



# HL IB Biology



## Classification & Cladistics

### Contents

- \* Biological Classification (HL)
- \* Cladistics (HL)
- \* Reclassification (HL)
- \* Classification System (HL)



Your notes

## Biological Classification (HL)

### Classifying Organisms

#### Why is biological classification needed?

- The diversity of life on Earth is vast, and is known as **global biodiversity**
  - The extent of global biodiversity is such that **scientists can only estimate** the total number of species present on Earth, and it is likely that there are many species yet to be discovered
- For biologists to make sense of the huge array of species, **organising them into logical groups** is essential; this process of putting organisms into groups is known as **classification**
  - The science of classification is known as **taxonomy**, and scientists working in the field of taxonomy are **taxonomists**
  - Classifying an organism involves deciding which biological group, or **taxon** (plural **taxa**), it fits into best, and then **naming it** according to its taxon
  - Historically an organism's biological group was determined on the basis of its **observable characteristics**, and today this information is combined with **genome sequencing** data for more accurate classification
- Correct classification allows scientists to:
  - Accurately determine the **number of known species**
    - Without classification there would be a risk of recording species more than once, or incorrectly grouping multiple species together as one
  - Learn about the **evolution** of species
    - Accurate classification allows scientists to determine the evolutionary relationships between species, i.e. which species share common ancestry, and how recently
  - Ensure that **conservation** is carried out when needed
    - It is not possible to conserve a species if we don't know that it exists, e.g. if several species are incorrectly classified as one species, then the extinction of what might be thought of as one local population could actually result in the extinction of an entire species
  - Carry out **medical research** more quickly
    - E.g. if one plant species is known to have medicinal properties, then it is possible to quickly identify others by looking at its close relatives; this is much faster than just selecting plants at random for research
  - **Identify and treat new diseases** more quickly
    - E.g. the virus that causes COVID-19 was quickly classified as being a coronavirus, providing medics with information about how it might affect the body, and how to go about developing a vaccine



Your notes

## Classifying Organisms: Challenges

### The traditional hierarchy of taxa

- Biological classification involves putting organisms into groups, or **taxa** (singular **taxon**)
- The taxa form a **hierarchy**
  - A hierarchical system is one in which **larger groups contain smaller groups with no overlap** between groups
- The taxonomic hierarchy contains the following taxonomic groups in descending order of size:
  - Domain
  - Kingdom
  - Phylum
  - Class
  - Order
  - Family
  - Genus
  - Species

### Examples of classification table

Taxonomic Rank	Wolf	Hibiscus
Domain	Eukaryote	Eukaryote
Kingdom	Animalia	Plantae
Phylum	Chordata	Angiospermae
Class	Mammalia	Dicotyledonae
Order	Carnivora	Malvales
Family	Canidae	Malvaceae
Genus	Canis	Hibiscus
Species	lupus	rosa—sinensis

### Difficulties with classification



Your notes

There are multiple challenges when it comes to accurately classifying organisms into the hierarchy of taxa described above; such difficulties include:

## Morphology

- Historically, organisms have been classified on the basis of their morphology, but this can lead to errors; similarities in observable characteristics do not always mean that organisms share a recent common ancestor
  - E.g. dolphins and sharks could in theory be grouped together as they are both groups of aquatic animals that share a similar body shape, but they in fact belong to different classes
    - Dolphins are mammals and sharks are fish
    - Their streamlined body shapes evolved separately rather than originating in one common ancestor
- One solution to this difficulty is to use **genome sequencing** data; this helps to avoid the difficulties with misleading morphology, but does not solve all of the challenges

## Taxonomic rank

- In the hierarchy system described above, each level of classification fits into an established taxonomic rank, i.e. kingdom, phylum, class etc.
- Classification can be complicated if a group of organisms falls **across taxa**, or needs to be **moved from one taxon** to another
  - Plant species in **distant taxa** can sometimes breed together to produce **fertile hybrids**; the resulting offspring will technically be a new species, but will be very difficult to classify under the hierarchy of taxa
  - Moving a group of organisms between taxa runs the risk of needing to shunt all of the groups currently in a taxon into a different rank to make room for the new grouping

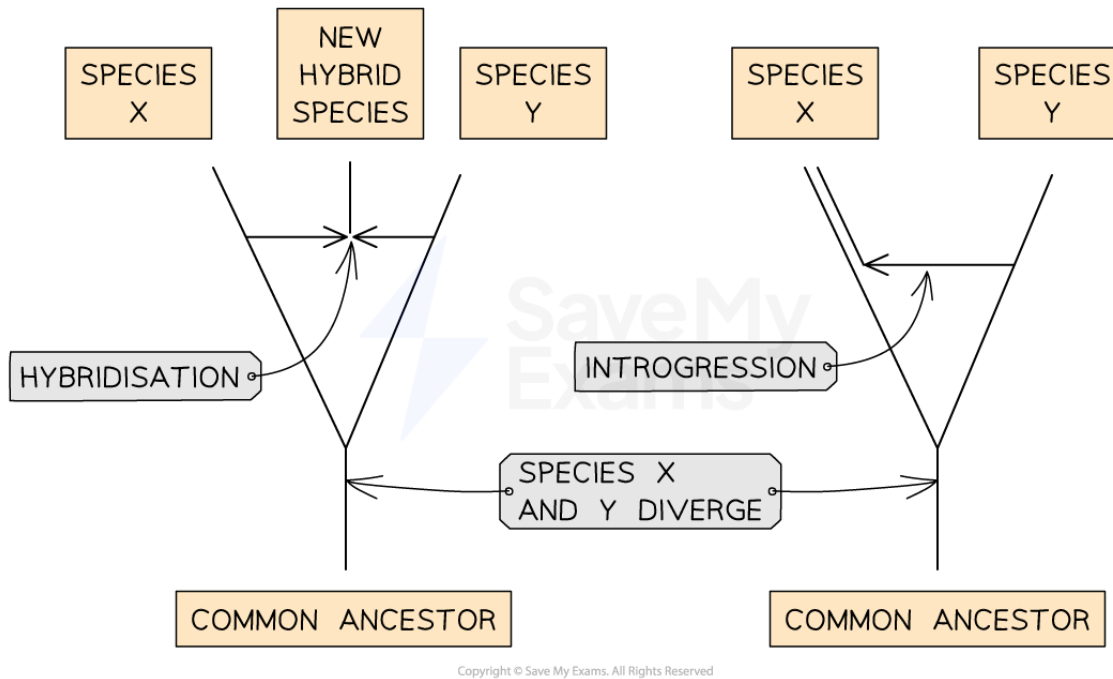
## Species

- The point at which two populations are classified as different species can be **highly subjective**
- The fertile offspring of a cross between two species may then go on to breed only with members of *one* of its parent species; this is known as **introgression**
  - Introgression demonstrates how difficult it can sometimes be to neatly apply species classification; the resulting offspring after several generations **do not fit completely into either species**, but **neither does it seem to make sense to classify them as a new species**
  - E.g. hundreds of thousands of years ago an early human bred with a Neanderthal and the offspring of this cross then went on to breed only with early humans; the result of this is that some groups of modern humans have some Neanderthal genes in their genomes

### Introgression diagram



Your notes



*Several generations after an interbreeding event has occurred, the species X individuals with some genes from species Y do not fit perfectly into either species X or Y, but cannot correctly be considered a new species; this is introgression*

### NOS: A fixed ranking of taxa is arbitrary because it does not reflect the gradation of variation

- The hierarchy of taxa described here has been **arbitrarily** set up by humans because it is a neat way of organising life into groups, not because it fits with the patterns that we see in the natural world
- In order to fall neatly into the taxonomic ranks of the traditional classification system, **species need to be clearly distinct** from each other, and they **need to obey reproduction rules** such as not interbreeding, and not producing fertile offspring; this is far from the reality of the biological world where the **differences between organisms are on a gradual scale**, and **neat breeding rules do not always apply**
- A successful classification system needs to **follow the evidence** rather than seeking to fit the natural world into a human-designed system
  - New evidence comes from **genome sequencing**, and often leads to the reclassification of species
  - The newer system of cladistics uses **unranked groups** based on **evolutionary relationships** alone to produce evolutionary trees

 **Examiner Tip**

The use of a mnemonic can help to recall the order of the taxonomic groups; an example might be:

- **D**ear (**d**omain)
- **K**ing (**k**ingdom)
- **P**hillip (**p**hylum)
- **C**ame (**c**lass)
- **O**ver (**o**rders)
- **F**or (**f**amily)
- **G**reat (**g**enus)
- **S**oup! (**s**pecies)



Your notes



Your notes

## Cladistics (HL)

### Classifying by Evolutionary Relationship

- Historically organisms would have been classified on the basis of morphology, which often led to organisms being classified into groups that were not all close relatives
- The development of **DNA sequencing technology** means that classification can now be carried out on the basis of **evolutionary relationship**
  - Organisms grouped together using this method of classification form groups known as **clades**; every member of a clade shares a recent **common ancestor**
    - A common ancestor is a shared ancestor, e.g.
      - The most recent common ancestor of siblings is their parents
      - The most recent common ancestor of cousins is their grandparents
  - Clades are **monophyletic** groups, meaning that they contain **all of the descendants of a common ancestor**

### Advantages of classification by evolutionary relationship

- Classifying organisms correctly **according to their clade** ensures that groups of organisms are **close evolutionary relatives** rather than arbitrary groups that happen to look similar
  - The characteristics within a clade are often **inherited from a common ancestor**, so are likely to be shared
    - This means that scientists need to be careful not to accidentally place organisms together on the basis of analogous characteristics; not all species with shared characteristics are closely related
- The use of DNA sequencing has allowed some organisms to be **reclassified** into more accurate groups
  - Some species have been reclassified into **different groups** of organisms
  - Some groups of organisms have been **split**
  - Some groups have been **merged**



Your notes

## Clades

### Cladistics

- **Cladistics** is the branch of science in which scientists put organisms into **clades**
- The term **clade** can be defined as
  - **A group of organisms that have all descended from a common ancestor**
- Clades can include both **living** and **extinct** species
  - Some of the descendants of a common ancestor may have gone extinct
  - The common ancestor species itself may have gone extinct
- Clades can be **large** or **small** depending on the common ancestor being studied
- While taxonomy is about classifying and then **naming organisms**, cladistics is about **identifying evolutionary relationships** between organisms
  - A **taxon** is a group of organisms that have been given a group name by taxonomists **on the basis on their shared features**
  - A **clade** is a group of organisms classified together **on the basis of their shared descent** from a common ancestor
- If taxonomy is carried out correctly then **all of the members of a taxon should form a clade**, but due to historical errors this is not always the case

### Identifying members of a clade

- In cladistics, it is important that species are placed into **true clades**; this avoids mistakes such as:
  - Some descendants of a common ancestor being placed in a different clade to each other
  - Some organisms that descend from a different ancestor being included in the same clade

### Sequence data

- The most objective method for grouping species into clades is the use of **sequence data** from:
  - DNA bases
  - mRNA bases
  - Amino acids in polypeptides
- For all types of sequence data, it can be said that **the more similar the sequences, the more closely related the species**
  - Two groups of organisms with very similar sequences have **separated into separate species more recently** than two groups with less similarity in their sequences
  - Species that have been separated for longer have had **a greater amount of time to accumulate mutations and changes** to their DNA, mRNA and amino acid sequences

### Morphology

- Species that share a recent common ancestor are more likely to **share similar morphologies** than species that diverged a long time ago, so identifying members of a clade can sometimes be done on the basis of morphology
- This method is considered to be more subjective, and has led to **classification errors** in the past



- Similar morphology can be a sign that convergent evolution has occurred rather than a sign of recent common ancestry
- When classifying organisms in this way it can be **difficult to decide on which characteristics classification should be based**
  - E.g. should we classify bats along with birds on the basis of their wings, or with mammals on the basis of their fur?
- In order to choose the most important characteristic, scientists need to decide which characteristics are **primitive**, and which are **derived**
  - Primitive traits evolve **early in the lineage of a clade** and are found in all clade members, e.g.
    - All vertebrates have spinal cords
    - All insects have six legs
  - Derived traits **evolve later** and can differ between clade members
    - E.g. within the vertebrate clade, birds have feathers while mammals have fur
  - Note that a derived trait in one clade could be an ancestral trait in another
    - E.g. fur is a primitive trait within the mammal clade, but a derived trait within the vertebrate clade
- More **closely related species** will share a **larger number of derived traits**



Your notes



Your notes

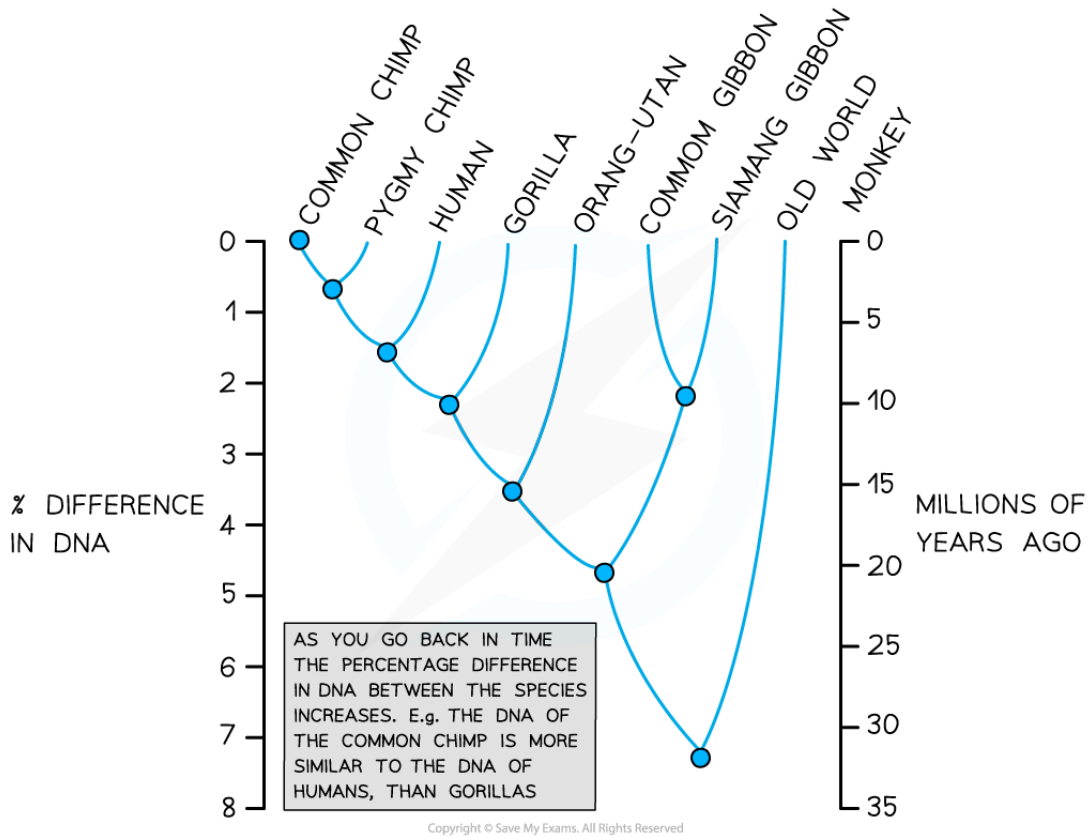
## The Molecular Clock

- The **evolutionary relationships** between species can be determined by analysing **sequence data** from, e.g. **DNA bases**, **mRNA bases**, or **amino acids in polypeptides**
- The **number of differences** between sets of sequence data provides information on **how closely related** two species are
  - The **more differences** there are between the sequences, the **longer ago the species diverged**, and vice versa
  - The number of differences between sequences can be determined using a method known as **DNA hybridisation**
    - Sections of **single-stranded DNA** from corresponding parts of the genome are taken from two species
    - The two **complementary strands** are allowed to bind to each other, or **hybridise**
    - The points on the DNA strand that **do not bind** show where bases are **different to each other**
    - The **number of differences** are recorded
- The differences between sequence data can also be used to produce a **quantitative estimate** for **how long ago** two species diverged from each other
  - Differences in sequence data come about due to mutations in the DNA
  - Evidence suggests that **mutations occur at a fairly constant rate**
  - This means that the **number of mutations** that have occurred gives an indication of the **amount of time that has passed** since two species diverged
    - Scientists refer to the constant rate of mutation as the **molecular clock**
      - It is worth noting that the assumed rate of mutations does not always match with the actual rate at which mutations occur, so **the molecular clock provides estimates** rather than exact time periods
      - The rate at which mutations accumulate can be affected by, e.g.
        - Generation time
        - Population size
        - Selection pressures
- Analysing the differences in sequence data allows evolutionary biologists to determine the **order** in which different species **diverged from a common ancestor**, and therefore **how closely related** species are

### Evolutionary tree diagram



Your notes



*The relatively consistent rate at which mutations occur in DNA provides scientists with a molecular clock. Differences in DNA sequence can therefore show how much time has passed since species diverged from each other.*

## Constructing Cladograms

- **Evolutionary relationships** between species can be represented visually using a diagram called a **cladogram**
- Cladograms are **evolutionary trees** that show probable **order of divergence from ancestral species** and therefore probable **relationships between species**
  - The point at which two branches separate is known as a **node**
  - Nodes represent **common ancestor** species
- The information used to build cladograms most often comes from **base** or **amino acid sequence data** due to difficulties in the use of observable characteristics
  - Observable characteristics can be misleading as they are not always the result of shared ancestry
- Sequence data can provide information about **how different species are** from each other, as well as **how much time has passed** since divergence from a common ancestor took place
  - The constant rate at which mutations accumulate can be used as a **molecular clock**
- The **reliability of a cladogram may vary** depending on the **amount of sequence data** used to construct it
  - A cladogram based on the sequencing of one gene will be less reliable than a cladogram based on the sequencing of several genes
- **Computers** use the information from sequence data to **build the most likely cladogram**
  - This is done using the principle of **parsimony**, which states that the simplest explanation is preferred
    - The computer builds the **shortest possible cladogram** with the **smallest number of divergence events** to fit the available data
  - We say that cladograms show the **most probable** divergence times and relationships rather than providing **definite** conclusions
- Cladograms **may change** in the future if **new evidence** comes to light

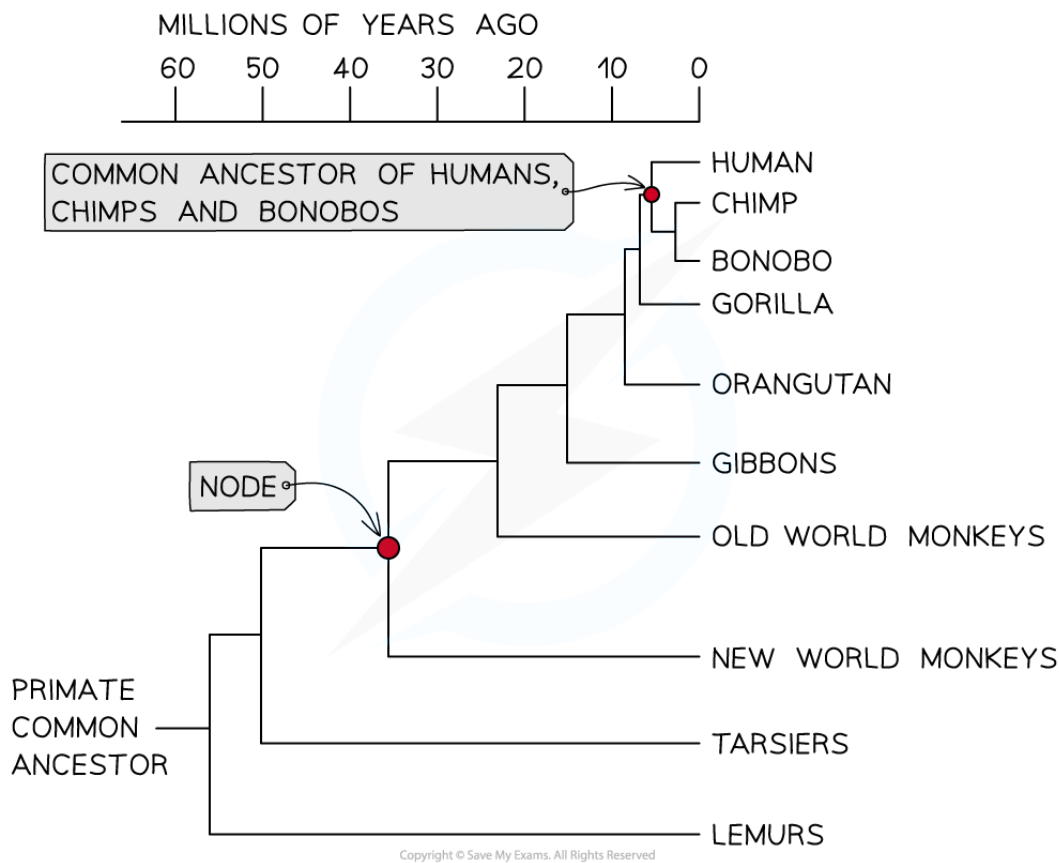
### Primate cladogram diagram



Your notes



Your notes



**A cladogram for primates shows that humans are most closely related to chimps and bonobos, and that the next closest relatives are gorillas. Humans are thought to have diverged from chimps and bonobos between 5–7 million years ago.**

### NOS: Different criteria for judgement can lead to different hypotheses

- Cladograms are constructed using **available evidence** and on the basis of the principle of **parsimony**
  - Cladograms show the **probable relationships** between species, so they form a **hypothesis**
- The principle of parsimony states that the **simplest explanation is the most likely**
  - E.g. it is more likely that the characteristic of **fur evolved once** than twice, so cladograms place mammals as the descendants of one common ancestor, rather than appearing in several places in the cladogram
- The principle of parsimony **provides the criteria** for the construction of cladograms
  - Without this criteria, cladograms might look different

## Analysis of Cladograms

- Analysis of a cladogram can provide several pieces of information
  - The point at which two branches separate is known as a **node**, and represents **common ancestor** species
    - A node **immediately adjacent** to a pair of clades indicates that these two clades share a **recent common ancestor**
      - This shows that the two clades are **more closely related to each other than they are to any other clade** in the cladogram
    - If **several nodes need to be traced back** before two clades can be joined, this indicates a **more distant relationship** between two clades
  - The **root** of a cladogram is found at its base, and represents the common ancestor of all of the organisms within the cladogram
    - The root of a cladogram will represent organisms that were present a long way back in **evolutionary history**
  - The **terminal branch** represents the **most recent** species in an evolutionary lineage
  - Cladograms sometimes show **numbers** along the branches; these indicate the **number of base or amino acid changes** that have occurred between one node and the next or between a node and an emerging clade or species
    - The constant rate at which mutations accumulate means that these numbers can be used as a **molecular clock** to calculate how much time has passed
  - Some cladograms have a **time scale** to show how many millions have years have passed

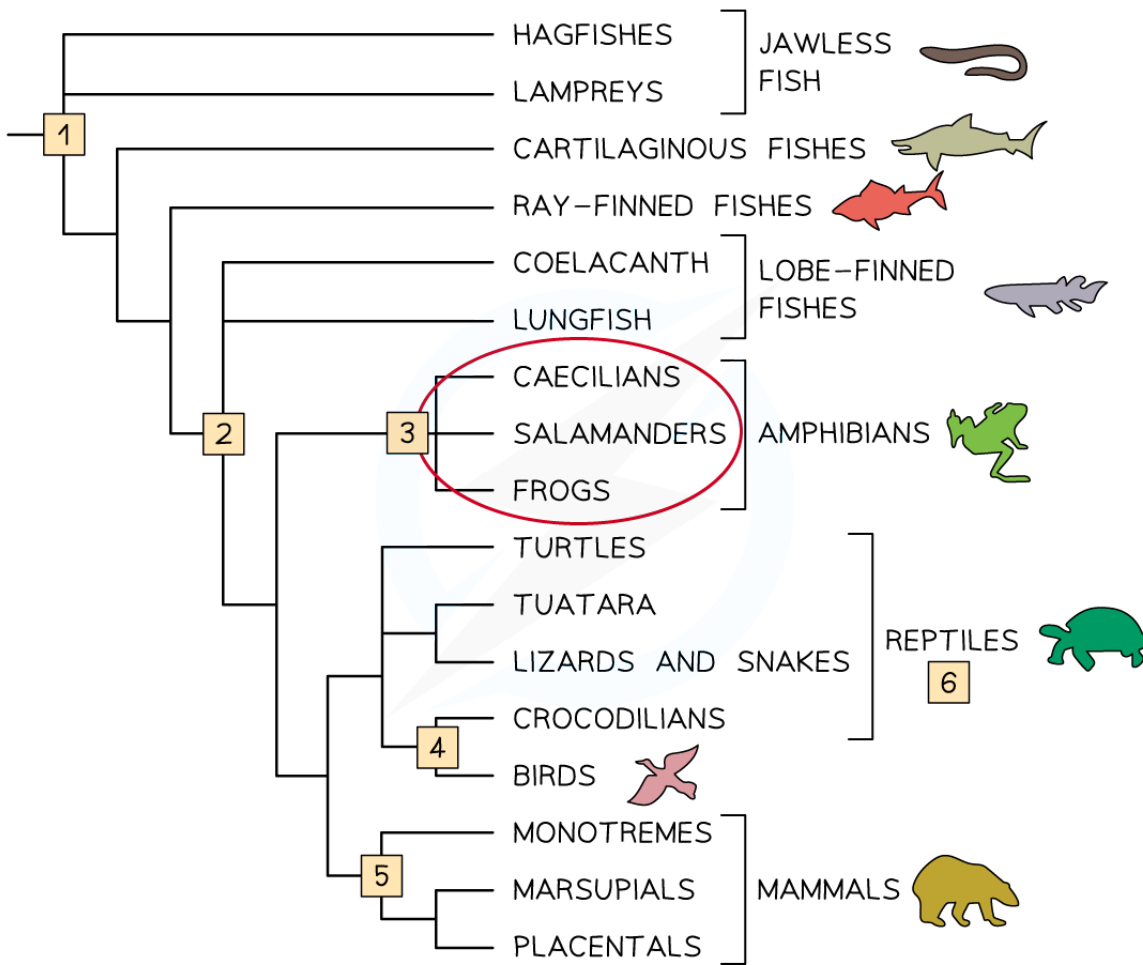
### Cladogram analysis diagram



Your notes



Your notes



Copyright © Save My Exams. All Rights Reserved



Your notes

- 1 NODE REPRESENTING COMMON ANCESTOR OF ALL VERTEBRATES
- 2 FIVE NODES NEED TO BE TRACED BACK TO LINK PLACENTAL MAMMALS WITH COELACANTHS, INDICATING A MORE DISTANT RELATIONSHIP
- 3 THREE CLADES DIRECTLY LINKED TO ONE COMMON ANCESTOR INDICATES THAT THIS CLADOGRAM WAS BUILT WITHOUT ENOUGH DATA TO DETERMINE WHICH OF THESE THREE CLADES ARE MORE CLOSELY RELATED
- 4 ONLY ONE NODE NEEDS TO BE TRACED BACK TO LINK CROCODILIANS AND BIRDS, SO THEY ARE MORE CLOSELY RELATED TO EACH OTHER THAN TO ANY OTHER CLADE
- 5 NODE REPRESENTING COMMON ANCESTOR OF ALL MAMMALS
- 6 NOTE THAT THE REPTILES ARE NOT A TRUE CLADE, AS THEY DO NOT INCLUDE THE BIRDS, WHICH SHARE A COMMON ANCESTOR WITH THEM

Copyright © Save My Exams. All Rights Reserved

**Multiple conclusions can be drawn from a large cladogram such as this vertebrate cladogram.**

**Note that this cladogram contains no numbers or time scale, so it does not show the number of base or amino acid changes that have occurred between one node and the next, or how much time has passed between nodes.**





Your notes

## Reclassification (HL)

### Cladistics & Reclassification

- The development of **DNA sequencing technology** means that classification can now be carried out on the basis of **evolutionary relationship**
  - Organisms classified in this way are grouped into clades
- Historically, organisms would have been classified on the basis of **observed traits**, which often led to organisms being classified into groups that were **not true clades**
- The use of DNA sequencing to classify organisms has led to discoveries of classification errors and the need to **reclassify** organisms
  - Some species have been reclassified into **different groups** of organisms
  - Some groups of organisms have been **split**
  - Some groups have been **merged**
- Classifying organisms correctly **according to their clade** is important to ensure that groups of organisms are **close evolutionary relatives**

### Reclassification of the figwort family

- The **figwort** (pronounced 'figwurt') **family**, also known as the **Scrophulariaceae**, was at one point the 8th largest family in the **angiosperm** phylum
- When the group was originally classified in the late 1700s it contained 16 genera, which later expanded to 275 genera
  - The classification of the figworts was based on **observable traits** such as a tube-shaped flower structure
  - Examples of members of the original figwort family include foxgloves and yellow rattle
- When **DNA sequence analysis** began, plant scientists discovered that the shared features of the figwort family were in fact not evidence of **shared ancestry**
  - Three chloroplast **genes** were analysed
  - It was discovered that the original figwort family was **not a true clade**
    - The old figwort family was shown to be **paraphyletic**, meaning that the group did not contain all descendants of a common ancestor
      - The group contained plant species that **should have been placed on separate branches** of a cladogram
  - The figworts were found to contain **several separate plant families**
    - **New families** were created
    - Several genera were moved into other **existing families**
    - The remaining genera were grouped together with two previously missed genera to form the **new figwort family**, still known as the *Scrophulariaceae*
- The new figwort family is **less than half** its original size

### Figwort family diagram



Your notes



FOXGLOVES ARE NOW IN THE PLANTAGINACEAE FAMILY



YELLOW RATTLE IS NOW IN THE OROBANCHACEAE FAMILY

Copyright © Save My Exams. All Rights Reserved



FIGWORT REMAINS IN THE NEW SCROPHULARIACEAE FAMILY

**DNA sequencing data led to the reclassification of many species within the original figwort family**

### NOS: Theories and other scientific knowledge claims may eventually be falsified

- A **theory** is an **explanation of observed phenomena** that is **supported by evidence**
- This means that when **new evidence** is found that **no longer supports the theory**, it needs to be **changed** to take the new evidence into account
- **Reclassification** of organisms on the basis of DNA sequencing data is a good example of this
  - Scientists **theorise** that a group of plants should be classified together on the basis of **existing evidence**; historically this would have been the **evidence of observable traits** such as flower shape
  - **New evidence from DNA sequence analysis** shows that these plant groups are not true clades, and so are **not descendants of a common ancestor**
    - The observable similarities between these plants are instead due to convergent evolution
  - The historical classification **theory has therefore been falsified** and needs to be changed to take the new evidence into account

- **New evidence** from **DNA** and **computer analysis** is used to calculate the most likely plant evolutionary relationships, and this **evidence** is used to form a **new theory** regarding the classification of plants
- Analysis of new DNA data continues all the time, and if **evidence** is found that doesn't fit with current classification theories then more **falsification** and reclassification will take place

### **Examiner Tip**

Note that you are **not** expected to memorise the specific details of the figwort case study shown here; the case study is provided to help develop understanding of the reclassification process.



Your notes



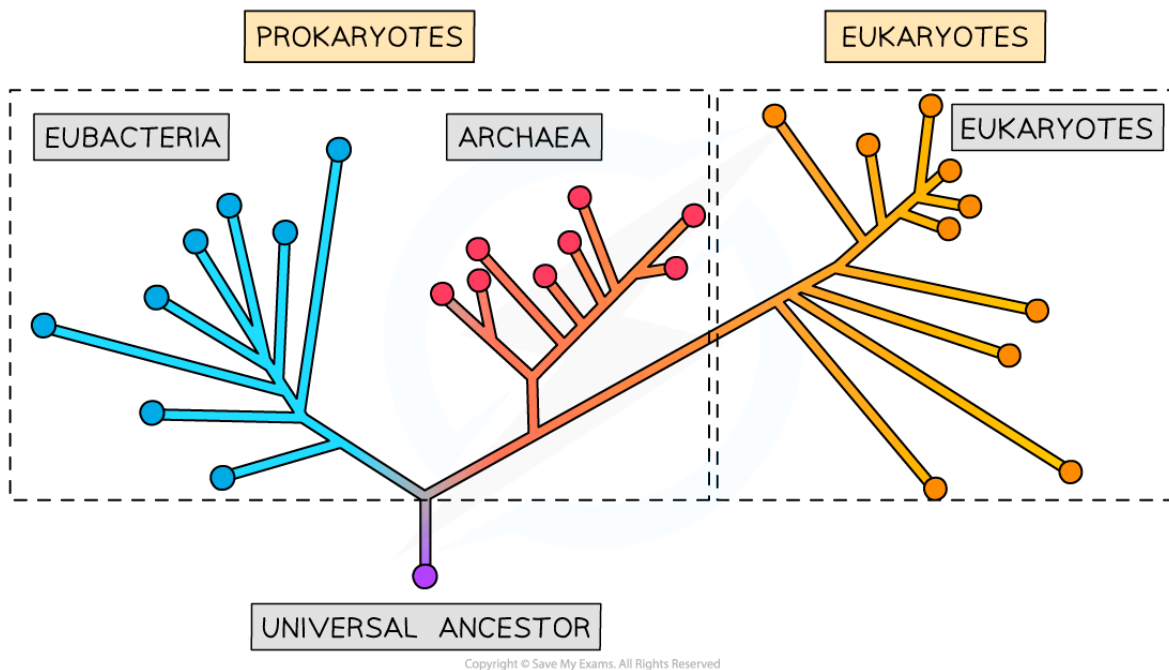
Your notes

## Classification System (HL)

### Domains

- The science of taxonomy has **frequently changed** to match the latest discoveries about the features of organisms
  - Historically, the largest taxonomic groups were the **plant** and **animal kingdoms**, then **fungi** were discovered and added (incorrectly) to the plant kingdom
  - **Microscopes** led to the discovery of **prokaryotes** and **eukaryotes**, and the taxa were later divided into five kingdoms; **plants, animals, fungi, protists**, and **prokaryotes**
    - The protists are eukaryotic, primarily single-celled, organisms
  - rRNA **analysis** has recently shown that there are **two distinct groups of prokaryotes**, leading to a **shift in taxonomic thinking** and the beginnings of the **three domain system**
- The **largest taxonomic group** is now known as a **domain**
- There are three domains, which are:
  - Archaea (prokaryotes)
  - Eubacteria (prokaryotes)
  - Eukaryotes (eukaryotes)

The three domains diagram



**The three domains are the eubacteria, archaea, and eukarya. The archaea are thought to be more closely related to the eukaryotes than to the prokaryotes.**



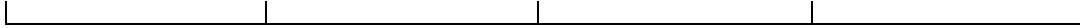
Your notes

## Classifying the Archaea

- The Archaea are **single-celled organisms** that live in a range of habitats
  - Some live in extreme environments, e.g.
    - Thermophiles live in hot springs
    - Methanophiles metabolise methane
- The Archaea were originally classified with the rest of the bacteria in one taxon due to **sharing several features** with them
  - Prokaryotic cell structure
  - Circular chromosome
  - Presence of a cell wall
  - 70S ribosomes
- Closer analysis of the Archaea, much of which was carried out by Carl Woese in 1977, showed that **some of their features were distinct** from the rest of the prokaryotes
  - Their **cell walls are made of a different material** from the cell walls of the rest of the bacteria
  - Their **cell membranes are made of a distinct type of lipid**
  - The **small subunit** of their **ribosomes** is **more similar to eukaryotic ribosomes** than to the ribosomes of the rest of the prokaryotes
  - Multiple **DNA sequences** present in bacteria **are absent** in methanogenic archaeans
- These discoveries led to a **change in the classification of the archaea**, which became their own domain

The three domains features table

Feature	Archaea	Eubacteria	Eukaryotes
Cell type	Prokaryotic	Prokaryotic	Eukaryotic
Chromosome	Circular	Circular	Linear (and circular mitochondrial and chloroplast DNA)
Cell membrane lipids	Glycerol-ether lipids	Glycerol-ester lipids	Glycerol-ester lipids
Ribosomes	70S with a small subunit that is more similar to eukaryotic ribosomes	70S	80S in cytoplasm and 70S in mitochondria and chloroplasts
Cell walls	Always present and without peptidoglycan	Always present and with peptidoglycan	Sometimes present and without peptidoglycan
Histones	Yes	No	Yes
Introns	Sometimes	Rarely	Yes



Your notes