Bioinformatics and Computational Biology

Professor Tandy Warnow
- Understanding of the sequence-structure-function relationship
- Machine learning algorithms for biological network integration
- Prediction of gene/protein function from heterogeneous datasets
- Role of protein mutations in human diseases
- Graphical models, statistical modeling and structured prediction
Saurabh Sinha: gene regulation, big data to knowledge

Two broad areas:
How is information about us encoded in our DNA?
How do we bring the latest and greatest in machine learning and graph mining to the biologist’s desktop computer?

Research questions:

– **Gene regulation**: How are genes turned on and off in precisely orchestrated ways?
– **Regulatory evolution**: Can we model evolution of regulatory sequences?
– **Genomics of behavior**: How does DNA encode animal behavior?
– **Cancer pharmacogenomics**: Can a person’s DNA predict the best drug treatment?
– **Big Data To Knowledge (BD2K)**: Build a “Knowledge Engine for Genomics”.

http://www.sinhalab.net/
Tandy Warnow
The Tree of Life: Multiple Challenges

NP-hard problems
Large datasets:
100,000+ sequences
10,000+ genes
“BigData” complexity

Large-scale statistical phylogeny estimation
Ultra-large multiple-sequence alignment
Estimating species trees from incongruent gene trees
Supertree estimation
Genome rearrangement phylogeny
Reticulate evolution
Visualization of large trees and alignments
Data mining techniques to explore multiple optima
Classes you might take

• BIOE 298 (Introduction to bioinformatics for bioengineering students) – Spring 2017
• CS 173 (Discrete Structures)
• CS 225 (Data Structures)
• CS 361 (Probability and Statistics for CS)
• CS 412 (Data Mining)
• CS 466 (Introduction to Bioinformatics)
After graduation

• Graduate programs in CS, studying computational biology (e.g., MIT, Stanford, UIUC)
• Graduate programs in computational biology (e.g., CMU, Univ Washington, UCLA)
• Pharmaceutical companies (e.g., drug design)
• Government (NIH, NIST, FDA, etc.)