

Intro to Bioinformatics – Midterm Review

Chapter 1: pages 1-12 and assigned reading

The Genetic code

DNA – G A T C

RNA – G A U C

Transcription

DNA → RNA

Translation

RNA → Protein

Codon

Start codon – AUG

Stop codon – UAA, UAG, UGA

Open reading frame – long run of codon with no stop codons

Introns and Exons – Exons are expressed

Splicing

Chapter 2

Dot Plots

Simple Alignments

Gap Penalties

Origination and Length Penalties

Scoring Matrices

Transition

Purine substitution or Pyrimidine substitution – less severe

Transversion

Purine → Pyrimidine or Pyrimidine → Purine substitution – more severe

Dynamic Programming

Needleman-Wunsch – Global alignment

Smith-Waterman – Local Alignment

Semi-Global Alignment – avoid penalizing for terminal gaps

BLAST – break query into words, search for word matches, extend match

Chapter 3

Substitution Patterns

Mutation Rates

Functional Constraint

Synonymous/Nonsynonymous Substitution

Indels – bias against because of reading frame dependency

Pseudogenes

Mutation versus Substitution and the concept of Fixation

Models of evolution

Jukes-Cantor Model

Kimura's Two-parameter Model

Other Models w/more parameters

Chapter 4

Phylogenetics Basics

Tree Terminology

Character versus Distance data

Distance Matrix Methods

UWGMA

Estimation of Branch Lengths

Chapter 5: pages 97 - 108

Character-Based Methods

Parsimony

Informative and uninformative sites

Unweighted parsimony

Weighted parsimony

Advantage of parsimony: Inferred Ancestral Sequences

Branch and Bound – algorithmic strategy for improving the time required for identifying most parsimonious trees

Additional Things to Study

Regular Expression in Perl – see practice sheet on web page