NextGen Sequencing Network Workshop

A. Malcolm Campbell

DAVIDSON

Juniata College, PA
July 11, 2011

Outline of Presentation

1. History of **GCAT**
2. Expansion to Meet Objectives
3. Moving to Synthetic Biology
4. Other Sequencing Models
5. Lessons and Advice

Davidson College

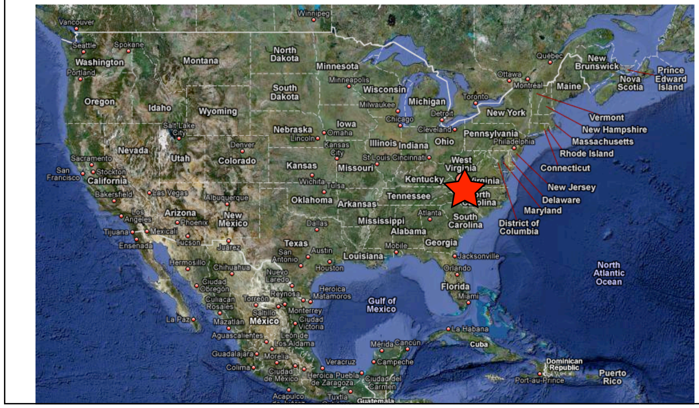
Davidson, NC USA

Liberal Arts College

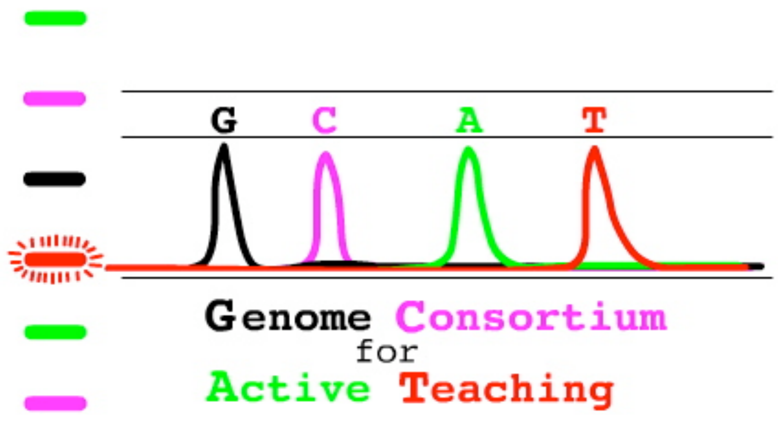
1800 Undergraduates

48 States

36 Countries

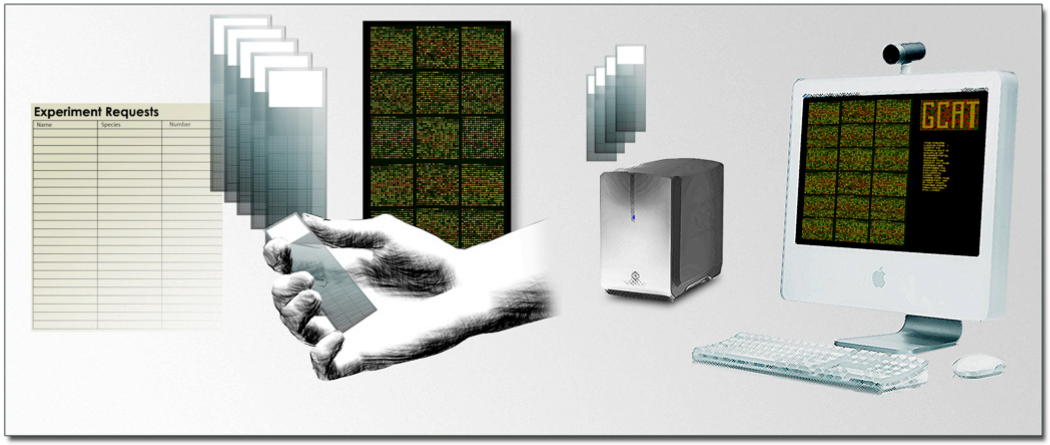


11 Year Collaboration
Three Countries



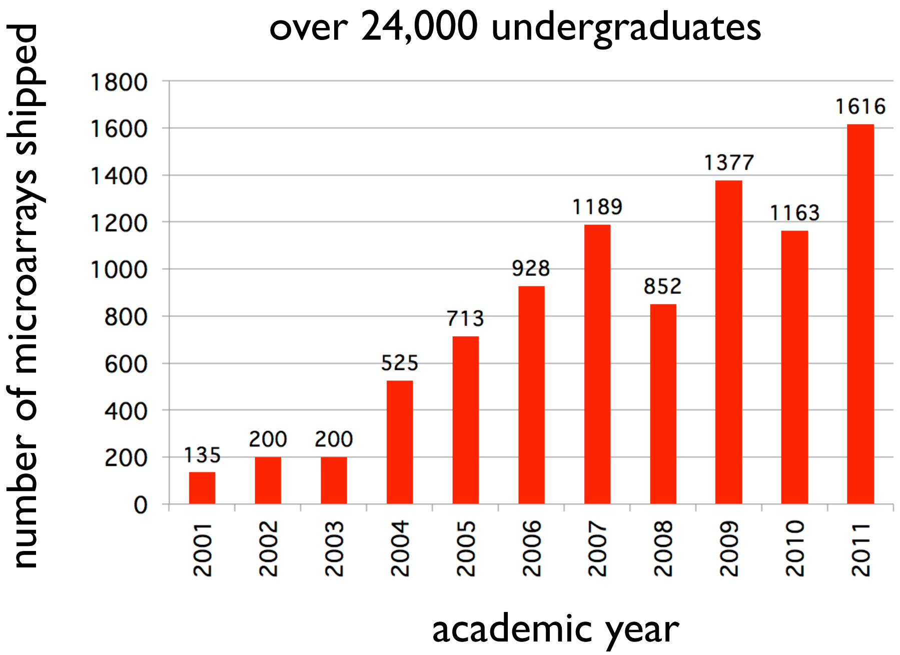
www.bio.davidson.edu/GCAT

Genome Consortium for Active Teaching



Started in 2000
Focused on DNA microarrays

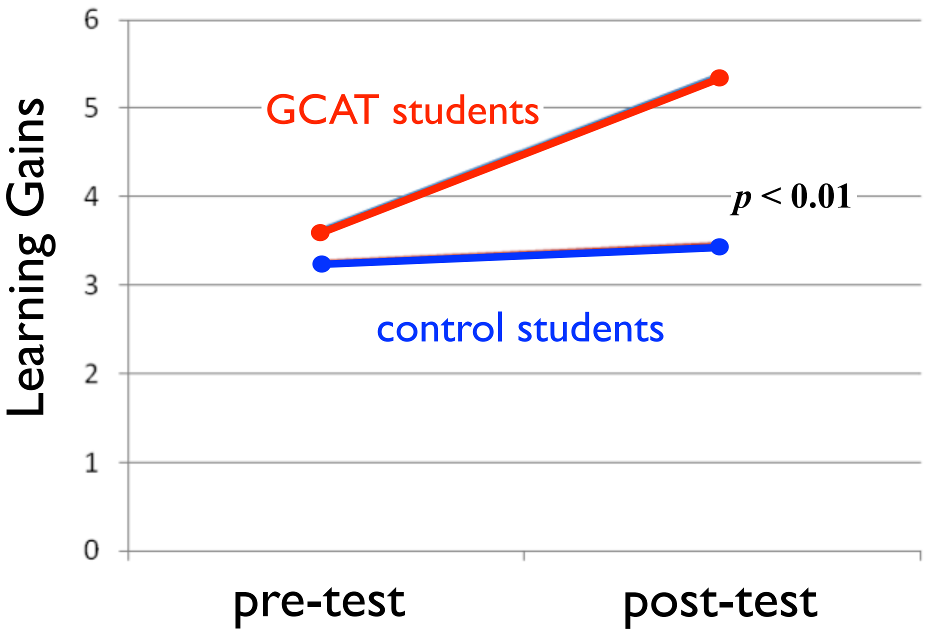
Genome Consortium for Active Teaching



Genome Consortium for Active Teaching

- Puerto Rico, Alaska, Hawaii, HBCUs, HSIs, community colleges and large state institutions. In 37 states plus Washington DC.
- Cited by BIO2010, Mathematical Association of America, *Nature Medicine*, HHMI, NSF, *Science*, & *Genetic Engineering News*
- **16 basic research papers** with **86 undergraduate coauthors**.
- **19 educational papers** with **3 undergraduate coauthors**.
- Protocols in English and Spanish, translated by workshop alumni.

Genome Consortium for Active Teaching



Student Learning Outcomes

Question	Topic	Increase (%)
1.	Microarray experimental error–dye bias	+ 36.2*
2.	Microarray experimental error–gradient	+ 10.5*
3.	Microarray negative controls	+ 10.3*
4.	Microarray experimental design	+ 38.2*
5.	Gene expression ratios using a graph	+ 5.8*
6.	Gene expression–probability	+ 0.2
7.	Gene expression–gene clusters	+ 22.3*
8.	Gene expression–regulatory cascade	+ 14.9*
9.	Gene expression–gene circuit graphs	+ 11.8*
10.	Interpreting microarray results	+ 19.0*
11.	Diagnosis with microarrays	+ 12.5*

* indicates $p < 0.05$; N = 409

Student Learning Outcomes

Question	Topic	Increase (%)
1.	Microarray experimental error–dye bias	+ 36.2*
2.	Microarray experimental error–gradient	+ 10.5*
3.	Microarray negative controls	+ 10.3*
4.	Microarray experimental design	+ 38.2*
5.	Gene expression ratios using a graph	+ 5.8*
6.	Gene expression–probability	+ 0.2
7.	Gene expression–gene clusters	+ 22.3*
8.	Gene expression–regulatory cascade	+ 14.9*
9.	Gene expression–gene circuit graphs	+ 11.8*
10.	Interpreting microarray results	+ 19.0*
11.	Diagnosis with microarrays	+ 12.5*

* indicates $p < 0.05$; N = 409

GCAT

Faculty Development

NSF funded

GCAT

Faculty Development

9 workshops, 360 faculty

Faculty Appreciate **GCAT** Resources

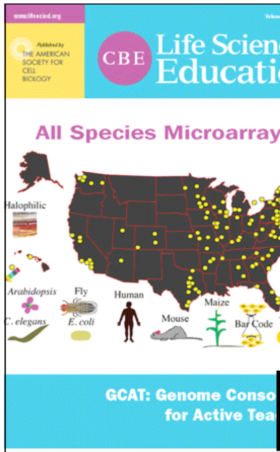
	Mean	SD
Access to microarray technology without **GCAT**	1.5	0.75
Online **GCAT** protocols useful	4.4	0.69
The **GCAT** -Listserv helpful	4.2	1.0
GCAT network significant factor	4.2	0.79
Positive experience using **GCAT**	4.6	0.60
I would use **GCAT** again in the future	4.7	0.63

1 = strongly disagree
5 = strongly agree

Faculty Development

"You have awakened parts of my brain that have been dormant since my last stats course. The only reason I have gone over the manual so carefully is that this is my first time teaching microarrays, or even using them, for that matter. **GCAT** has been remarkably helpful to me. In fact I don't think I would have undertaken this new module in my lab course without the tools **GCAT** makes available."

GCAT



www.lifescied.org

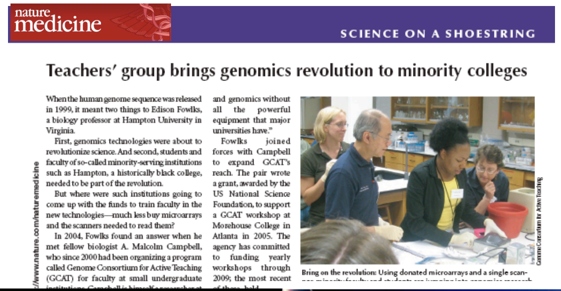
Volume 6, Summer 2007

Published by THE AMERICAN SOCIETY FOR CELL BIOLOGY

CBE Life Sciences Education

All Species Microarrays

GCAT: Genome Conso
for Active Tea



nature medicine

SCIENCE ON A SHOESTRING

Teachers' group brings genomics revolution to minority colleges

When the human genome sequence was released in 1999, it meant two things to Edison Fowlks, a biology professor at Hampton University in Virginia.

First, genomics technologies were about to revolutionize science. And second, students and faculty of so-called minority-serving institutions such as Hampton, a historically black college, needed to be part of the revolution.

But where were such institutions going to come up with the funds to train faculty in the new technologies—much less buy microarrays and the scanners needed to read them?

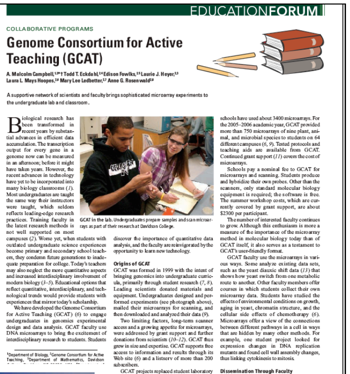
In 2004, Fowlks found an answer when he met fellow biologist A. Malcolm Campbell, who since 2000 had been organizing a program called Genome Consortium for Active Teaching (GCAT) for faculty at small undergraduate

and genomics without all the powerful equipment that major universities have."

Fowlks joined forces with Campbell to expand GCAT's reach. The pair wrote a grant, awarded by the US National Science Foundation, to support a GCAT workshop at Morehouse College in Atlanta in 2005. The agency has committed to funding yearly workshops through 2009; the most recent

Bring on the revolution: Using donated microarrays and a single scan-

http://www.nature.com/naturemedicine



EDUCATIONFORUM

COLLABORATIVE PROGRAMS

Genome Consortium for Active Teaching (GCAT)

A. Malcolm Campbell,[1,2*†] Todd T. Eckdahl,[2,4] Edison Fowlks,[2,5] Laurie J. Heyer,[2,3] Laura L. Mays Hoopes,[2,6] Mary Lee Ledbetter,[2,7] Anne G. Rosenwald[2,8]

A supportive network of scientists and faculty brings sophisticated microarray experiments to the undergraduate lab and classroom.

Biological research has been transformed in recent years by substantial advances in efficient data accumulation. The transcription output for every gene in a genome now can be measured in an afternoon; before it might have taken years. However, the recent advances in technology have yet to be incorporated into many biology classrooms (*1*). Most undergraduates are taught the same way their instructors were taught, which seldom reflects leading-edge research practices. Training faculty in the latest research methods is not well supported on most campuses (*2*). Worse yet, when students with outdated undergraduate science experiences become primary and secondary school teachers, they condemn future generations to inadequate preparation for college. Today's teachers may also neglect the more quantitative aspects and increased interdisciplinary involvement of modern biology (*3–5*). Educational options that reflect quantitative, interdisciplinary, and technological trends would provide students with experiences that mirror today's scholarship.

We have developed the Genome Consortium for Active Teaching (GCAT) (*6*) to engage undergraduates in genomics experimental design and data analysis. GCAT faculty use DNA microarrays to bring the excitement of interdisciplinary research to students. Students discover the importance of quantitative data analysis, and the faculty are reinvigorated by the opportunity to learn new technology.

GCAT in the lab. Undergraduates prepare samples and scan microarrays as part of their research at Davidson College.

Origins of GCAT

GCAT was formed in 1999 with the intent of bringing genomics into undergraduate curricula, primarily through student research (*7*, *8*). Leading scientists donated materials and equipment. Undergraduates designed and performed experiments (see photograph above), mailed their microarrays for scanning, and then downloaded and analyzed their data (*9*).

Two limiting factors, long-term scanner access and a growing appetite for microarrays, were addressed by grant support and further donations from scientists (*10–12*). GCAT thus grew in size and expertise. GCAT supports free access to information and results through its Web site (*6*) and a listserv of more than 200 subscribers.

GCAT projects replaced student laboratory methods less prevalent in today's research, such as cloning and sequencing a gene and Northern blotting.

Rapid Growth

GCAT is committed to enabling any institution to adopt the use of microarrays in its undergraduate curriculum at affordable prices. To date, about 5000 undergraduates from 120 schools have used about 3400 microarrays. For the 2005–2006 academic year, GCAT provided more than 750 microarrays of nine plant, animal, and microbial species to students on 64 different campuses (*6*, *9*). Tested protocols and teaching aids are available from GCAT. Continued grant support (*11*) covers the cost of microarrays.

Schools pay a nominal fee to GCAT for microarrays and scanning. Students produce and hybridize their own probes. Other than the scanners, only standard molecular biology equipment is required; the software is free. The summer workshop costs, which are currently covered by grant support, are about $2300 per participant.

The number of interested faculty continues to grow. Although this enthusiasm is more a measure of the importance of the microarray method in molecular biology today than of GCAT itself, it also serves as a testament to GCAT's user-friendly format.

GCAT faculty use the microarrays in various ways. Some analyze existing data sets, such as the yeast diauxic shift data (*13*) that shows how yeast switch from one metabolic route to another. Other faculty members offer courses in which students collect their own microarray data. Students have studied the effects of environmental conditions on growth, aging in yeast, chromatin structure, and the cellular side effects of chemotherapy (*6*). Microarrays offer a view of the connections between different pathways in a cell in ways that are hidden by many other methods. For example, one student project looked for expression changes in DNA replication mutants and found cell wall assembly changes, thus linking cytokinesis to mitosis.

Dissemination Through Faculty Development

GCAT has sponsored data generation (wet lab) and data analysis (dry lab) workshops in various settings (*14*). Wet and dry lab sessions work best when they run 2 and 3 days, respectively. Participants learn data analysis using MAGIC Tool freeware (*15*). MAGIC Tool works on any computer platform and is designed to enhance student understanding of

---

[1] Department of Biology, [2] Genome Consortium for Active Teaching, [3] Department of Mathematics, Davidson

ncemag.org SCIENCE VOL 311 24 FEBRUARY 2006 1103

Published by AAAS

JMBE

Journal of Microbiology & Biology Education

May 2007
Volume 8

In This Issue:

Microarrays and Data Analysis
David Kushner

Introductory Biology Discussion Group Evaluation
Marcy Peteroy-Kelly

Online Versus Onsite Bioinformatics Instruction
Kristina Obom and Patrick Cummings

Microbial Mats as Educational Tools
Carlos Rios-Velazquez, Lilliam Casillas-Martinez, and Pieter T. Visscher

Published by the: AMERICAN SOCIETY FOR MICROBIOLOGY

www.MicrobeLibrary.org

NSF Current

Monthly Highlights of Research and Education Sponsored by the National Science Foundation

In This Issue: Latest Plant Genome Awards • Teachers Jockey Genes
RNA Plays Novel Role • New Nanotechnology Centers • NSF's 2005 Facility Plan

October 2005

Awards 19 New Plant Genome Research Projects

genomes of economically important plants are often large and plex, but through in-depth studies scientists uncover information that be translated into new and improved agricultural products and ices.

National Science Foundation (NSF) made 19 new awards totaling million in the eighth year of its Plant Genome Research Program RP). The two- to five- year awards, ranging from $622,000 to $7.7 n, fund research and tools to reveal information in the genomes of omically important crop plants, such as wheat and soybean, as well as crease understanding of the genetics underlying plant processes ding disease resistance, flavor development, seed growth and wood ation.

NSF's press release and the list of 2005 PGRP awards for more nation.

NSF made 19 new awards in the eighth year of its Plant Genome Research Program. The awards will support genomics research in major crop plants such as soybean and also in trees including the loblolly pine and poplar. Credit: N. Rager Fuller, NSF.

ructors from Minority-Serving Institutions Learn to Teach Microarray Technology

Microarray technology, one of the hottest techniques in biological research, simultaneously measures the expression levels of tens of thousands of genes. Performing DNA microarray experiments and analyzing the mounds of resulting data are generally thought to be beyond the reach of all but a small number of undergraduates working in top research labs. However, the Genome Consortium for Active Teaching (GCAT), composed of faculty from over 120 primarily undergraduate institutions, has allowed over 4,000 undergraduates to conduct research using DNA microarrays.

Research grants funded

Published basic research

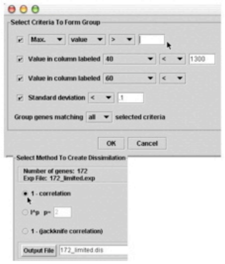
Published learning gains

Teaching awards

Expansion to Meet Objectives

Open Source and Free Software

Choosing Search Criteria



Select Criteria To Form Group

☑ Max. value >

☑ Value in column labeled 40 < 1300

☑ Value in column labeled 60 <

☑ Standard deviation < .1

Group genes matching all selected criteria

Select Method To Create Dissimilation

Number of genes: 172
Exp File: 172_limited.exp

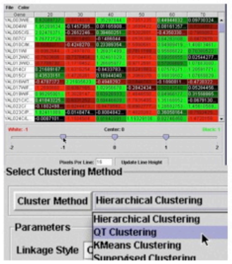
● 1 - correlation

○ l^p p= 2

○ 1 - (jackknife correlation)

Output File 172_limited.dis

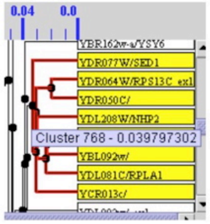
Exploring ratios

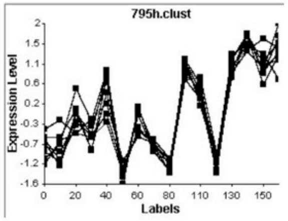


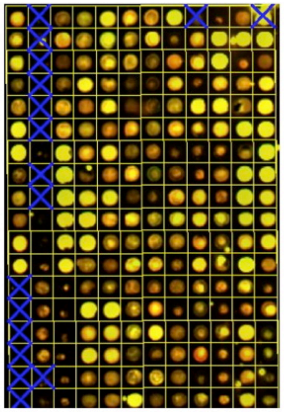
Select Clustering Method

Cluster Method: Hierarchical Clustering

MAGIC Tool

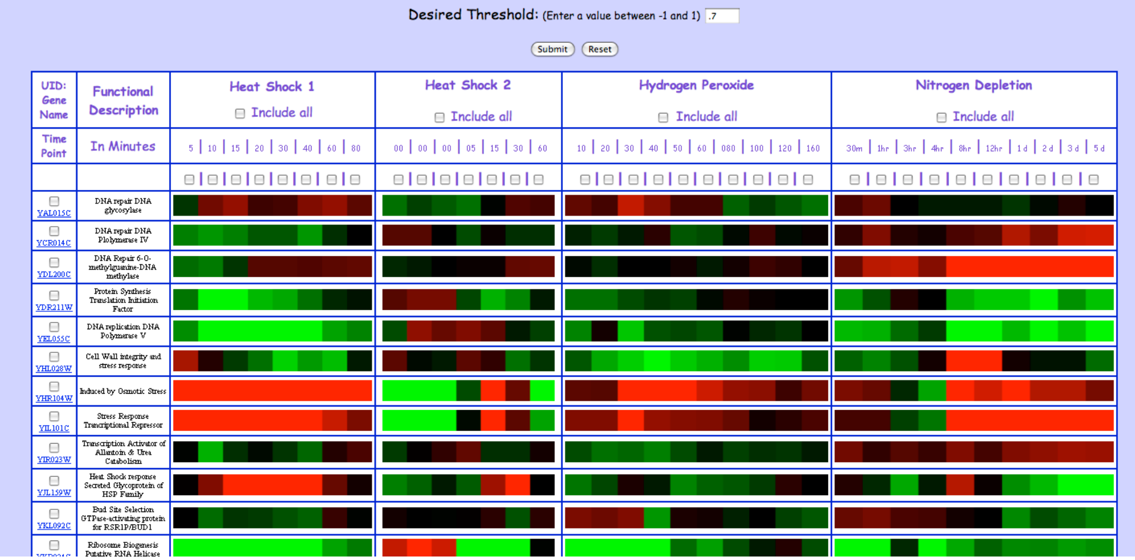






www.bio.davidson.edu/MAGIC

Online Clustering

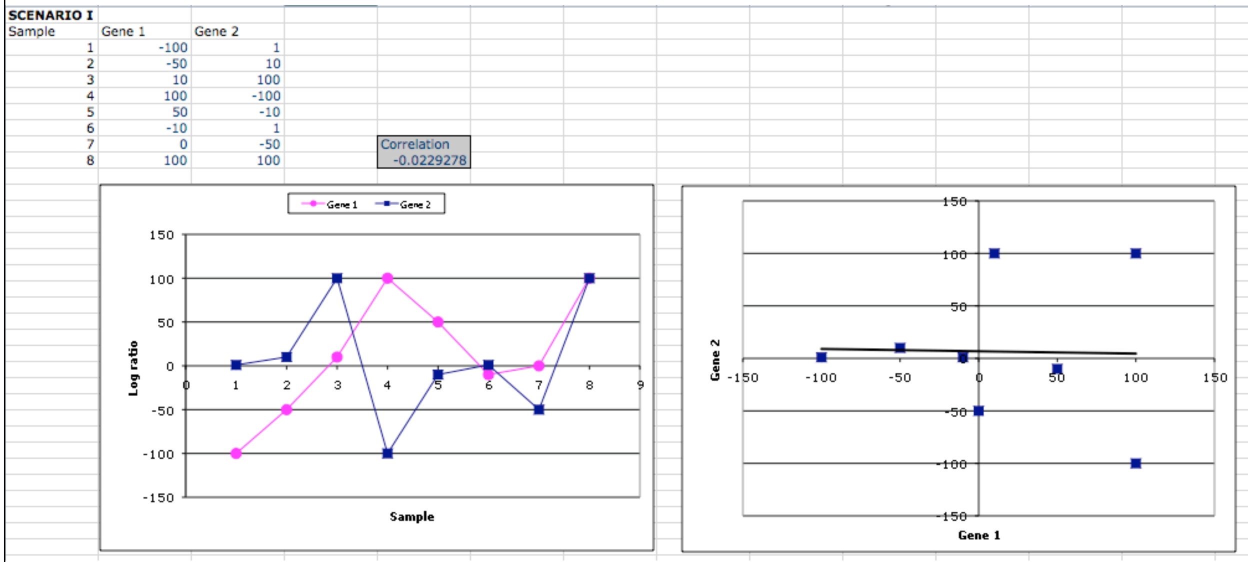


Desired Threshold: (Enter a value between -1 and 1) .7

UID: Gene Name	Functional Description
☐ YAL015C	DNA repair DNA glycosylase
☐ YCR014C	DNA repair DNA Plolymerase IV
☐ YDL200C	DNA Repair 6-O-methylguanine-DNA methylase
☐ YDR211W	Protein Synthesis Translation Initiation Factor
☐ YEL055C	DNA replication DNA Polymerase V
☐ YHL028W	Cell Wall integrity and stress response
☐ YHR104W	Induced by Osmotic Stress
☐ YIL101C	Stress Response Trancriptional Repressor
☐ YIR023W	Transcription Activator of Allantoin & Urea Catabolism
☐ YJL159W	Heat Shock response Secreted Glycoprotein of HSP Family
☐ YKL092C	Bud Site Selection GTPase-activating protein for RSR1P/BUD1
☐	Ribosome Biogenesis Putative RNA Helicase

Interactive clustering web page to learn about clusters and threshold values
http://gcat.davidson.edu/DGPB/clust/clustering.htm

Correlation Exercise



SCENARIO I

Sample	Gene 1	Gene 2
1	-100	1
2	-50	10
3	10	100
4	100	-100
5	50	-10
6	-10	1
7	0	-50
8	100	100

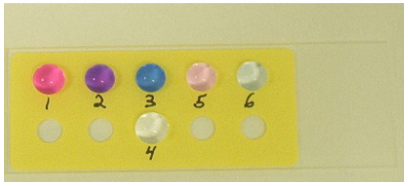
Correlation
-0.0229278

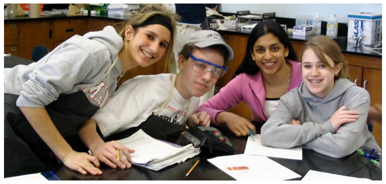
Four interactive scenarios to learn about correlation coefficients
http://gcat.davidson.edu/Online_Genomics/Online.html

Introduced Microarrays Early

1 minute

Ben Kittinger ‘05





Wet-lab microarray simulation kit - fast, cheap, works every time.
www.bio.davidson.edu/projects/gcat/HSChips/HSchips.html

GCAT Develops Commercial Product

Home > Search Term: 21-1520 > DNA Chips: Genes to Disease Lab Kit

DNA Chips: Genes to Disease Lab Kit

Advanced - For experienced high school and college classes; requires some technical skill.

- Simple microarray simulation
- No special equipment
- Comprehensive 2-to-4 day lesson plan
- 45-minute experiment
- Affordable lab option

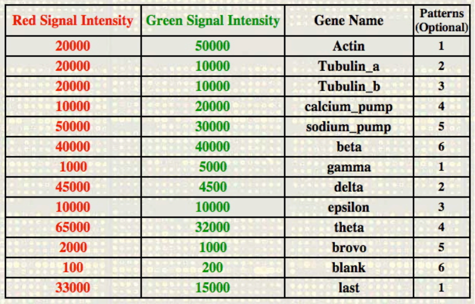
This microarray simulation provides an inexpensive, hands-on lab to introduce students to microarrays, the complexities of gene expression, and the role of gene expression in cancer. Using simulated microarray technology, students compare the relative expression levels of 6 different genes in healthy lung cells and lung cancer cells.

Students then discuss the significance of the relative expression levels with respect to the genes' roles in causing cancer. **Teacher's manual includes instructions for a bonus dry lab.**

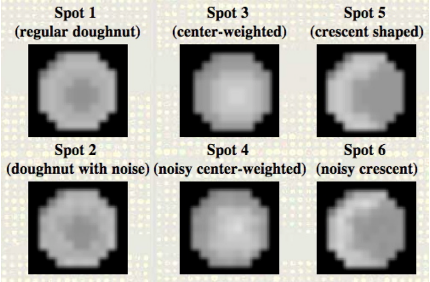
Lab Kit includes:

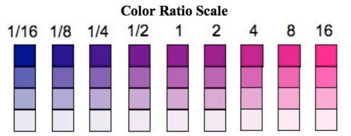
- Glass Slides
- Simulated Hybridization Solution
- 6 Simulated cDNA Solutions
- Instructions with Student Masters
- Pack of 30 Student Worksheets

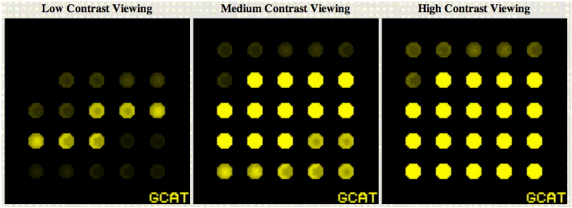
Enable Students to Practice

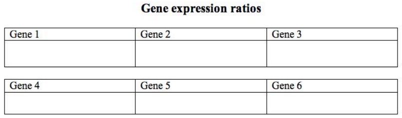


Red Signal Intensity	Green Signal Intensity	Gene Name	Patterns (Optional)
20000	50000	Actin	1
20000	10000	Tubulin_a	2
20000	10000	Tubulin_b	3
10000	20000	calcium_pump	4
50000	30000	sodium_pump	5
40000	40000	beta	6
1000	5000	gamma	1
45000	4500	delta	2
10000	10000	epsilon	3
65000	32000	theta	4
2000	1000	brovo	5
100	200	blank	6
33000	15000	last	1









Gene expression ratios

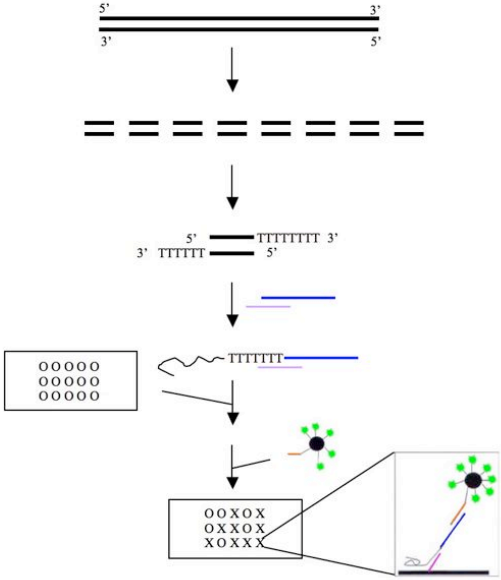
Gene 1	Gene 2	Gene 3

Gene 4	Gene 5	Gene 6

www.bio.davidson.edu/projects/GCAT/Spot_synthesizer/Spot_synthesizer.html

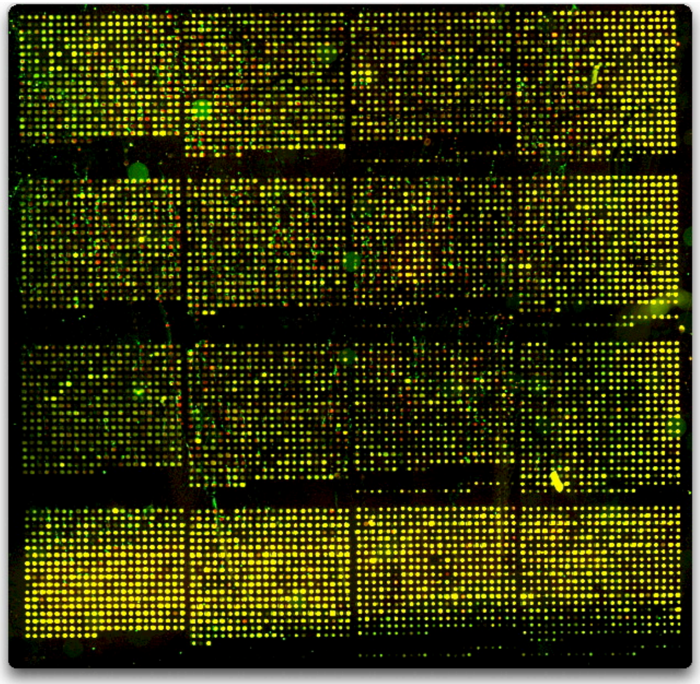
What Else Can Chips Do?

Jackie Ryan ‘05



www.bio.davidson.edu/projects/gcat/CGH/CGH.html

Comparative Genome Hybridizations

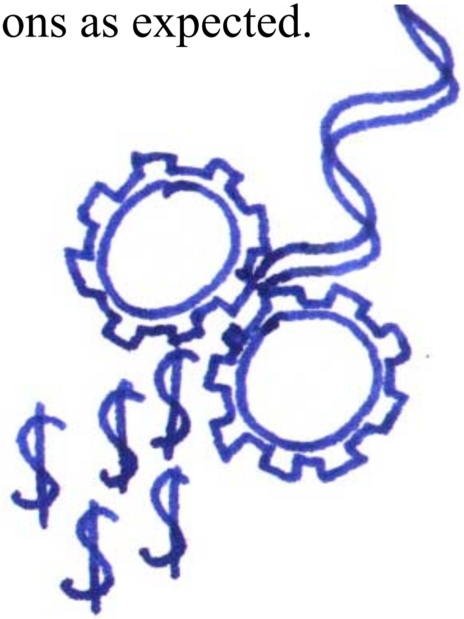


Moving into Synthetic Biology

www.bio.davidson.edu/projects/gcat/Synthetic/synthetic.html

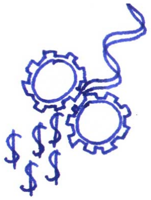
Synthetic Biology: Win-Win Research

Win #1: your design functions as expected.

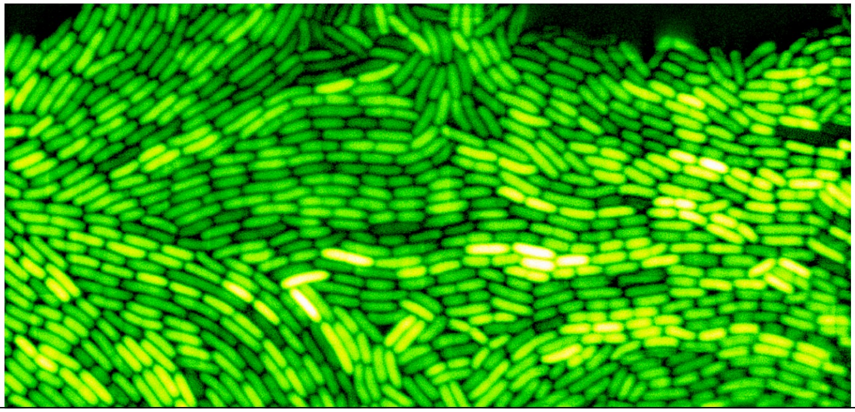


Synthetic Biology: Win-Win Research

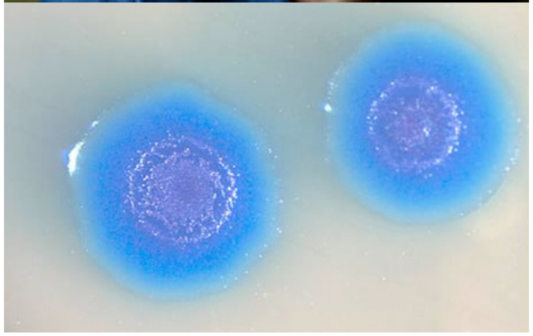
Win #1: your design functions as expected.



Win #2: your design fails but you uncover basic biology



How is Synthetic Biology Different?



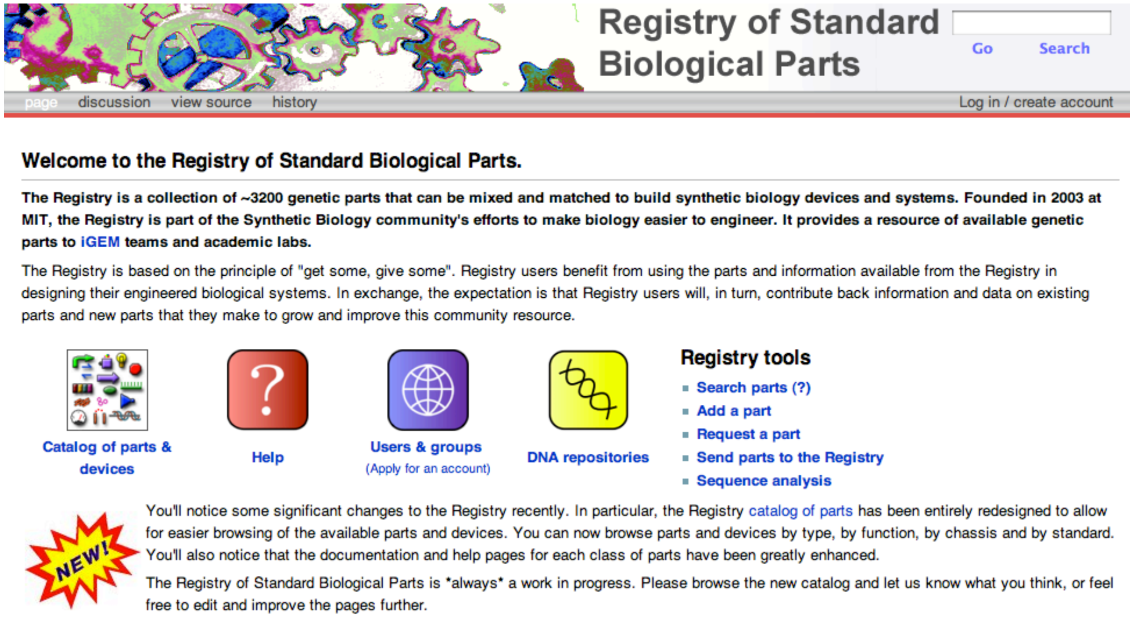
Abstraction

Modularity

Standards

Designing and modeling

Standardized & Modular DNA



Welcome to the Registry of Standard Biological Parts.

The Registry is a collection of ~3200 genetic parts that can be mixed and matched to build synthetic biology devices and systems. Founded in 2003 at MIT, the Registry is part of the Synthetic Biology community's efforts to make biology easier to engineer. It provides a resource of available genetic parts to iGEM teams and academic labs.

The Registry is based on the principle of "get some, give some". Registry users benefit from using the parts and information available from the Registry in designing their engineered biological systems. In exchange, the expectation is that Registry users will, in turn, contribute back information and data on existing parts and new parts that they make to grow and improve this community resource.

Registry tools

- Search parts (?)
- Add a part
- Request a part
- Send parts to the Registry
- Sequence analysis

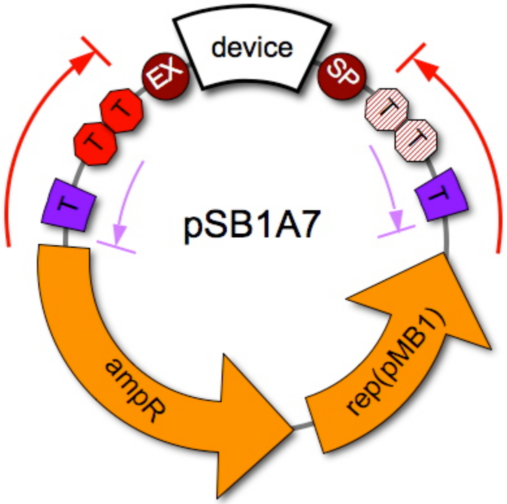
You'll notice some significant changes to the Registry recently. In particular, the Registry catalog of parts has been entirely redesigned to allow for easier browsing of the available parts and devices. You can now browse parts and devices by type, by function, by chassis and by standard. You'll also notice that the documentation and help pages for each class of parts have been greatly enhanced.

The Registry of Standard Biological Parts is *always* a work in progress. Please browse the new catalog and let us know what you think, or feel free to edit and improve the pages further.

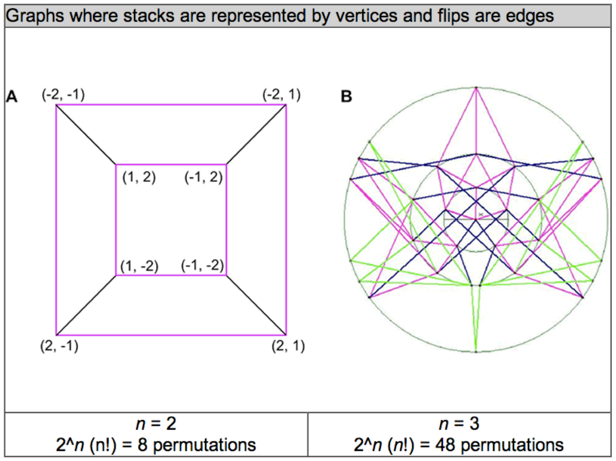
http://partsregistry.org/Main_Page

Think of Radio Shack for DNA parts.

Abstraction and Modeling



2^*n* (*n*!)

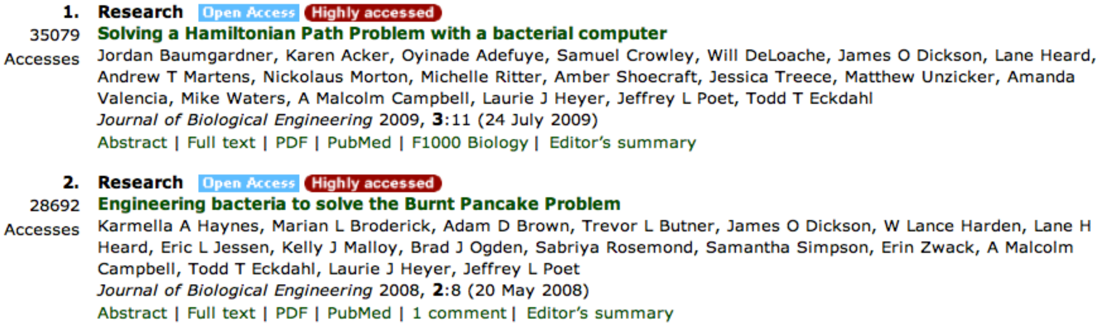

Graphs where stacks are represented by vertices and flips are edges

n = 2 2^*n* (n!) = 8 permutations	*n* = 3 2^*n* (*n*!) = 48 permutations

SynBio Papers Published

JOURNAL OF BIOLOGICAL ENGINEERING

Top 10 most accessed articles of all time

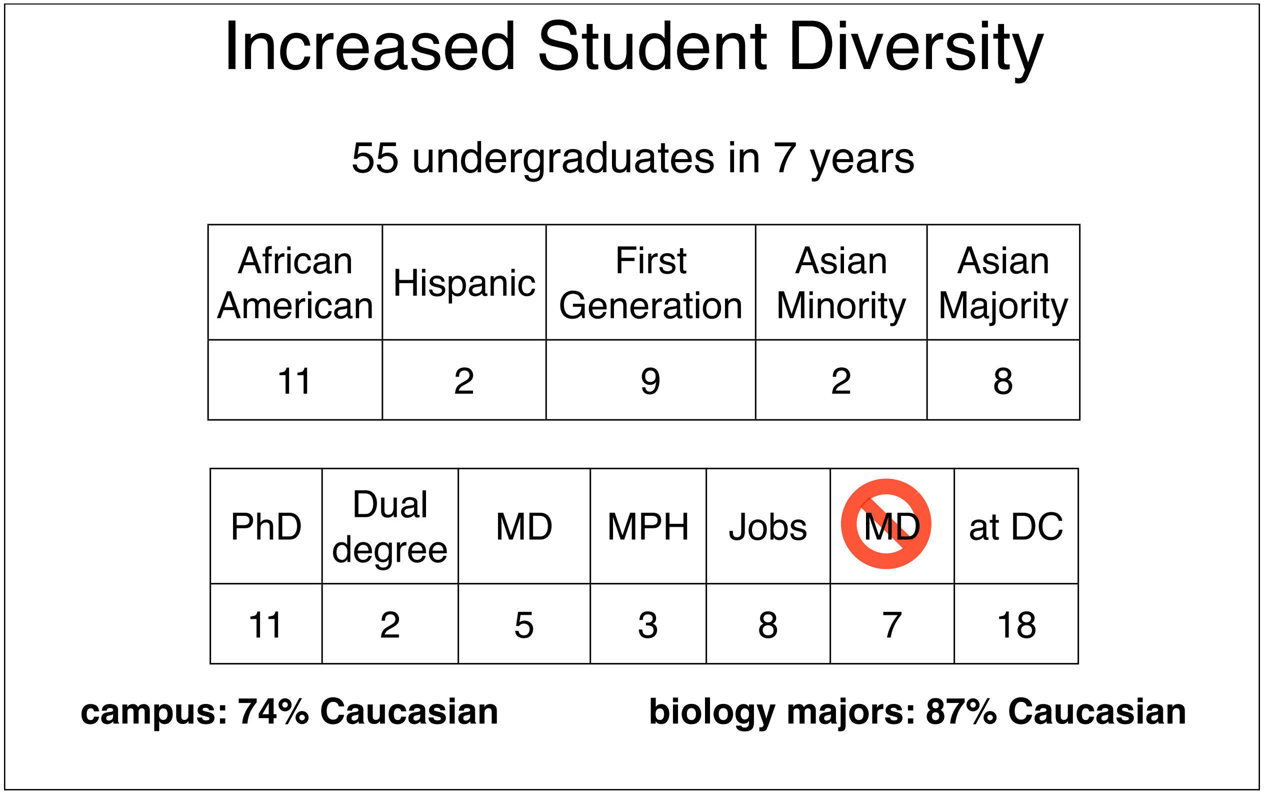


1. **Research** Open Access Highly accessed
35079 Accesses
Solving a Hamiltonian Path Problem with a bacterial computer
Jordan Baumgardner, Karen Acker, Oyinade Adefuye, Samuel Crowley, Will DeLoache, James O Dickson, Lane Heard, Andrew T Martens, Nickolaus Morton, Michelle Ritter, Amber Shoecraft, Jessica Treece, Matthew Unzicker, Amanda Valencia, Mike Waters, A Malcolm Campbell, Laurie J Heyer, Jeffrey L Poet, Todd T Eckdahl
Journal of Biological Engineering 2009, **3**:11 (24 July 2009)
Abstract | Full text | PDF | PubMed | F1000 Biology | Editor's summary

2. **Research** Open Access Highly accessed
28692 Accesses
Engineering bacteria to solve the Burnt Pancake Problem
Karmella A Haynes, Marian L Broderick, Adam D Brown, Trevor L Butner, James O Dickson, W Lance Harden, Lane H Heard, Eric L Jessen, Kelly J Malloy, Brad J Ogden, Sabriya Rosemond, Samantha Simpson, Erin Zwack, A Malcolm Campbell, Todd T Eckdahl, Laurie J Heyer, Jeffrey L Poet
Journal of Biological Engineering 2008, **2**:8 (20 May 2008)
Abstract | Full text | PDF | PubMed | 1 comment | Editor's summary

27 undergraduate coauthors

Having the number 2 paper of all time for the journal is nice, but having the number 1 paper is better.



Increased Student Diversity

55 undergraduates in 7 years

African American	Hispanic	First Generation	Asian Minority	Asian Majority
11	2	9	2	8

PhD	Dual degree	MD	MPH	Jobs	🚫 MD	at DC
11	2	5	3	8	7	18

campus: 74% Caucasian **biology majors: 87% Caucasian**

8 graduated in May 2011

More Interest in BioMath

[Davidson College](#)

4:00-5:00 p.m. Eastern
Chambers 3146

Be a Part of the Puzzle

See how Mathematics and Biology Connect

Join Bio-Math Connections meetings every Friday

[Missouri Western State University](#)

3:00-4:00 p.m. Central
Agenstein Hall 215

Fall 2009 Meetings

September 4

Introductions
NSF Grant Summary
Journal Search

September 11

How do you answer big questions?

September 18

This is an exceptional leaf!

September 25

DNA Microarrays

October 2

Phylogenetic tree pics

Phylogenetic tree sequences

No BMC Oct 9-30

Homo evolutis - rebooting our species.
18 min. talk at the 2009 TED conference

iGEM2009 Shared Wiki Space

iGEM 2009 Team Page

Spring 2010 Meetings

DRI proposals due Jan. 29
for Davidson Students

HHMI applicaitons due Feb 15 for Davidson Stud

a dozen students
every Friday 4 pm
2 campuses
video conference

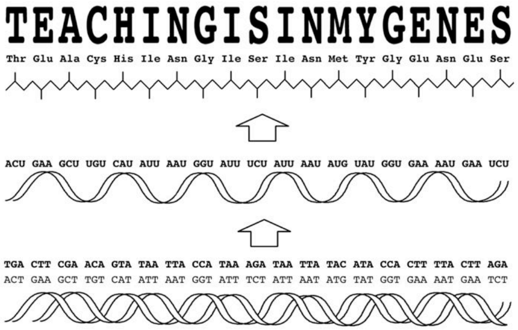
gcat.davidson.edu/sybr-u/bmc.html

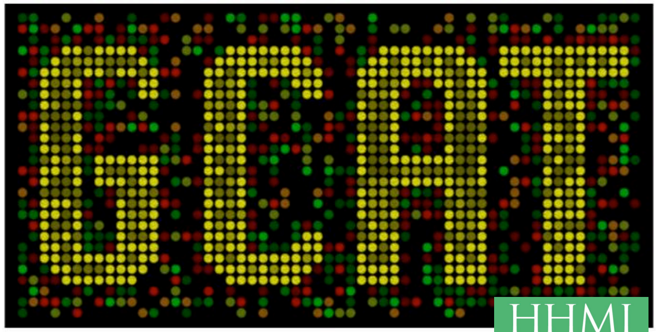
GCAT Faculty Workshop
Synthetic Biology

15 pairs of faculty

1 Bio + 1 Other

Summers 2010, 2011 + 3 More!

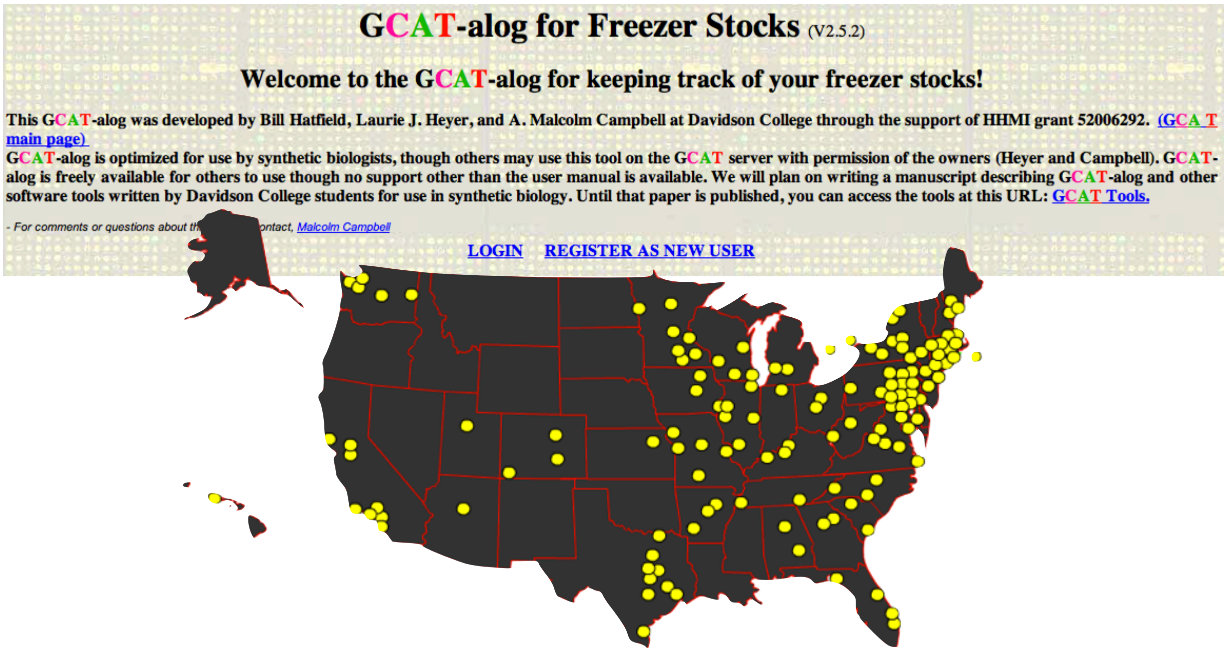




HHMI
HOWARD HUGHES
MEDICAL INSTITUTE

www.bio.davidson.edu/projects/gcat/workshop_2010/workshop_2010.html

GCAT Distributed Freezer Stocks



GCAT-alog for Freezer Stocks (V2.5.2)

Welcome to the GCAT-alog for keeping track of your freezer stocks!

This GCAT-alog was developed by Bill Hatfield, Laurie J. Heyer, and A. Malcolm Campbell at Davidson College through the support of HHMI grant 52006292. (GCAT main page)

GCAT-alog is optimized for use by synthetic biologists, though others may use this tool on the GCAT server with permission of the owners (Heyer and Campbell). GCAT-alog is freely available for others to use though no support other than the user manual is available. We will plan on writing a manuscript describing GCAT-alog and other software tools written by Davidson College students for use in synthetic biology. Until that paper is published, you can access the tools at this URL: GCAT Tools.

- For comments or questions about th ontact, Malcolm Campbell

LOGIN **REGISTER AS NEW USER**

Other Sequencing Models

www.bio.davidson.edu/courses/Bio343/LabMethods.html

Cheryl Kerfeld at Joint Genome Institute

Undergraduate Genomics Research Initiative

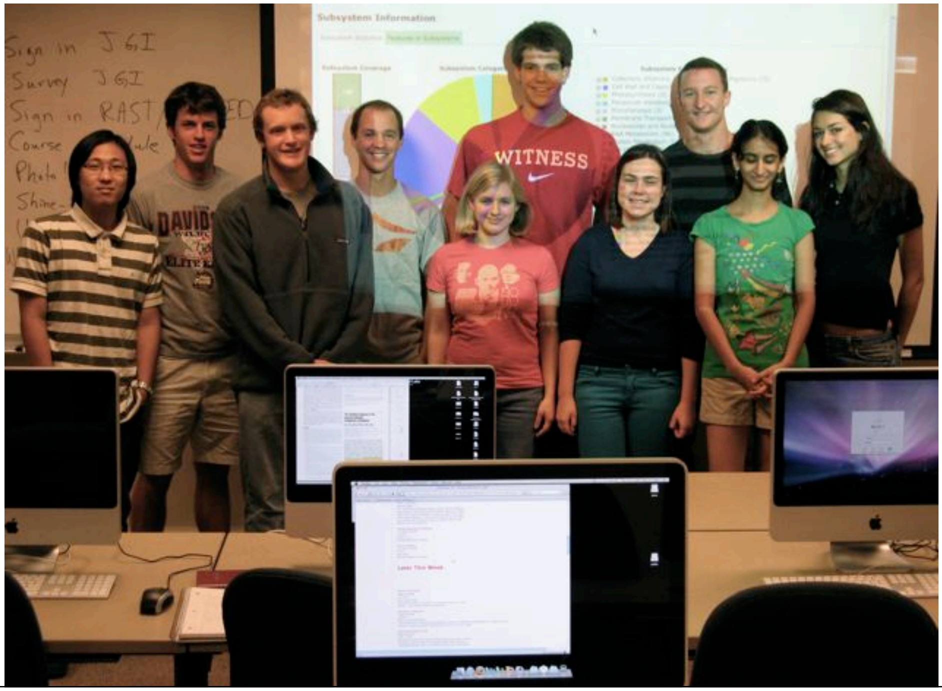
Adopt a Genome for Education

www.jgi.doe.gov/education/genomeannotation.html

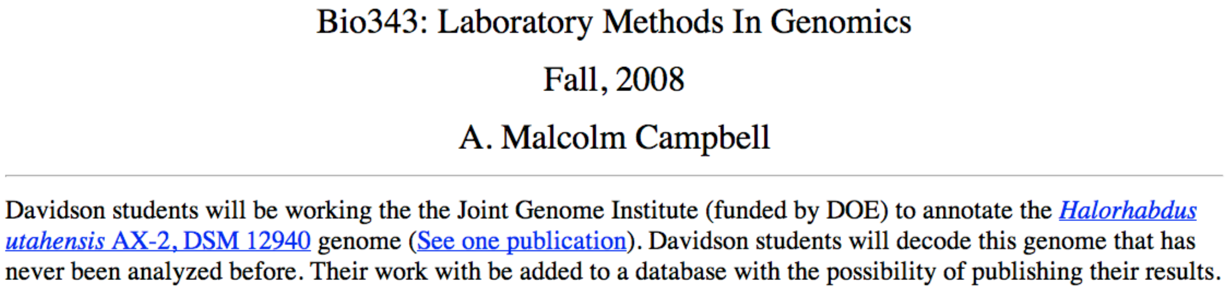
Office of Science
U.S. DEPARTMENT OF ENERGY

DOE GENOMICS:GTL
SYSTEMS BIOLOGY FOR ENERGY AND ENVIRONMENT
OFFICE OF SCIENCE
U.S. DEPARTMENT OF ENERGY

Laboratory Methods in Genomics



Laboratory Methods in Genomics



Bio343: Laboratory Methods In Genomics

Fall, 2008

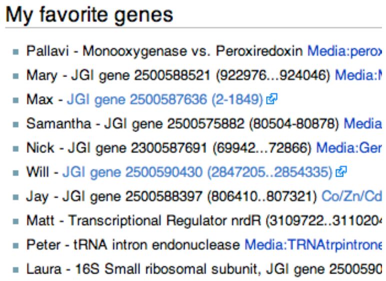
A. Malcolm Campbell

Davidson students will be working the the Joint Genome Institute (funded by DOE) to annotate the *Halorhabdus utahensis* AX-2, DSM 12940 genome (See one publication). Davidson students will decode this genome that has never been analyzed before. Their work with be added to a database with the possibility of publishing their results.

Learning Outcomes

Understand what a gene is through in depth analysis of a genome.
Determine how genomes are organized.
Generate species-specific metabolic maps.
Recognize automated annotation is imperfect & judgment calls are necessary.
Evaluate evolutionary paths as revealed in novel genomes.
Gain a real research experience and all that comes with it.
Develop computer skills used in modern genomics.
Excel in collaborative learning and research.

Laboratory Methods in Genomics



My favorite genes

- Pallavi - Monooxygenase vs. Peroxiredoxin Media:perox
- Mary - JGI gene 2500588521 (922976...924046) Media:
- Max - JGI gene 2500587636 (2-1849)
- Samantha - JGI gene 2500575882 (80504-80878) Media
- Nick - JGI gene 2300587691 (69942...72866) Media:Ger
- Will - JGI gene 2500590430 (2847205..2854335)
- Jay - JGI gene 2500588397 (806410..807321) Co/Zn/Cd
- Matt - Transcriptional Regulator nrdR (3109722..311020
- Peter - tRNA intron endonuclease Media:TRNAtrpintrone
- Laura - 16S Small ribosomal subunit, JGI gene 2500590

My Favorite Pathways

Pallavi - Carbohydrate Metabolism, specifically glycolysis/gl

Jay - Media:Jay's_Favorite_Pathway.ppt

Will - RBS Consensus and Alternative Start Codons

Max -RBS/Shine-Dalgarno Part B

Peter Bakke - Origin of Replication

Samantha - Purine Metabolism!!! Media:Purines.ppt

Laura - Amino Acid Biosynthesis

Nick - Pentose Phosphate Pentose Phosphate Pathway

Matt - Chitin Metabolism Media: ChitinMetabolism.ppt

Mary - Citric acid cycle Media: Citric acid cycle.ppt

Malcolm - protein export Protein Secretion

My Favorite Term Paper

Pallavi - Media:genomicsfinalpallavi.pdf

Samantha - Media:GenomicsLabFinalSS.pdf

Mary - Media:Gearing final paper.doc

Nick - Media: Nick Carney Final Paper.pdf

Peter - Media:GenomicsBakke.pdf

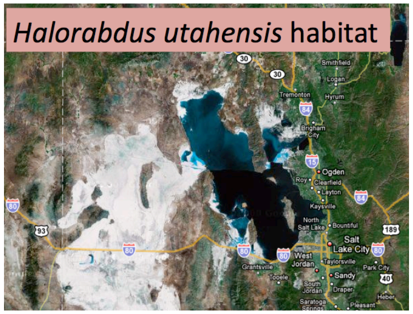
Matt - Media:LotzFinalPaper2.pdf

Will - Final Paper

Max -Final Paper

Jay - Media:McNairFinalPaper.pdf

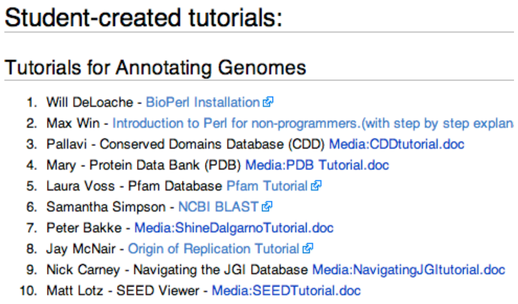
Laura - Voss_Paper.doc



Halorabdus utahensis habitat

http://gcat.davidson.edu/GcatWiki/index.php/Halorhabdus_utahensis_Genome

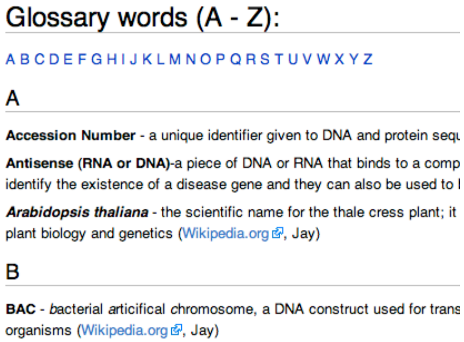
Laboratory Methods in Genomics



Student-created tutorials:

Tutorials for Annotating Genomes

1. Will DeLoache - BioPerl Installation
2. Max Win - Introduction to Perl for non-programmers.(with step by step explan
3. Pallavi - Conserved Domains Database (CDD) Media:CDDtutorial.doc
4. Mary - Protein Data Bank (PDB) Media:PDB Tutorial.doc
5. Laura Voss - Pfam Database Pfam Tutorial
6. Samantha Simpson - NCBI BLAST
7. Peter Bakke - Media:ShineDalgarnoTutorial.doc
8. Jay McNair - Origin of Replication Tutorial
9. Nick Carney - Navigating the JGI Database Media:NavigatingJGItutorial.doc
10. Matt Lotz - SEED Viewer - Media:SEEDTutorial.doc



Glossary words (A - Z):

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

A

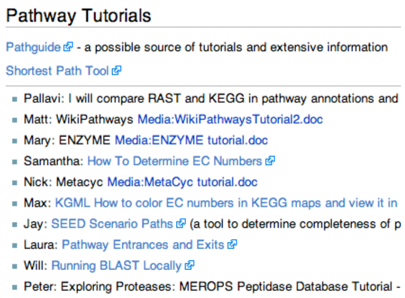
Accession Number - a unique identifier given to DNA and protein sequ

Antisense (RNA or DNA)-a piece of DNA or RNA that binds to a comp
identify the existence of a disease gene and they can also be used to l

Arabidopsis thaliana - the scientific name for the thale cress plant; it
plant biology and genetics (Wikipedia.org, Jay)

B

BAC - *b*acterial *a*rtificial *c*hromosome, a DNA construct used for trans
organisms (Wikipedia.org, Jay)

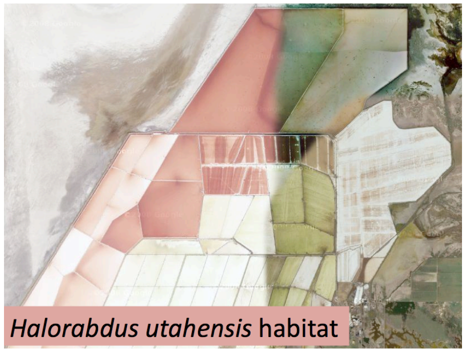


Pathway Tutorials

Pathguide - a possible source of tutorials and extensive information

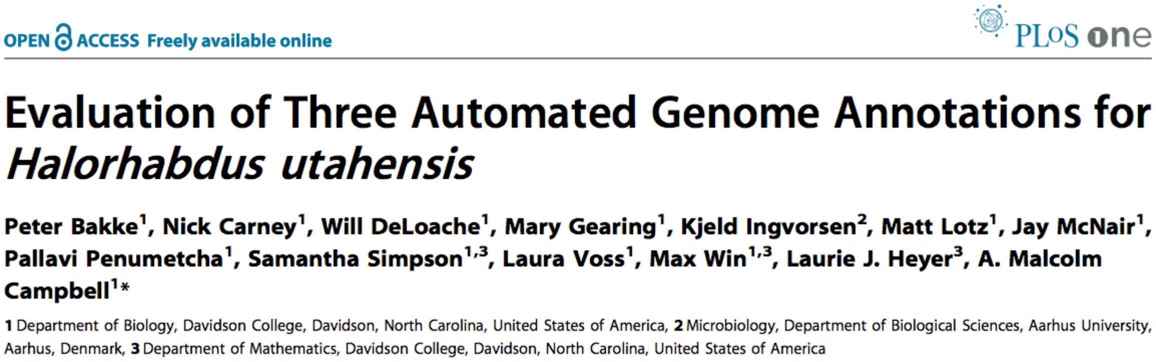
Shortest Path Tool

- Pallavi: I will compare RAST and KEGG in pathway annotations and
- Matt: WikiPathways Media:WikiPathwaysTutorial2.doc
- Mary: ENZYME Media:ENZYME tutorial.doc
- Samantha: How To Determine EC Numbers
- Nick: Metacyc Media:MetaCyc tutorial.doc
- Max: KGML How to color EC numbers in KEGG maps and view it in
- Jay: SEED Scenario Paths (a tool to determine completeness of p
- Laura: Pathway Entrances and Exits
- Will: Running BLAST Locally
- Peter: Exploring Proteases: MEROPS Peptidase Database Tutorial -



Halorabdus utahensis habitat

Undergraduates Publish Results 7/09



OPEN ACCESS Freely available online

PLoS one

Evaluation of Three Automated Genome Annotations for *Halorhabdus utahensis*

Peter Bakke[1], Nick Carney[1], Will DeLoache[1], Mary Gearing[1], Kjeld Ingvorsen[2], Matt Lotz[1], Jay McNair[1], Pallavi Penumetcha[1], Samantha Simpson[1,3], Laura Voss[1], Max Win[1,3], Laurie J. Heyer[3], A. Malcolm Campbell[1]*

1 Department of Biology, Davidson College, Davidson, North Carolina, United States of America, **2** Microbiology, Department of Biological Sciences, Aarhus University, Aarhus, Denmark, **3** Department of Mathematics, Davidson College, Davidson, North Carolina, United States of America

10 undergraduate coauthors, written by students

4 out of 5 star rating by independent reader

www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0006291

Limits of JGI

- no publication process
- no public database connectivity
- metagenomics limited access

Office of Science
U.S. DEPARTMENT OF ENERGY

DOE GENOMICS:GTL
SYSTEMS BIOLOGY FOR ENERGY AND ENVIRONMENT
OFFICE OF SCIENCE
U.S. DEPARTMENT OF ENERGY

Adopt a Genome for Education

Sarah Elgin at Washington University

Genome **E**ducation **P**artnership

http://gep.wustl.edu/

- Students finish and annotate genome sequences
- Support staff online
- Free workshops in St. Louis
- Growing number of schools participating

Limits of GEP

Home > Community > GEP Publications

Publications

The GEP has published two peer reviewed scientific articles and two education research articles. Names of undergraduate authors are underlined.

- Slawson EE, Shaffer CD, Malone CD, Leung W, Kellmann E, Shevchek RB, Craig CA, Bloom S, Bogenpohl J II, Dee J, Morimoto ETA, Myoung J, Nett AS, Ozsolak F, Tittiger ME, Zeug A, Pardue ML, Buhler J, Mardis E, and Elgin SCR., Comparison of dot chromosome sequences from *D. melanogaster* and *D. virilis* reveals an enrichment of DNA transposon sequences in heterochromatic domains, 2006. Genome Biology, 7:R15. **(Full Text)**
- Lopatto D, Alvarez C, Barnard D, Chandrasekaran C, Chung H-M, Du C, Eckdahl T, Goodman AL, Hauser C, Jones CJ, Kopp OR, Kuleck GA, McNeil G, Morris R, Myka JL, Nagengast A, Overvoorde PJ, Poet JL, Reed K, Regisford G, Revie D, Rosenwald A, Saville K, Shaw M, Skuse GR, Smith C, Smith M, Spratt M, Stamm J, Thompson JS, Wilson BA, Witkowski C, Youngblom J, Leung W, Shaffer CD, Buhler J, Mardis E, Elgin SCR., Education Forum: Genomics Education Partnership, 2008, Science 322, 684-5. **(Full Text)**
- Shaffer CD, Alvarez C, Bailey C, Barnard D, Bhalla S, Chandrasekaran C, Chandrasekaran V, Chung HM, Dorer DR, Du C, Eckdahl TT, Poet JL, Frohlich D, Goodman AL, Gosser Y, Hauser C, Hoopes LL, Johnson D, Jones CJ, Kaehler M, Kokan N, Kopp OR, Kuleck GA, McNeil G, Moss R, Myka JL, Nagengast A, Morris R, Overvoorde PJ, Shoop E, Parrish S, Reed K, Regisford EG, Revie D, Rosenwald AG, Saville K, Schroeder S, Shaw M, Skuse G, Smith C, Smith M, Spana EP, Spratt M, Stamm J, Thompson JS, Wawersik M, Wilson BA, Youngblom J, Leung W, Buhler J, Mardis ER, Lopatto D, Elgin SCR., The Genomics Education Partnership: Successful Integration of Research into Laboratory Classes at a Diverse Group of Undergraduate Institutions, 2010, CBE Life Sci Educ 9(1): 55-69. **(Full Text)**
- Leung W, Shaffer CD, Cordonnier T, Wong J, Itano MS, Slawson Tempel EE, Kellmann E, Desruisseau DM, Cain C, Carrasquillo R, Chusak TM, Falkowska K, Grim KD, Guan R, Honeybourne J, Khan S, Lo L, McGaha R, Plunkett J, Richner JM, Richt R, Sabin L, Shah A, Sharma A, Singhal S, Song F, Swope C, Wilen CB, Buhler J, Mardis ER, Elgin SCR., Evolution of a Distinct Genomic Domain in Drosophila: Comparative Analysis of the Dot Chromosome in *Drosophila melanogaster* and *Drosophila virilis*, 2010. Genetics, Vol. 185, 1519-1534. **(Full Text)**

© Copyright 2006-2011, Washington University in St. Louis. **Legal Notices** Last Update: 06/16/2011.

- web page and 2 publications
- limited by species
- VERY resource intensive
- how often can you publish this?

Phage Genome Initiative
Science **E**ducation **A**lliance

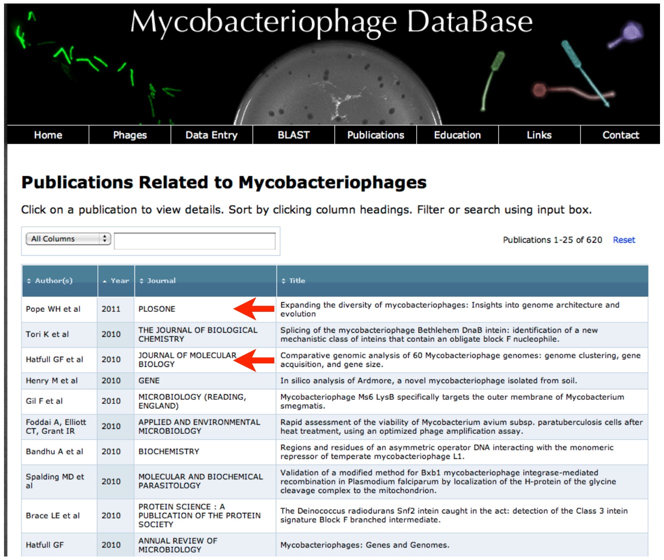
HHMI
HOWARD HUGHES
MEDICAL INSTITUTE

- Students isolate phage
- Students purify phage DNA; Sequenced at JGI
- Students annotate and compare genomes
- National experiment to examine phage variation
- Free workshop and reagents

www.hhmi.org/grants/sea/

science education alliance
HOWARD HUGHES MEDICAL INSTITUTE

Limits of SEA



Publications Related to Mycobacteriophages

Click on a publication to view details. Sort by clicking column headings. Filter or search using input box.

Author(s)	Year	Journal	Title
Pope WH et al	2011	PLOSONE	Expanding the diversity of mycobacteriophages: Insights into genome architecture and evolution
Tori K et al	2010	THE JOURNAL OF BIOLOGICAL CHEMISTRY	Splicing of the mycobacteriophage Bethlehem DnaB intein: identification of a new mechanistic class of inteins that contain an obligate block F nucleophile.
Hatfull GF et al	2010	JOURNAL OF MOLECULAR BIOLOGY	Comparative genomic analysis of 60 Mycobacteriophage genomes: genome clustering, gene acquisition, and gene size.
Henry M et al	2010	GENE	In silico analysis of Ardmore, a novel mycobacteriophage isolated from soil.
Gil F et al	2010	MICROBIOLOGY (READING, ENGLAND)	Mycobacteriophage Ms6 LysB specifically targets the outer membrane of Mycobacterium smegmatis.
Foddai A, Elliott CT, Grant IR	2010	APPLIED AND ENVIRONMENTAL MICROBIOLOGY	Rapid assessment of the viability of Mycobacterium avium subsp. paratuberculosis cells after heat treatment, using an optimized phage amplification assay.
Bandhu A et al	2010	BIOCHEMISTRY	Regions and residues of an asymmetric operator DNA interacting with the monomeric repressor of temperate mycobacteriophage L1.
Spalding MD et al	2010	MOLECULAR AND BIOCHEMICAL PARASITOLOGY	Validation of a modified method for Bxb1 mycobacteriophage integrase-mediated recombination in Plasmodium falciparum by localization of the H-protein of the glycine cleavage complex to the mitochondrion.
Brace LE et al	2010	PROTEIN SCIENCE : A PUBLICATION OF THE PROTEIN SOCIETY	The Deinococcus radiodurans Snf2 intein caught in the act: detection of the Class 3 intein signature Block F branched intermediate.
Hatfull GF	2010	ANNUAL REVIEW OF MICROBIOLOGY	Mycobacteriophages: Genes and Genomes.

- uncertain future
- limited species
- must follow their model
- publications into the future?

science education alliance
HOWARD HUGHES MEDICAL INSTITUTE

Lessons & Advice

- Define network’s **goals and limits** first
- Get a good name and logo - branding is important
- Conduct assessment each year and share results
- Provide faculty training and target MSIs (haves and have nots)
- Allow academic freedom - support THEIR goals
- Get publicity and publish meetings, pedagogy and research
- Build web page that provides resources and FAQs
- Distribute network’s responsibilities - redundancies are good
- Adapt to community needs but don’t compromise goals
- Get clerical and logistical support for leadership
- Mike - make sure you are still enjoying this
- Each year, ask whether you network should continue or not
- Be careful what you wish for
- Naiveté can carry you a long way

Questions for You

- Will you want/require ESTs? If so, how many per project?
- Is automated annotation sufficient? Which program(s) will you use?
- How much manual annotation do you expect/want?
- Will you require finishing? Block finishing?
- How will future projects be funded?
- Who determines which projects are selected?
- What is transformative about these projects?
- Will you allow any one with $ to use your consortium?
- Will every project be independent, or will you have a central repository/interface?
- When will the sequence be made public?
- Can high schools participate? Graduate students?

Acknowledgements

Faculty: Laurie Heyer, Jeff Poet, Todd Eckdahl, Karmella Haynes, Pat Sellers, Mark Barsoum

Students: Romina Clemente, Clif Davis, A.J. Grant, Mary Gearing, Kin Lau, Olivia Ho-Shing, Shamita Punjabi, Eric Sawyer, Ashley Schooner, Siya Sun, Shashank Suresh, Bryce Szczepanik, Leland Taylor, Annie Temmink, Alyndria Thompson, Will Vernon, Oyinade Adefuye, Will DeLoache, Jim Dickson, Andrew Martens, Amber Shoecraft, Mike Waters, Jordan Baumgardner, Tom Crowley, Lane Heard, Nick Morton, Michelle Ritter, Karen Acker, Bruce Henschen, Jessica Treece, Matt Unzicker, Amanda Valencia, Lance Harden, Sabriya Rosemond, Samantha Simpson, Erin Zwack, Marian Broderick, Adam Brown, Trevor Butner, Lane Heard, Eric Jessen, Kelley Malloy, Brad Ogden, Kelly Davis, Alicia Allen, James Barron, Robert Cool, Kelly Davis, Will DeLoache, Erin Feeney, Andrew Gordon, John Igo, Aaron Lewis, Kristi Muscalino, Madeline Parra, Pallavi Penumetcha, Karlesha Roland, Max Win, Xiao Zhu, Kristen DeCelle, Matt Gemberling, Oscar Hernandez, Andrew Drysdale, Nick Cain, Tamar Odel, and Jackie Ryan.

The Duke Endowment, NSF, HHMI
Genome Consortium for Active Teaching (**GCAT**)
Davidson College James G. Martin Genomics Program
MWSU SGA, Foundation & Summer Research Institute