Sample Topics

These are the general areas that students should be prepared to address in the written exam.

I. Computation and Statistics

A. Computer Science

a. Basic data structures (lists, stacks, queues, trees, hash tables)
b. Basic complexity analysis (growth function, NP and NP-complete)
c. Basic database queries and propositional logic
d. Recursion and mathematical induction
e. Basic Algorithms (sorting, Minimum Spanning Trees, Shortest Paths, Graph Traversal, Numerical Optimization)
f. Algorithm Design Principles (Divide and Conquer, Greedy, Approximation and Heuristics, Dynamic Programming)

B. Computational Biology

a. Algorithms used in bioinformatics
b. Basic string matching
c. Sequence alignment
d. Probabilistic String Generative Models (HMMs, stochastic context free grammar)
e. BLAST and other DB search algorithms
f. Phylogeny construction
g. Machine Learning for bioinformatics (Clustering, Support Vector Machines, Neural Nets)
h. Markov Chains

C. Statistics

a. Probability distributions for discrete and continuous random variables.
b. Means, general expectations and variances, conditional expectations, Bayes' rule.
c. Hypothesis testing and confidence intervals for means and proportions.
d. Multiple testing corrections.
e. Linear regression, logistic regression, ANOVA, chi-square goodness-of-fit tests.
f . Permutation tests and simple simulation strategies.
g. Bootstrap methods.
h. Simple clustering techniques (k-means, hierarchical, etc).
i. Introductory stochastic processes, in particular Markov chains and Poisson processes.

II. Genetics/Genomics

A. Molecular Genetics, Biochemistry and Cell Biology

a. Nucleic acids: structure and function
b. Proteins: structure, domain, reactions
c. Molecular basis of gene expression, translation, and regulation
d. Subcellular organelles: structure and functions
e. Signal transduction principles
f. Biochemical pathways

B. The Structure and Transmission of Genetic Information

a. Chromatin and chromosome orgranization
b. Meiosis and mitosis
c. Genetic pathway and analysis/epistasis

C. Genetic Variation and Mapping

a. Polymorphic markers
b. Heterozygosity
c. Meiosis, crossing over, recombination, and genetic maps
d. Quantitative Trait Loci
e. "Forward" and "reverse" genetics
f. Linkage in pedigrees, LOD score, Linkage disequilibrium (LD)
g. Haplotype blocks
h. Mapping by LD

D. DNA Sequencing, Genome Projects and Comparative Sequence Analysis

a. Sequence analysis and databases
b. Genome sequencing and assembly strategies
c. Experimental organisms and human

E. Functional Genomics

a. Genome-wide gene expression technology and analysis.
b. Basic methods and principles of proteomics
c. High-throughput screens

F. Molecular Evolution of Genomes

a. Evolutionary processes (mutation, drift, natural selection)
b. Neutral theory of evolution
c. Comparative genomics (multiple-species sequence comparison, functional inference, gene family evolution)
d. Phylogeny reconstruction