



Classification and cladistics

Revision notes · IB Biology A3.2 (HL): why we classify, the traditional hierarchy, clades, molecular clocks, building and analysing cladograms, and the three domains.

For IB Diploma Biology, Standard Level and Higher Level (first assessment 2025). **HL only** is additional higher level content, for HL students only.

Classification and cladistics **Spec A3.2** **HL only**

Why classify, and problems with the traditional hierarchy **Spec A3.2.1, A3.2.2, A3.2.3**

The need for classification

Classification is needed because of the **immense diversity of species**: well over a million have been named and many more remain undescribed. Sorting species into groups makes this diversity manageable. **After classification is completed, a broad range of further study is facilitated**: researchers can find information about related species, identify new specimens, compare groups and plan conservation.

The traditional hierarchy of taxa

The traditional hierarchy, using the lion as an example	
Taxon	Lion
Kingdom	Animalia
Phylum	Chordata
Class	Mammalia
Order	Carnivora
Family	Felidae
Genus	<i>Panthera</i>
Species	<i>leo</i>

This hierarchy **does not always correspond to patterns of divergence generated by evolution**. Evolution produces branching with gradual variation, not a fixed number of equal levels. Deciding whether a group is a family or an order is a matter of judgement, and taxa at the same rank in different groups are not equivalent in age or diversity.

Some traditional groups do not include all the descendants of their common ancestor. The class Reptilia excludes birds, yet birds evolved from within the reptiles: birds are more closely related to crocodiles than crocodiles are to lizards.

Exam tip: Nature of science: a fixed ranking of taxa is arbitrary because it does not reflect the gradation of variation. Cladistics replaces it with unranked clades, an example of a paradigm shift in scientific theories.

Advantages of classifying by evolutionary relationships

- In the ideal classification, **all members of a group have evolved from a common ancestor**.

- Characteristics of organisms in such a group **can be predicted** because they are shared within a clade. If one species in a clade makes a useful chemical, related species may make it too.
- The classification reflects real evolutionary history rather than human judgement about which traits matter.
- It can be tested and updated as new sequence evidence appears.

Clades and the molecular clock Spec A3.2.4, A3.2.5

Clade	A group of organisms with common ancestry and shared characteristics: a common ancestor and all of its descendants.
Molecular clock	A method of estimating when clades diverged from a common ancestor, using the number of differences in base or amino acid sequences.

Evidence for clades

The **most objective evidence** for placing organisms in the same clade comes from **base sequences of genes** or **amino acid sequences of proteins**. Sequences can be compared position by position and the differences counted, so the evidence is quantitative.

Morphological traits can also be used to assign organisms to clades, but they are less reliable. Similar structures can evolve separately in unrelated groups (convergent evolution), as in the wings of bats and birds.

The molecular clock

After two lineages split, mutations build up in each independently. Differences in sequence therefore **accumulate gradually**: the more differences between two species, the longer ago they shared a common ancestor. If the rate of change is known, for example calibrated using a fossil of known age, the number of differences can be used to estimate when the clades diverged.

The molecular clock can **only give estimates**, because mutation rates are affected by:

- the length of the **generation time** (more generations per year means more DNA replication and more chances of mutation);
- the **size of the population**;
- the **intensity of selective pressure** on the gene;
- other factors, such as differences between genes in how freely their sequence can change.

Using a molecular clock

Two species differ at 18 positions in the amino acid sequence of a protein. A fossil-calibrated rate for this protein is 1 difference per 4 million years (a practice value). Estimate when the species diverged.

1. Time since divergence = number of differences × time per difference
2. = 18×4 million years
3. = 72 million years

Answer: About 72 million years ago. This is only an estimate, as the rate of change may not have been constant.

Exam tip: Linking question: how can similarities between distantly related organisms be explained? By convergent evolution, which is why sequence data is preferred to morphology for defining clades.

Constructing and analysing cladograms Spec A3.2.6, A3.2.7

Cladogram	A tree diagram that shows the most probable sequence of divergence in clades.
Root	The base of the cladogram, representing the common ancestor of all the organisms in it.
Node	A branching point, representing a hypothetical common ancestor that split into two lineages.
Terminal branch	A branch that ends at a species (or group) being compared, at the tips of the cladogram.

Constructing a cladogram from sequences

Base sequences of genes or amino acid sequences of proteins from the organisms are aligned and compared. Organisms with fewer differences are placed closer together. Many trees could fit the data, so parsimony analysis selects the most probable cladogram: the one in which the observed sequence variation is accounted for with the smallest number of sequence changes.

Aligned sample sequences (practice data); O is an outgroup used to root the tree						
Organism	Base 1	Base 2	Base 3	Base 4	Base 5	Base 6
O	A	C	G	T	A	C
W	A	C	G	T	A	G
X	T	C	G	A	A	G
Y	T	C	C	A	A	G
Z	T	C	C	A	T	G

Choosing the most parsimonious cladogram

Use the sequences to deduce the most parsimonious cladogram for W, X, Y and Z.

1. Base 6 (C→G) is shared by W, X, Y and Z: one change at the node joining all four.
2. Bases 1 and 4 (A→T, T→A) are shared by X, Y and Z: two changes at the node joining X, Y and Z.
3. Base 3 (G→C) is shared by Y and Z: one change at the node joining Y and Z.
4. Base 5 (A→T) is found only in Z: one change on the terminal branch to Z.
5. Total = 5 changes, one per variable base, which is the minimum possible. Any other tree, for example one grouping W with Z, needs extra changes.

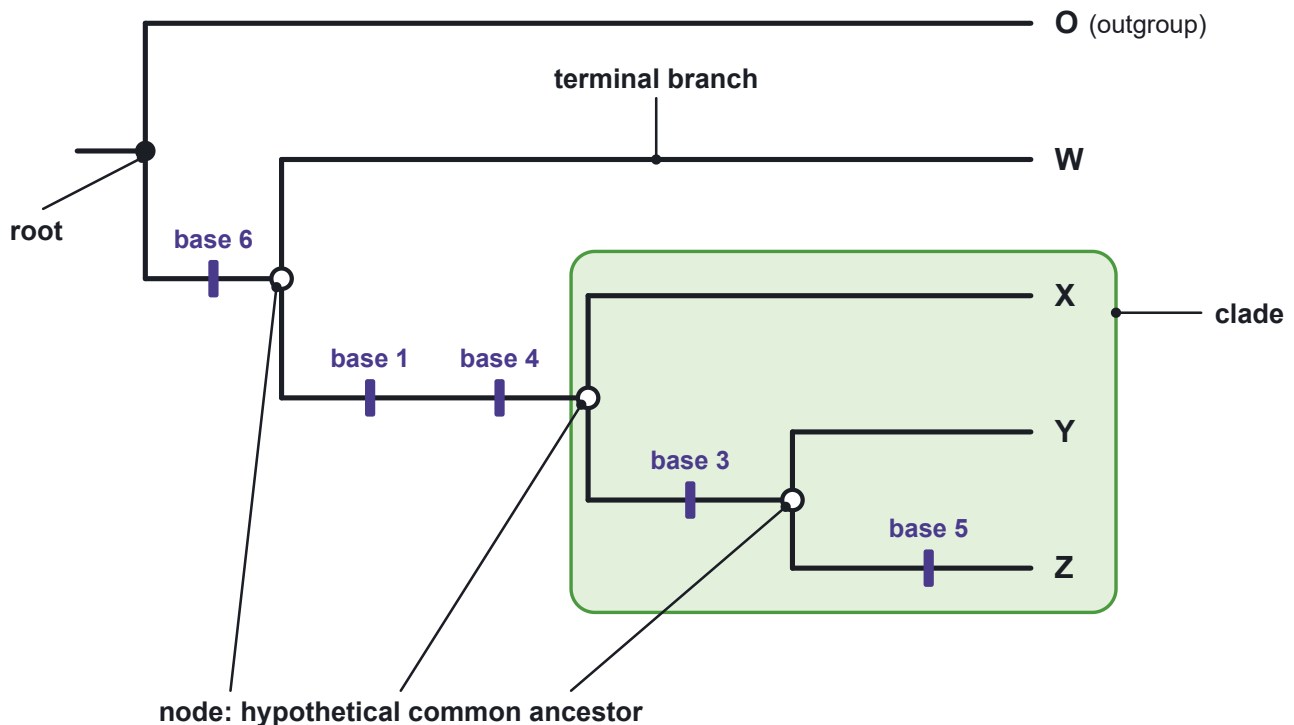
Answer: O branches off first, then W, then X; Y and Z are the closest relatives (sister species).

Exam tip: Nature of science: different criteria for judgement can lead to different hypotheses. Parsimony is one criterion, so a cladogram is a hypothesis that can change with more data.

Analysing cladograms

- Each node represents a hypothetical common ancestor; the root is the common ancestor of every organism shown.
- A clade is any node together with everything that branches from it.
- Two species are most closely related if they share the most recent common ancestor (the node nearest the tips).

- Branches can be rotated about a node without changing the relationships, so the order of species along the tips does not matter.
- If branch lengths are drawn to scale, they show the amount of sequence change or the estimated time since divergence.



[\(opens full size in a new tab\)](#)

The most parsimonious cladogram: **five sequence changes**, one per variable base. Y and Z share the most recent common ancestor.

Common mistake: Species next to each other at the tips are not necessarily closest relatives. Trace back to the node they share.

Testing classifications and the three domains Spec A3.2.8, A3.2.9

Using cladistics to test classifications

If a traditional group is a true clade, the species in it should form one branch on a cladogram built from sequence data. If they are spread across different branches, the classification does **not correspond to evolutionary relationships** and should be revised.

The figwort family (**Scrophulariaceae**) is an example of the **transfer of plant species between families**. It was traditionally defined by flower and fruit structure and held about 275 genera and over 5,000 species. Cladograms based on chloroplast gene sequences (published around 1995-2001) showed it was not a single clade but several separate ones, so many genera were moved: snapdragons, foxgloves and speedwells to the plantain family (Plantaginaceae), and many parasitic genera to the broomrape family (Orobanchaceae); some went to new or other families. A few genera from other families, such as *Buddleja*, were moved in. The figwort family is now much smaller, with roughly 60 genera. The details of this case do not need to be memorised.

Exam tip: Nature of science: knowledge claims may eventually be falsified. Similar morphology due to convergent evolution, not common ancestry, suggested a classification that cladistics has shown to be false.

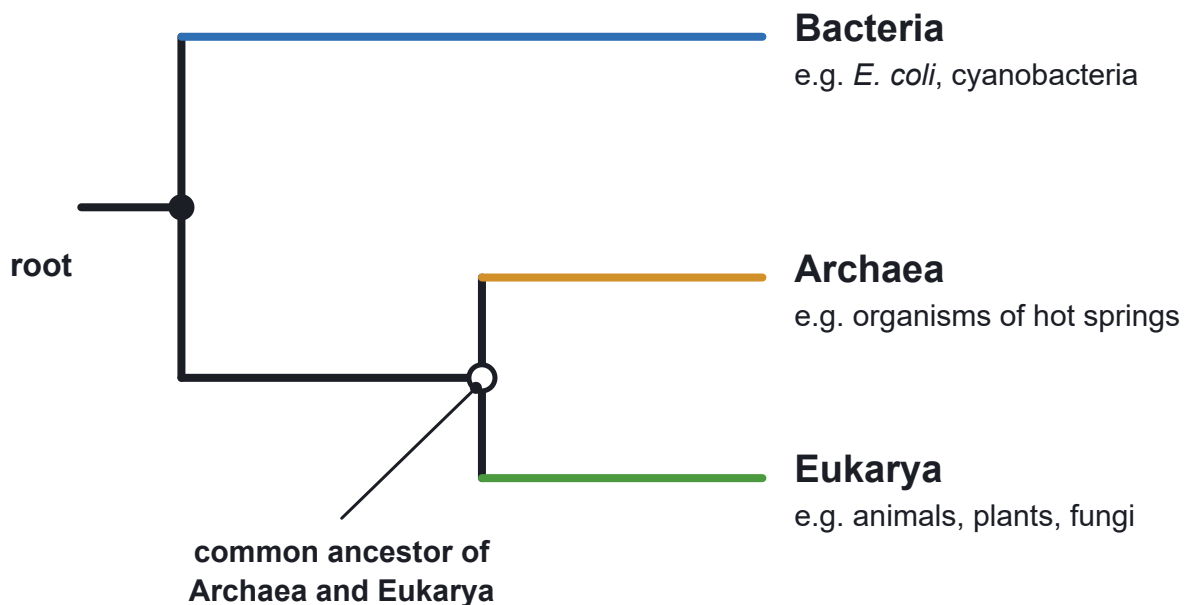
The three domains

Living things used to be split into prokaryotes and eukaryotes. In 1977, base sequences of ribosomal RNA (rRNA) were compared across many organisms. rRNA is found in all cells and changes slowly, so it is suited to comparing very distantly related organisms. The results showed that prokaryotes form two groups as different from each other as either is from eukaryotes. Woese and Fox proposed this in 1977; the name "domain" for the new top level was introduced later, in 1990.

This led to the revolutionary reclassification of all organisms into three domains, an extra taxonomic level above kingdoms:

The three domains		
Domain	Cell type	Examples
Bacteria	prokaryotic	<i>Escherichia coli</i> , cyanobacteria
Archaea	prokaryotic	many found in extreme habitats, such as hot springs and very salty water, and also in soil and oceans
Eukarya	eukaryotic	animals, plants, fungi and protists

rRNA sequences also suggest that archaea are more closely related to eukaryotes than to bacteria.



[\(opens full size in a new tab\)](#)

The three domains, based on rRNA base sequences (1977 onwards): archaea are more closely related to eukaryotes than to bacteria.

Quick check

1. List the traditional hierarchy of taxa from largest to smallest.
2. What does a node on a cladogram represent?
3. What is parsimony analysis?
4. Name the three domains.
5. Why can a molecular clock only give estimates?

Quick check answers

Classification and cladistics

- 1 . Kingdom, phylum, class, order, family, genus, species.
- 2 . A hypothetical common ancestor.
- 3 . Choosing the cladogram that accounts for the observed sequence variation with the smallest number of sequence changes.
- 4 . Bacteria, Archaea and Eukarya.
- 5 . Mutation rates vary with generation time, population size, intensity of selective pressure and other factors.