

SYMBIOSIS: An Integration of Biology, Math and Statistics at the Freshman Level: Walking Together Instead of on Opposite Sides of the Street[†]

*
Karl H. Joplin , Dept. of Biological Sciences
Edith, Dept. of Mathematics & Statistics
Anant Godbole, Dept. of Mathematics & Statistics
Michel Helfgott, Dept. of Mathematics & Statistics
Istvan Karsai, Dept. of Biological Sciences
Darrell Moore, Dept. of Biological Sciences
Hugh A. Miller III, Dept. of Biological Sciences
all of East Tennessee State University

Name of Institution: East Tennessee State University	
Size	about 15,000 students
Institution Type	regional state institution with graduate programs
Student Demographic	freshman curriculum for all biology majors
Department Structure	Mathematics and Statistics, and Biological Sciences are individual departments in College of Arts and Sciences

Abstract

SYMBIOSIS is a novel three-semester curriculum that teaches biology, statistics and mathematics in an integrated curriculum at the introductory level for freshmen. It was developed by faculty in the Departments of Biological Sciences and Mathematics and Statistics. We describe the goals, organization, and aims of this project and processes used to establish it and we discuss the pedagogical and cultural barriers between these disciplines that needed to be addressed.

Course Structure

- Weeks per term: 15 weeks
- Classes per week/type/length: M (Lec-2 hrs), T (Lab-2 hrs), W (Lec-2 hrs), Th (Lec-2 hrs), F (Lec-2 hrs)
- Labs per week/length: one 2-hr lab/wk
- Average class size: 16 students in one section
- Enrollment requirements: Students supported by our NSF STEP grant
- Faculty/dept per class, TAs: One biology and one mathematics instructor, two TAs
- Next course: IBMS 1200, Integrated Biology and Calculus
- Website: <http://www.etsu.edu/cas/symbiosis/default.aspx>

[†] supported by HHMI grant#52005872
*joplin@etsu.edu

Introduction

Picture a busy thoroughfare through a city with cars speeding by and people standing on the sidewalk. Viewpoints of the people on the sidewalk depend on which side of the street they are on. This is the state of affairs in biology and mathematics education, with biologists standing on one side, mathematicians and statisticians standing on the other side, and little connection between them. This situation was addressed in 2003 by the National Academies in the publication *BIO 2010* (National Resource Council, 2003), an analysis and set of recommendations calling for the integration of biology and mathematics for academic development and pre-professional training. *BIO 2010* has been followed by *Math & BIO 2010: Linking Undergraduate Disciplines* (Steen, 2005) and by federally and privately funded initiatives. Some programs have been started in response to these reports, but most of them have focused on introducing mathematical topics into biology classes, usually at the upper undergraduate and graduate levels.

East Tennessee State University (ETSU) is a regional university of approximately 15,000 students and 700 faculty. It is primarily an undergraduate teaching institution with masters programs in biology and mathematics (<http://www.etsu.edu>). Faculty of the departments of Biological Sciences and Mathematics at ETSU have a history of interdepartmental cooperation in biological research. This led to the creation of the Institute of Quantitative Biology (IQB) in 2003 to enhance interdepartmental integration. Two groups of faculty drawn from both departments applied for and received an NSF-UBM grant, an NSF-funded STEP grant in 2005, and a curriculum grant funded by the Howard Hughes Medical Institute (HHMI) in 2006. These programs are connected but represent different aspects of our approach to undergraduate biology and mathematics education. The STEP program is meant to recruit students and introduce them to research and the goal of the HHMI grant is to design and implement an integrated curriculum of mathematics and biology.

The design and implementation of the HHMI-supported curriculum change has been and continues to be a major undertaking requiring rethinking of the pedagogy of both disciplines. This paper describes the process used for this project and the resulting curriculum model. Other aspects of SYMBIOSIS are described in an accompanying paper (Moore, et al., 2012).

Description

Our HHMI-funded curriculum grant is titled SYMBIOSIS: An Introductory Integrated Mathematics and Biology Curriculum. The award was to create an integrated curriculum that would count as three semesters of introductory biology for majors, one semester of statistics, and one semester of calculus. The four-year grant was funded in the fall of 2006 and SYMBIOSIS I was taught for the first time in Fall 2007.

SYMBIOSIS is our response to the *BIO 2010* report, which calls for creation of integrated courses. Most responses to this call have taken the form of mathematical modules added to existing biology courses, biological applications added to existing mathematics courses, or integrated research projects for upper division students and mainly directed at mathematics content.

We have taken a different approach with SYMBIOSIS by integrating statistics, calculus, and biology in a three-semester course at the introductory level. We describe the material used to teach SYMBIOSIS I, which combines the topics in General Biology I for majors and the introductory probability and statistics course. During the development of the course, we realized

that this approach of teaching biology with statistics has added to the conceptual richness of biology instruction while providing a biological context for statistics instruction.

The purposes of the SYMBIOSIS curriculum are to introduce a quantitative viewpoint into the introductory biology curriculum, develop mathematical concepts using biological applications, and investigate biological phenomena using analytical tools. The use of an integrative method rather than a juxtaposition method pedagogy (Jean and Iglesias, 1990) means that students see the relevance of a quantitative approach to biology. The problem with the traditional juxtaposition method is that it treats mathematics and biology as separate subjects, with students in one major viewing the other course as nothing more than a general education requirement.

The integrative method is based on the observation “Biology students are prepared to receive the mathematical concepts once they see their applications” (Riego, 1983). The same can be said of mathematics students, in that they are not taught the applicability of mathematics to biology. By presenting biology and mathematics as an integrated subject, we hope to overcome the reluctance of biology students to consider mathematical methods as essential to full understanding of biological processes.

The material for SYMBIOSIS I was developed by a year-long collaborative effort between mathematics, statistics, and biology faculty. The process required that each group develop an understanding for the approach that they use to develop their material. As one faculty member said, “You biologists don’t use mathematics like a physicist does.” Much of the work in integrating the material revolved around this realization. Some of the biology faculty were envious of the mathematicians’ ability to develop a full lecture showing relationships and applications on the board. Math faculty were surprised that biologists needed so much illustration to show the three-dimensional structure of biology components or the effects of change over time. However, we found that each group could adjust its approach and that the disciplines could be presented in an integrated manner.

The course was first taught to a cohort of students from the NSF-funded STEP program. The students had a summer bridge program before their freshman year in which they were exposed to biological research activity and mathematical concepts. In future years, the course will be open to biology and mathematics majors who have been previously advised about its nature, and no additional prerequisites will be required. The lectures are team-taught by biology and mathematics faculty. Biology is used to introduce each module and to define the topic; this is followed by the statistics or mathematics concepts and tools that address the biological issues. Although the organization of the course is based on biological considerations, we still wanted the mathematical and statistical topics to be presented completely and in a logical order. These goals required us to decide what biological and mathematical components can be covered (see discussion below). Lectures are taught using Powerpoint and class notes are also available to the students through the university’s D2L platform. The labs are taught by graduate teaching assistants with participation and overview by faculty. Typically there are two labs for each module, and experimental lab and a lab for data analysis and preparation of presentations or lab reports. Minitab and R are used in lectures and in labs to analyze data. Students complete two projects involving analysis of datasets and prepare posters of the results. The initial projects were on bird allometry and analysis of DNA sequence patterns.

We have found that statistics and biology are easy to pair, both conceptually and operationally. They have a long history together, since many statistical methods that appeared at the end of the nineteenth century and beginning of the twentieth century were developed by statisticians, such as R.A. Fisher, working in genetics and agricultural research and motivated by

the need of tools to analyze the data they produced. Recent advances in genetics and bioinformatics and the acquisition of large data sets and high speed computers are again challenging the discipline of statistics with the need of tools for analysis.

The development of the material for the first semester depends on the contextual needs of biology and the developmental needs of statistics. Statistics, as with much of mathematics, depends on a logical development of concepts. Thus, both the statistics development and the biology content were considered in developing the framework for the modules. We believe that modern biology pedagogy is based too much on a pseudo-logical framework of going from small-to-big and is based on what biology has done historically and not why or how it is done. An examination of modern biology textbooks supports this contention, because they are encyclopedic in content and there is little carryover of material from chapter to chapter (Moore, et al., 2012). There is little quantitative methodology, with at most two or three equations presented in an entire book. Graphs commonly lack statistical information such as error bars, which are important because they demonstrate the variation in a population that is the basis of evolutionary change, which is basic to biology. Thus, students are presented with a collection of facts that have no logical connection to the whole. Instructors observe that this approach produces students who do not know the introductory material needed for upper level courses.

We are attempting to address this concern by presenting students with “5 Themes of Biology” in the introductory modules, and to address each theme explicitly in each of the subsequent modules. The themes that we focus on in SYMBIOSIS are energy utilization, homeostasis, growth and reproduction, adaptation, and evolution. So when we present material on cells, we also examine how physical properties, such as surface-to-volume ratio, affect cell size and transport of material in and out of the cell. Mathematical functions can be used to show how these properties affect the energy, homeostasis, and growth and reproduction themes. In the same module, the number of erythrocytes of humans living at different altitudes (Spector, 1956) permits us to talk about the adaptation and evolution themes.

We use “module” to denote a unit of class content or chapter. Our modules define the biology and mathematics or statistical components of the semester. Each module consists of ten hours of lectures, a two-hour “wet” experimental lab, and a two-hour “dry” analytical lab.

The modules developed for SYMBIOSIS I include:

Introduction and the scientific method. A biologist’s viewpoint of the scientific method and the role that statistics and mathematics plays in developing models and testing hypotheses. The binomial distribution is introduced to test hypotheses about population proportions and the randomization test is introduced to test hypotheses about the equality of means.

The cell. Cellular functions are a logical topic for the introduction of biological concepts. When we study a certain type of cell, such as an erythrocyte, its form can be classified as normal or abnormal. Counting the number of red blood cells in a sample and measuring cell dimensions provides student-generated data we can use to introduce descriptive statistics, correlation, and statistical graphs. Students are shown how to go beyond descriptive statistics and take the step toward inference. Estimation (by bootstrapping) and tests of hypotheses (randomization test) are used to arrive at conclusions based on experimental data. The biological implications of the surface to volume ratio of a cell are discussed as well as the strategies of cells to increase their surface area.

Size and scale. The concepts of scaling and allometry are used to study relationships among variables. Differences between isometric and allometric scaling are introduced, as are fractal branching for surface area and volume problems. Slope as a rate of change of scaling and log-log plots and the power law are also discussed. Exponential functions, the normal distribution, linear regression, and transformations are used to describe biological processes.

Mendelian genetics. Genetics provides an ideal motivation for the study of probability, including conditional probability, independence, and tests of independence. Mendel's original data are used to draw conclusions based on probability and to discuss the basics of Mendelian genetics. Meiosis is discussed as the biological basis of genetic probability and the rationale of why Punnett's square and probability trees demonstrate how the probability of allelic combination represents the meiosis process. Mendel's actual experimental data are used to perform goodness of fit tests for a coin-based model of genotype and phenotype. Conditional probability, Bayes rule, Poisson and normal approximation to the binomial distribution and an introduction to sampling methods are statistical topics of this module. In the Mendelian genetics module, biology and statistics integrate very well; biology provides a motivation for statistics and probability helps to understand the random nature of inheritance. The binomial distribution has always been useful in discussing the probability of each phenotype. The situation in which the sample size is large and the probability of success is small serves as a motivation for introducing the Poisson distribution as an approximation to the binomial.

DNA genetics and the genome is the natural topic to follow Mendelian genetics. DNA replication and sequence analysis are discussed and provide the opportunity to apply probability and hypothesis testing to new problems such as calculating the probability of palindromes, specific sequences of nucleotides, and specific palindromes related to enzyme restriction sites, the probability of matches, and so on. DNA databases from the internet allow us to use real data to discuss nucleotide frequency, GC content, non-independence in the two letters of a dinucleotide, presence of palindromes, and distances between palindromes. Terms and tools that can be useful later in the understanding of topics in bioinformatics are introduced in this module, including random walks, transition probabilities, matrices, and transition probability graphs. Classic topics of statistical inference (confidence interval estimation, test of hypotheses for proportions using large samples, the t -tests) that are part of our introductory statistics course are also included in this module. Examination of genomes and genome sequences for defined elements are used to statistically describe mitochondrial DNA sequences of insect species. Students compare the analysis of their sequence with another insect mitochondrial sequence analyzed by another pair of students, and both groups compare their sequence with the *Drosophila* mitochondria as a reference. They become aware of the differences between species at the DNA level and how to use statistical tools for this analysis. Data bases and free software available in the internet such as NCBI, Genomatrix (Genomatrix Software Suite, 2012), and ClustalW (European Molecular Biology Laboratory/European Bioinformatics Institute, 2012) are used.

Evolution is taught using probability functions to demonstrate changes in gene frequency over generational times. The Hardy-Weinberg equation is derived from allelic frequency data and the underlying assumptions are discussed. Its application as a standard against situations in which the assumptions are violated allows us to use the chi-square test.

In SYMBIOSIS II, the mathematical topics belong mainly to calculus, but with a statistical component that includes nonlinear estimation and the use of the periodogram. The biological topics of the models are populations, ecology, behavioral ecology, structured populations, chronobiology, energy, and enzymes.

In SYMBIOSIS III, the emphasis of the quantitative component is on additional topics of calculus, matrices, graph theory, some multivariate statistical methods, and an introduction to bio-informatics. The biological topics are neurons, membranes, photosynthesis, development, bioinformatics and evolution. For a further description of these modules see our Symbiosis website (East Tennessee State University, 2012).

Discussion

The reason for developing this course is simple and straightforward: to present the statistical concepts that biology and mathematical students need to do modern biology. The development and implementation have not been as simple as the goal. Each module was developed by a team consisting of at least one biologist and at least one mathematician or statistician. Members of the team need to be comfortable working together for the development process to succeed, and this is easiest for teams that have a prior working relationship, as did many of ours. The process benefited from the unqualified support of our respective chairs and college and university administration. The process is ongoing, but the results to date are encouraging. We have taken SYMBIOSIS I as the basis of a successful Governor's School program (a summer enrichment program for gifted high school students).

There are still problems at various levels of the university system, such as policies for transferring SYMBIOSIS credits to other institutions, accepting credit from transfer students, and obtaining acceptance from committees that determine appropriate credentials for admission to professional schools.

An important course design issue was to look for the best matching of biology topics and statistics topics to form the modules, but this posed some pedagogical difficulties. It is natural to start a course for future scientists with the scientific method, which implies that we need an early introduction to statistical hypotheses testing. How could we do that in the first week of the course when sampling distributions had not yet been covered? Fortunately, randomization methods (permutations test) can be included at this time to test hypotheses about the means of two populations. To test hypotheses about a population proportion, the exact test can be used; basics of probability and the binomial distribution are introduced and the exact test appears as a simple application of the binomial distribution.

The criteria that have guided the development of the statistical component of the first semester of SYMBIOSIS are as follows: do an early introduction of inference in order to be able to answer research questions from the beginning, give examples of the sequence rationale-algorithm-computer program, use a problem oriented approach presenting statistical methods when they are needed, give importance to the study of variability, take a multivariate view whenever possible, and include topics at the elementary level that can serve as a preparation for understanding the language and methods of biostatistics/bioinformatics, including exposure to R (The R Project for Statistical Computing, 2012)¹ Active learning and critical thinking are promoted through class discussion and activities, homework, and poster assignments. The statistical analysis of real biological data drives the students into discovering facts instead of

¹R is free software, similar to Matlab, that is used extensively in bioinformatics and biostatistics; we use it to perform simple calculations and plots.

listening to facts in a passive way. We have created a textbook that we are using in our statistics sections (Seier and Joplin, 2011)

Suggestions

The development of this course has led us to examine many of the assumptions that traditional pedagogy makes for the biology and mathematics curricula. This is not easy and requires good communication between the mathematics and biology developmental teams. All faculty have to be willing to eliminate or restructure material. For example, our biologists agreed to move much of the biochemistry from the first to the third semester, at which point the mathematics is advanced enough to begin modeling chemical kinetics, and to eliminate time spent on biological groups (phyla and class descriptions), while the mathematicians agreed to reduce the number of Calculus II topics that lack important biological applications.

What is required is a rethinking of undergraduate education in this day of advancing biology and increasing computational power. Students must be taught to think conceptually from a critical viewpoint and this is not served by the traditional approaches.

Conclusion

SYMBIOSIS I covers somewhat more than the traditional introductory statistics course since it introduces classical parametric methods and randomization methods for inference. The biology component focus on topics that can be used to illustrate the statistics topics but nevertheless covers much of the standard first-semester biology curriculum, with the exception of atomic and molecular biochemical structure and metabolism, which are covered in SYMBIOSIS III after the mathematics has been developed in SYMBIOSIS II.

In SYMBIOSIS II, the mathematics portion is primarily concerned with the introduction and development of calculus. Thus, we are creating an introductory mathematics curriculum that emphasizes a biological, instead of the traditional physics-engineering, approach to calculus. In our collaboration, we are beginning to realize how different this approach really is.

Although the integrated approach works for small classes, implementation for large (300+) lecture sections introduces additional obstacles, since mathematics courses are typically limited to sections of 50-75. We are in the process of adapting the material for our introductory biology curriculum to address this difference. One approach is to create a co-requisite requirement so that students who enroll in Biology I must also enroll in one of the sections of the probability and statistics course that use biology for material development. Interestingly the statisticians feel that these sections would be considered more rigorous than their standard statistics course. Thus, this is not a 'baby biology' mathematics approach, but is a vigorous approach to an integrated curriculum.

In summary, because the material has been team-taught, we have noticed that both the students and the faculty have become more comfortable with the two subjects, creating a *symbiotic* approach to both biology and statistics that is developing as we walk together on the same side of the street.

Acknowledgements

This program was supported in part by grants to East Tennessee State University from the Howard Hughes Medical Institute (#52005872) through the Undergraduate Science Education Program and NSF-STEP program, Talent Expansion in Quantitative Biology (#0525447).

References

- East Tennessee State University, cited 2012: Symbiosis, the Whole is Greater than its Parts. [Available online at <http://www.etsu.edu/cas/symbiosis/default.aspx>].
- European Molecular Biology Laboratory/European Bioinformatics Institute, cited 2012: ClustalW [Available online at <http://www.ebi.ac.uk/Tools/clustalw/index.html>].
- Genomatrix Software Suite, cited 2012. [Available online at <http://www.genomatix.de/cgi-bin/tools/tools.pl>].
- Govett, A., H.A. Miller III, D. Moore, E. Seier, K.H. Joplin, A. Godbole, M. Helfgott, I. Karsai, 2012: Adventures in assessment: How to evaluate a new integrated quantitative biology program. This volume, xx-yy.
- Jean, R. V., and A. Iglesias, 1990: The juxtaposition vs. the integration approach to mathematics in biology. *Zentralblatt fur Didaktik der Mathematik*, **22**, 147-153.
- Moore, D., M. Helfgott, A. Godbole, K.H. Joplin, I. Karsai, H.A. Miller, III, E. Seier, 2012: Creating Quantitative Biologists: The immediate future of SYMBIOSIS. This volume, xx-yy.
- National Research Council, 2003: *Bio 2010: Transforming Undergraduate Education for Future Research Biologists*. National Academies Press, Washington, D.C.
- National Center for Biotechnology Information, cited 2012 [Available online at <http://www.ncbi.nlm.nih.gov/>].
- Riego L., 1983: Mathematics and the Biological Sciences – Implications for teaching. *Proceedings of the Fourth International Congress on Mathematical Education*. Zweng, M., T. Green, J. Kilpatrick, H. Pollak, M. Suydam. Birkhauser Boston, 207.
- Seier E. and K.H. Joplin, 2012: *Introduction to STATISTICS in a biological context*. Createspace.
- Spector, W., 1956: *Handbook of Biological Data*. National Academy of Science., Philadelphia and London: W.B. Saunders Company.
- Steen L., Ed., 2005: *Math & Bio 2010: Linking Undergraduate Disciplines*. The Mathematical Association of America, Washington, D.C.
- The R Project for Statistical Computing, cited 2012 [Available online at <http://www.r-project.org/>].