

Organisms (Unity and diversity) revision notes

This topic asks what a species is and how the variety of living organisms is named, compared and classified. It covers species concepts, chromosome numbers, karyograms and genomes, then (HL) dichotomous keys, environmental DNA and the cladistic methods used to classify organisms by evolutionary relationships.

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

1 Diversity of organisms Spec A3.1

Variation, species and binomial names Spec A3.1.1, A3.1.2, A3.1.3

Variation between organisms is a defining feature of life. No two individuals are identical in all their traits, even identical twins or clones differ slightly. The patterns of variation are complex, and they are the basis for naming and classifying organisms.

Some differences are small and found within a group (eye colour in humans). Others are large and consistent and separate one group from another (the number of legs in insects and spiders). Biologists use these consistent patterns to recognise species.

Species as groups with shared traits

The original idea of a species, used by Linnaeus in the 18th century, was **morphological**: a species is a group of organisms with shared traits. Organisms that look alike in their structure (morphology) are placed in the same species, and organisms with consistently different traits are placed in different species.

This works well for many organisms, but appearance can mislead. Males and females of one species can look very different, and different species can look almost identical.

The binomial system

Every species has a two-part scientific name, its **binomial**. The first part identifies the **genus**. The second part distinguishes the **species** within that genus. Species in the same genus have similar traits.

- The genus name has an initial capital letter; the species name is all lowercase: *Homo sapiens*, *Pan troglodytes*.
- Binomials are printed in italics or underlined when handwritten.
- After first use, the genus can be shortened to its initial: *H. sapiens*.
- Binomials are international, so scientists everywhere know exactly which organism is meant, whatever its local common name.

Two species in the genus *Panthera*

Binomial	Common name	Genus	Species name
<i>Panthera leo</i>	lion	<i>Panthera</i>	<i>leo</i>
<i>Panthera tigris</i>	tiger	<i>Panthera</i>	<i>tigris</i>

Common mistake: Writing homo Sapiens or Homo Sapiens loses the mark. Only the genus has a capital letter.

Exam tip: Linking question: how do species show both continuous and discontinuous variation? Traits vary continuously within a species, yet species themselves are discrete groups.

The biological species concept and speciation Spec A3.1.4, A3.1.5

Biological species concept	A species is a group of organisms that can breed and produce fertile offspring.
Population	A group of organisms of the same species living in the same area at the same time.
Speciation	The splitting of one species into two or more.

The biological species concept focuses on **interbreeding** rather than appearance. A horse and a donkey can mate, but their offspring, the mule, is **infertile**, so horses and donkeys are separate species.

Challenges to the biological species concept

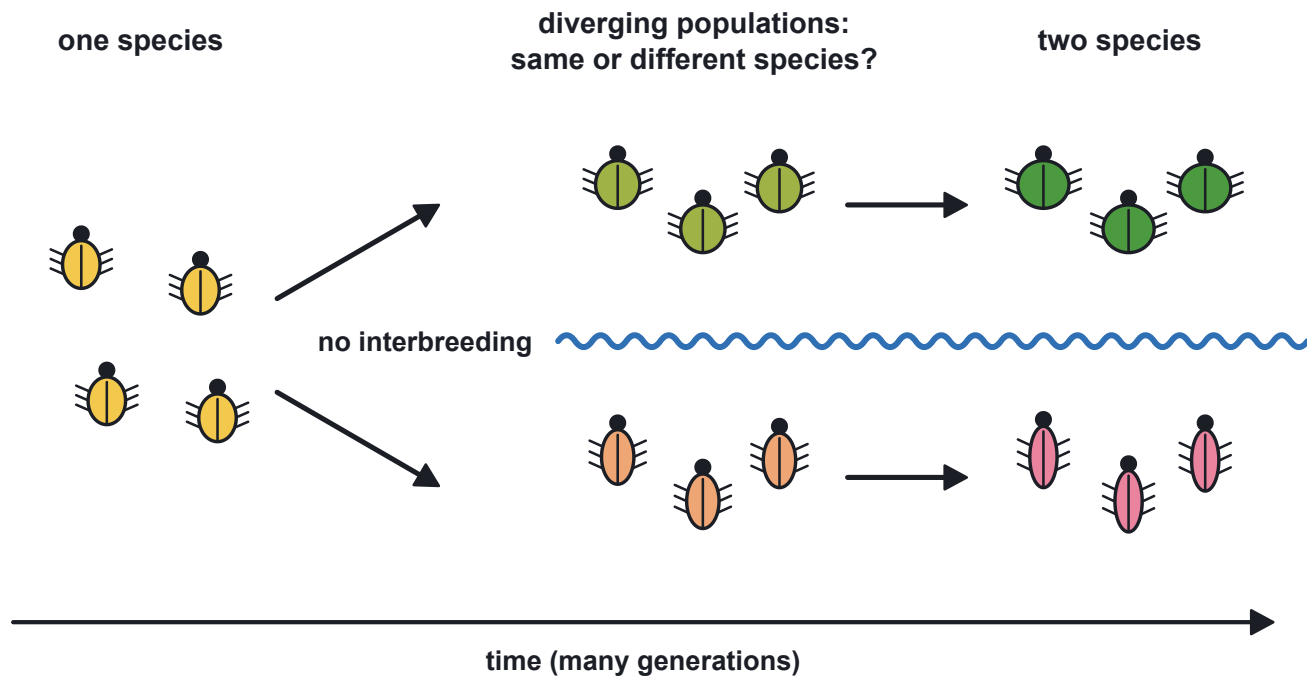
- Extinct organisms known only from fossils cannot be tested for interbreeding.
- Populations that are geographically separated never meet, so it is not known whether they would interbreed in the wild.
- Some organisms reproduce asexually, so interbreeding does not apply to them.
- Some closely related species occasionally produce fertile hybrids.
- Testing interbreeding in captivity may not reflect what happens in the wild.

Because of these problems, **competing species definitions** exist. Examples include the morphological concept (shared traits), an ecological concept (a shared niche) and a phylogenetic concept (a group sharing a unique common ancestry, often judged from DNA sequences). Biologists choose the concept that suits the organisms they study.

Why populations and species are hard to tell apart

Speciation usually happens **gradually** rather than by a single act. When populations stop interbreeding, each one changes independently, so they become more and more different in their traits. There is no clear moment when two populations become two species.

At an intermediate stage, populations may be partly different, with some interbreeding still possible. It can therefore be an **arbitrary** decision whether two populations are regarded as the same species or different species.



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

Speciation is **gradual**: once populations stop interbreeding they diverge, and there is no clear moment when they become **two species**.

Chromosome numbers and karyograms Spec A3.1.6, A3.1.7

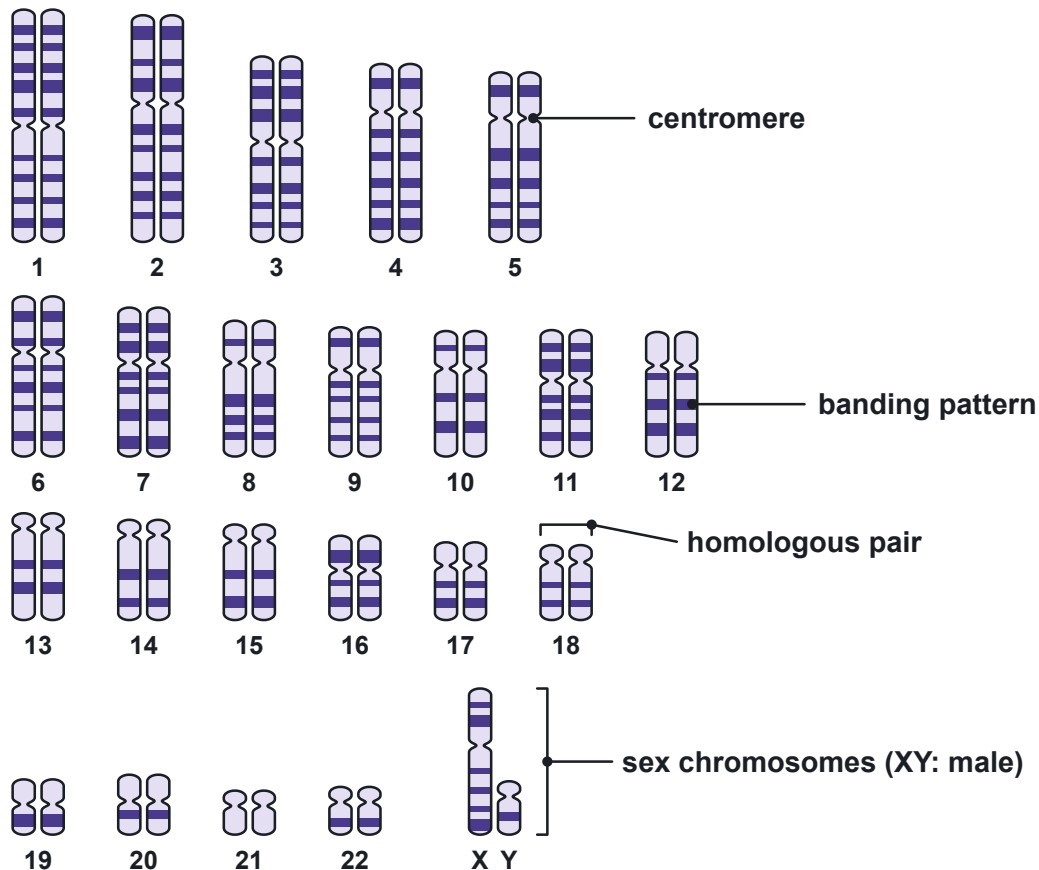
Plant and animal species show great **diversity in chromosome number**. Humans have **46** chromosomes and chimpanzees have **48**. Diploid cells normally have an **even number**, because their chromosomes are in homologous pairs, one from each parent.

Karyotype	The number and type of chromosomes present in the nucleus of a cell of an organism.
Karyogram	An image of the chromosomes of a cell arranged in homologous pairs, in order of decreasing length.
Centromere	The region where the two sister chromatids are held together; its position along the chromosome is used to identify chromosomes.

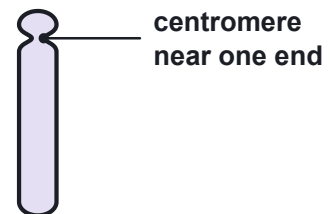
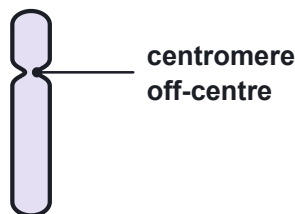
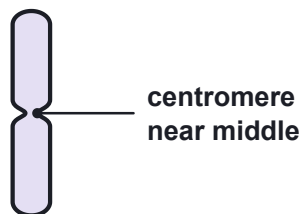
Making a karyogram

1. Take cells that are dividing and stop them in metaphase of mitosis, when chromosomes are most condensed.
2. Spread the chromosomes on a slide and stain them, which gives each chromosome a pattern of light and dark bands.
3. Photograph the chromosomes under a microscope.
4. Pair up homologous chromosomes and arrange the pairs by length, longest first, with the sex chromosomes last.

Chromosomes are classified by three features: **length**, **centromere position** (near the middle, off-centre or near one end) and **banding pattern**. A karyogram shows the sex of the individual (XX or XY) and any unusual chromosome number, such as three copies of chromosome 21.



centromere position



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

A male karyogram (XY): homologous pairs arranged by **length**, **centromere position** and **banding pattern** (bands drawn schematically).

The chromosome 2 fusion hypothesis

Humans and chimpanzees share a common primate ancestor. The hypothesis is that human **chromosome 2** arose from the end-to-end **fusion** of two chromosomes inherited from that shared ancestor. These two chromosomes are still separate in chimpanzees: the IB guide calls them chromosomes 12 and 13, from an older numbering, and today they are usually called chromosomes 2A and 2B. The fusion happened in the human lineage after it split from the chimpanzee lineage, and it reduced the diploid number from 48 to 46. The evidence:

- The banding pattern of human chromosome 2 matches that of chimpanzee chromosomes 12 and 13 (2A and 2B) placed end to end.
- Telomere sequences, normally found only at chromosome ends, occur near the middle of human chromosome 2, at the proposed fusion point.

- Human chromosome 2 has remains of a second, inactive centromere, in the position expected for the centromere of the second ancestral chromosome.
- Genes on human chromosome 2 are largely in the same order as on the two chimpanzee chromosomes.

The evidence strongly supports the hypothesis but the fusion itself was never observed. The match is not perfect in every detail, because each lineage has had some smaller chromosome changes, such as inversions, since the split.

Practical skill: When classifying chromosomes in a karyogram, compare length first, then centromere position, then banding pattern. When evaluating the fusion hypothesis, say what each piece of evidence predicts and whether it is observed.

Exam tip: Nature of science: the origin of chromosome 2 is a testable hypothesis, because it makes predictions (telomere sequences mid-chromosome, a second centromere) that could have been shown false. A statement that makes no testable prediction is not scientific.

Unity and diversity of genomes Spec A3.1.8, A3.1.9, A3.1.10

Genome	All the genetic information of an organism.
Single-nucleotide polymorphism (SNP)	A position in the base sequence of a gene or genome where individuals of a species differ in a single base.
Genome size	The total amount of DNA in one set of an organism's chromosomes, usually given in base pairs.

Unity and diversity within a species

Organisms in the same species **share most of their genome**: they have the same genes in the same order on the same chromosomes. This is the unity. Diversity comes from small differences in base sequence, mainly **single-nucleotide polymorphisms** (SNPs). Different versions of a gene (alleles) often differ in only one or a few bases. Any two humans have about 99.9% of their base sequence in common.

Diversity of eukaryote genomes

- Genomes vary in **overall size**, which is determined by the total amount of DNA.
- Genomes vary in **base sequence**, including which genes are present.
- Variation **between species** is much larger than variation within a species.

Comparing genome sizes

Approximate haploid (1C) genome sizes, rounded, from genome databases such as NCBI Genome and the Plant DNA C-values database		
Organism	Group	Genome size / million base pairs
<i>Escherichia coli</i>	bacterium (prokaryote, for comparison)	4.6
<i>Saccharomyces cerevisiae</i> (yeast)	fungus	12
<i>Arabidopsis thaliana</i>	flowering plant	135
<i>Drosophila melanogaster</i> (fruit fly)	insect	180
<i>Homo sapiens</i> (human)	mammal	3,100
<i>Triticum aestivum</i> (bread wheat)	flowering plant	16,000
<i>Paris japonica</i>	flowering plant	149,000

Very simple organisms tend to have small genomes, but there is **no close correlation between genome size and complexity**. Some plants have genomes many times larger than the human genome. Much of a large genome is non-coding DNA, including repeated sequences, rather than extra genes. Some plants, including bread wheat and *Paris japonica*, are also polyploid, with several whole sets of chromosomes in each nucleus.

Comparing genome sizes

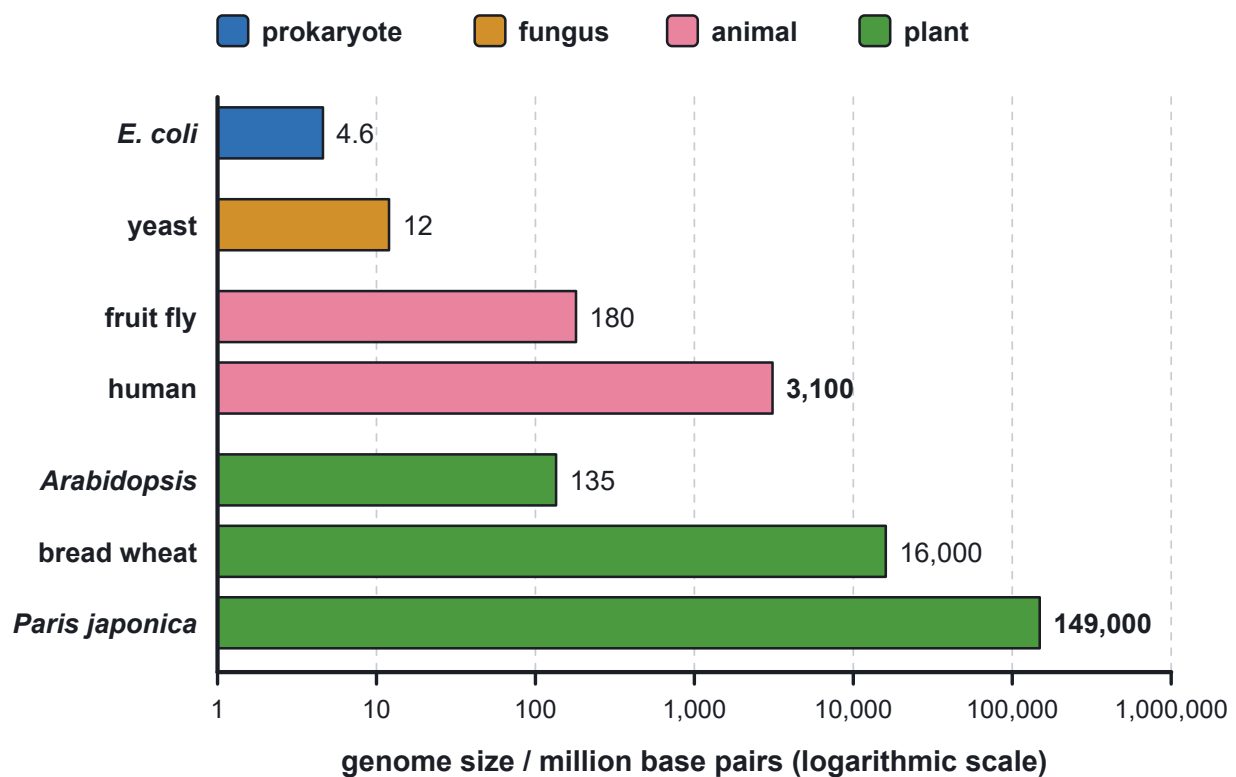
Using the table, calculate how many times larger the genome of bread wheat is than the human genome.

1. Ratio = wheat genome size ÷ human genome size

2. = 16,000 ÷ 3,100

3. = 5.16

Answer: The wheat genome is about 5.2 times larger than the human genome.



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

Genome sizes from the table on a **logarithmic scale**: *Paris japonica* has about 50 times more DNA than a human, so size does **not** track complexity.

Practical skill: Tools: genome sizes span many orders of magnitude, so plot them on a logarithmic scale. When extracting data from a database, record the source and use the same measure (haploid size in base pairs) for every organism.

Whole genome sequencing Spec A3.1.11

Whole genome sequencing means determining the complete base sequence of all the DNA of an organism. The Human Genome Project produced the first human genome sequence, completed in 2003 after about 13 years and at a cost of billions of US dollars.

Since then, automated high-throughput sequencers have given **increasing speed and decreasing costs**. A human genome can now be sequenced in around a day for roughly a thousand US dollars or less. This

means thousands of genomes, from many individuals and species, can be compared.

Uses of whole genome sequencing	
Use	What it involves
Current: research into evolutionary relationships	Comparing whole genomes of different species shows how many base differences separate them, which indicates how closely related they are and when they shared a common ancestor.
Current: research into variation within species	Sequencing many individuals reveals SNPs and links some of them to traits or diseases.
Potential future: personalised medicine	A patient's own genome could be used to predict disease risk and to choose the drugs and doses that work best for that person.

Personalised medicine raises issues too: genome data is personal, and knowing a genetic risk can affect insurance, employment and family members.

Exam tip: For whole genome sequencing, the guide expects three points: speed up and cost down, evolutionary relationships as a current use, and personalised medicine as a possible future use.

Limits of the biological species concept Spec A3.1.12, A3.1.13 HL only

Asexual reproduction

The biological species concept depends on **interbreeding**. Organisms that reproduce only **asexually** produce offspring from a single parent, so there is no interbreeding to test. Every individual is reproductively isolated, so by this definition each clone could count as a separate species, which is not useful.

Bacteria and horizontal gene transfer

Bacteria reproduce asexually by binary fission. They can also gain genes from other bacteria, even of other species, by **horizontal gene transfer**: for example plasmids passed from cell to cell. Genes therefore move between groups that the concept would treat as separate species, so the boundaries between bacterial species are blurred. Bacterial species are usually defined by similarity of base sequences instead.

Chromosome number as a shared trait

Members of a species have the **same chromosome number**. Cross-breeding between closely related species is **unlikely to produce fertile offspring** if the parents have different chromosome numbers.

- A horse (64 chromosomes) and a donkey (62) can produce a mule, with 63 chromosomes.
- In meiosis, the mule's chromosomes cannot all form homologous pairs, so meiosis fails and few or no viable gametes form.
- The mule is therefore infertile, which confirms that horses and donkeys are different species.

Exam tip: Linking question: what might cause a species to persist or go extinct? Reproductive isolation, such as a difference in chromosome number, keeps a species' gene pool separate.

Dichotomous keys and environmental DNA Spec A3.1.14, A3.1.15 HL only

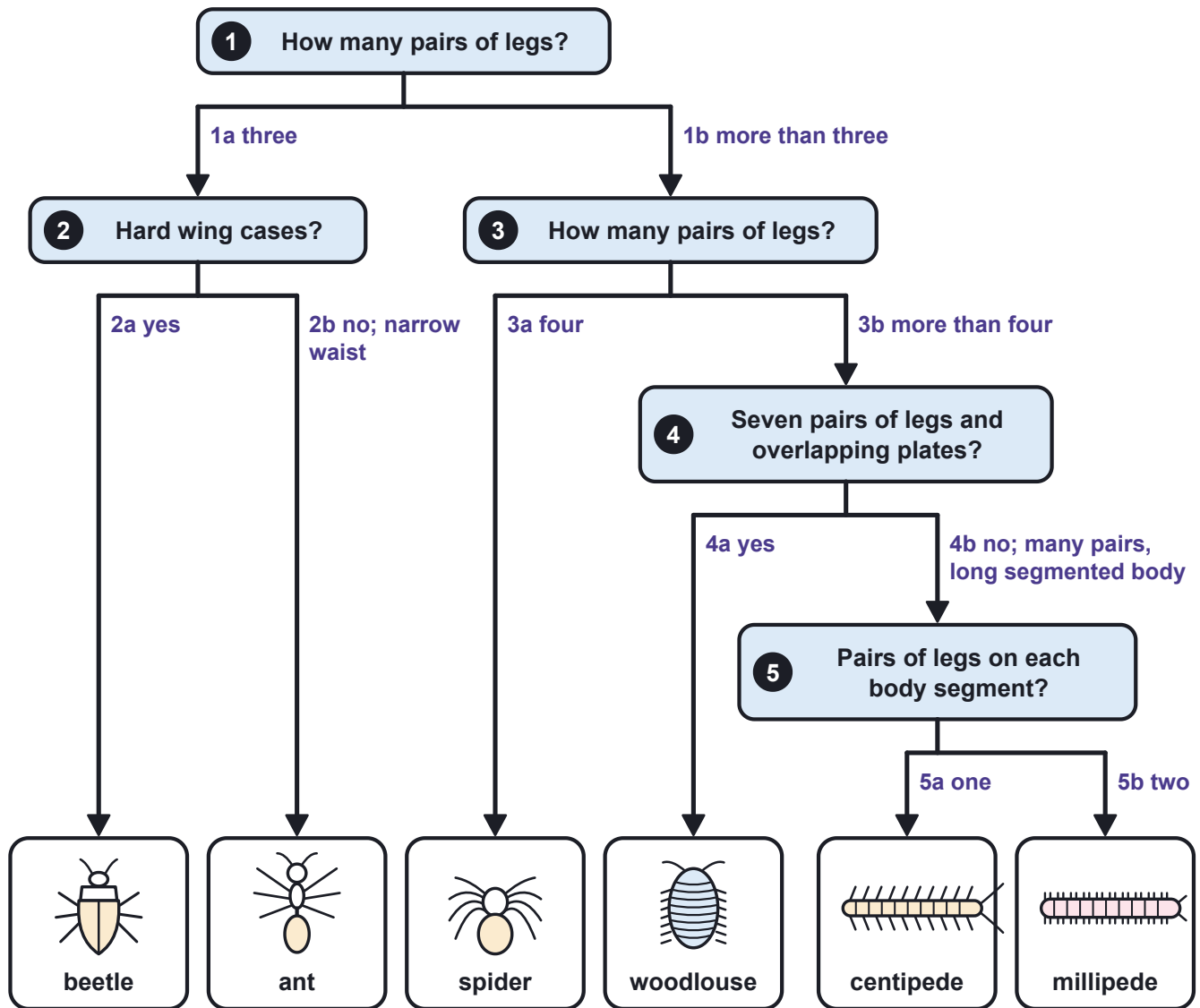
Dichotomous keys

A **dichotomous key** is a series of numbered steps, each with **two contrasting descriptions**. Choosing the description that fits leads either to another step or to the name of the organism.

A key to common invertebrates found in leaf litter

Step	Description	Go to
1a	Three pairs of legs	2
1b	More than three pairs of legs	3
2a	Hard wing cases covering the abdomen	beetle
2b	No wing cases; narrow waist between thorax and abdomen	ant
3a	Four pairs of legs	spider
3b	More than four pairs of legs	4
4a	Seven pairs of legs; body covered by overlapping plates	woodlouse
4b	Many pairs of legs; long, segmented body	5
5a	One pair of legs on each body segment	centipede
5b	Two pairs of legs on each body segment	millipede

Practical skill: Build your key from local species you can observe. Use features that are easy to see and do not vary within a species, such as leg number or leaf arrangement, not size or colour. Test the key on specimens and revise any step that gives the wrong name.



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

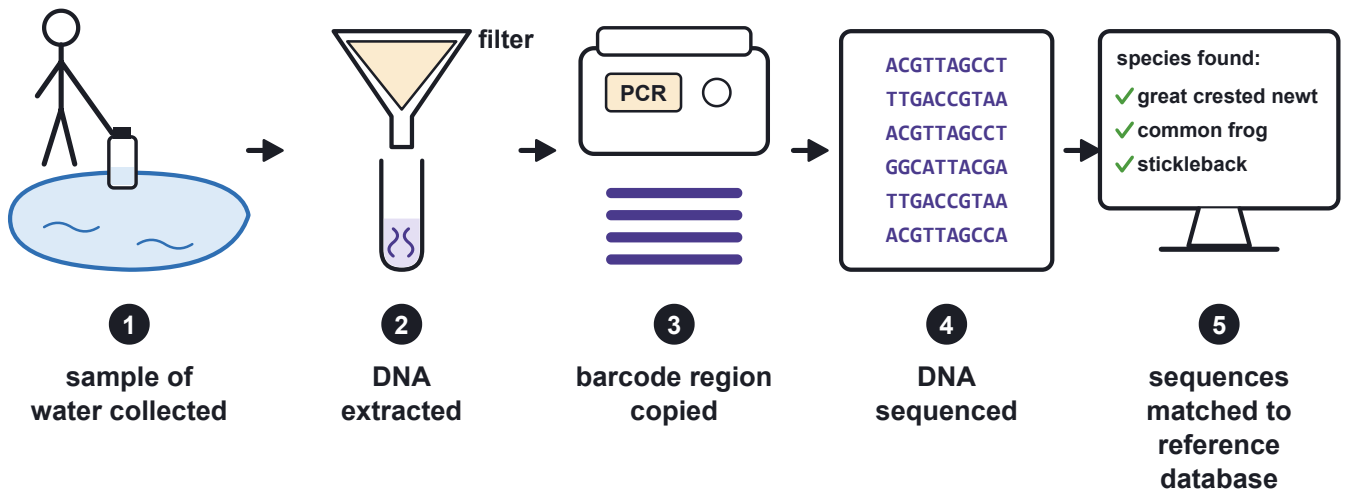
The leaf-litter key as a flow chart: each numbered step has **two contrasting choices**.

Environmental DNA and barcodes

Organisms constantly shed cells, mucus, faeces and gametes, so soil and water contain their DNA. This is **environmental DNA** (eDNA). A **DNA barcode** is a short, standard section of the genome that differs between species, for example part of a mitochondrial gene in animals or a chloroplast gene in plants.

1. Collect a sample of water, soil or sediment from the habitat.
2. Extract the DNA from the sample.
3. Copy the barcode region by PCR, using primers that bind to sequences shared by a whole group of organisms.
4. Sequence the copied DNA.
5. Compare each sequence with a reference database of barcodes from identified species to find which species are present.

Using barcodes and eDNA allows the **biodiversity of habitats to be investigated rapidly**. It can detect rare, hidden or microscopic species without catching or even seeing them, for example great crested newts in a pond from a water sample. Limits: a species can only be named if its barcode is in the database, and DNA can drift in from elsewhere or persist after the organism has gone.



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

Using **eDNA** and barcodes: sample, extract, copy by PCR, sequence, then match to a **reference database**.

Quick check

1. In a binomial, which part has a capital letter?
2. State the biological species concept.
3. How many chromosomes do humans and chimpanzees have?
4. What is a genome?
5. **HL only** Why is a mule infertile?

2 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.

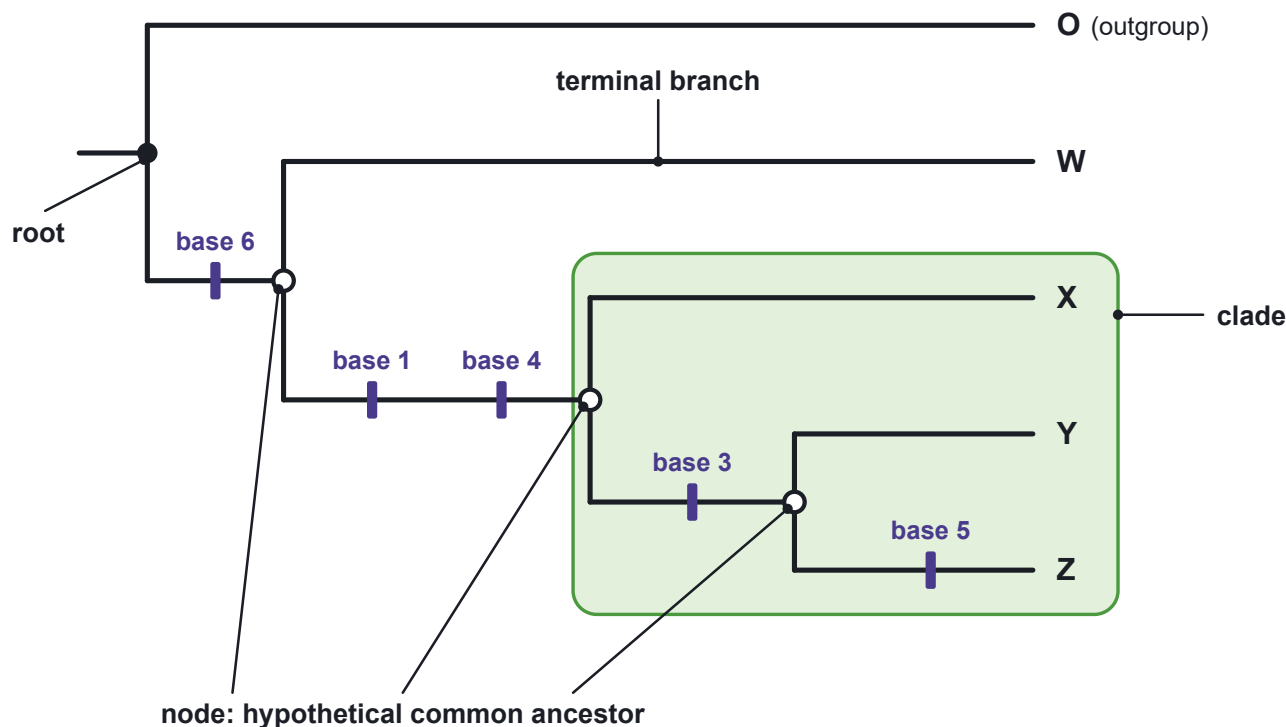
Answer: O branches off first, then W, then X; Y and Z are the closest relatives (sister species).

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.

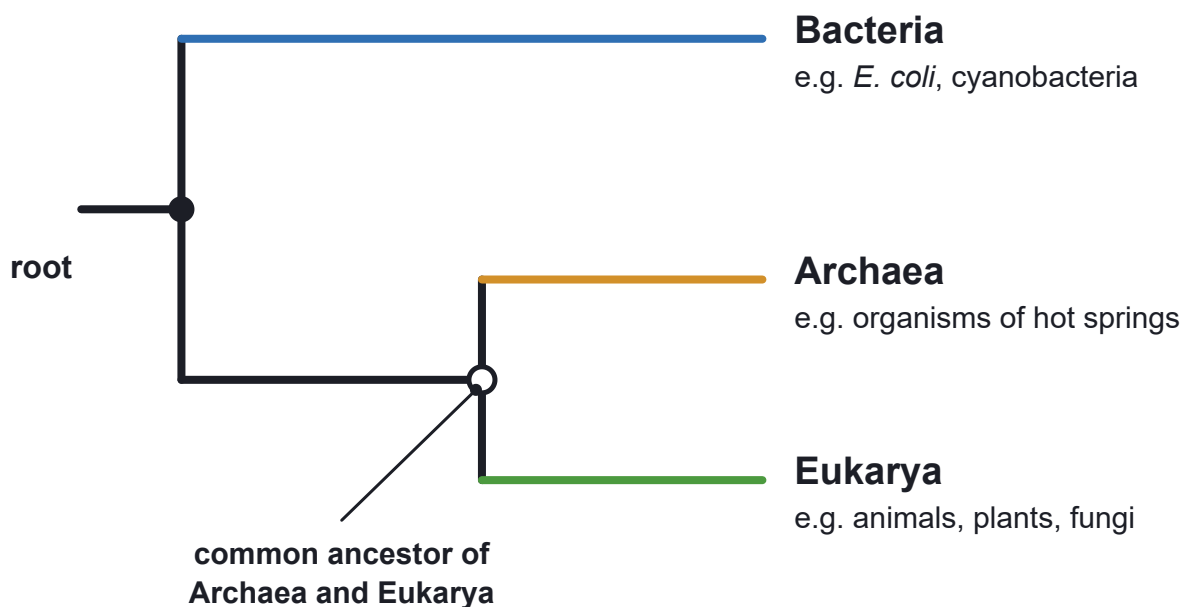
The three domains

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

Diversity of organisms

1. The genus name (first part). The species name is all lowercase.
2. A species is a group of organisms that can breed and produce fertile offspring.
3. Humans 46, chimpanzees 48.
4. All the genetic information of an organism.
5. Its parents have different chromosome numbers, so its chromosomes cannot all pair in meiosis and viable gametes are not produced.

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.