

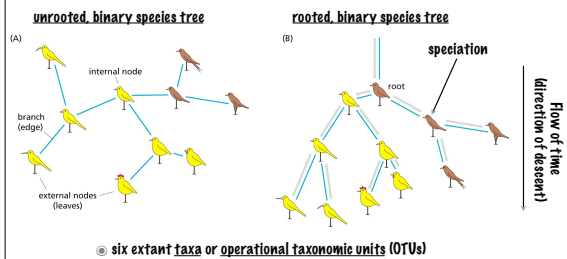
Phylogenetics: Recovering Evolutionary History

COMP 571
Luay Nakhleh, Rice University

1

The Structure and Interpretation of Phylogenetic Trees

2



The Structure and Interpretation of Phylogenetic Trees

3

The Structure and Interpretation of Phylogenetic Trees

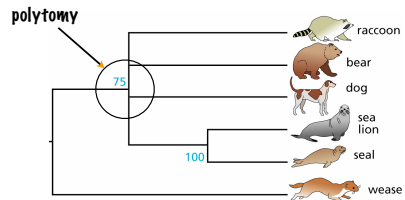
4

- * In a binary tree on n taxa, how many nodes, branches, internal nodes and internal branches are there?
- * How many unrooted binary trees on n taxa are there? How many rooted binary trees on n taxa are there?

• six extant taxa or operational taxonomic units (OTUs)

The Structure and Interpretation of Phylogenetic Trees

5



Non-binary Multifurcating Partially resolved Polytomous

• six extant taxa or operational taxonomic units (OTUs)

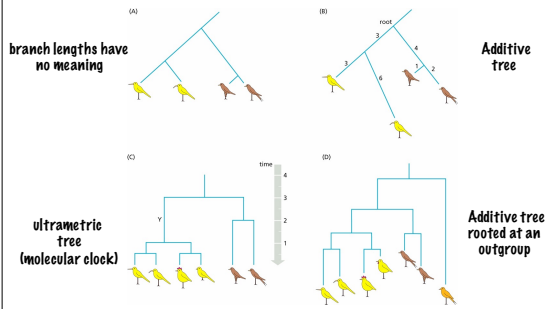
The Structure and Interpretation of Phylogenetic Trees

6

- * A polytomy in a tree can be resolved (not necessarily fully) in many ways, thus producing trees with higher resolution (including binary trees)
- * A binary tree can be turned into a partially resolved tree by contracting edges
- * In how many ways can a polytomy of degree d be resolved?
- * Compatibility between two trees guarantees that one can back and forth between the two trees by means of node refinement and edge contraction

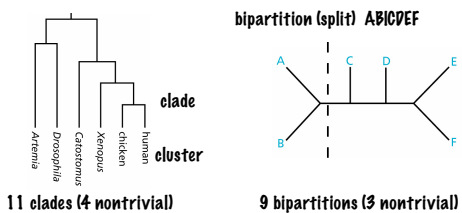
The Structure and Interpretation of Phylogenetic Trees

7



The Structure and Interpretation of Phylogenetic Trees

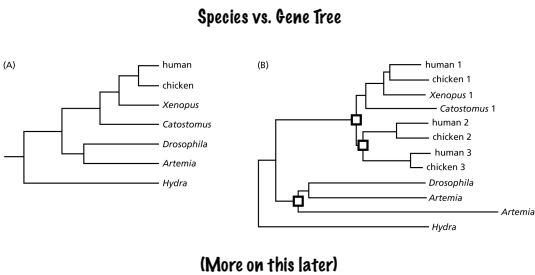
8



How many nontrivial clades are there in a binary tree on n taxa?
How many nontrivial bipartitions are there in a binary tree on n taxa?
How many possible nontrivial clusters of n taxa are there?

The Structure and Interpretation of Phylogenetic Trees

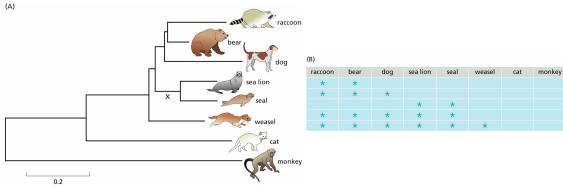
9



The Structure and Interpretation of Phylogenetic Trees

10

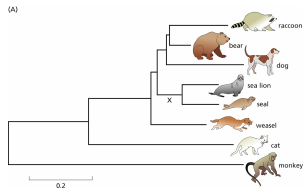
A tree can be represented as a set of splits



The Structure and Interpretation of Phylogenetic Trees

11

The Newick Format

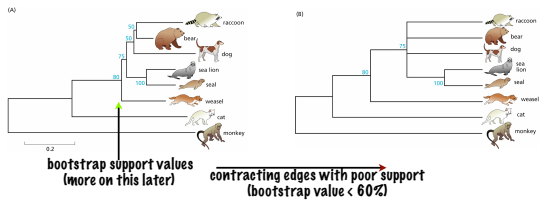


`(monkey,(cat,(weasel,((seal,sea_lion),(dog,(bear,raccoon))))))`;

The Structure and Interpretation of Phylogenetic Trees

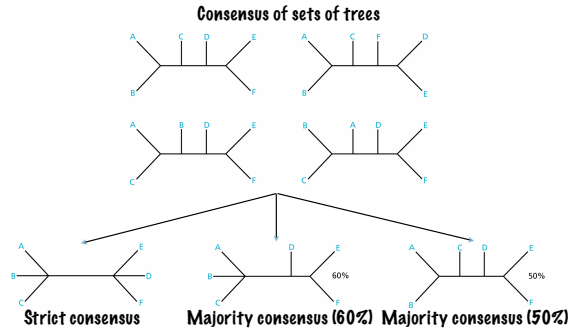
12

Branch support and condensed trees



The Structure and Interpretation of Phylogenetic Trees

13



Molecular Evolution and Its Consequences

14

Molecular Evolution and Its Consequences

15

- * The Darwinian concept of evolution by natural selection concentrates on the consequences of evolutionary changes for the fitness of the organism: its ability to survive and transmit its genes to the next generation by producing offsprings
- * Fitness depends on the properties of the organism as a whole, and thus change at the DNA sequence level will be constrained by considerations of how it affects protein expression and function, and how these affect cellular properties and whole-organism physiology and behavior

Molecular Evolution and Its Consequences

16

- * In the case of sequence-based phylogenetic reconstruction, the changes that occur in genomic DNA are the main focus and, where relevant, their effects on amino acid sequence
- * Changes may affect one or a few nucleotides, entire genes, and even whole genomes

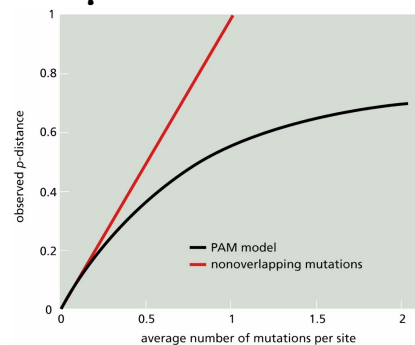
Multiple Mutations at a Site

17

- * Most related sequences have many positions that have mutated several times
- * The p distance between two sequences is the fraction of nonidentical alignment positions
- * The p distance is almost always an underestimate of the number of mutations that actually occurred
- * This distance is, therefore, corrected to reflect the correct evolutionary distance

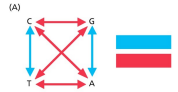
Multiple Mutations at a Site

18

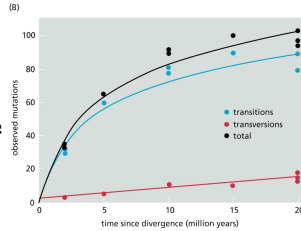


Transitions vs. Transversions

- * The rate of accepted mutation is usually not the same for all types of base substitutions



- Transition mutations have little effect on the DNA structure, and hence are much more commonly observed



- The transition/transversion ratio, R , is defined as the number of transitions per transversion during the evolution of the sequences being studied (more on this later)

19

Synonymous vs. Nonsynonymous Substitutions

- * Another factor that affects the acceptance rates of mutations in protein-coding sequences is the effect of the mutation on the amino acid sequence, and thus, potentially, on the function of the protein
- * Nucleotide mutations that do not change the encoded amino acid are called synonymous mutations (most changes at the third codon position are synonymous)
- * Synonymous mutations are generally considered to be neutral
- * Nucleotide mutations that alter the encoded amino acid are called nonsynonymous mutations

20

- * Evolutionary models exist that can accommodate variations in mutation rates between transitions and transversions (by setting the value of R) as well as variations in mutation rates at different sites (usually by using the gamma correction)
- * In analyzing a data set that involves long evolutionary timescales, it may be useful to remove the third codon sites from the data before further analysis

21

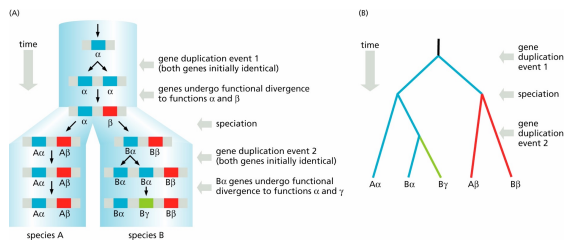
Homology

22

- * The term was introduced by Richard Owen in 1843 to designate "the same organ in different animals under every variety of form and function"
- * The distinction between orthology and paralogy was introduced by Walter Fitch in 1970

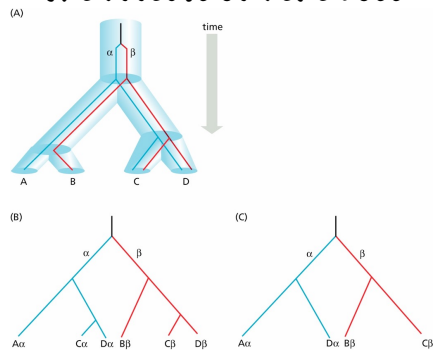
Orthology vs. Paralogy

23



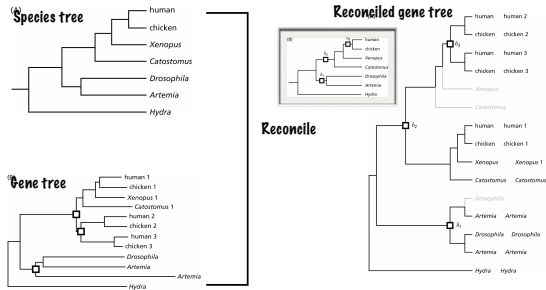
The Effects of Gene Loss

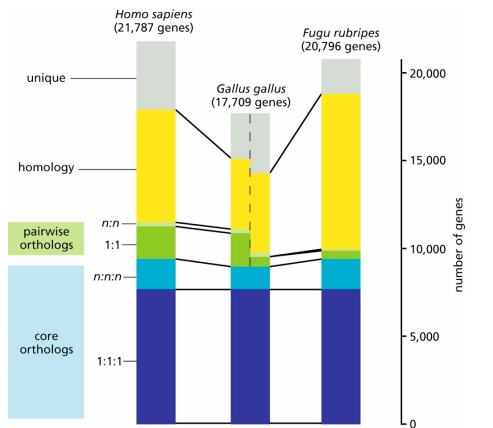
24



Species/Gene Tree Reconciliation

25

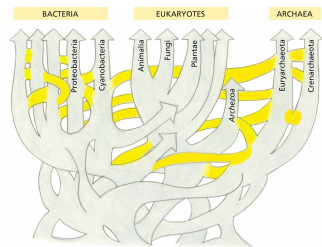




26

Horizontal Gene Transfer

27



The genes in the donor and recipient species are called xenologues

28

* **Common practice:** Only orthologous genes should be used to construct species phylogenetic trees

* **Caveats:**

* Sometimes even the gene tree of orthologous genes disagrees with the species tree (e.g., incomplete lineage sorting)

* Events such as gene duplication, loss, and HGT do not necessarily result in gene trees that disagree with the species tree

* Under certain evolutionary scenarios (such as hybrid speciation), a tree does not exist; rather, the species phylogeny is a directed, acyclic graph (known as phylogenetic network)

29

Questions?
