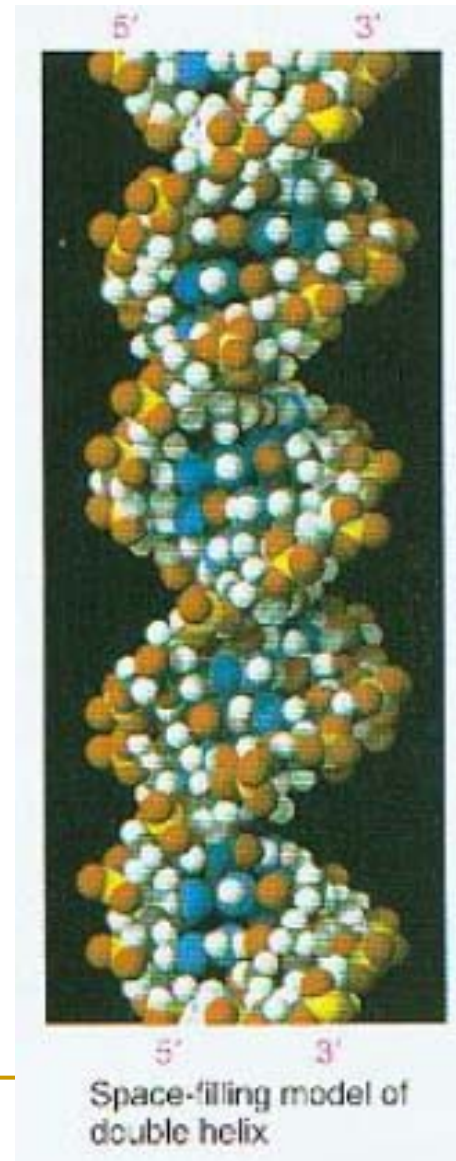


Outline

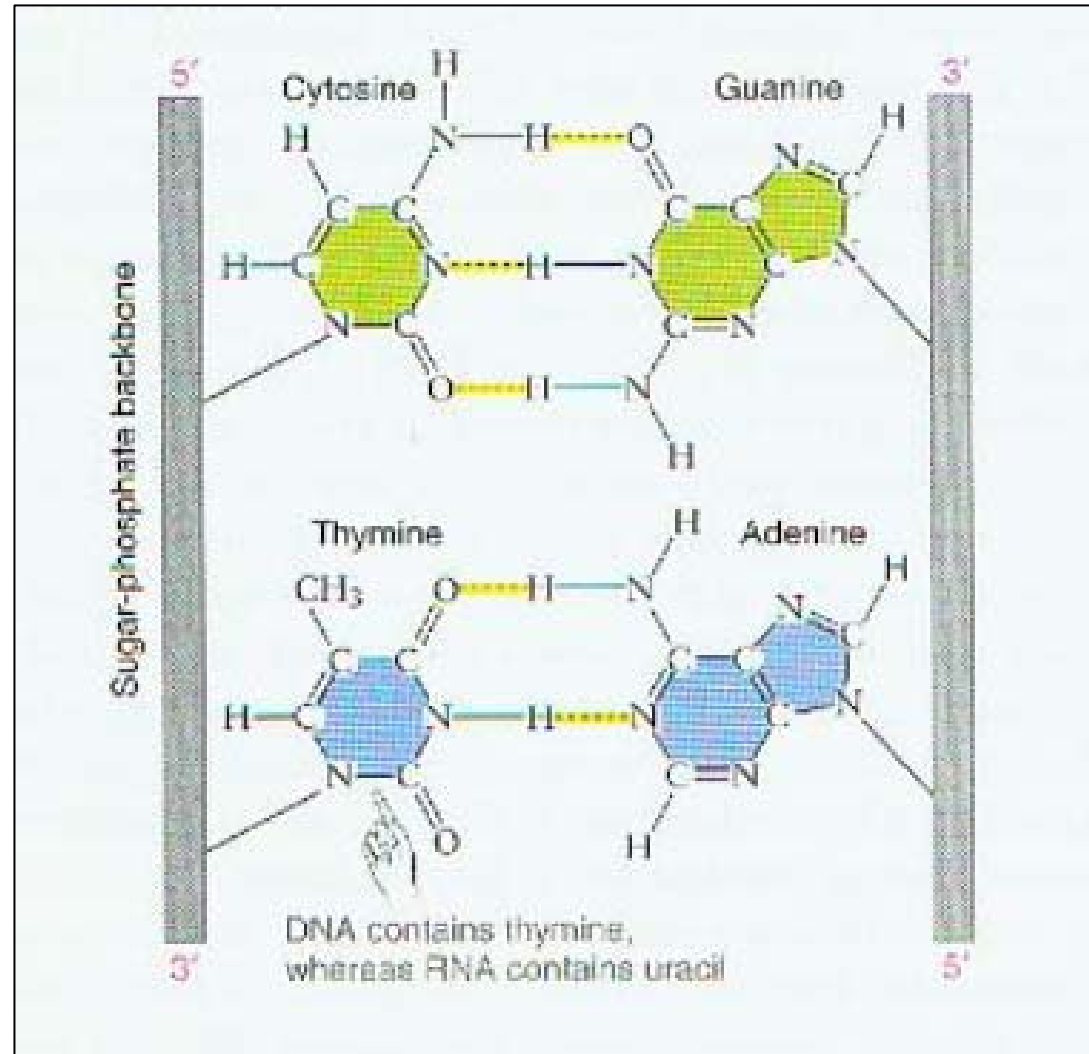
- Why is Biology so BIG? – Tracy
- Biology at MIT – Louisa
- Collections and Biology – Louisa
- Informatics – Erja
- Academic Libraries and Biology – Howard
- What's Next – Howard
- Questions and Discussion

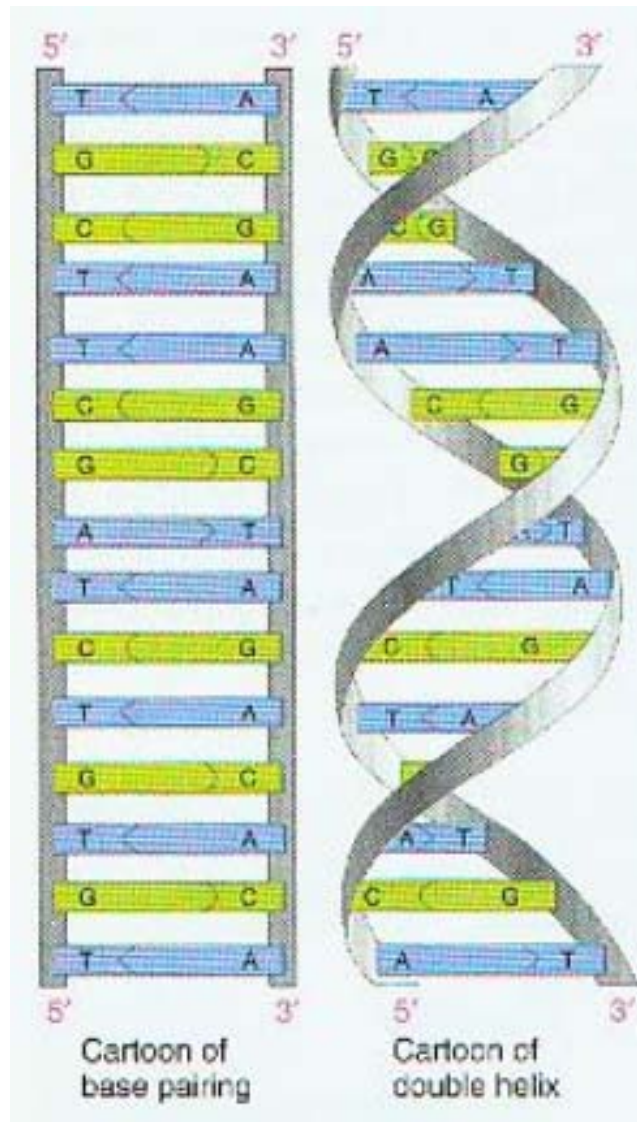
Why is biology so BIG?

DNA!



Nucleic Acids – 4 of them





3 Nucleic Acids = 1 Amino Acid

		Second base			
		U	C	A	G
First base	U	UUU } Phenyl- UUC } alanine UUA } Leucine UUG }	UCU } UCC } Serine UCA } UCG }	UAU } Tyrosine UAC } UAA } Stop codon UAG } Stop codon	UGU } Cysteine UGC } UGA } Stop codon UGG } Tryptophan
	C	CUU } CUC } Leucine CUA } CUG }	CCU } CCC } Proline CCA } CCG }	CAU } Histidine CAC } CAA } Glutamine CAG }	CGU } CGC } Arginine CGA } CGG }
	A	AUU } AUC } Isoleucine AUA } AUG } Methionine start codon	ACU } ACC } Threonine ACA } ACG }	AAU } Asparagine AAC } AAA } Lysine AAG }	AGU } Serine AGC } AGA } Arginine AGG }
	G	GUU } GUC } Valine GUA } GUG }	GCU } GCC } Alanine GCA } GCG }	GAU } Aspartic GAC } acid GAA } Glutamic GAG } acid	GGU } GGC } Glycine GGA } GGG }

Genomes

- Human = 3,000,000,000 (3×10^9) nucleic acid bases
- Mouse = 2,700,000,000 bases
- Yeast = 120,000,000 bases

Simple!

Central Dogma



Biology/Computational metaphors

- Nucleic Acids (base 4) → bits (base 2)
- Amino Acids (3 Nucleic Acids) → Bytes (8 bits)
- Proteins (lots of Amino Acids) → Assembly language
- Genes → subroutines
- Genome → a program

Biology/Library metaphors

- Nucleic Acids (base 4) → letters
- Amino Acids (3 Nucleic Acids) → words
- Proteins (lots of Amino Acids) → sentences
- Genes → chapters/articles
- Genome → book
- Many Genomes → a library

Life Sciences Overview

Life Sciences Departments – 100% life sciences research

- Biology (67 faculty)
 - Center for Cancer Research
 - Broad Institute (4 faculty plus many research scientists)
 - Computational and Systems Biology (CSBi)
 - Whitehead Institute – all principal investigators are MIT biology faculty even though Whitehead is separately administered
- Brain & Cognitive Sciences – BCS (40+ faculty)
 - McGovern Brain Institute
 - Picower Center for Learning and Memory
- Harvard-MIT Division of Health Sciences & Technology – HST (200+ faculty)
- Division of Biological Engineering – BE (36 faculty)
- Clinical Research Center – CRC
- Division of Comparative Medicine

Note that there are joint appointments among all departments.

Life Sciences Overview

Departments Across Institute Involved in Biomedical Research

Aeronautics	4.7%
Chemical engineering	53.3%
Chemistry	55.0%
Civil and Environmental Engineering	22.2%
EAPS	20.0%
EECS	25.8%
Materials science	27.0%
Mathematics	6.0%
Mechanical engineering	15.1%
Nuclear Engineering	30.4%
Ocean Engineering	4.2%
Physics	5.0%

Life Sciences Overview

Quantitative Studies Across Disciplines

Bioinformatics	Biology, EECS, chemistry, HST, math
Medical informatics	Biology, medicine, health planning, HST, math
Computational biology	Biology, EECS, BE, math (CSBi)
Computational neurosciences	BCS, biology, physics, math (CSBi)
Genomics and proteomics	Biology, chemistry, EECS, math
Gene sequencing	Biology, EECS, math
Systems biology	Biology, EECS, physics, chemistry, BE, math (CSBi)
Statistical models of biomedical applications	Biology, BCS, EECS, biophysics, chemistry, math (probability and statistics)

Life Sciences Overview

Interdisciplinary Centers, Labs and Programs

- Biotechnology Process Engineering Center
- Broad Institute
- Center for Biomedical Engineering
- Center for Cancer Research
- Center for Environmental Health Sciences
- Center for Materials Science and Engineering
- Center for Space Research
- Clinical Research Center
- Computational and Systems Biology
- George Russell Harrison Spectroscopy Laboratory
- Institute for Soldier Nanotechnologies
- McGovern Institute for Brain Research
- Microsystems Technology Laboratories
- Picower Center for Learning and Memory
- Program in Polymer Science and Technology
- Program on the Pharmaceutical Industry

Life Sciences Overview

THE INTELLECTUAL CHALLENGE

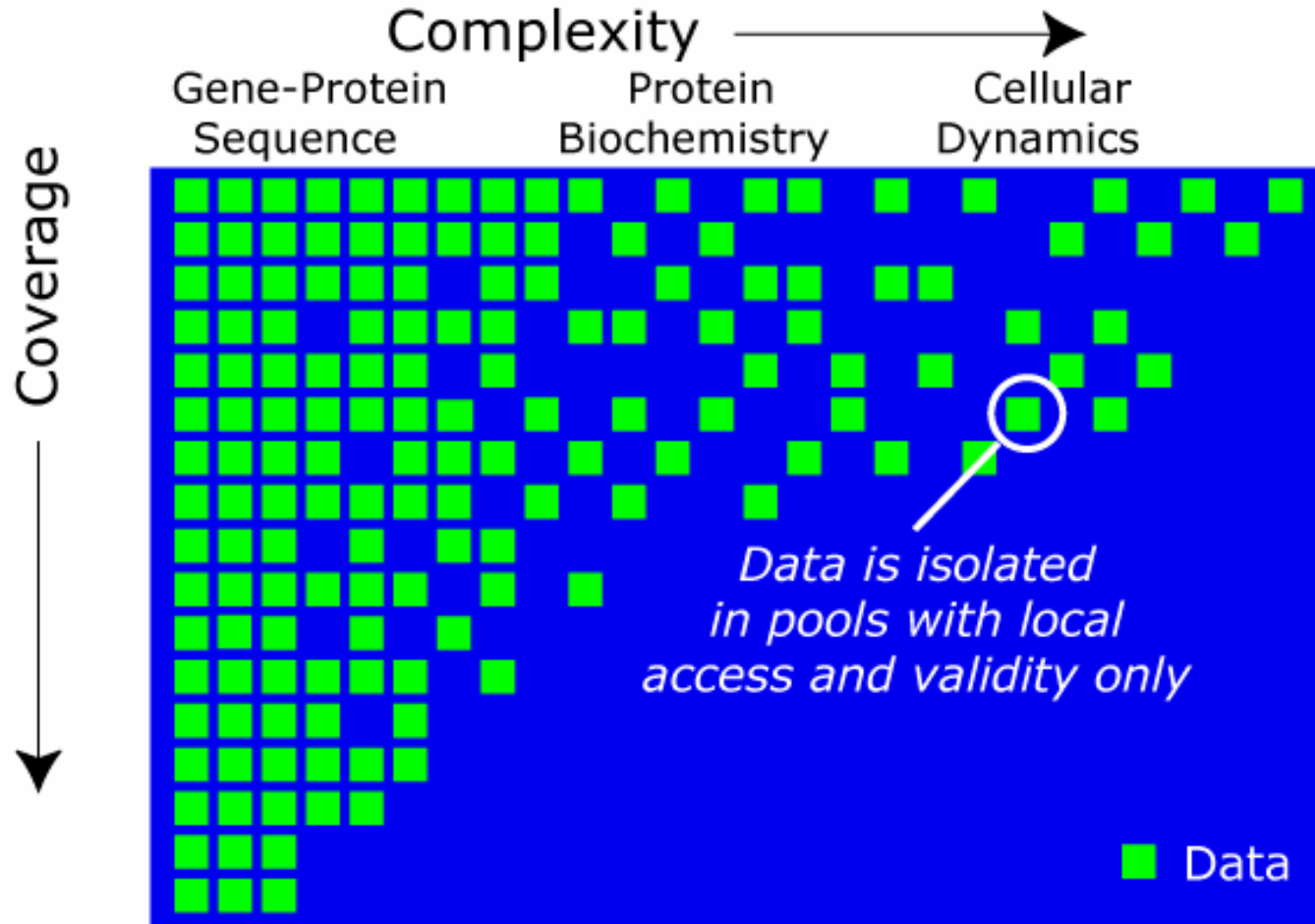
- “Biological systems are extremely complex. Over the past 50 years, the challenge of biological complexity has been addressed by studying one molecule at a time. Now, at the start of the 21st century, the limit of component-by-component analysis is becoming apparent.

It is now recognized that dealing with biological complexity is not simply a matter of increasing the scale of research. Rather, tackling complexity requires the development of new analytical tools and experimental strategies that lie at the interface between biology, engineering and computation.”

CSBi: “A Proposalfor Support of the CSBi Technology Platform”, 2003

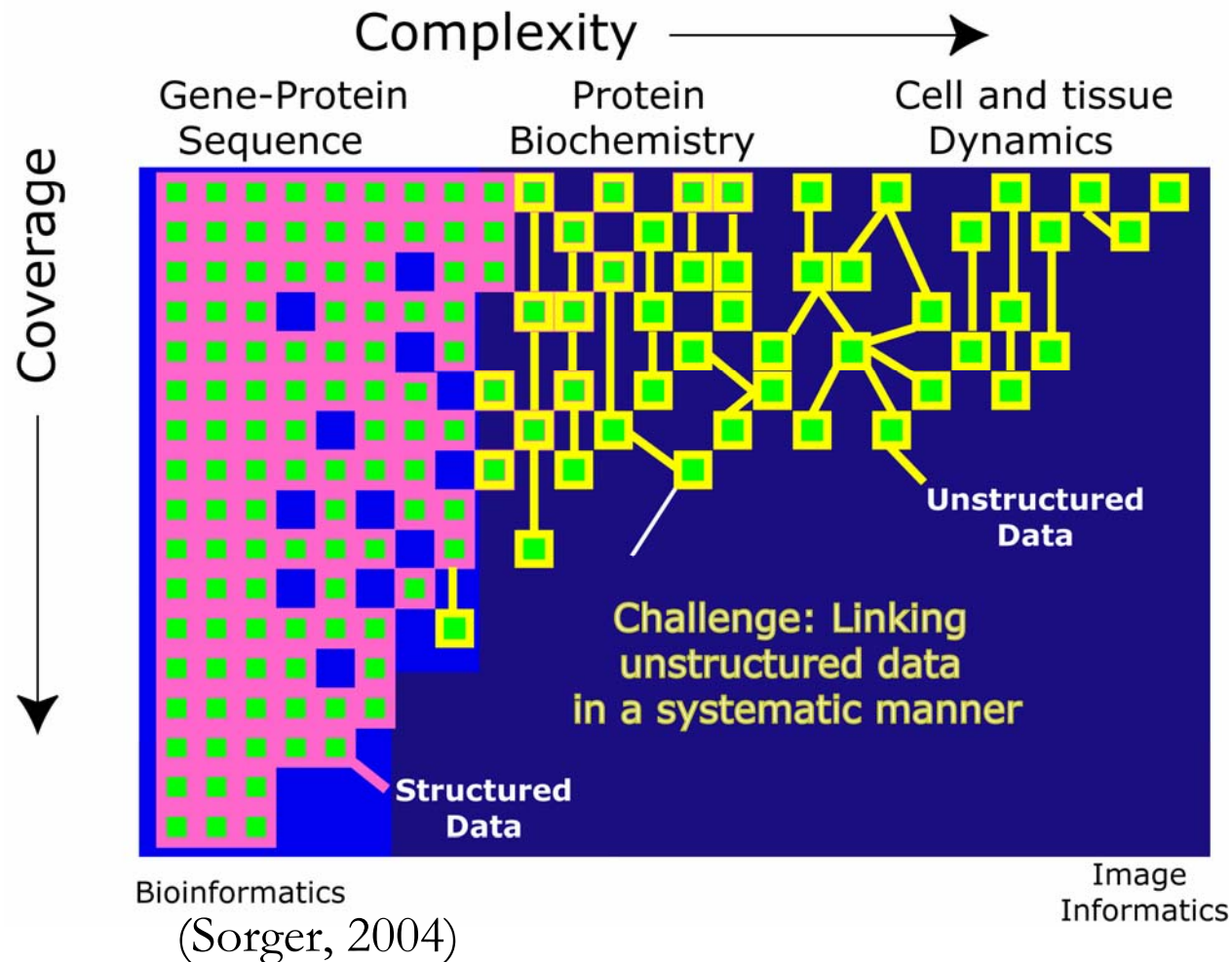
- **We are talking about a paradigm shift in biological research!**

Data problem: Biology is data poor in systematically acquired sets of complex data



(Sorger, 2004)

Data challenge: Integrating complex unstructured data represents a fundamentally new challenge



Collections

WHAT ARE SOME ISSUES INVOLVED WITH INTERDISCIPLINARY RESEARCH?

- Researchers need basic background to apply to other fields
- Researchers often need to get up to speed quickly in another field
- Techniques and methods are applied to other fields
- Interdisciplinary Centers to promote interchange of techniques and skills

Collections

Issue: Researchers need basic background to apply to other fields

- RLE uses basic neurosciences research for developing artificial neural networks
- Mech Eng needs human anatomy and physiology to see how the body works and moves in order to create products, as cardiac implants, artificial skin and valves

Issue: Researchers often need to get up to speed quickly in another field

- Mech Eng professor needed to read neurosurgery journals for a period of time as background for his work on nerve regeneration
- Biologists are taking mini-courses on UNIX, PERL, and other computational tools for gene sequencing and related work. Courses sponsored by department

Collections

Issue: Techniques and methods are applied to other fields

- Computer programming and analysis, as well as chemistry techniques, are used for programming of biological data sets of genes, proteins, genomes, etc. Called bioinformatics.
- Mathematics and statistics used to analyze data sets in life sciences. Head of Broad Institute (genome sequencing) is a mathematician, not a biologist.

Collections

Issue: Interdisciplinary Centers to promote interchange of techniques and skills.

- Computational and Systems Biology initiative (CSBi) – Biologists, bioengineers, chemists, computer scientists, biostatisticians
- Computational programs established in life sciences departments, as computational neuroscience in BCS, computational biology in Biology, Bioengineering

Collections

How Do Libraries Address Interdisciplinary Needs?

Current Methods

- Selectors share information
- Pattern analysis and publishing studies
- Provide access to data banks as NCBI (Entrez), other data sets and gene transcription data bases. Templates for use with data in labs.
- Collection manager's role
- Increased budget

Collections

How Do Libraries Address Interdisciplinary Needs?

Ideas to Pursue

- Benchmarking in other libraries
- More consistent analysis of interlibrary loan requests
- Buy different kinds of resources, as data sets
- Buy software in conjunction with data sets
- Increased budget

Outline - *informatics*

- Definitions
- Findings from surveys and interviews.
- Plans for additional data collection.
- Preview of possible recommendations.

Informatics-

- - “is the discipline of science which investigates the structure and properties (not specific content) of scientific information, as well as the regularities of scientific information activity, its theory, history, methodology and organization”. *OED*

Examples

From Wikipedia - <http://en.wikipedia.org/wiki/Informatics>

- Bioinformatics – Informatics in biology, especially genomics.
- Cheminformatics – Informatics in chemistry and drug discovery.
- Ecoinformatics – Informatics in ecology and environmental science.
- Legal informatics - Informatics in the legal field.
- Medical informatics – Informatics in medicine and healthcare.
- Neuroinformatics – Informatics in neuroscience.
- Hydroinformatics – Informatics as a tool in resolving water-related problems.
- Veterinary informatics – Informatics in veterinary medicine.
- ...

Bioinformatics -

- - “is the field of science in which biology, computer science, and information technology merge into a single discipline...” *NCBI* -<http://www.ncbi.nlm.nih.gov/>
- important sub-disciplines
 - development of new algorithms and statistics with which to assess relationships among members of large data sets
 - the analysis and interpretation of various types of data including nucleotide and amino acid sequences, protein domains, and protein structures
 - the development and implementation of tools that enable efficient access and management of different types of information.“

Examples – “*bioinformatist*”

- Knows...
 - biology and/or biomedicine
 - IT networking & database design
- Is skilled in
 - searching specialized databases
 - research study design
 - developing services tailored to users
 - instruction & outreach
 - licensing & copyright

Bioinformatics @ MIT

- Whitehead
- Broad Institute
- Departments: Biology, EECS, Chemistry, Biological Engineering Division,...

NCBI overview

- Established in 1988 as a national resource for molecular biology information for the better understanding of molecular processes affecting human health and disease
- Creates public databases, conducts research in computational biology
- Develops software tools for analyzing genome data, and disseminates biomedical information
- <http://www.ncbi.nlm.nih.gov/>
- <http://www.ncbi.nlm.nih.gov/gquery/gquery.fcgi>

Examples – *Blast* (*Basic Local Alignment Search Tool*)

- Finds regions of local similarity between sequences.
- Compares nucleotide or protein sequences to sequence databases (e.g. GenBank) and calculates the statistical significance of matches.
- Used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.

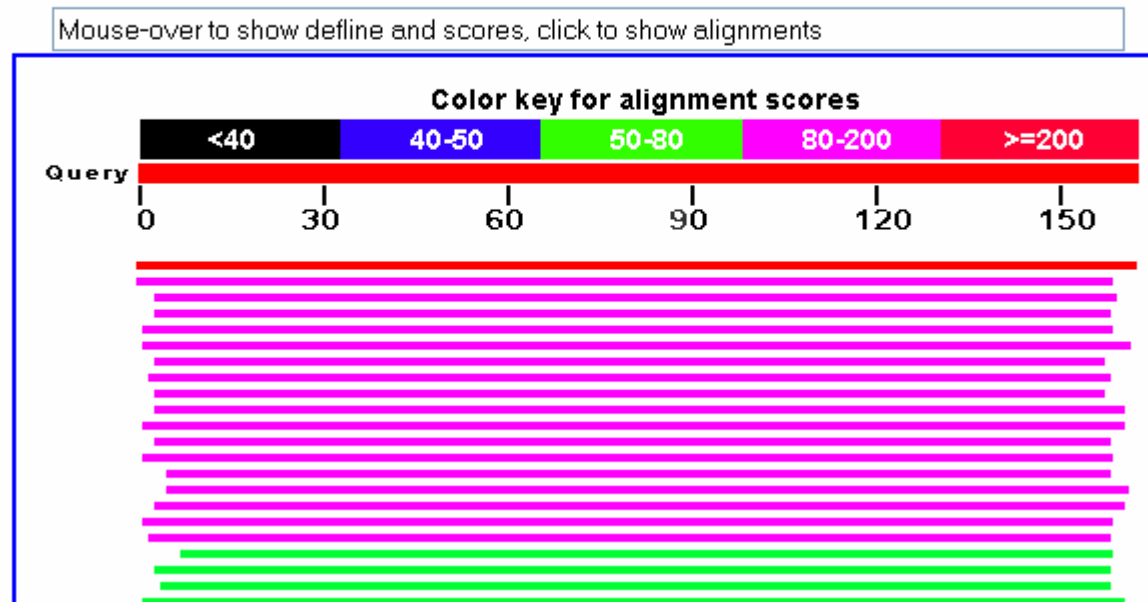
Examples – *Blast* (2)

■ Search

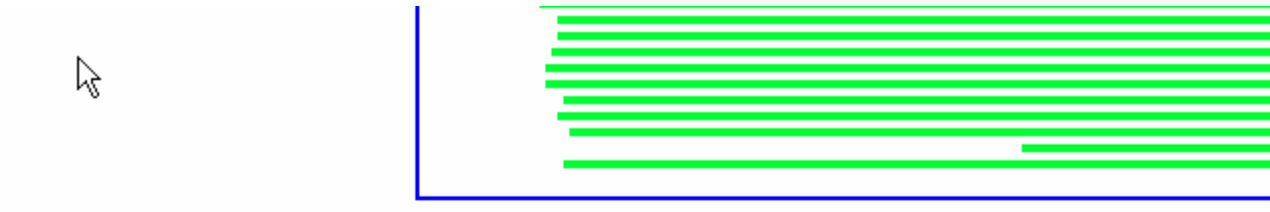
“SVMYKKILYPTDFSETAEIALKHKAFKTLKAE
EVILLHVIDEREIKKRDIFSLLLGVAGLNKSVEEF
ENELKNKLTEEAKNKMENIKKELEDVGFKVKDII
VVGIPHEEIVKIAEDEGVDIIMGSHGKTNLKEIL
LGSVTENVIKKSNKPVLVVKRKNS”

Examples – *Blast* (3)

Distribution of 560 Blast Hits on the Query Sequence



Examples – *Blast* (4)



Related Structures

Sequences producing significant alignments:			Score (Bits)	E Value	
gi 15668757 ref NP_247556.1 	hypothetical protein MJ0577 [Met...	_311	5e-84	G	
gi 14590690 ref NP_142758.1 	hypothetical protein PH0823 [Pyr...	_106	3e-22	G	
gi 15668711 ref NP_247510.1 	hypothetical protein MJ0531 [Met...	90.5	1e-17	G	
gi 19918032 gb AAM07295.1 	universal stress protein [Methanos...	88.6	6e-17	G	
gi 18893697 gb AAL81681.1 	hypothetical protein [Pyrococcus f...	88.2	7e-17	G	

Examples – Blast (4)

Entrez Gene

My NCBI
[Sign In] [Registered]

PubMed Nucleotide Protein Genome Structure PMC Taxonomy Books OMIM

for 2622094[PUID] Go Clear Save Search current records only

Limits Preview/Index History Clipboard Details

Display Summary Show 20 Send to

All: 1 Genes Genomes: 1 SNP GeneView: 0

1: [MTH993](#)

[Links](#)

hypothetical protein [*Methanothermobacter thermautotrophicus str. Delta H*]

Other Aliases: MTH993

GeneID: 1471401

Examples - *visualization*

- Structure data is integrated into Entrez, so that literature searches can lead to MMDB structure summaries.
- For example, go to [Entrez](#), select "Structure" in the "Search" pull-down menu, then type "PTEN" in the input box and hit "Go" to do the search. (show result at desktop)




What done so far...

- Literature search – health science libraries
- Few discussions within MIT
- Survey results of other academic libraries
(handout)



Survey – Q6

- Do you get any bioinformatics and/or cheminformatics questions at your library?
- Typical questions –
 - How find/use NCBI databases?
 - How to find protein or nucleotide sequences?
 - How to use BLAST?
 - Help id'ing SNPs?



Survey – Q8

What kind of support does your library offer for 8. bioinformatics/cheminformatics related questions?		Number of Responses	Response Ratio
Basic - at least some reference staff is able to locate resources, programs and databases.		24	60%
Intermediate - a subject specialist(s) able to offer specialized in-depth instruction for some resources.		14	35%
Expert - you have an informatics specialist handling analytical questions, planning experiments and in-depth assistance with data analysis.		2	5%
Total		40	100%



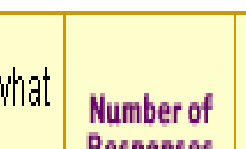

Survey – Q10

Does your library offer training in the use of bioinformatics (or cheminformatics) resources?		Number of Responses	Response Ratio
Yes		15	38%
No		24	62%
Total		39	100%

Survey – Q11

Do you have any bioinformatics/cheminformatics related web pages (e.g. 11. subject resources, training, etc.)		Number of Responses	Response Ratio
Yes		10	27%
No		27	73%
Total		37	100%

Survey – Q12

12. If you have a bioinformatics (cheminformatics) specialist on campus, what kind of position is it?		Number of Responses	Response Ratio
Located in the library - go to #13.		6	15%
Located outside the library (e.g. in a department) - go to #15.		16	40%
Do not have one - go to #16.		12	30%
VIEW Other?		6	15%
Total		40	100%

Survey – Q13

- Do you know what kind of education your informatics specialist has? Librarian with extra training? Informatics degree? Other?
 - MS/PhD – biology, microbiology, genetics, biochemistry,...
 - Computer science + biology/chemistry degrees
 - Some also MLS/MLIS

 - Position funded mostly by libraries or libraries/depts

What done so far... other

- Peter Murray-Rust visit (3/17-18)
- Bioinformatics - David Osterbur (Harvard) as a part of *Interdisciplinary Research -Tools for finding Biological Information* (4/26)
- Jeff Loo up-coming (7/26)
- Lonelyss Charles (7/26)

What services could we offer?

- ❑ Basic - *at least some reference staff is able to locate resources, programs and databases*
- ❑ Intermediate - *a subject specialist able to offer specialized in-depth instruction for some resources*
- ❑ Expert - *you have an informatics specialist handling analytical questions, planning experiments and in-depth assistance with data analysis*

Academic Libraries and Biology

Questions

- Has the organization of the library changed?
- Is there a staff development strategy?
- Are there changes in services and collection practices?
- Has the library developed new partnerships, or modified its research agenda?

Academic Libraries - Research

- Conferences & Listserves
- Literature
- Interviews/Survey
 - Selected organizations: UC San Diego, Cornell.....

What's Next?

- How are we organized?
- What do we do for our users?
- Input in collection building
- Build staff knowledge
- Engage ESL
- Host NLM Fellow

Questions and Discussion



Feedback welcome!

- bteam@mit.edu
- [Wiki](#)



Schedule

- How and why did life sciences become so important, and why B-team was charged: Tracy ~10 minutes
- B@MIT - Louisa - Overview and quick review of departmental findings ~10 min
- Collections - Louisa - questions we developed, types of data, how it helps build collections, how it can inform service and outreach, new money for biol. ~15 min
- Informatics - Erja - What is it? Findings from surveys and interviews. Plans for additional data collection. Preview of possible recommendations. ~15 min
- Academic libraries - Howard - Findings from literature, web and interviews. ~5 minutes
- Questions and Discussion - ESL ~ 20 minutes
- What's next? - Howard 2 minutes