Recommendations from the BCB Graduate Curriculum Committee[1]

Vasant Honavar, Volker Brendel, Karin Dorman, Scott Emrich,
David Fernandez-Baca, and Steve Willson

April 10, 2006

Background

The current BCB curriculum suffers from:

- Too much overlap between some courses
- No coverage of many important topics
- Lack of adequate training in computational and statistical methods
- Poor match between courses offered and courses needed

There is a need for a rationally designed curriculum that adequately covers a set of topics in Bioinformatics – as part of a required set of core courses for all BCB students.

Charge to the Committee

The curriculum committee was asked to develop and propose a suitable core course sequence to provide sufficient breadth and depth of coverage of Bioinformatics.

Challenges

Designing an effective curriculum requires addressing several challenges:

- Background and preparation that can be reasonably assumed of incoming students
- Need to cover a core set of topics with sufficient rigor without assuming a long chain of prerequisites in several disciplines
- Need to ensure the offering of relevant non BCB courses that cover the relevant prerequisites (e.g., in biological sciences, computer science, mathematics, statistics)
- Need to package the necessary coursework into a small number of courses that can be taken within the first 2 years of the graduate program (at least by a majority of students)
- Need to staff the courses with faculty with the relevant expertise as they juggle their commitments to their respective departments and the BCB program

Proposed Solution

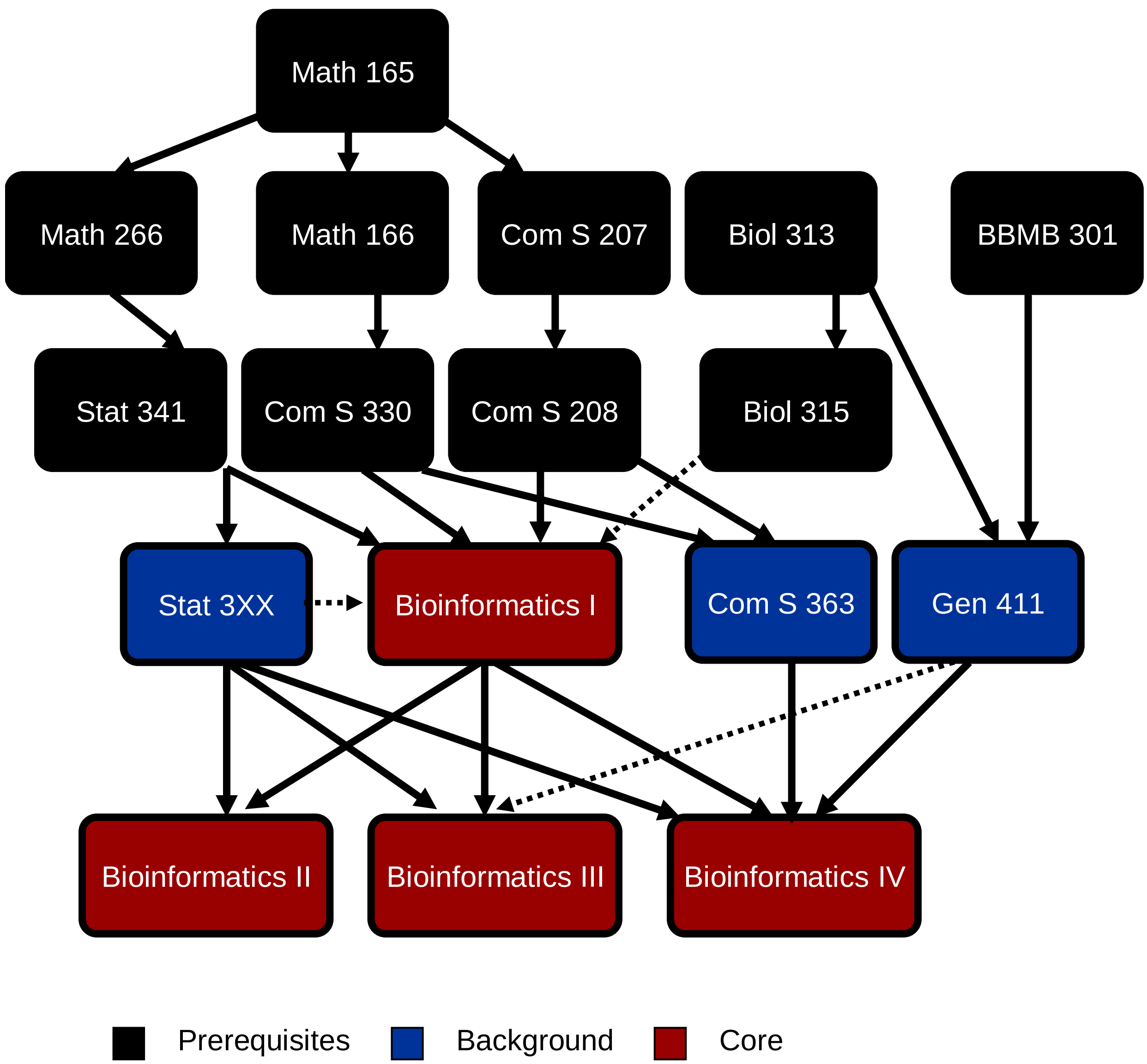
The curriculum committee recommends:

- A rational (and realistic) set of prerequisites and background courses
- A required set of 4 core courses covering major topics in bioinformatics

[1] The committee is grateful to extremely valuable input from the BCB supervisory committee, and several BCB faculty members especially – Drena Dobbs, Bob Jernigan, Chris Tuggle, and Dan Voytas, Jackie Shanks, Steve Proulx, Dennis Lavrov, John Mayfield, and Gustavo Mackintosh.

- POS recommended electives instead of advanced group requirements

Course Requirements for BCB Ph.D. Students



Required BCB Core Courses

Bioinformatics I (Fundamentals of Genome Informatics) (BCB 548, COM S 548, CprE 548) rerequisites: Com S 208, Com S 330, Credit or enrollment in Biol 315, Stat 341, credit or enrollment in Stat 3xx. Offered in Fall semester, Potential Instructors: Srinivas Aluru; David Fernandez-Baca; Oliver Eulenstein.

Biology as an information science. Review of algorithms and information processing: design of algorithms; space and time complexity analysis of algorithms; basic search algorithms; branch and bound search; dynamic programming. Generative models for sequences: multinomial models; Markov models. String algorithms: exact string matching; suffix trees and suffix arrays; approximate string matching (k mismatches, k differences). Pairwise sequence alignment: amino acid substitution scoring matrices; local and global alignment. Multiple sequence alignment: progressive alignment; word-based methods; local multiple alignment (sequence profiles and motifs). Sequence database search: dot matrix methods; heuristic methods; statistics of database searches; Introduction to genome sequence assembly.

Bioinformatics II (Advanced Genome Informatics) (BCB 594, GDCB 594, Stat 594)
Prerequisites: BCB 548, Stat 3xx. Biol 315. Offered in Spring semester, Potential Instructors: Volker Brendel; Karin Dorman; Xun Gu.

Applications of sequence models: codon usage; discrete and continuous models of nucleotide substitution; synonymous and nonsynonymous nucleotide substitutions. Basic methods in molecular phylogeny: phylogentic trees; distance matrix methods; maximum parsimony methods; maximum likelihood methods. Advanced sequence models: Random walks; score-based sequence analysis; Interpolated Markov Models; Markov Random Fields; applications to genome annotation; genome rearrangements. Hidden Markov Models: theory; training; applications to gene structure annotation, sequence alignment, and protein classification. DNA and protein motifs: weight matrices; word-based methods; EM algorithm, Gibbs sampling, and simulated annealing; Bayesian methods. Introduction to gene expression analysis, mRNA and protein expression data analysis, multiple comparisons.

Bioinformatics III (Structural Genome Informatics) (BCB 5xx, Com S 5xx, BBMB 5xx)
Prerequisites: BCB 548, BBMB 301, Biol 315, Credit or enrollment in Gen 411. Offered in Fall Semester, Potential Instructors: Bob Jernigan, Guan Song, Zhijun Wu

Algorithmic and statistical approaches in structural genomics including: Protein, DNA and RNA structure; Protein and Nucleic acid databases; Computational problems in structure determination including structure representation, transformation between coordinate systems, structure comparison (using RMS and distance matrix based methods) and visualization, structure determination with NMR derived distances, Distance-based structure modeling, energy minimization methods for structure refinement, protein structure modeling using threading and homology based methods. Analysis and prediction of protein secondary structure and tertiary structure, ordered and disordered regions, structural domains, 3-dimensional structural motifs, protein cores and surfaces, structural classes, protein function from primary, secondary, or tertiary structure, protein-protein, protein-RNA and protein-DNA interfaces; analysis and prediction of RNA structure.

Bioinformatics IV (Computational Functional Genomics and Systems Biology) (ComS 5ZZ, Stat 5ZZ, GDCB 5ZZ). Prerequisites: BCB 548, Biol 315, Gen 411, Stat 3xx, Com S 363. Offered in Spring Semester: Potential Instructors: Vasant Honavar, Karin Dorman, Steve Proulx

Algorithmic and statistical approaches in computational functional genomics and systems biology; Biological Information Integration – Knowledge (ontology) driven and statistical approaches; Qualitative, probabilistic, and dynamic network models; Modeling, analysis, simulation and inference of transcriptional regulatory modules and networks, protein-protein interaction networks; metabolic networks; cells and systems.

Recommended Prerequisites for Admission to the BCB Ph.D. program

COURSE PREREQUISITES FOR ADMISSION TO BCB Courses that should be taken ideally *prior to enrollment*		
Category I. Mathematics and Statistics		
Math 165 or equiv	Differential calculus, applications of the derivative, introduction to integral calculus.	4 cr. – F S SS*
Math 166 or equiv	Integral calculus, applications of the integral, infinite series.	4 cr. – F S SS
Math 265 or equiv	Analytic geometry and vectors, differential calculus of functions of several variables, multiple integrals, vector calculus	4 Cr. – F S SS
Stat 341 or equiv	Probability; distribution functions and their properties; classical discrete and continuous distribution functions; moment generating functions, multivariate probability distributions and their properties.	3 cr. – F S SS
Category II. Biological Sciences		
Biol 313 or equiv.	Introduction to the principles of transmission and molecular genetics of plants, animals, and bacteria. Recombination, structure and replication of DNA, gene expression, cloning, quantitative and population genetics.	3 cr. F S
BBMB 301 or equiv.	A survey of biochemistry: structure and function of amino acids, proteins, carbohydrates, lipids, and nucleic acids; enzymology; metabolism; biosynthesis; and selected topics.	3 cr. F S
Biol. 315	The mechanisms of evolution. Topics in microevolution: population genetics, natural selection, genetic variation, and adaptation. Macroevolution: speciation, extinction, phylogeny, and major evolutionary patterns.	3 cr. F S
Category III. Computer Science		
Com S 207 or equiv	An introduction to computer programming using an object-oriented programming language. Emphasis on the basics of good programming techniques and style. Extensive practice in designing, implementing, and debugging small programs. Use of abstract data types. Interactive and fild I/O. Exceptions/error-handling.	3 cr. – F S SS
Com S 208 or equiv	Intermediate-level programming techniques. Emphasis on designing, writing, testing, debugging, and documenting medium-sized programs. Data structures and their uses. Dynamic memory usage. Inheritance and polymorphism. Algorithm design and efficiency: recursion, searching, and sorting. Event-driven and GUI programming. The software development process.	3 cr. – F S SS
Com S 330 or equiv	Concepts in discrete mathematics as applied to computer science. Logic, proof techniques, set theory, relations, graphs,	3 Cr. F S SS

	combinatorics, discrete probability and number theory.	

Required Background Coursework (typically taken in year 1 unless similar coursework was completed prior to joining the BCB program)

BCB BACKGROUND COURSEWORK REQUIREMENTS Courses (or equiv.) that should be taken p*rior to enrollment or during first year*)		
Category I. Mathematics and Statistics		
Stat 3xx or equiv (**new course** that combines relevant material from Stat 432 and some material from Stat 401)	Review of probability and random variables. Data, sampling, and basic statistical Inference, Classical Estimation and Hypothesis testing, Elementary experiment design and ANOVA. Stochastic Processes - Poisson processes and Markov Chains	3 cr – F
Category II. Biological Sciences		
(**revised**) Gen 411 or equiv Note: Dan Voytas to work with BBMB/GDCB on the necessary changes	Molecular biology and cellular biochemistry with focus on systems-level analyses and high-throughput technologies. Review of basic cell structure and function; Principles of molecular genetics; Regulation of gene expression; Principles of cellular and developmental regulation, molecular evolution; Methods for high-throughput genomic, transcriptomic, metabolomic, structural genomic, and proteomic analyses.	3 cr. – S
Category III. Computer Science		
Com S 363 (**new requirement** replacing Com S 311)	Relational, object-oriented, and semistructured data models and query languages. SQL, ODMG, and XML standards. Database design using entity-relationship model, data dependencies and object definition language. Application development in SQL-like languages and general purpose host languages with application program interfaces. Information integration using data warehouses, mediators and wrappers. Programming Projects.	3 cr. – F S

* F = Fall semester; S = Spring semester; SS = Summer Session
** must be taken as Gen 495 to apply toward graduate credit

Required Advanced Electives

A total of 12 graduate credits at the 500-level or above (excluding seminars, journal clubs, rotation credits, workshop credits, 1-credit lab or tools courses, or other courses that are graded on a pass / fail or satisfactory / unsatisfactory basis) recommended by the POS committee. These courses should ideally provide some depth in computer science, statistics, mathematics, or biological sciences.

Additional Course Requirements

Required Rotations, Seminars, Workshops, Ethics Modules, etc. are unchanged.

New courses or Significant Updates Proposed to Existing Courses

Stat 3XX Probability and Statistical Inference for the Biological Sciences.
Offered in Fall Semester. IPotential Instructors: Karin Dorman or other Statistics faculty. Prerequisites: STAT 101 or 104 or 226 or equivalent; STAT 341 or equivalent.

Introduction to stochastic processes and statistical inference with applications to the biological sciences. Review of probability and random variables. Basic statistical inference, estimation, hypothesis testing. Elementary experimental design and ANOVA. Stochastic processes - Poisson processes, Markov chains.

Gen 411 Molecular Genetics. (To be revised – Dan Voytas)
Offered in Spring Semester. Prerequisites: Biol 314.

Molecular biology and cellular biochemistry with focus on systems-level analyses and high-throughput technologies. Review of basic cell structure and function; Principles of molecular genetics; Regulation of gene expression; Principles of cellular and developmental regulation, molecular evolution; Methods for high-throughput genomic, transcriptomic, metabolomic, structural genomic, and proteomic analyses.

Recommended (not required) Courses

It is recommended that all BCB graduate students who have not had laboratory experience in biological sciences take at least two 1-credit modules of BCB 542 (Introduction to Molecular Biology Techniques). Similarly, BCB graduate students who come in with a biology undergraduate degree take at least two modules of Introduction to Bioinformatics Tools (including modules on Sequence Analysis, Microarray Data Analysis, Protein Structure Analysis, Phylogenetics etc. to be developed and offered by the Baker Center).

Sample Course Plans

Well-prepared students

Year I	Fall	Spring
	Bioinformatics I Stat 3xx Rotations	Bioinformatics II Gen 411 Rotations
Year 2	Fall	Spring
	Bioinformatics III	Bioinformatics IV

Year I	Fall	Spring
	Bioinformatics I Stat 3xx Rotations	Bioinformatics II Com S 363 Rotations
Year 2	Fall	Spring
	Bioinformatics III	Bioinformatics IV

Year I	Fall	Spring
	Bioinformatics I Stat 3xx Rotations	Bioinformatics II Gen 411 Rotations
Year 2	Fall	Spring
	Com S 363 Bioinformatics III	Bioinformatics IV

Students Missing Biology Prerequisites, but with strong Computer Science background

Year I	Fall	Spring
	Biol 313 BBMB 301 Stat 3xx Rotations	Gen 411 Com S 363 Rotations
Year 2	Fall	Spring
	Bioinformatics I Biol 315 Bioinformatics III	Bioinformatics II Bioinformatics IV

Students Missing Computer Science Prerequisites, but with strong Biology background

Year I	Fall	Spring
	Com S 207 Stat 3xx Rotations	Com S 208 Com S 330 Gen 411 Rotations
Year 2	Fall	Spring
	Bioinformatics I Com S 363	Bioinformatics II Bioinformatics IV
Year 3	Bioinformatics III	

Year I	Fall	Spring
	Com S 208 Stat 3xx Rotations	Com S 330 Gen 411 Rotations
Year 2	Fall	Spring
	Bioinformatics I Com S 363	Bioinformatics II Bioinformatics IV
Year 3	Bioinformatics III	

Students Missing Statistics Prerequisites AND Computer Science but with strong Biology background

Year I	Fall	Spring
	Com S 207 Stat 341 Rotations	Com S 208 Gen 411 Rotations
Year 2	Fall	Spring
	Stat 3xx Com S 330	Com S 363
Year 3	Bioinformatics I Bioinformatics III	Bioinformatics II Bioinformatics IV