Phylogenetic tree

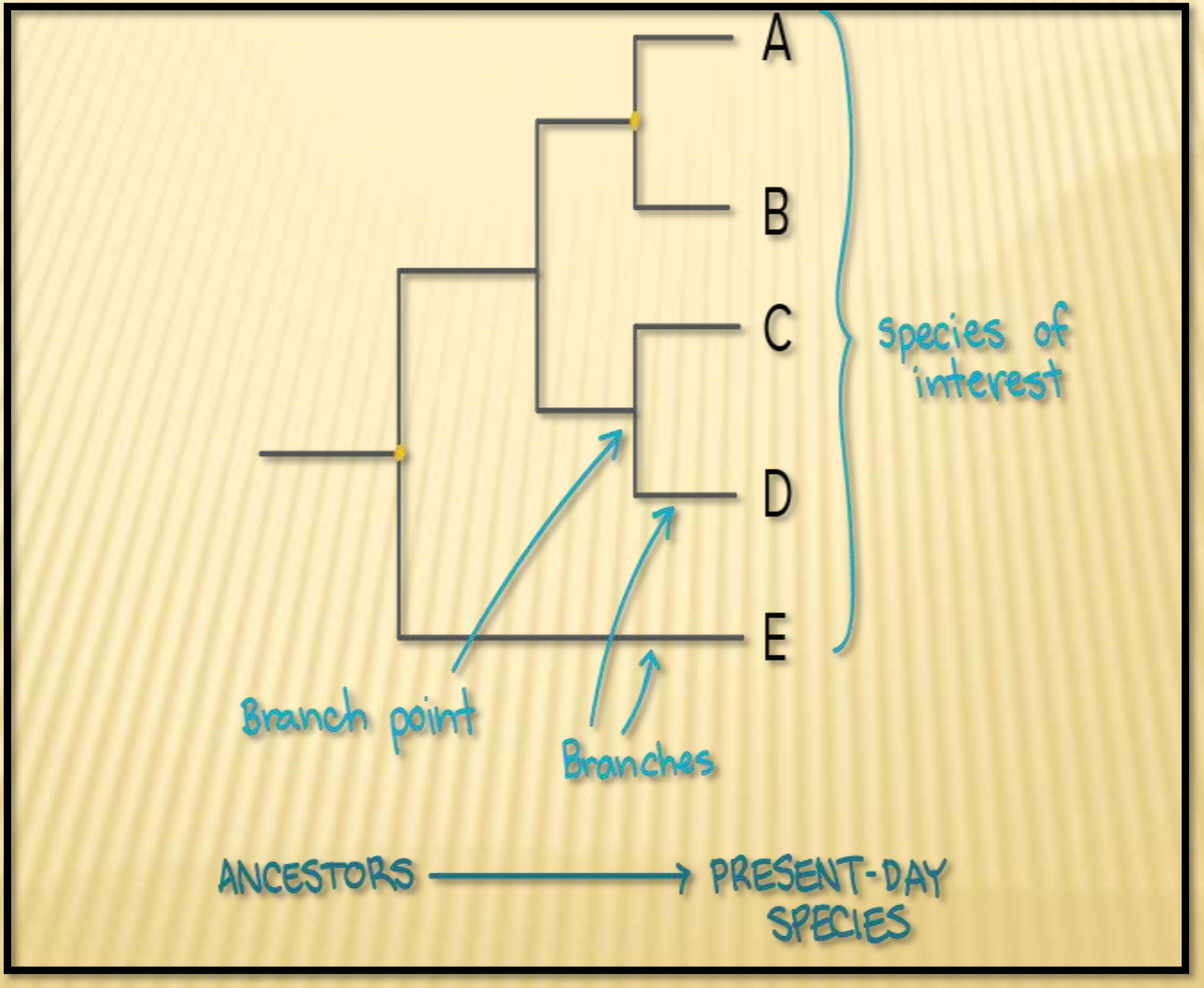
- A **phylogenetic tree** or **evolutionary tree** is a branching diagram or "tree" showing the evolutionary relationships among various biological species or other entities—their **phylogeny** ---based upon similarities and differences in their physical or genetic characteristics. All life on Earth is part of a single phylogenetic tree, indicating common ancestry.

- A **phylogenetic tree** is a diagram that represents evolutionary relationships among organisms. **Phylogenetic trees are hypotheses, not definitive facts.**
- The pattern of branching in a phylogenetic tree reflects how species or other groups evolved from a series of common ancestors.

- In trees, two species are **more related** if they have a more recent common ancestor and **less related** if they have a less recent common ancestor.
- Phylogenetic trees can be drawn in various equivalent styles. Rotating a tree about its branch points doesn't change the information it carries.

ANATOMY OF A PHYLOGENETIC TREE

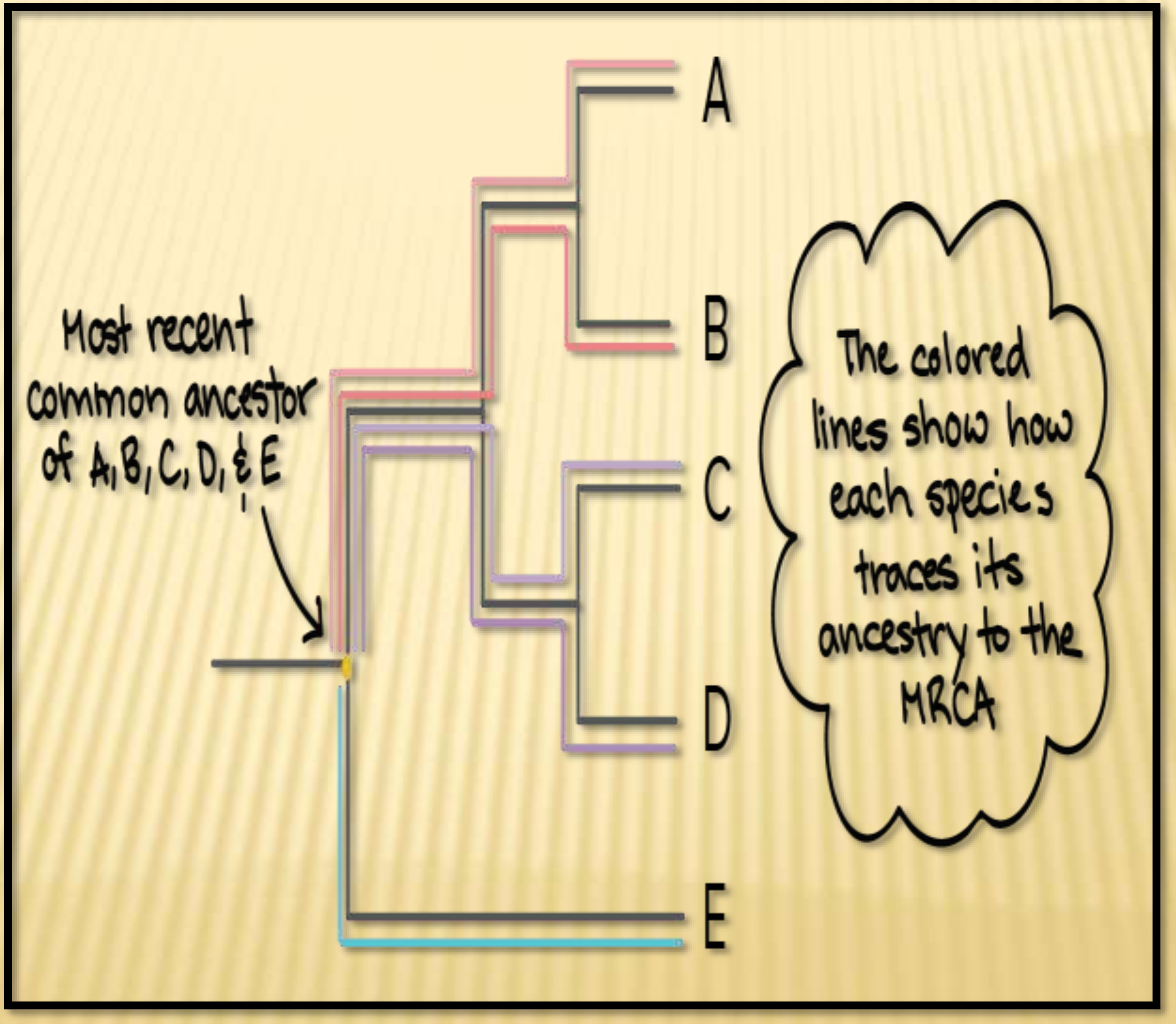
- In a phylogenetic tree, the species or groups of interest are found at the tips of lines referred to as the tree's **branches.** For example, the phylogenetic tree below represents relationships between five species, A, B, C, D, and E, which are positioned at the ends of the branches:


A
B
C
D
E
species of interest
Branch point
Branches
ANCESTORS
PRESENT-DAY SPECIES

- The pattern in which the branches connect represents our understanding of how the species in the tree evolved from a series of common ancestors. Each branch point (also called an **internal node**) represents a **divergence** event, or splitting apart of a single group into two descendant groups.

- At each branch point lies the **most recent common ancestor** of all the groups descended from that branch point. For instance, at the branch point giving rise to species A and B, we would find the most recent common ancestor of those two species.

- At the branch point right above the root of the tree, we would find the most recent common ancestor of all the species in the tree (A, B, C, D, E).
- The diagram below shows how each species in the tree can trace its ancestry back to the most recent common ancestor at the branch point above the root:


A
B
C
D
E
Most recent common ancestor of A, B, C, D, & E
The colored lines show how each species traces its ancestry to the MRCA

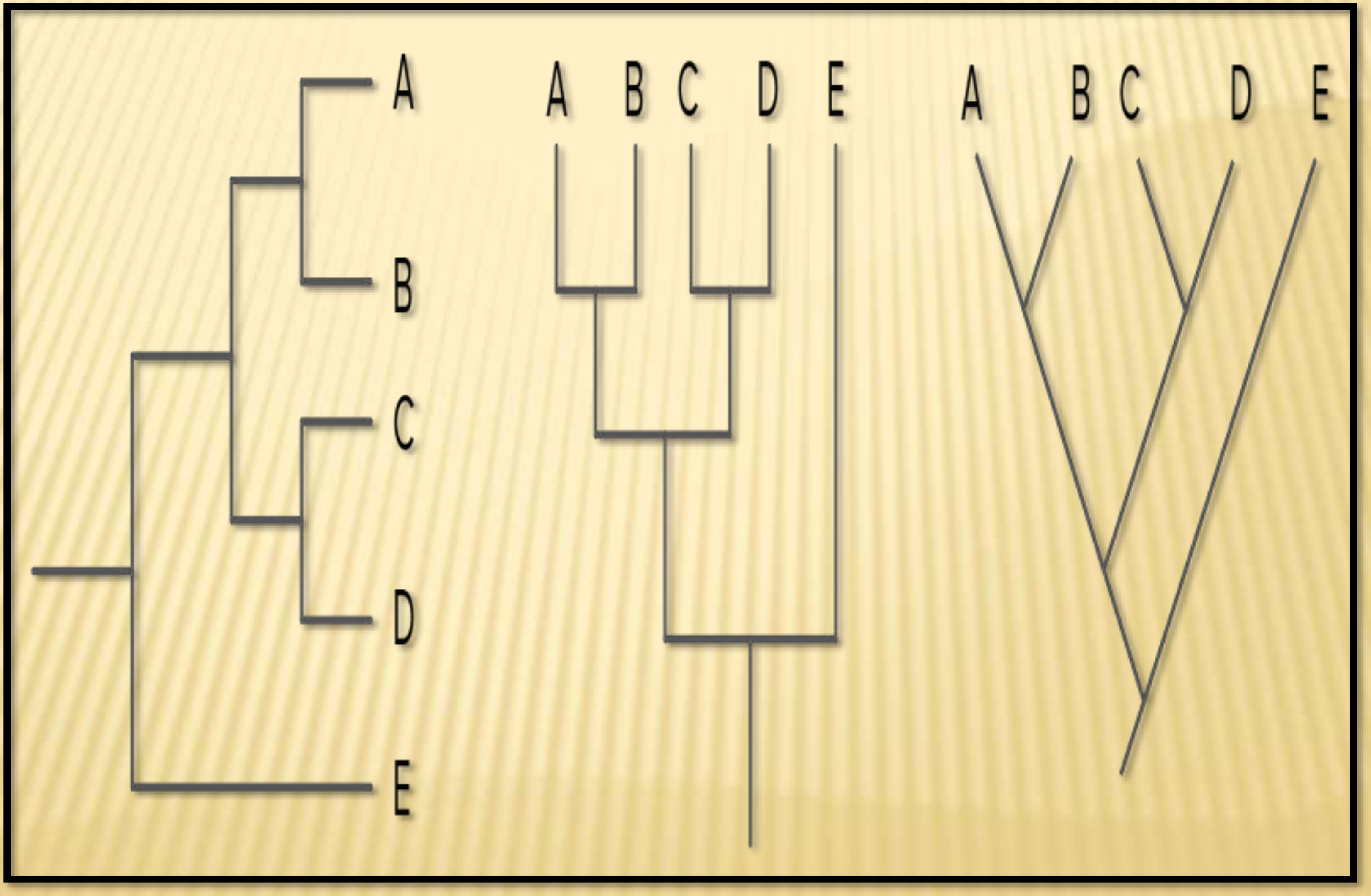
- Each **horizontal line** in our tree represents a series of ancestors, leading up to the species at its end. For instance, the line leading up to species E represents the species' ancestors since it diverged from the other species in the tree. Similarly, the root represents a series of ancestors leading up to the most recent common ancestor of all the species in the tree

WHICH SPECIES ARE MORE RELATED?

- In a phylogenetic tree, the **relatedness** of two species has a very specific meaning. Two species are *more* related if they have ***a more recent* common ancestor,** and *less* related if they have a *less recent* common ancestor.
- We can use a pretty straightforward method to find the most recent common ancestor of any pair or group of species. In this method, we start at the branch ends carrying the two species of interest and “walk backwards” in the tree until we find the point where the species’ lines converge.

- For instance, suppose that we wanted to say whether A and B or B and C are more closely related. To do so, we would follow the lines of both pairs of species backward in the tree. Since A and B converge at a common ancestor first as we move backwards, and B only converges with C after its junction point with A, we can say that A and B are more related than B and C.

Some tips for reading phylogenetic trees



- You may see phylogenetic trees drawn in many different formats. Some are blocky, like the tree at left below. Others use diagonal lines, like the tree at right below. You may also see trees of either kind oriented vertically or flipped on their sides, as shown for the blocky tree.

- The three trees above represent identical relationships among species A, B, C, D, and E. You may want to take a moment to convince yourself that this is really the case – that is, that no branching patterns or recent-ness of common ancestors are different between the two trees. The identical information in these different-looking trees reminds us that it's the branching pattern (and not the lengths of branches) that's meaningful in a typical tree.

- Another critical point about these trees is that if you rotate the structures, using one of the branch points as a pivot, you don't change the relationships. So just like the two trees above, which show the same relationships even though they are formatted differently, all of the trees below show the same relationships among four species:

- In a *rooted* phylogenetic tree, each node with descendants represents the inferred most recent common ancestor of those descendants, and the edge lengths in some trees may be interpreted as time estimates. Each node is called a taxonomic unit. Internal nodes are generally called hypothetical taxonomic units, as they cannot be directly observed. Trees are useful in fields of biology such as bioinformatics, systematics, and phylogenetics.
- *Unrooted* trees illustrate only the relatedness of the leaf nodes and do not require the ancestral root to be known or inferred.

TYPES

- A rooted phylogenetic tree is a directed tree with a unique node — the root — corresponding to the (usually imputed) most recent common ancestor of all the entities at the leaves of the tree. The root node does not have a parent node, but serves as the parent of all other nodes in the tree.

- The root is therefore a node of degree 2 while other internal nodes have a minimum degree of 3 (where "degree" here refers to the total number of incoming and outgoing edges).
- The most common method for rooting trees is the use of an uncontroversial outgroup—close enough to allow inference from trait data or molecular sequencing, but far enough to be a clear outgroup.

- **Unrooted tree**
- Unrooted trees illustrate the relatedness of the leaf nodes without making assumptions about ancestry. They do not require the ancestral root to be known or inferred.
- Unrooted trees can always be generated from rooted ones by simply omitting the root. By contrast, inferring the root of an unrooted tree requires some means of identifying ancestry.]

- This is normally done by including an outgroup in the input data so that the root is necessarily between the outgroup and the rest of the taxa in the tree, or by introducing additional assumptions about the relative rates of evolution on each branch, such as an application of the molecular clock hypothesis.

- **Bifurcating tree**
- Both rooted and unrooted phylogenetic trees can be either bifurcating or multifurcating, and either labeled or unlabeled. A rooted bifurcating tree has exactly two descendants arising from each interior node (that is, it forms a binary tree), and an unrooted bifurcating tree takes the form of an unrooted binary tree, a free tree with exactly three neighbors at each internal node. In contrast, a rooted multifurcating tree may have more than two children at some nodes and an unrooted multifurcating tree may have more than three neighbors at some nodes.

- A labeled tree has specific values assigned to its leaves, while an unlabeled tree, sometimes called a tree shape, defines a topology only. The number of possible trees for a given number of leaf nodes depends on the specific type of tree, but there are always more multifurcating than bifurcating trees, more labeled than unlabeled trees, and more rooted than unrooted trees. The last distinction is the most biologically relevant; it arises because there are many places on an unrooted tree to put the root.

concepts

- A <u>dendrogram</u> is a general name for a tree, whether phylogenetic or not, and hence also for the diagrammatic representation of a phylogenetic tree.
- A <u>cladogram</u> only represents a branching pattern; i.e., its branch spans do not represent time or relative amount of character change, and its internal nodes do not represent ancestors.
- <u>A phylogram</u> is a phylogenetic tree that has branch spans proportional to the amount of character change.
- <u>A chronogram</u> is a phylogenetic tree that explicitly represents time through its branch spans.

- A spindle diagram (often called a Romerogram after the American palaeontologist Alfred Romer) is the representation of the evolution and abundance of the various taxa through time, but is not an evolutionary tree.
- A Dahlgrenogram is a diagram representing a cross section of a phylogenetic tree
- A phylogenetic network is not strictly speaking a tree, but rather a more general graph, or a directed acyclic graph in the case of rooted networks. They are used to overcome some of the limitations inherent to trees

Computational phylogenetics

- Phylogenetic trees composed with a nontrivial number of input sequences are constructed using computational phylogenetics methods. Distance-matrix methods such as neighbor-joining or UPGMA, which calculate genetic distance from multiple sequence alignments, are simplest to implement, but do not invoke an evolutionary model. Many sequence alignment methods such as ClustalW also create trees by using the simpler algorithms (i.e. those based on distance) of tree construction.

- Maximum parsimony is another simple method of estimating phylogenetic trees, but implies an implicit model of evolution (i.e. parsimony). More advanced methods use the optimality criterion of maximum likelihood, often within a Bayesian Framework, and apply an explicit model of evolution to phylogenetic tree estimation.[6] Identifying the optimal tree using many of these techniques is NP-hard,] so heuristic search and optimization methods are used in combination with tree-scoring functions to identify a reasonably good tree that fits the data.

- **Tree-building methods can be assessed on the basis of several criteria:[**
- efficiency (how long does it take to compute the answer, how much memory does it need?)
- power (does it make good use of the data, or is information being wasted?)
- consistency (will it converge on the same answer repeatedly, if each time given different data for the same model problem?)
- robustness (does it cope well with violations of the assumptions of the underlying model?)
- falsifiability (does it alert us when it is not good to use, i.e. when assumptions are violated?)
- Tree-building techniques have also gained the attention of mathematicians. Trees can also be built using T-theory.

Questions