DUE-0836869

Integration of Bioinformatics into a Biology Curriculum

A Course, Curriculum and Laboratory Improvement Grant from the National Science Foundation to Muhlenberg College

Project Director:
Dr. Bruce Wightman
Biology Department
Muhlenberg College
2400 Chew Street
Allentown, PA 18104

wightman@muhlenberg.edu

484-664-3254

Co-PIs:
Dr. Amy Hark
Dr. Keri Colabroay
Dr. Marten Edwards
Dr. Clifton Kussmaul

PROJECT OVERVIEW

The proposed project will improve education in quantitative aspects of biology by enhancing bioinformatics instruction at multiple levels of the curriculum into existing biology and biochemistry courses. The goals of the project will be supported by the addition of a core genomics, proteomics, and bioinformatics laboratory, which will include real-time PCR and 2-D gel electrophoresis equipment and computers. Students will learn basic computational techniques in an introductory biology course, and will then build on that basis by participating in multiple week-long investigative laboratory experiences in intermediate and advanced level courses. Faculty development, in the form of a local bioinformatics course for biology, biochemistry, and mathematics faculty, will help to increase faculty sophistication and collaboration.

RESULTS FROM PRIOR NSF SUPPORT

Bruce Wightman is the PI on a current NSF Research at Undergraduate Institutions (RUI) grant, IOS-0640483, entitled "RUI: Genetic Analysis of the *C. elegans* Tailless Gene *nhr-67*." The total amount of the award is $370,000, with a duration of March 1, 2007 through February 28, 2010. The project focuses on a molecular genetic analysis of the role of the *nhr-67* gene of *C. elegans* in regulating the development of the uterus. Results to date strongly support our hypothesis that *nhr-67*, which is the ortholog of the evolutionarily-conserved *tailless* gene, functions downstream of a *Notch*-based signaling pathway to specific execution of particular uterine cell fates. Current student projects focus determining the nature of hypomorphic promoter mutations and testing cell-specific function of *nhr-67*. This project is an extension of a previously funded project, IBN-0234716, entitled "RUI: Function of NR2E Nuclear Receptors in *C. elegans*" (August 2003- July 2007). To date we have published two papers based on this work, both with undergraduate co-authors (underlined):

Wightman, B., N. Carmean, B. Ebert, K. Weber, and S. Clever, 2005, The *C. elegans* nuclear receptor gene *fax-1* and homeobox gene *unc-42* coordinate interneuron identity by regulating the expression of glutamate receptor subunits and other neuron-specific genes, *Developmental Biology*, **287**: 74-85.

DeMeo, S., Lombel, R., Snowflack, D., Smith, E., Reinert, K., Cronin, M., Clever, S., and B. Wightman, 2008, Specificity of DNA-binding by the FAX-1 and NHR-67 nuclear receptors of *Caenorhabditis elegans* is partially mediated via a subclass-specific P-box residue, *BMC Molecular Biology*, **9**:2.

Marten Edwards is a co-PI on an NSF C-RUI grant, CRUI-0442049, entitled “Ecological determinants of a plant-insect interaction.” The project is directed by Richard Niesenbaum of the Muhlenberg Biology Department and funds a total of $1,060,000 from 2004-2009. The overall objective of this project is to determine how abiotic factors and plant genotype interact to influence patterns of herbivory in a natural system consisting of the forest understory plant, *Lindera benzoin* and its primary lepidopteran herbivore, *Epimecis hortaria.* The project is a collaborative effort involving a plant ecologist, analytical chemist, computer scientist and an insect molecular biologist. It involves 11-16 undergraduates each summer and has already culminated in a total of five peer-reviewed publications. Dr. Edward’s role in the project has been to develop informative microsatellite loci for spicebush plants and use them to determine

the genetic structure of spicebush populations. The results of Dr. Edwards component of the project were published in 2007, with a second paper currently in preparation:

Edwards M.J. and Niesenbaum R.A., 2007, Eleven polymorphic microsatellite loci in *Lindera benzoin,* Lauraceae. *Molecular Ecology Notes* **7**: 1302–1304.

Bruce Wightman was also a co-PI on a successful CCLI grant that relates more directly to the proposed project. DUE-0126027, entitled “Integrating Technology into Introductory Biology Laboratories," was written in collaboration with Elizabeth McCain and Chrysan Cronin of the Muhlenberg Biology Department. The grant awarded $112,195 over the period June 1, 2002 to July 31, 2004. The central goal of this initiative was to expand the use of computer technology and the culture of science to students enrolled in introductory biology laboratories. These courses introduce the students to all elements of scientific inquiry through investigative exercises and doing independent research. The students used computers, software, and peripherals to 1) record data from their short-term investigative and long-term independent experiments, 2) do statistics and graphing, 3) place data on the web and use web-based technology to access molecular databases, and 4) present their long-term projects in the format of a poster and/or oral presentation. These goals were achieved by purchasing laptop computers, statistical software, digital cameras, Vernier gas and heart monitors and associated peripherals to support investigative projects in a traditional introductory laboratory setting.

The 2002-2004 CCLI project was implemented successfully during the grant period. Although not part of the original proposal, project instruments were also employed by Mary Byrne of the Muhlenberg Biology Department for a two-day outreach project that helped to train 50 Allentown City School District teachers in science instruction. Assessment of student outcomes were designed and administered by Elizabeth McCain and the data evaluated with the statistical assistance of Laura Snodgrass, Professor of Psychology at Muhlenberg. The indirect assessment tools demonstrated a significant increase in student confidence in learning specific content and skills over the course of the both semesters. The effect was most pronounced on concepts and skills that students were least confident of at the beginning of the course, indicating the general soundness of the approach. Student responses on course evaluations supported the conclusions that the technology added to the laboratory aided their learning and improved their scientific self-confidence. The results of this study were disseminated via the Online Evaluation Resource Library (http://oerl.sri.com/).

PROJECT DESCRIPTION

The twentieth century saw the fractionation of the natural sciences in general, and biology in particular, into multiple disciplines that have allowed for a progressively more detailed and sophisticated understanding of the mechanisms of natural processes (Mayr, 1982; Wilson, 1998). However, this disciplinary emphasis has come at a cost; it has become increasingly difficult for scientists to communicate with each other across disciplinary boundaries (Bialek and Botstein, 2004). In particular, many biologists lack a sophisticated understanding of the mathematical algorithms that power important computational tools they use (May, 2004). Over the last decade, many scientists and educators have begun to emphasize the need for integration among the natural sciences and mathematics (Gazzaniga, 1998; Bio2010,

2003). In highly interdisciplinary fields such as genomics, the need for effective integration of biology and mathematics is especially acute. This level of integration demands that scientists of the future are able to synthesize ideas and approaches from different disciplines. This is also true for physicians, who face an increasingly quantitative and technologically complex profession. Moreover, an emphasis on synthesis in science has been appreciated as effective pedagogy: students learn better if they can apply familiar concepts to various problems introduced within the context of different disciplines (Garfalo and LoPresti, 1993; Dahir, 1995; Anderson, 2007). By establishing connections between mathematics and biology, students' understanding of each discipline is improved.

Bioinformatics and genomics education provide an ideal opportunity to emphasize quantitative approaches to solving biological problems (Altman, 1998; Jackson, 2005; Campbell et al., 2007a). Bioinformatics, in particular, has grown out of the combined efforts of computational scientists and geneticists. Thus these disciplines provide an opportunity for students to explore the more familiar qualitative aspects of biology that interest most of them, such as cancer and human disease, in the context of contemporary computational approaches.

Jackson (2005) has outlined the challenges and opportunities for undergraduate education in bioinformatics. Among his many suggestions are: 1) developing introductory biology courses that feature mathematical applications, 2) introducing substantially more mathematics into existing biology courses, 3) offer a computer-based course as an applied laboratory in bioinformatics, and 4) invest in faculty development and interdisciplinary collaboration. Our proposed project seeks to implement these suggestions while simultaneously improving undergraduate education in genomics by systematically introducing bioinformatics into the undergraduate biology and biochemistry curricula at multiple levels. This approach is an adaptation of a similar approach of "bioinformatics across the curriculum" taken by the University of Wisconsin- La Crosse (Hydorn et al., 2005); instead of creating specialized "bioinformatics" courses that will attract only a small number of students, quantitative approaches and bioinformatics are systematically woven into multiple biology courses.

Students at Muhlenberg College will be introduced to central computational and genomic concepts such as databases, search algorithms and microarray analysis in introductory biology. Sophomores and Juniors will apply these core concepts and applications as they use the resources of a new genomics lab facility in intermediate and upper-level elective lab courses. Students in new advanced genomics and biochemistry lab courses will explore more sophisticated applications, including computational phylogenetics, DNA microarray projects, computer structural modeling, and proteomics. Finally, faculty development will help advance the mathematical sophistication of existing biology faculty and involve a computer scientist in bioinformatics applications and instruction.

Biology and Mathematics at Muhlenberg College

Founded in 1848, Muhlenberg College is an independent liberal arts college of 2221 students located in Allentown, PA. The mission of the college is to "develop independent critical thinkers who are intellectually agile…and prepared for lives of leadership and service" (Mission Statement). This goal is accomplished in part by integrating a traditional liberal arts curriculum with programs in pre-professional study, including pre-medical, business, and education. In addition, the College mission emphasizes the importance of "… educating the whole person

through experiences within and beyond the classroom," in part via the "… pedagogical and intellectual importance of research." The College has embraced this mission by investing heavily in the natural sciences and mathematics. Over the past three years, the College has completed the construction of a new 50,000 sq. ft. science facility and the complete renovation of the existing 36,400 sq. ft. 1970's era biology building at a cost of over $27 million. Furthermore, three additional new tenure-track faculty positions in the natural sciences and mathematics have been approved in the past two years.

The Departments of Biology, Chemistry, and Mathematics and Computer Science offer academic programs in biochemistry, biology, chemistry, computer science, environmental science, neuroscience, and mathematics. The strength of the natural science program at Muhlenberg is revealed in part by the popularity of natural science majors. In 2006, there were 272 students majoring in one of the natural science disciplines (24% of all declared majors), with biology and neuroscience constituting 13% of all student majors (Fall 2006 statistics). In 2007, biology was the third most popular major among graduates, and neuroscience was the fastest-growing major on campus. The current curriculum offers a rigorous introduction to the sciences, with ample opportunity for upper-level exploration in the student's major field.

The last five years have also seen the development of new and revised interdisciplinary science programs. In 2004, a new Neuroscience major was approved that includes required coursework in biology, chemistry, mathematics, philosophy and psychology. In 2006, the College approved a revised interdisciplinary Biochemistry major with two co-directors, one from the Biology Department and one from the Chemistry Department. These new curricular efforts are among the first steps toward increasing the interdisciplinary nature of instruction and institutional organization.

Faculty and student research activity has also reached new levels over the past five years. Muhlenberg science faculty members have received over $1.9 million in federal and corporate research funding between 2002 and 2007. These efforts have led to faculty publications, some with undergraduate student co-authors, in major peer-reviewed journals, including *Journal of Biology*, *Journal of Biological Chemistry*, *BMC Molecular Biology*, *Developmental Biology*, *Oecologia*, and *Molecular Pharmacology.* In total, the 30 tenure/tenure-track faculty members of the natural science and mathematics departments have published 89 peer-reviewed research papers over the last five years, which included 37 undergraduate co-authors. Some of these activities have been interdisciplinary in nature and have involved quantitative approaches to studying biology. Richard Niesenbaum of the Biology Department was awarded a $1,060,000 grant from the NSF (2005-2009) under the C-RUI (cross-disciplinary research at undergraduate institutions) program, one of only four awards made nationally under the C-RUI program that year. This research program links ecology, leaf chemistry and genetics to study plant herbivory and is a collaborative effort with Marten Edwards of the Biology Department, Christine Ingersoll of the Chemistry Department, and George Benjamin of the Mathematics Department.

As is the case at many liberal arts colleges, a significant fraction (over 75%) of first-year science students at Muhlenberg express an interest in pursuing a clinical career. The typical first year pre-med student is highly motivated and bright, but relatively unsophisticated about science. They are also often wary of mathematics, and may sometimes direct their studies toward more qualitative scientific subjects such as anatomy. Thus a major curricular challenge is to bring mathematics into the biology classroom and laboratory such that students can continuously engage in mathematical thinking within the context of learning about biology. Introductory

textbooks generally do not do a good job of bringing much mathematical rigor to the subject, leaving the task largely to individual instructors and individual programs.

The Biology, Biochemistry, and Neuroscience majors at Muhlenberg all require 1-2 semesters of Calculus and completion of the three-course core *Principles of Biology I-III* (BIO 150, 151, and 152), in addition to cognates in chemistry and physics. The BIO 150/151/152 core sequence is organized around a "big to small" reductionist sequence: the first semester covers ecology, evolution and diversity, the second semester organismal biology, and the last semester cell and molecular biology. This sequence requires that students master the basics of chemistry and mathematics (through cognate courses) before attempting molecular biology. While this sequence is the reverse of that found in most introductory textbooks, we have found that it works quite well, is popular with students, and allows the third semester to explore molecular biology at much higher level of complexity than would otherwise be possible. After completion of the core sequence, students are free to choose among various upper-level elective courses. All three majors also require capstone courses for juniors and seniors that are integrative, seminar-style, and primary literature-based. Course descriptions can be found as webpages citations in the References section of this proposal.

Specific Aim 1: Introduce quantitative approaches to genomics in introductory biology

We will improve quantitative integration into the introductory biology curriculum by expanding the bioinformatics component of BIO 152, the third course in the core biology sequence. BIO 152 is offered every fall semester and has three components: lecture, recitation, and lab. Students enroll in one of two lecture sections of 35-50 students per section, taught in rotation by co-PIs Bruce Wightman, Amy Hark, and Marten Edwards. Students concurrently enroll in one of six "wet" laboratory sections of approximately 16 students per section (three hours per week). The students also must enroll in one of six recitation sections of 10-18 students per section (one hour per week), taught by the lecture instructors. These recitation sections provide an opportunity for problem-solving, group projects, discussion, and peer-tutoring. This proposal will focus on expanding the bioinformatics experience in recitation sections.

The instructors of BIO 152 have used different approaches to teaching basic bioinformatics over the past six years. The first iteration, initiated in 2003, utilized a web-based assignment designed by a former student that led students through a series of didactic exercises via NCBI's utilities during lab sections. While the approach was successful in that it introduced students to the NCBI interface and the general nature of genomic databases, many students also complained that it was too directed and did not provide them with much a sense of what is going on "under the hood." In addition, the need for more time in the "wet lab" for bench work encouraged the instructors to move the bioinformatics component from the lab to recitation sections. We used Young (2008) as a starting place to develop a series of recitation exercises that led students through the basics of databases (Genbank, Prosite), information retrieval (OMIM, PubMed, FASTA format, structural/PDB Viewer) and local alignment (BLAST). In the course of this analysis, students learn about alignment scores and E values, realizing some of the quantitative goals of Jackson (2005). A major limitation to the recitation experience is that students do not necessarily have their own laptop computer. Therefore, time in recitation is limited to instructor demonstration, with most of the active learning taking place while students work on assignments on their own. Students reported on course evaluations that the experience

was empowering and helped to improve their sense of "what a gene is" and "what protein domains look like" (Griffin et al., 2003). We have not yet formally assessed the contribution of these experiences to their sense of comfort with mathematics in biology, but will do so as part of this proposed project (see *Project Assessment and Dissemination*, below).

For this CCLI proposal, we will build on this initiative in BIO 152 through three major project objectives: 1) adding microarray data analysis to the bioinformatics exercises, 2) providing enough laptop computers for all students in a recitation section to have their own, and 3) creation of a local NCBI web portal to provide a side window of help information to improve the accessibility of NCBI utilities.

The BIO 152 bioinformatics curriculum currently emphasizes comparisons *across* genomes (BLAST). We propose to expand the experience to have students look *within* an entire genome by assessing genome-wide transcriptional outputs. While experience with an actual "wet-lab" microarray experiment would be ideal, time and cost make this prohibitive for such a large cohort. Instead, we will adapt a scenario for investigating the role of genes in cancer to introduce students to microarray technology using existing microarray data (Campbell, 2006). The new BIO 152 microarray component will begin with an overview of the basics of two-color microarray technology (Campbell et al., 2007a). In recitation sections, students will utilize sample data from existing microarray expression data using one of several methods developed through the Genome Consortium for Active Teaching (GCAT; Campbell et al., 2007a, 2007b). The focus of the students' work will be analyzing and interpreting the data. They will determine expression ratios, explore the value of log transformation and other data manipulation, and consider the variability observed in their simulated experiment (Heyer and Campbell, 2006). Thus the focus of the student experience in recitation will be on the quantitative methods to evaluate data, rather than the technical details of how microarrays are constructed and probed (the basics of these concepts are covered in lecture sections). While the exercise will be directed, to the extent that students will be led toward quantitative aspects of the data analysis, it will be sufficiently open-ended to allow students to make choices about which data sets to consider and which choices to make when evaluating their data. Thus we aim to strike a balance between leading them to specific pedagogical goals and giving them the freedom to explore the databases.

The availability of one laptop computer per student in each recitation section will allow us to quickly and easily get students working on bioinformatics projects via the NCBI website and the new microarray analysis exercise. In particular, microarray data analysis software such as MAGIC Tool (Heyer et al., 2005) have steep memory requirements and must be installed on each CPU, making reliance on students' own computers impractical. Each laptop computer will connect with the internet via Ethernet cable ports or the wireless system active throughout the new science facilities. The computers may also be used in laboratory sections of BIO 151 and 152, as needed.

The final new component of bioinformatics education for BIO 152 will be the development of a new local web portal for NCBI. As is the case for many new skills, students may master them at the introductory level, but then have difficulty recalling what they learned a year or two later when it is needed in an upper-level course. They may become confused when they return to a website many months after learning it only to discover the design of the webpage has changed (NCBI updates its webpages frequently). This resource will provide a simple and direct avenue for basic bioinformatics investigations by students in BIO 152 recitations and be useful as a common interface as students move into different upper-level electives. Thus they will be able to return to a familiar instructional resource when they need to utilize BLAST or

other utilities in upper-level elective courses or independent research. The local web resource will be maintained and modified as NCBI's web interface changes over time and will be accessible to anyone on the internet. The new web-based portal will be developed in conjunction with Muhlenberg Library and Office of Information Technology personnel by Amy Hark of the Biology Department.

Specific Aim 2: Quantitative real-time PCR in intermediate laboratory courses

While the introductory experience will provide students with some basic tools for inquiry in bioinformatics and genomics, our most intensive laboratory experiences come in more disciplinary, focused intermediate-level electives. Our next step in incorporating quantitative approaches to genomic study into the curriculum is the development of longitudinal, inquiry-oriented laboratory experiences for mid-level (primarily sophomore-junior) undergraduates. These laboratory experiences will take advantage of new laptop computers dedicated to these courses, and a new genomics core laboratory facility that includes a real-time PCR instrument and associated peripherals. The addition of real-time PCR technology provides a flexible platform for quantitative investigations in our 200-level lab electives, including courses in Genetics, Biochemistry, and Cell Biology. The common experience in the different courses will insure that nearly all life science majors will confront amplification curves, melting curve analysis, relative quantitations, and log transformations of data. In addition, the instrument interface provides an opportunity for students to experience simple programming. These three courses have a total annual enrollment of approximately 90 students, including all biochemistry majors, nearly all biology majors, and most neuroscience majors (68%). The consistent incorporation of this technology in all these courses insures that the vast majority of life science majors at Muhlenberg will obtain a more advanced treatment of quantitative approaches to studying gene regulation and basic bioinformatics techniques.

<u>Genetics</u>
BIO 215 (see References for web citations of course descriptions) is taught annually by Bruce Wightman; it covers Mendelian, molecular and population genetics, with an emphasis on the practice of genetics, especially gene mapping and analytical techniques. The laboratory experience focuses on a limited number of longitudinal experiments that allow students to explore different model systems and different genetic concepts. For example, students follow *Drosophila* populations over ten weeks while modeling the effect of different evolutionary forces on allele frequency. In another experimental series, students perform a mutagenesis of *C. elegans* to recover a new mutation and then map their mutation using a transposon insertion physical mapping strategy (Williams, 1992). The laboratory sequence is by its nature already very quantitative, with repeated experiences in probability (mapping and allele frequencies), applied algebra (population biology), and graphing logarithmic functions (gel electrophoresis).

The addition of real-time PCR allows for new quantitative extensions of existing laboratory experiences in genetics. In one series of experiments, students use yeast mutations in the *ade1* and *ade2* genes to explore complementation analysis. Their observations about colony color (the mutants form red colonies under certain conditions due to intermediate accumulation) lead them to propose individual experiments to test how a variable influences the formation of color (genotype-environment interactions). Students then design and execute their experiment.

Most students are led to the conclusion that the end-product (adenine) reduces the amount of red color, which allows them to connect the biochemical concept of feedback inhibition to their genetics experience. In this case, however, feedback also occurs at the level of gene regulation (Daignan-Fornier and Fink, 1992; Rebora et al., 2001). Students' exploration of the literature in conjunction with their self-designed experiment should lead most of them to recognize the need to test gene regulation in their yeast strain under different conditions (those who don't will be directed there by gentle nudging from the instructor). The addition of real-time PCR will allow them to test their hypothesis directly by measuring RNA levels in their mutant strain under different conditions. Because all students will be examining the same targets, we can use optimally-designed reagents such as *TaqMan* probes (Applied Biosystems) to minimize trouble-shooting and increase the probability of success.

In a later laboratory experience, students will use Wormbase (www.wormbase.org) and NCBI's database utilities to identify predicted nuclear hormone receptors in the genomes of non-*Caenorhabditis elegans* nematodes. Surprisingly, the genomes of *Caenorhabditis* nematodes have over 250 different predicted nuclear receptors in stark contrast to the approximately two dozen found in most animals, including the nematode *Brugia* (Sluder et al., 1999; Maglich et al., 2001; Robinson-Rechavi et al., 2005; Ghedin et al., 2007). The gene class is also very plastic evolutionarily, with about a third of *C. elegans* nuclear receptors not found in the closely-related species *C. briggsae* (and vice versa). Other studies have shown that large expanded gene families in *C. elegans* may harbor a large fraction of pseudogenes (Denver et al., 2004). Students in BIO 215 will test the possibility that some annotated nuclear receptors in the complete *C. briggsae* and *C. elegans* genomes may actually be pseudogenes. Students will use the database to identify a predicted, but unstudied, nuclear receptor from a nematode genome and design primers for PCR amplification of the predicted cDNA product. They will then use instructor-prepared mixed-stage nematode cDNA as a template for real-time PCR quantitation of transcript levels of their target gene. Products that do not show measurable expression are candidate pseudogenes. This study will allow them to apply the data analysis skills they learned earlier in the semester in quantitating yeast RNA levels by real-time PCR. In addition, students in BIO 215 will have the opportunity to explore a genuine research question, grapple with issues of primer design and optimization, and profit from the possibility of making a potential contribution to scientific knowledge.

Biochemistry

BIO 220 is taught annually by Amy Hark; it covers basic biochemical topics including structure and function of nucleic acids and proteins; an introduction to enzyme kinetics and regulation; and aspects of metabolism and signal transduction. The laboratory component of this course is designed to introduce students to experimental approaches and techniques used to assay nucleic acids and proteins in modern biochemical research, through both application (hands-on) and discussion.

The coverage of nucleic acids includes reviewing the chemistry that drives structure and considering this when discussing the many experimental applications that involve nucleic acid hybridization. The course currently employs a PCR-based laboratory (detecting genetic modification of plants through amplification of conserved regulatory sequences) as one way to stimulate students' thinking about nucleic acid polymerization and hybridization as well as gene transcription. Real-time technology would greatly enhance the mathematical treatment of the

underlying PCR reactions; this could be further extended in ways that would encourage faculty and students to work in a cross-disciplinary manner; for example, biology and mathematics students could use modeling software to illustrate aspects of the polymerase chain reaction. As another example of the potential to increase quantitative applications, in an experiment assaying food samples to detect the presence of genetic modification, students would now be able to design and conduct experiments to determine relative amounts of plant DNA vs. genetically modified organism DNA recovered rather than just assess presence or absence of signal (BIO-RAD GMO Kit Application Note). Students may also wish to design primers to assay additional sequences uncovered from the literature as being diagnostic for plant genetic modification or to perform multiplex PCR.

In an extension of this current laboratory, students would further explore plant genetic engineering through engaging in quantitation of transgene expression levels using reverse-transcriptase PCR. For example, one focus in plant biotechnology is on plants' ability to survive cold or freezing temperatures. Basic research in model species such as *Arabidopsis thaliana* has revealed the role of CBF proteins as master regulators in promoting freezing tolerance (Thomashow, 2001). Students will isolate RNA from plants endogenously expressing CBF mRNA under normal or low temperature growing conditions, from transgenic plants expressing CBF constitutively or heterologously (Jaglo-Ottosen et al. 1998), and from plants harboring mutations that might be implicated in affecting the cold response pathway (Vlachonasios et al., 2003). Students will hypothesize what they expect in terms of the relative amounts of CBF mRNA in each sample, and following reverse transcriptase reactions, real-time PCR will be utilized to quantitate differences in expression levels that relate to physiological responses. This hands-on approach will complement journal-club style discussions of primary literature in which students encounter various methods for quantitation of RNA levels.

<u>Cell Biology</u>

BIO 205 is taught annually by Marten Edwards; it focuses on cell signaling, cell cycle control, apoptosis, and cellular ultrastructure. The BIO 205 lab experience takes advantage of murine hybridoma cells. The cells are easy to culture and they strongly express an immunoglobulin that can be readily detected. Keyhole Limpet Hemacyanin (KLH) is frequently employed as a carrier protein for small or poorly immunogenic antigens. Hybridoma cells that express anti-KLH IgG antibodies are a by-product of monoclonal antibody production facilities, since mice are much more likely to mount an immune response to KLH than the molecule that it is conjugated to. Several anti-KLH secreting hybridoma cell lines have been donated to the Muhlenberg College by Lampire Biologicals Inc (Pipersville, PA).

Students in BIO 205 learn the methods of cell culture by seeding a flask with a known quantity of cells that they estimate using a hemacytomer. After each day of culture, students take an aliquot of the culture medium that contains secreted anti-KLH IgG. Students then use this medium as the "primary antibody" in an enzyme-linked immunosorbent assay (ELISA) for KLH. Due to its common use as a lab reagent, purified KLH is readily available and relatively inexpensive. Students first determine the concentration of an unknown KLH sample using the Bradford protein assay, and then calculate the concentrations of protein that need to be arrayed on a microtiter plate in preparation for the ELISA. Anti-KLH IgG is detected using an alkaline phosphatase labeled rabbit anti-mouse secondary antibody. Plates are read using a microtiter plate reader.

A major advantage of this ELISA exercise is that it is highly quantitative. Students must use a standard curve that they generate to determine the concentration of KLH and then use this information to calculate how they will set up dilutions to array onto an ELISA plate. It provides students with an opportunity to determine the exact quantities of protein, while allowing them to gain hands-on experience with cell culture methods. Students respond very well to the fact that the accuracy of their work is revealed in the form of a color reaction that can be readily quantified on a microtiter plate reader.

The addition of real-time RT-PCR would allow this experience to be taken to a deeper level, and would allow students to integrate what they learn in other courses. Methods have been developed to amplify mouse immunoglobulin variable regions using universal degenerate primers that work for all murine IgGs (Wang et al., 2000). Thus, the production of IgG by the hybridoma cells can be assayed not only at the level of protein expression (by ELISA), but also at the transcript level. Different anti-KLH hybridoma cell lines that are known to show different levels of IgG expression will be compared. RT-PCR products can be readily cloned and sequenced, and bioinformatics methods can be used to determine the differences in amino acids that are present in the variable regions of different antibodies that recognize the KLH antigen.

Specific Aim 3: Genomics and proteomics in advanced courses

Advanced courses, taken mostly by juniors and seniors, will explore more sophisticated and, in some cases, more technically demanding applications of genomics and bioinformatics. These courses will benefit from a new suite of laptop computers, shared among several classes, and from new genomics/proteomics technology including real-time PCR and 2-D gel electrophoresis. Each of these courses typically enroll 8-16 students each year.

<u>Genomes and Gene Evolution</u>

Genomes and Gene Evolution (GGE) is a new course designed and taught by Amy Hark and Bruce Wightman for the first time during the Spring 2008 semester. This upper-level, seminar-style course covers evolution and development (Evo-Devo), molecular evolution and systematics, and comparative, regulatory, and functional genomics. GGE focuses on the tools used for comparative evolutionary studies, such as global alignment, distance matrices, phylogenetic algorithms and tree-building, and for regulatory and functional genomics, especially microarrays. As such, the course involves the most intensive bioinformatics content of any course currently offered at Muhlenberg College. In the laboratory component of the course, students identify a gene family using NCBI tools including BLAST, and then perform an analysis of predicted evolutionary relationships among family members using CLUSTAL W (Thompson et al. 1994), and phylogenetic tree-building using MEGA 4.0 (Tamura et al. 2007). We discuss different algorithms for tree-building, such as UPGMA and Neighbor-joining, comparing strengths and weaknesses of each approach, and computational techniques such as bootstrapping and rooting. The second portion of the laboratory involves a microarray experiment using yeast chips supplied and scanned by GCAT at Davidson College (Campbell et al., 2007a). In these experiments students use two-color microarray analysis to compare gene expression between mutant and wild-type yeast strains, which forces them to consider experimental design and technical obstacles. Students also begin their analysis with raw data (.tif files) from scans of the microarrays and conduct data analysis using MAGIC Tool (Heyer et al.,

2005), forcing them to consider many of the parameters and choices involved in this type of data analysis. They must make decisions about log transformation of data, background subtraction, and other issues that require quantitative analysis of how data manipulation influences results.

We propose to support this course by: 1) providing laptop computers with sufficient memory for each student, thus overcoming a major logistical and technical hurdle, 2) bringing computer scientist Clifton Kussmaul into the course for a guest lecture on algorithms in bioinformatics, and 3) the use of real-time PCR to confirm expression analysis from microarray experiments. The first iteration of this course depended on students own laptop computers and additional laptops borrowed from the Computer Science Department. While the Biology Department does own a suite of laptop computers, they are several years old and lack sufficient memory and processor speed to run software for this course. Clifton Kussmaul, Associate Professor of Computer Science, will contribute to the next iteration of the course by providing a guest lecture on algorithms, with a specific focus on how BLAST works, why heuristics are employed, and related topics. Real-time PCR will allow students in the course to confirm one of the genes identified as differentially-regulated by two color microarray analysis.

Experimental Biochemistry

BCM 341, *Experimental Biochemistry*, is taught by Keri Colabroy of the Chemistry Department. This lab-intensive course focuses on techniques and applications in modern experimental biochemistry, including protein based bioinformatics, protein mass spectrometry, X-ray crystallography, protein NMR, and transient-state enzyme kinetics. Students explore BLAST, multiple sequence alignment (CLUSTAL W), and protein structure modeling (Swiss-PDB) as a means for understanding amino acid conservation and therefore protein function. Currently, this analysis must be performed independently (with some instructor input) by students as part of a lab exercise. The laptop computers requested as part of this proposal would allow whole classroom instruction and participation in these activities. In addition, while students in this advanced class gain valuable experience with a range of biochemical techniques, they currently have little opportunity to explore proteomics applications in the laboratory. Support for this CCLI proposal will enable us to secure equipment to allow new proteomics experiments to be incorporated into the lab experience in this class. In one proposed series of investigations, BCM 341 students will examine the proteome of *B. subtilis* when grown in the presence or absence of the thiamine. Most students would predict that thiamine biosynthetic genes would be upregulated in the absence of exogenous thiamine. Six 2-D gel electrophoresis, requested as part of this proposal, will allow separation of proteins extracted from experimental samples. Once proteins have been separated, the gels of two different proteomes will be compared to ascertain which protein spots changed under the different conditions. This comparison can be made with the aid of Flicker freeware using the new laptop computers requested as part of this proposal. Promising spots could be cut from the gel and digested with trypsin as a first step towards identifying the protein. Some peptide species identified by this approach could be analyzed by MALDI-TOF using a Mass Spectrometer at nearby Moravian College on a fee-for-service basis. In this way, advanced biochemistry students will gain exposure to important experimental approaches in proteomics. In addition, they will be exposed to quantitative analyses such as log transformations of data and digital image evaluations.

Structural Biology and Biophysics

This new upper-level course for science majors, currently under development by Jeremy Teissere of the Biology Department, focuses on biophysical approaches to understanding the structure and function of biologically-relevant molecules. As such it effectively bridges biology, chemistry, mathematics and physics. Course topics will include macromolecular structure, receptor-ligand interactions, allosteric signaling, channel kinetics, and membrane dynamics. Dr. Teissere's research focuses on neurotransmitter receptor structure, function, and biophysics; thus he is well-versed in various approaches to studying structural biology. The laboratory portion of the course will take advantage of the new upper-level laptop computers to allow students to use various freeware packages, such as Protein Explorer, FirstGlance, and Consurf 3.0, to view, manipulate, and measure macromolecular structures.

Specific Aim 4: Faculty Development

Successful integration of mathematics and biology into the curriculum requires biologists to become more familiar with new quantitative approaches and mathematicians and computer scientists to become familiar with the basic empirical needs of biologists. One way to address this problem is to attempt to hire a faculty member with expertise in both areas—a computational biologist or bioinformatics specialist. Some undergraduate colleges have attempted this, but with mixed success. Such individuals are in high demand and liberal arts colleges have difficulty hiring the most talented individuals. Individual scientists may expand their training by attending enrichment courses, for example co-PI Amy Hark has attended an NSF-funded GCAT faculty development course, but the impact of these activities is limited to the individual instructor. Accordingly, we propose to improve expertise and encourage collaboration among current faculty through an on-site week-long course for science and math faculty that focuses on bioinformatics.

We will develop new faculty expertise in bioinformatics by sponsoring an intensive on-campus bioinformatics mini-course during the summer of 2009. The course will bring together members of the Biology, Chemistry, and Mathematics and Computer Science Departments for a week of instruction on algorithms, tree construction, modification of parameters in bioinformatics tools, and sequence and microarray data analysis. Based on current faculty interest, we expect that the course would be attended by a dozen or more faculty members, including ecologists, mathematicians, computer scientists, and introductory biology lab instructors, in addition to the co-PI's of this proposal. Holding the course on-campus has the value of insuring a shared experience among faculty from different departments, and be considerably less expensive than sending all interested faculty to a course offered elsewhere. The common experience will help foster communication and new collaborative teaching or research between mathematicians and biologists. Short-term outcomes include the enrichment of all the biology and biochemistry courses described in this proposal through the increased depth of faculty sophistication. Long-term outcomes may include new courses, such as a team-taught Computational Biology course, and possible research collaborations between biologists and mathematicians. Malcolm Campbell, director of the GCAT program at Davidson College (Campbell et al., 2007a), has suggested several possible instructors.

Supporting undergraduate research

The addition of real-time PCR, 2-D electrophoresis, and improved computational tools will also have a significant effect on supporting undergraduate student research at Muhlenberg. As evidenced on the accompanying c.v.'s, the PI's have active research programs, with regular publications including undergraduate student co-authors. The labs of Bruce Wightman, Amy Hark, and Marten Edwards all focus on aspects of gene regulation in nematodes, plants, and insects, respectively. Thus students from all three laboratories will profit from having access to real-time PCR technology. Keri Colabroy's lab studies biosynthesis of antibiotics and mechanistic enzymology in bacteria. Therefore, her research students will be able to employ the power of 2-D electrophoresis and mass spectrometry in their studies of natural product biosynthetic pathways. The computational tools supported by this proposal will allow all research groups to perform more sophisticated evolutionary comparisons among members of gene families.

Project assessment and dissemination

Assessment of project goals will be achieved in the context of routine course assessment using evaluation materials similar to those used by Griffin et al. (2003). Each course that is part of the proposal will administer a pre-course survey and post-course survey that will measure student comfort and knowledge about quantitative goals. Examples of possible indirect and direct assessment questions are show below in Table 1. Surveys will be administered by course instructors during the first week of each course and during the last week of each course in order to allow us to assess student learning. The results will be compiled and statistically analyzed by a trained student statistician working under the supervision of Kathy Harring, Muhlenberg Associate Dean of Assessment and Professor of Psychology.

	Don't know	Strongly disagree	Weakly disagree	Weakly agree	Strongly agree
1. I am confident in my ability to calculate logarithmic functions.					
2. A heuristic is a "rule of thumb."					

Table 1. Sample questions for assessment of student mathematical confidence (#1) and comprehension (#2).

The results of our project will be disseminated through presentations at professional conferences, contribution of assessment information to a public database, the Muhlenberg College webpage, and publication to an educational journal. Bruce Wightman has made four educational presentations at professional conferences over the past decade (for example, Wightman, 2005). He will be able to make a similar presentation on the results of this project in 2011. Amy Hark has made professional educational presentations and published science education articles (see c.v.) and would present results of this project at meetings of the American Society for Biochemistry and Molecular Biology. Other Co-PI's will make educational presentations at conferences as opportunities present themselves. At the completion of the project, we will contribute our project assessment data to the on-line public database at the Online Evaluation Resource Library (http://oerl.sri.com/) and the National Science Digital

Library (http:/nsdl.org/). Bruce Wightman, with the assistance of the Muhlenberg College Office of Information Technology, will establish and maintain a webpage linked to the Muhlenberg Biology Department's webpage the provides an overview of project goals, links to detailed project protocols, and assessment results (when complete). Finally, assuming successful completion of project goals, we will submit an account of the bioinformatics laboratory experiences and assessment data to an educational journal such as *CBE-Life Sciences* for review and publication.

Project roles

Bruce Wightman will serve as project coordinator and will oversee implementation, assessment, and dissemination; he will also directly participate in the execution of all project aims since he is an instructor for courses included in each category. He has 12 years of experience as a faculty member at Muhlenberg and has continually directed funded research projects over that time. One funded project included a microarray component, and his current research includes phylogenetic analysis of nuclear hormone receptors. Amy Hark and Marten Edwards, as instructors of BIO 152, will also participate in the execution of Aim 1. Aims 2 and 3 will be performed by the relevant instructors for each course: Amy Hark, Marten Edwards, Keri Colabroy, and Jeremy Teissere. (Dr. Teissere's c.v. is not included because of his less extensive role in the project and NSF limits on the number of co-PI's on the proposal.) Amy Hark will develop the local bioinformatics web portal for the NCBI website in collaboration with the Office of Information Technology and Muhlenberg College Library. Clifton Kussmaul of the Mathematics and Computer Science Department will function as the local computational consultant, participate in Aim 3 through a guest lecture in the GGE course, and be a critical participant in the local faculty development program (Aim 4).

Project timeline

Spring 2009	Purchase equipment for genomics core and upper-level computers
	Begin implementation into Spring semester courses
Summer 2009	Plan implementation of BIO 152 recitation
	Purchase computers for BIO 152
	Faculty development mini-course
Fall 2009	Initial implementation of Aim 1 (BIO 152)
	Implementation for Fall courses (Cell Biology, Experimental Biochemistry, Structural Biology and Biophysics)
	First assessment of Fall courses
Spring 2010	Full implementation for Spring courses (Genetics, Biochemistry, GGE)
	First assessment of Spring courses
Summer 2010	Development and implementation of local NCBI web portal
Fall 2010	Second assessment of Fall courses
Spring 2011	Second assessment of Spring courses
Summer 2011	Project assessment and dissemination
Fall 2011	Complete project assessment and dissemination

REFERENCES

Altman, R.B., 1998, A curriculum for bioinformatics: the time is ripe, *Bioinformatics* **14**: 549-550.

Anderson, J., 2007, Enriching the teaching of biology with mathematical concepts, *American Biology Teacher* **69**: 205-209.

Bialek, W., and Botstein, D., 2004, Introductory science and mathematics education for 21st-century biologists, *Science* **303**: 788-790.

Bishop, J. M., and Knoll, A. H., 2003, How science should be taught; humans' next move, *Chronicle of Higher Education* (8/8/2003) **49**: B6.

BIO2010: Transforming Undergraduate Education for Future Research Biologists, 2003, National Academy of Sciences Press.

Campbell, A.M. 2006, DNA microarray wet lab simulation brings genomics into the high school curriculum. *CBE-Life Sciences Education* **5**: 332-339.

Campbell, A. M., et al., 2007a, Genome consortium for active teaching: meeting the goals of BIO2010, *CBE Life Sci Educ.* **6**:109-118.

Campbell, A.M., et al. 2007b, Make microarray data with known ratios. *CBE-Life Sciences Education* **6**:196-197.

Dahir, M., 1995, Integrating the science curriculum: Standards created by the National Academy of Sciences, *Technology Review* **98:** 22-23.

Denver, D. R., Morris, K., Lynch, M., and Thomas, W. K., 2004. High mutation rate and predominance of insertions in the *Caenorhabditis elegans* nuclear genome. *Nature* 430:679-82.

Garfalo, F. and LoPresti, V., 1993, Evolution of an integrated college freshman curriculum, *J. Chem. Educ.* **70:** 352-359.

Gazzaniga, M., 1998, How to change the university, *Science* **282:** 237.

Ghedin, E. et al. 2007. Draft genome of the filarial nematode parasite Brugia malayi. *Science (New York, N.Y.)* 317:1756-60.

Griffin, V., McMiller, T., Jones, E., and Johnson, C. M., 2003, Identifying novel helix-loop-helix genes in *Caenorhabditis elegans* through a classroom demonstration of functional genomics, *Cell Biol. Educ.* **2**: 51-62.

Guterman, L., 2007, What good is undergraduate research, anyway? *The Chronicle of Higher Ed.* **53(50):** A12.

Heyer, L.J., et al. 2005. MAGIC Tool: integrated microarray data analysis. *Bioinformatics* **21**: 2114-2115.

Heyer, L.J. and Campbell, A.M. 2006. Value added: blending math into a high school genomics lab. http://www.bio.davidson.edu/projects/GCAT/HSChips/hs_kit_math_module_v2.pdf (accessed 7 May 2008)

Hydorn, D., Baker, S., and Boats, J., 2005, Quantitative initiatives in college biology: profiles of projects at undergraduate institutions, In *Math and Bio2010: Linking Undergraduate Disciplines*, L. A. Steen, Ed., Mathematical Association of America.

Jackson, J. H., 2005, Bioinformatics and genomics, In *Math and Bio2010: Linking Undergraduate Disciplines*, L. A. Steen, Ed., Mathematical Association of America.

Jaglo-Ottosen, K R, S J Gilmour, D G Zarka, O Schabenberger, and M F Thomashow, 1998, Arabidopsis CBF1 overexpression induces COR genes and enhances freezing tolerance. *Science* **280**:104-6.

Laws, P. W., 1999, New approaches to science and mathematics teaching at liberal arts colleges, *Daedalus***128**: 217-239.

Maglich, J. M. et al. 2001. Comparison of complete nuclear receptor sets from the human, *Caenorhabditis elegans* and *Drosophila* genomes. *Genome Biology* 2: 29.

May, R. M., 2004, Uses and abuses of mathematics in biology, *Science* **303**: 790.

Mayr, E., 1982, *The Growth of Biological Thought*, Harvard University Press.

Reiss, M. J., Millar, R., and Osborne, J., 1999, Beyond 2000: science/biology education for the future, *J. Biol. Educ.* **33**: 68-70.

Robinson-Rechavi, M., Maina, C. V., Gissendanner, G. R., Laudet, V., and Sluder, A., 2005, Explosive lineage-specific expansion of the orphan nuclear receptor HNF4 in nematodes. *Journal of Molecular Evolution* 60:577-86.

Russell, S. H., Hancock, M. P., and McCullough, J., 2007, Benefits of undergraduate research experiences, *Science* **316**: 548-549.

Sluder, A. E., Mathews, S.W., Hough, D., Yin, V. P., and C. V. Maina, 1999, The nuclear receptor superfamily has undergone extensive proliferation and diversification in nematodes. *Genome Research* 9:103-120.

Thomashow, M F., 2001, "So what's new in the field of plant cold acclimation? Lots!" *Plant Physiology* **125**:89-93.

Thompson J.D., Higgins D.G., and Gibson T.J., 1994, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**:4673-4680.

Tamura K., Dudley J., Nei M., and Kumar S., 2007, MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol.* **24**:1596-1599.

Vlachonasios, K E, Thomashow, M F, and S J Triezenberg, 2003, Disruption mutations of ADA2b and GCN5 transcriptional adaptor genes dramatically affect Arabidopsis growth, development, and gene expression. *The Plant Cell* **15**:626-638.

Wang, Z. et al., 2000, Universal PCR amplification of mouse immunoglobulin gene variable regions: the design of degenerate primers and an assessment of the effect of DNA polymerase 3' to 5' exonuclease activity. *J. Immunological Methods*, **233**:167-177.

Wightman, B., 2005, Mutant screening and gene mapping in fifteen weeks, abstracts of the 15th International *C. elegans* Meeting, Los Angeles, CA, http://www.wormbase.org/db/misc/paper?name=WBPaper00026208;class=Paper (accessed 05/15/08).

Williams, B. D., Schrank, B., Huynh, C., Shownkeen, R., and Waterston, R.H., 1992, A genetic mapping system in *Caenorhabditis elegans* based on polymorphic sequence-tagged sites. *Genetics*. **131**:609-624.

Wilson, E. O., 1998, *Consilience: The Unity of Knowledge*, Knopf.

Young, P. G., 2008, *Web-based bioinformatics tutorials: exploring genomes*, WH Freeman.

MUHLENBERG BIOLOGY COURSE DESCRIPTIONS:

http://www.muhlenberg.edu/depts/biology/program/coursesM.html

MUHLENBERG BIOCHEMISTRY COURSE DESCRIPTIONS:

http://www.muhlenberg.edu/depts/biochem/overview.html
http://www.muhlenberg.edu/depts/biochem/bcm341.html

MUHLENBERG NEUROSCIENCE COURSE DESCRIPTIONS:

http://www.muhlenberg.edu/depts/neuroscience/coursework.html