



MEGA
LECTURE

IB DP BIOLOGY

Theme A: Unity and Diversity

A3.2 Classification and Cladistics

Revision Notes · Standard and Higher Level

Fahad H. Ahmad

+92 323 509 4443 | Megalecture.com

What the syllabus requires

By the end of A3.2 you should be able to work confidently with each of the following. Cladistics is largely Higher Level, so items marked **(HL)** are assessed only at HL. Use this list as a final checklist before the exam.

Understanding	You should be able to...
The hierarchy of taxa	Name the taxa from domain to species in order and place an organism within them.
Binomial nomenclature	State and apply the rules for writing a scientific name correctly.
The three domains	Identify Archaea, Bacteria and Eukarya and outline the evidence for separating them.
Natural classification	Explain why classification should reflect evolutionary relationships (common ancestry).
Cladistics (HL)	Define a clade and explain how shared derived characteristics and molecular data reveal ancestry.
Cladograms (HL)	Read nodes and branch points, and relate base-sequence differences to time since divergence.
Reclassification (HL)	Explain how new molecular evidence can overturn groups based on morphology.
Analogy vs homology	Distinguish convergent evolution from homologous structures and its effect on classifying.

Exam note: The core hierarchy, binomial rules, three domains and natural classification are common to SL and HL. The mechanics of cladistics and cladograms are HL only, but the underlying idea — that classification should mirror evolution — is examinable at both levels.

1. Why we classify: taxonomy and the hierarchy of taxa

There are millions of known species and many more undescribed. Without an agreed system, biologists working in different countries and languages could not be sure they were discussing the same organism.

Taxonomy is the science of naming, describing and grouping organisms; **classification** is the act of assigning an organism to those groups.

Classification serves several practical purposes:

- It gives every organism a single, universally understood name, avoiding the confusion of local common names.
- It organises biodiversity so that information can be stored, retrieved and shared efficiently.
- It lets biologists predict shared features: if a new species is placed in a familiar group, we can anticipate much of its biology.
- A well-built system reveals evolutionary relationships, showing how life has diversified from common ancestors.

The hierarchy of taxa

Organisms are placed in a nested hierarchy of groups called **taxa** (singular: taxon). Each level is contained within the one above it, so a small, tightly related group sits inside progressively larger, more inclusive groups. The principal ranks, from most inclusive to least, are:

Rank	Meaning	Example (humans)
Domain	Broadest division of life, based on cell type and biochemistry	Eukarya
Kingdom	Major group within a domain	Animalia
Phylum	Body plan / major lineage	Chordata
Class	Large group sharing key features	Mammalia
Order	Group of related families	Primates
Family	Group of related genera	Hominidae
Genus	Group of closely related species	<i>Homo</i>
Species	Organisms that interbreed to give fertile offspring	<i>Homo sapiens</i>

As you move down the hierarchy the groups become smaller, contain fewer organisms, and those organisms share more features in common. As you move up, groups become larger and more inclusive but their members share fewer characteristics.

Memory aid: To recall Domain, Kingdom, Phylum, Class, Order, Family, Genus, Species in order, use the sentence **Dear King Philip Came Over For Good Soup**. The initials give the eight ranks from the broadest to the narrowest.

Binomial nomenclature

Every species is given a two-part Latinised name — the **binomial** — first standardised by Carl Linnaeus. The first part is the **genus**, the second is the **species epithet**. Together they identify the species uniquely. Strict rules govern how a binomial is written:

Rule	Correct	Incorrect
Genus name begins with a capital letter	<i>Panthera</i>	<i>panthera</i>
Species epithet is entirely lower case	<i>leo</i>	<i>Leo</i>
The whole name is italicised (or underlined if handwritten)	<i>Panthera leo</i>	Panthera leo
After first use the genus may be abbreviated to its initial	<i>P. leo</i>	<i>Pan. leo</i>
Latin or Latinised, and the same worldwide	<i>Homo sapiens</i>	human / man

So the lion is *Panthera leo*, the tiger *Panthera tigris* and the leopard *Panthera pardus*: the shared genus name signals that these cats are closely related, while the differing epithets keep the species distinct.

Exam technique: Marks are routinely awarded for correct binomial format. In a written exam, always underline a scientific name (you cannot italicise by hand), capitalise the genus and keep the species name lower case.

2. The three domains of life

The broadest level of classification divides all cellular life into three **domains**: **Archaea**, **Bacteria** and **Eukarya**. This three-domain system, proposed by Carl Woese, replaced the older idea of a single prokaryote group once molecular evidence showed that archaea and bacteria, though both prokaryotic in appearance, are profoundly different at the molecular level.

Feature	Bacteria	Archaea	Eukarya
Cell type	Prokaryotic	Prokaryotic	Eukaryotic (nucleus present)
Membrane lipids	Unbranched fatty acids, ester bonds	Branched hydrocarbons, ether bonds	Unbranched fatty acids, ester bonds
Cell wall	Contains peptidoglycan	No peptidoglycan	Cellulose, chitin, or none
Histone proteins on DNA	Absent	Present (like eukaryotes)	Present
Example	<i>Escherichia coli</i>	Methanogens, halophiles	Plants, animals, fungi, protists

Evidence for the three domains

The separation is not based on appearance but on molecular and biochemical data:

- **Ribosomal RNA (rRNA) sequences.** The base sequence of the small-subunit rRNA gene is highly conserved yet differs consistently between the three groups. Comparing these sequences was the key evidence that archaea form a distinct lineage.
- **Membrane lipid chemistry.** Archaeal membranes use branched hydrocarbon chains linked to glycerol by ether bonds, whereas bacteria and eukaryotes use unbranched fatty acids linked by ester bonds — a fundamental biochemical difference.
- **Other biochemistry.** Archaea lack peptidoglycan in their cell walls, wrap their DNA around histone-like proteins, and have RNA-polymerase and ribosome features that in several respects resemble eukaryotes more than bacteria.

How Eukarya are organised into kingdoms

Within the domain Eukarya, organisms are traditionally divided into **kingdoms**. The commonly taught kingdoms are Animalia (multicellular heterotrophs, no cell wall), Plantae (multicellular autotrophs with cellulose walls and chloroplasts), Fungi (heterotrophs that absorb nutrients, with chitin walls) and the diverse, mostly single-celled Protista. Molecular studies continue to refine these groups, and the protists in particular are now known not to form a single natural group; for IB purposes the key point is that kingdoms sit *within* the domain Eukarya.

Watch out: Archaea and Bacteria look similar under a microscope — both are small prokaryotes — but their molecular differences are as great as the difference between either of them and eukaryotes. Appearance alone would misclassify them; molecules reveal the true picture.

3. Natural versus artificial classification

Classifications can be built in two very different ways, and the distinction is a favourite exam theme.

	Artificial classification	Natural classification
Basis	One or a few convenient, observable features chosen for ease of use	Evolutionary relationships — shared ancestry
Aim	Practical identification	To reflect the true history of life
Example	Grouping all flying animals (birds, bats, insects) together	Grouping organisms by descent from a common ancestor
Stability	May group unrelated organisms; often revised for convenience	Groups reflect real lineages; revised as evidence improves

An **artificial** system groups organisms by superficial similarity — for example, lumping together everything that lives in water, or everything that flies. Such groups are easy to use but can be biologically misleading: a bat, a bird and a butterfly all fly, yet they are only distantly related.

A **natural** classification instead aims to group organisms according to their evolutionary relationships, so that each taxon consists of organisms descended from a **common ancestor**. The modern goal of taxonomy is that every named group should be a natural group, reflecting a real branch of the tree of life. This is why classifications are revised when new evidence — especially DNA and protein sequence data — reveals relationships that were previously hidden.

Key principle: A good classification should reflect evolution. Organisms are placed together because they share a common ancestor, not merely because they look alike. This principle underpins everything in the rest of this topic.

4. Cladistics (HL)

Cladistics is the method used to build a natural classification from evidence of common ancestry. It groups organisms into **clades**. A **clade** is a group consisting of a common ancestor *and all of its descendants* — a complete branch of the evolutionary tree.

Shared derived characteristics

Clades are recognised using **shared derived characteristics**: features that arose in a common ancestor and are inherited by all its descendants. Because they were passed down from that ancestor, they mark out the members of a clade. It is important to distinguish these from ancestral features shared more widely:

- A **shared derived character** is a novel feature that evolved in the ancestor of a clade (for example, feathers in the ancestor of birds). It defines the clade.

- A **shared ancestral character** is an older feature present before the clade arose (for example, a backbone, shared by all vertebrates). It cannot, on its own, distinguish smaller groups within.

Molecular evidence

Modern cladistics relies heavily on **molecular evidence** — the sequences of bases in DNA (or RNA) and of amino acids in proteins. The reasoning is simple: as species diverge from a common ancestor, mutations accumulate independently in each lineage. The **more differences** there are between the sequences of two species, the **longer** they have been evolving separately, and so the more distantly related they are. Closely related species have very similar sequences.

Molecular data are powerful because they are:

- Quantitative — differences can be counted objectively rather than judged by eye.
- Abundant — a genome provides an enormous number of characters to compare.
- Universal — the same molecules (e.g. rRNA, cytochrome c, haemoglobin) can be compared across very different organisms.

The molecular clock

The **molecular clock** is the idea that, for a given gene or protein, mutations accumulate at a roughly **constant average rate** over long periods of time. If the rate is known, the number of differences between two sequences can be used to estimate the **time since they diverged** from their common ancestor. In effect, the accumulated mutations act like the ticking of a clock. The clock is calibrated using events whose dates are known independently, such as fossils.

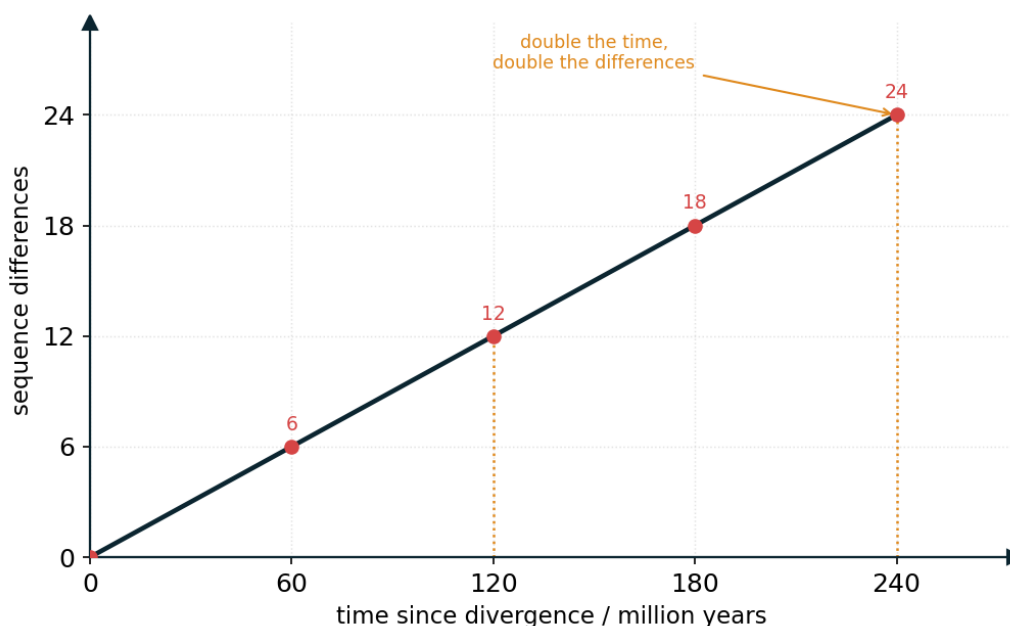


Figure 3. The molecular clock. If mutations accumulate at a roughly constant rate (here 1 difference per 10 million years), the number of sequence differences is proportional to the time since divergence, so doubling the time doubles the differences.

Caution: The molecular clock is an approximation, not a precise timer. Mutation rates differ between genes and between lineages, so estimates carry uncertainty. Different genes may also give slightly different trees, which is why biologists compare several molecules.

5. Reading and constructing cladograms (HL)

A **cladogram** is a tree diagram that shows the hypothesised evolutionary relationships between clades. Learning to read one is a core HL skill.

The anatomy of a cladogram

- The tips (ends of the branches) represent the organisms or groups being compared.
- A **node** is a branch point. It represents the **most recent common ancestor** of everything above it — the point at which one lineage split into two.
- The **branch points** mark divergence events. Two organisms that share a more recent node are more closely related than two whose nearest shared node lies deeper in the tree.
- A **clade** is any node together with all the branches and tips arising from it.

Reading relatedness is about *where lineages join*, not how close the tips happen to be printed on the page. Rotating the branches about a node does not change the relationships shown.

Sequence differences and divergence time

On many cladograms the length of a branch is proportional to the number of sequence differences that have accumulated along it, and therefore to the time elapsed. A larger number of base or amino-acid differences between two species implies a longer time since they diverged, so their branch point lies further back. This lets biologists both **construct** a cladogram from sequence data and **estimate** when lineages split.

A worked figure

The simple cladogram below shows four organisms. Read it from left (the oldest ancestor) to right (the present day). Each fork is a node; the labels 1, 2 and 3 mark the branch points.

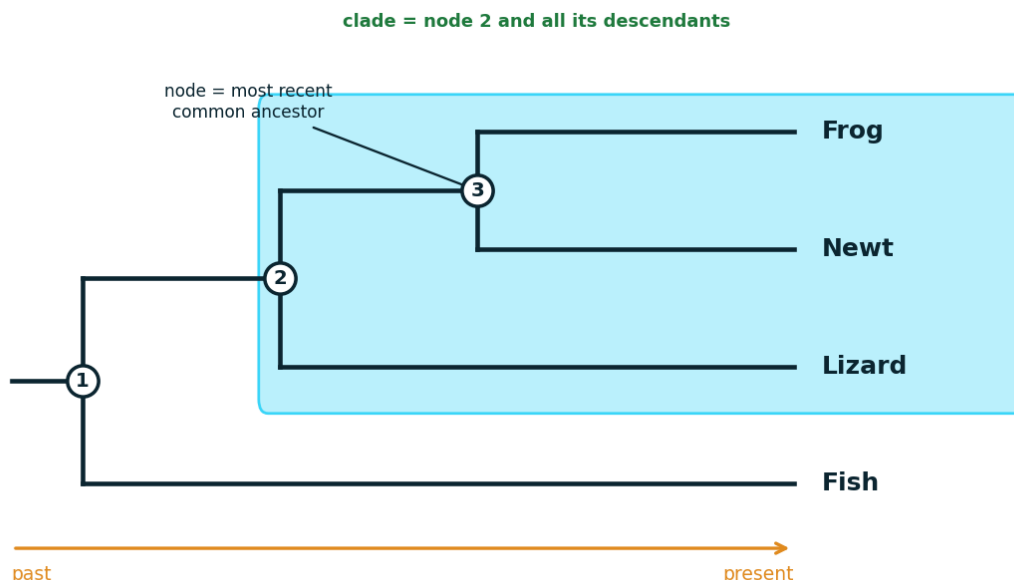
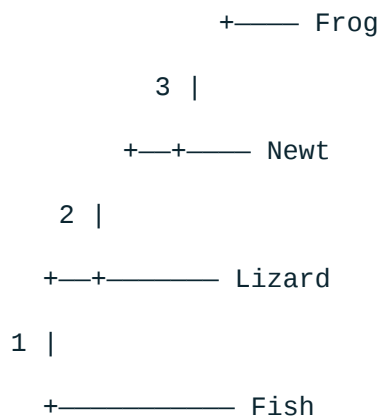


Figure 1. A cladogram of four vertebrates. Each numbered node is the most recent common ancestor of the tips above it: node 3 joins frog and newt (most closely related), node 2 adds the lizard, and node 1 (the deepest node) is where the fish lineage split off first. The shaded clade is node 2 together with all of its descendants.

Figure — a cladogram to read

Reading it: Node 3 is the most recent common ancestor of the frog and newt, so those two are the most closely related. Node 2 joins the lizard to the (frog, newt) clade. Node 1 is the deepest node, where the fish lineage split off first, so the fish is the most distantly related of the four. The frog, newt and lizard together form a clade (node 2 plus all its descendants); adding the fish gives the clade defined by node 1.

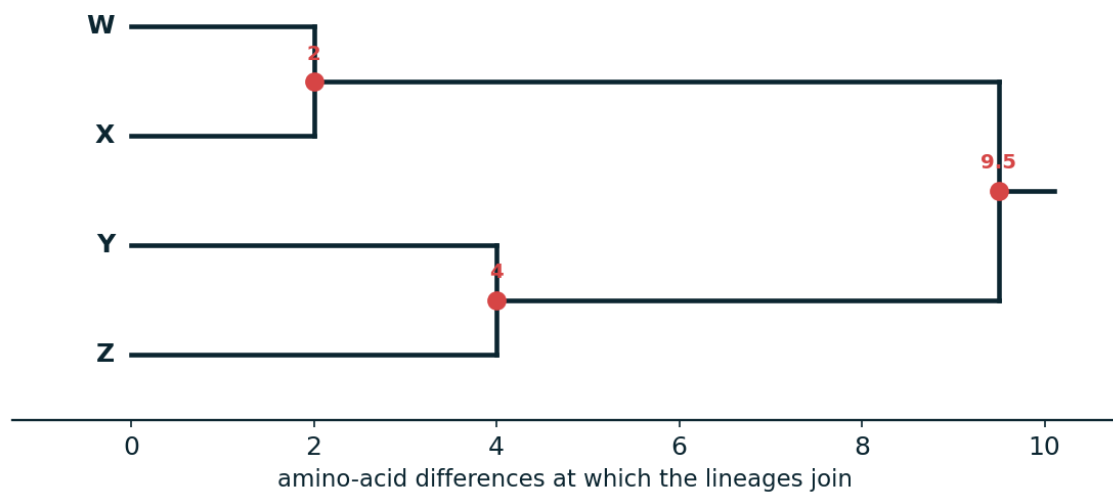


Figure 2. Building a tree from a distance matrix. Species cluster at the number of amino-acid differences between them: W and X join at 2, Y and Z at 4, and the two pairs join at their average cross-cluster difference of 9.5, giving the branching order ((W, X), (Y, Z)).

Worked example 1 — building a tree from sequence differences

Four species W, X, Y, Z are compared using the same protein. The table gives the number of amino-acid differences between each pair. Deduce which two are most closely related and sketch the branching order.

	X	Y	Z
W	2	9	10
X		9	10
Y			4

Step 1: find the smallest difference. W and X differ by only 2, so they are the most closely related — they share the most recent common ancestor.

Step 2: the next smallest is Y and Z (4 differences), so they form a second close pair.

Step 3: the two pairs differ from each other by 9–10, a much larger number, so the split between the (W, X) group and the (Y, Z) group is the deepest, oldest divergence.

Branching order: ((W, X), (Y, Z)). W and X join at a shallow, recent node; Y and Z join at another; those two nodes join at a deep node representing the last common ancestor of all four.

Worked example 2 — interpreting a cladogram

A cladogram of five vertebrates has this branching order, reading the nodes from the base outward:

(Lamprey, (Shark, (Frog, (Lizard, Mouse))))

(a) Which species is most distantly related to the mouse?

The lamprey branches off at the deepest node, so it shares the most ancient common ancestor with the mouse — it is the **most distantly related**.

(b) Which species is the mouse's closest relative?

The lizard shares the most recent node with the mouse, so the **lizard** is its closest relative among those shown.

(c) Do the frog, lizard and mouse form a clade?

Yes. Their most recent common ancestor is the node from which the frog branch and the (lizard, mouse) branch arise, and the group includes that ancestor and all its descendants — the definition of a clade.

6. Reclassification when the evidence changes (HL)

A natural classification is a **hypothesis** about evolutionary history, and like any scientific hypothesis it is revised when better evidence appears. Historically, groups were defined mainly by **morphology** (visible structure). Molecular data now frequently reveal that morphology-based groups were wrong, prompting **reclassification**.

This happens because appearance can be misleading. Unrelated organisms may look alike through convergent evolution, while close relatives may look very different if they have adapted to different ways of life. DNA and protein sequences record ancestry more reliably than external form, so when the molecular evidence conflicts with the morphological grouping, the classification is redrawn to match the molecules.

A well-known example is the **figwort family** (Scrophulariaceae). For a long time this was a large family defined by shared floral morphology. When DNA sequence data were analysed, it emerged that the family as traditionally drawn was not a single clade — several of its members were more closely related to plants in other families. The group was therefore broken up and its genera redistributed among several families so that each family would once again represent a true clade. The morphological similarities that had held the old family together turned out to be shared ancestral or convergent features rather than evidence of recent common ancestry.

Nature of science: Reclassification is not a weakness of taxonomy but a strength: the system is self-correcting. As tools improve, classifications converge ever more closely on the real evolutionary tree. Molecular evidence commonly overturns groups founded on morphology alone.

7. A difficulty: convergent evolution versus homology

The greatest obstacle to building a natural classification is that similarity does not always mean relatedness. Two kinds of similar structure must be told apart.

	Homologous structures	Analogous structures
Origin	Inherited from a common ancestor	Evolved independently in unrelated lineages
Underlying cause	Common ancestry (divergent evolution)	Similar selection pressures (convergent evolution)
Internal structure	Same basic plan, different functions	Different plan, similar function
Example	Pentadactyl limb of human, whale, bat, horse	Wing of a bird and wing of an insect
Use in classification	Valid evidence of relatedness	Misleading — must be excluded

Homologous structures share the same fundamental anatomy because they were inherited from a common ancestor, even though they may now perform different functions. The pentadactyl (five-digit) limb is the classic case: the same arrangement of bones underlies a human arm, a whale flipper, a bat wing and a horse leg. Such structures are genuine evidence of common ancestry and are the basis of cladistic grouping.

Analogous structures perform a similar function and may look alike, but have different underlying structures and separate evolutionary origins. They arise through **convergent evolution**, in which unrelated organisms independently evolve similar solutions to similar environmental challenges. The streamlined body shape of a shark (a fish) and a dolphin (a mammal) is a striking example. Because analogous features

do not reflect shared ancestry, they must be identified and set aside when classifying — relying on them produces artificial groups.

Exam-defining point: Only **homologous** characters (including conserved molecular sequences) reveal true relationships. **Analogous** characters, produced by convergent evolution, are traps. Distinguishing the two is exactly why molecular evidence is so valuable.

8. More worked examples and common pitfalls

Worked example 3 — ordering the taxa

Place the following taxa for the domestic dog in the correct order, from most inclusive to least inclusive: Carnivora, Animalia, *Canis*, Chordata, Eukarya, Mammalia, Canidae, *Canis familiaris*.

Work down the hierarchy Domain → Kingdom → Phylum → Class → Order → Family → Genus → Species:

Eukarya (domain) → Animalia (kingdom) → Chordata (phylum) → Mammalia (class) → Carnivora (order) → Canidae (family) → *Canis* (genus) → *Canis familiaris* (species).

Notice the genus name is capitalised and italicised, and the full species name repeats the genus.

Worked example 4 — placing a species in a clade from data

Cytochrome c amino-acid differences from a human are: chimpanzee 0, rhesus monkey 1, horse 12, tuna 21. Rank these species by relatedness to humans and state which shares the most recent common ancestor.

Fewer differences means a more recent common ancestor. Ranking from closest to most distant: chimpanzee (0) → rhesus monkey (1) → horse (12) → tuna (21).

The **chimpanzee** shares the most recent common ancestor with humans (identical sequence), so it sits in the smallest, most recent clade with us; the tuna, with the most differences, branches off at the deepest node.

This ordering matches the accepted tree: primates cluster together, then the mammal (horse) joins, and the fish (tuna) is the outgroup.

Worked example 5 — spotting convergence in a data set

A student groups a shark, a dolphin and a trout together as one clade because all three have a streamlined body and fins. DNA comparison then gives these percentages of sequence similarity to the dolphin: shark 44%, trout 45%, cow 82%. Evaluate the student's grouping.

The molecular data show the dolphin is far more similar to the cow (82%) than to the shark or trout (both around 44–45%). This means the dolphin shares a much more recent common ancestor with the cow — both are mammals — than with either fish.

The student's group is therefore **artificial**. The streamlined shape and fins of sharks and dolphins are **analogous** structures produced by convergent evolution (similar selection pressures in water), not evidence of common ancestry. A natural classification places the dolphin with the cow and other mammals, and the shark and trout with the fish.

Common pitfalls

- Treating look-alikes as relatives. Analogous structures (bird wing vs insect wing) do **not** indicate common ancestry — they are convergent.
- Reading a cladogram by tip position. Relatedness is set by the **nodes** (where lineages join), not by how near two tips are drawn.
- Mis-writing a binomial. The genus is capitalised, the species is lower case, and the whole name is italicised or underlined.
- Confusing shared ancestral with shared derived features. Only **derived** characters define a clade; ancestral ones are too widespread to distinguish subgroups.
- Assuming molecular clocks are exact. They give estimates with uncertainty, because mutation rates vary.
- Thinking archaea are just unusual bacteria. Their membrane lipids and rRNA place them in a separate domain.

Quick reference

Term	One-line meaning
Taxon	Any named group in the classification hierarchy
Hierarchy	Domain → Kingdom → Phylum → Class → Order → Family → Genus → Species
Binomial	Two-part name: <i>Genus species</i> , italicised, genus capitalised
Three domains	Archaea, Bacteria, Eukarya — separated by rRNA and membrane-lipid chemistry
Natural classification	Groups reflect common ancestry (evolutionary relationships)
Clade (HL)	An ancestor and all of its descendants
Cladogram (HL)	Tree of relationships; nodes = common ancestors, branch points = divergences
Molecular clock (HL)	Mutations accumulate at a roughly constant rate, estimating divergence time
Homologous	Same structure, common ancestry — valid evidence

Term	One-line meaning
Analogous	Similar function, convergent evolution — misleading

9. Test yourself

Attempt these without notes; full worked answers follow. Questions marked (HL) are Higher Level only.

1. List the eight principal taxa in order from the most inclusive to the least inclusive.
2. State three rules for writing the scientific name of a species correctly, using *Panthera leo* as your example.
3. Name the three domains and give one piece of molecular evidence used to distinguish them.
4. Explain the difference between a natural and an artificial classification, with an example of each.
5. (HL) Define a clade and explain what a node on a cladogram represents.
6. (HL) Two species differ in 3 bases of a gene; two others differ in 18 bases of the same gene. What does this tell you about their relative divergence times, and what assumption underlies your reasoning?
7. (HL) Explain how molecular evidence can lead to the reclassification of a group previously defined by morphology.
8. Distinguish homologous from analogous structures and state why the distinction matters for classification.

Answers

1. Domain, Kingdom, Phylum, Class, Order, Family, Genus, Species (Dear King Philip Came Over For Good Soup). Groups get smaller and members share more features as you go down.
2. Any three of: the genus is capitalised (*Panthera*); the species epithet is lower case (*leo*); the whole name is italicised, or underlined when handwritten; after first use the genus may be abbreviated (*P. leo*). Correct form: *Panthera leo*.
3. Archaea, Bacteria and Eukarya. Molecular evidence: differences in small-subunit ribosomal RNA (rRNA) base sequences; also archaeal membrane lipids use branched chains with ether bonds while bacteria and eukaryotes use unbranched fatty acids with ester bonds; and archaea lack peptidoglycan in their cell walls.
4. A natural classification groups organisms by evolutionary relationship / common ancestry (e.g. grouping organisms descended from one ancestor into a clade). An artificial classification groups by convenient superficial features (e.g. lumping all flying animals – birds, bats, insects – together despite distant relationships). Natural systems reflect real evolutionary history; artificial ones need not.
5. (HL) A clade is a group consisting of a common ancestor together with all of its descendants. A node on a cladogram represents the most recent common ancestor of all the branches arising from it – the point at which one lineage split into two.
6. (HL) Fewer differences (3 bases) means less time has passed since those two species diverged, so they are more closely related; the pair differing by 18 bases diverged much earlier and are more distantly related. The reasoning assumes a molecular clock – that mutations accumulate at a roughly constant average rate, so the number of differences is proportional to divergence time.
7. (HL) Morphology can be misleading because unrelated species may converge on similar forms while relatives may diverge. DNA or protein sequences record ancestry more reliably. If sequence data show that

members of a morphology-based group are actually more closely related to organisms in other groups, the group is not a true clade and is split up or redrawn (as happened with the figwort family) so that each taxon again represents a single clade.

8. Homologous structures share the same basic anatomy inherited from a common ancestor (e.g. the pentadactyl limb of human, whale and bat), even if their functions differ; they arise by divergent evolution and are valid evidence of relatedness. Analogous structures have similar functions but different underlying structures and separate origins (e.g. bird and insect wings); they arise by convergent evolution and must be excluded from classification because they do not reflect common ancestry.