

Diversity of organisms

International Baccalaureate · IB Diploma · Biology · HL

Scope and prerequisites

Two organisms can look alike yet belong to different species. Two cells can contain very different amounts of DNA without the organism with more DNA being more complex. Diversity needs several kinds of evidence, with a clear definition of what is being compared.

This unit covers A3.1: variation, species concepts and names, chromosome numbers and karyograms, evidence for human chromosome 2 fusion, genome diversity and size, and whole-genome sequencing. SL understandings plus limits of species concepts, chromosome pairing, local identification keys and environmental DNA (marked AHL). Prerequisites: DNA, chromosomes, cell division and percentages. Worked datasets and chromosome/leaf diagrams are invented unless a named database or study is explicitly cited.

Variation and the problem of defining a species

Variation means differences among individuals. Some differences are inherited; others arise through environmental conditions or development. Leaf length in one species may vary with genotype, light, water and age. A measured difference therefore does not automatically identify a genetic difference or a different species. Record the character, measurement method, life stage and environmental conditions before comparing samples.

A species is a way of recognizing and naming groups in this diversity. Under a morphological species concept, organisms are grouped by shared structural traits. This is useful for field identification and fossils, where breeding tests or living DNA may be unavailable. Use several stable diagnostic characters rather than one colour or size. Young and adult forms, male and female forms, and environmental effects can make members of one species look different. Conversely, unrelated organisms can evolve similar features, and some distinct species look very similar.

Under the biological species concept, a species consists of populations whose members can interbreed and produce viable, fertile offspring, with reproductive isolation from other such groups. It concerns populations and their potential gene flow; it does not require every individual to mate with every other individual. Age, sex or temporary geographical separation of two individuals does not by itself define different species.

The concepts answer related but different questions: “Can I distinguish these forms?” and “Do their populations form a reproductive group?” There is no single simple test that works perfectly for every organism and every historical sample. State which concept a question uses and which evidence it makes available.

■ Worked example: one visible trait is insufficient

In an invented population, adults have two different colour forms and readily produce fertile offspring with one another under natural conditions. Colour alone is weak evidence for splitting the forms into separate biological species. If a second population looks similar but does not exchange genes because of consistent reproductive barriers, appearance alone could also hide a real species boundary. The conclusion depends on population-level evidence, not an assumption that similar means identical.

Scientific names identify the group precisely

A binomial name contains a genus name followed by a specific epithet. In *Homo sapiens*, *Homo* is the genus and the complete two-word name identifies the species. Capitalize the genus, use lowercase for the epithet, and italicize both when typing. When handwriting, underline the two parts separately. After first writing the full name, the genus may be abbreviated when unambiguous: *H. sapiens*. “Sapiens” alone is not the full species name.

Pan troglodytes and *Pan paniscus* share a genus but are different species. Sharing a genus does not mean sharing every feature or having the same chromosome sequence. Common names can vary between languages and regions, so a recorded scientific name improves communication and makes a database search more precise.

A **taxon** is a named classification group, such as a species or genus. A **clade** comprises a common ancestor and all its descendants. Naming a group and inferring ancestry are linked tasks, but a name alone is not evidence for one branching pattern. A modern species is not automatically the direct ancestor of a different modern species.

Speciation can make boundaries difficult to draw

Speciation is the splitting of an ancestral species into two or more species. Reduced gene flow allows separated populations to diverge through mutation, selection and other population processes. Differences in mating behaviour, breeding time or compatibility can develop gradually. Populations at an intermediate stage may be partly isolated rather than perfectly connected or perfectly separate.

A judgement about whether to name two populations as species may therefore depend on the evidence and the species concept. Consistent genetic differences do not automatically establish complete reproductive isolation; occasional hybridization does not automatically make all populations one species either. Compare fertility, patterns of gene flow, morphology and molecular relationships in context.

An observed sequence comparison supplies one piece of evidence. Six differences among 40 aligned sites give $6/40 = 0.15 = 15\%$. This is a difference proportion for that alignment, not a universal species threshold or an evolutionary date. Reliable comparison

needs homologous regions and appropriate alignment. Test whether additional independent characters or genes support the proposed grouping. On an unscaled branching diagram, rotating branches at a node does not change ancestry, and branch lengths must not be read as measured time.

Chromosome number, DNA amount and copying

Chromosome number varies among species. Typical human diploid body cells have 46 chromosomes, arranged as 23 pairs; chimpanzees typically have 48, arranged as 24 pairs. Human gametes normally have 23 chromosomes, while chimpanzee gametes normally have 24. More chromosomes does not automatically mean more genes, more total DNA or greater complexity.

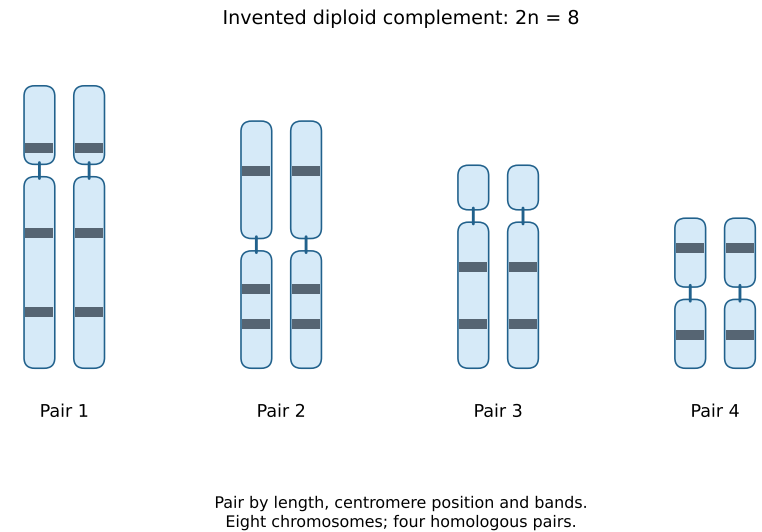
In the usual diploid model, homologous chromosomes occur in pairs, one inherited from each parent. They carry corresponding genes, although their alleles can differ. The paired model gives $2n$ chromosomes; gametes produced by meiosis have n . This describes the normal model, not a claim that chromosome-number abnormalities or unusual sex-chromosome systems cannot exist.

Count chromosomes by centromeres in a standard chromosome-count question. After DNA replication, a human cell still has 46 chromosomes, now with two sister chromatids each, giving 92 chromatids before their separation. Homologous partners are not the same thing as the two sister chromatids of one copied chromosome. The amount of DNA doubles during replication even though the chromosome count at that stage does not.

Build and interpret a karyogram

A karyogram displays chromosomes arranged into matching groups. A karyotype describes the chromosome complement, including number and appearance. In a typical preparation, cells are observed when chromosomes are condensed, stained or otherwise imaged, then chromosomes are identified and arranged. A spread of dark rods needs analysis; it is not already a labelled karyogram.

Match homologues using relative length, centromere position and banding pattern. Length alone is not enough, and orientation can be reversed in an image. In a copied metaphase chromosome, the two joined chromatids belong to one chromosome. Pairing homologues in a karyogram is a comparison made by the analyst, not an assertion that the photographed chromosomes were already physically paired along their lengths.



Four pairs mean eight chromosomes, not four. Each rod represents one chromosome; chromatids are not shown separately. This is not a real human karyogram.

■ Worked example: classify an invented complement

Suppose a sample contains two long chromosomes with matching bands, two medium chromosomes with a central centromere, and two short chromosomes with matching bands. The complement has three homologous pairs: $2n = 6$, so a normal haploid set has $n = 3$. After replication but before chromatid separation, there are still six chromosomes, with 12 chromatids. If the image instead contains a third copy of one type, identify the extra chromosome explicitly rather than forcing everything into pairs.

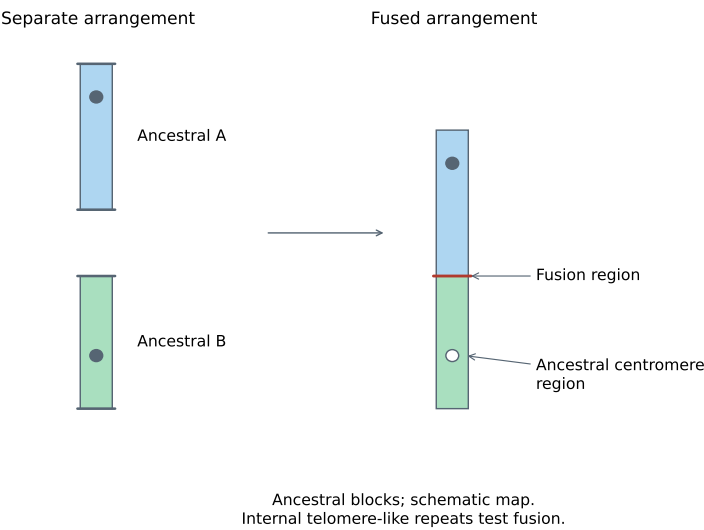
A karyogram can reveal large number changes and some structural changes. It usually cannot reveal a single-base substitution, distinguish every gene variant, or by itself prove that an individual will show a particular trait. Match the claim to the resolving power of the evidence.

Human chromosome 2: test a fusion hypothesis

The chromosome-number difference between humans and chimpanzees suggests a question, not the full answer. If two chromosomes present separately in an ancestral ape lineage fused into one chromosome in the human lineage, corresponding blocks of genes should lie in the appropriate order along human chromosome 2. This is a testable prediction from a historical hypothesis.

Human chromosome 2 corresponds to two chimpanzee chromosomes now called 2A and 2B, labelled 12 and 13 in older nomenclature. Modern chimpanzees are not our direct ancestors: the inference concerns changes in lineages after shared ancestry. Do not

describe a modern chimpanzee suddenly losing two chromosomes to become a human. Evidence includes correspondence of sequence and banding blocks, internal telomere-like repeats near the inferred fusion region, and evidence of an additional ancestral centromere region. Telomeres normally occur at chromosome ends, so oppositely oriented telomere-repeat arrays inside a chromosome are relevant to an end-to-end fusion hypothesis. The 1991 study linked below analysed such internal repeat sequences; later genomic comparisons add context.



The drawing models a testable relationship between chromosome blocks. It does not show a newly observed fusion event or an exact sequence map.

A chromosome count alone would be insufficient: other changes can alter counts. Combining independent observations makes a stronger explanation. The simple fusion model predicts a reduction of two in the diploid number when both homologous sets carry the fused arrangement: two pairs become one pair, while much of their sequence remains represented. It does not predict that half the genome disappears.

Genomes share information and contain variation

A genome is the complete genetic information of an organism. In eukaryotes, distinguish the nuclear genome from DNA in mitochondria or chloroplasts when a comparison specifies the nucleus. A genome includes non-coding regions as well as protein-coding genes. Genome size is an amount of DNA, not the number of chromosomes or genes.

Members of one species share much of their genome but can differ in base sequence. A single-nucleotide polymorphism (SNP) is variation at a particular nucleotide position in a population. Individuals may have different bases at that site. In a diploid indi-

vidual, a site can carry the same allele on both homologues or different alleles. Not every sequence difference changes a protein, and not every protein change produces a visible trait.

Between species, sequences and the total amount of DNA can differ substantially. For a selected comparison, state whether you counted all sites, only coding sites, one sequence region or an entire assembly. A reported similarity percentage without its denominator and method is incomplete. Similar sequence does not mean identical organisms; sequence difference does not by itself date divergence.

Compare genome sizes using a database

Use a named record and version so another student can repeat a comparison. The following values come from NCBI assembly records, rounded for teaching. They describe assembled sequence, not DNA mass measured in a diploid cell.

Organism / group	NCBI assembly accession	Reported total sequence length, rounded
<i>Escherichia coli</i> K-12 MG1655 / bacterium	GCF_000005845.2	4.64 million bp
<i>Saccharomyces cerevisiae</i> S288C / fungus	GCF_000146045.2	12.1 million bp
<i>Homo sapiens</i> / animal	GCF_000001405.40	3,100 million bp

To reproduce the comparison, open the linked NCBI record, record accession/version, organism/strain, assembly name and the statistic used. Distinguish “total sequence length” from “ungapped length”, and check which chromosome sequences and organelar sequences are included. A human reference can include both X and Y sequence types; its count of chromosome sequence types is not the 46 chromosomes in an ordinary diploid body cell. An incomplete assembly can underestimate sequence not yet assembled.

Compare across groups without ranking organisms by DNA quantity. Many differences in genome size involve repeated and other non-coding sequences, duplication and loss. A larger genome is not automatically a larger gene set or a more complex organism. As a plant counterexample, a published study measured about 160.75 billion base pairs per unreplicated nuclear complement in the fork fern *Tmesipteris oblancoolata*. That DNA-content estimate uses a different measurement approach from the assembly totals above; avoid treating every displayed number as exactly the same statistic.

■ **Worked example: compare stated quantities**

An invented plant sample has a haploid nuclear DNA amount of 150 million bp, while an invented animal sample has 3,000 million bp measured on the same basis. The ratio is $3000/150 = 20$. This says the animal sample has 20 times the DNA amount in the stated comparison. It does not show 20 times the complexity, 20 times the genes, or the chromosome number. If one sample were diploid and the other haploid, first correct the comparison rather than dividing unlike quantities.

Whole-genome sequencing: uses and limits

Whole-genome sequencing determines the base sequence across a genome rather than testing only one short marker. Many short or long reads must be checked, assembled or aligned, and interpreted. Improvements have generally increased throughput and reduced cost compared with early genome projects, but data generation and accurate interpretation remain different tasks.

Current research uses include comparing evolutionary relationships, finding variants, investigating inherited conditions, studying pathogen transmission and improving crop research. Genome-scale comparisons can test whether a relationship inferred from one gene is supported by many others. Different parts of a genome can have different histories, so careful analysis still matters.

Personalized medicine aims to use relevant genomic variation with clinical and environmental information to guide prevention or treatment. Sequencing does not guarantee that every detected variant has a known effect or that a person’s future can be predicted precisely. Some results are uncertain, and privacy, consent and access matter. A sequence record is evidence to interpret, not an automatic diagnosis or a complete account of an organism’s phenotype.

AHL: Limits of the biological species concept

Asexual organisms do not normally form reproductive groups by mating, so the biological species concept cannot be applied by a simple fertile-offspring test. Fossils also cannot be bred. Other evidence and species concepts are needed. Bacteria add a further difficulty: horizontal gene transfer can move genes between lineages outside ordinary parent-to-offspring inheritance. One transferred gene may therefore suggest a relationship different from the broader genome history.

Chromosome number is often shared within a species and can affect the success of crosses. During meiosis, homologous chromosomes must align and separate appropriately. Different numbers or structural arrangements can interfere with pairing, giving unbalanced gametes and reduced fertility in a hybrid. Equal chromosome number does not guarantee compatibility, and a count difference is not an infallible universal species test. Combine the mechanism with evidence from the particular populations.

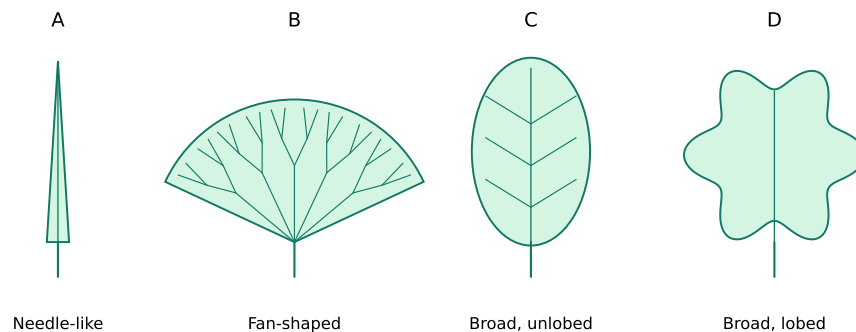
AHL: Build a key from local observations

A dichotomous key identifies a specimen by a sequence of paired alternatives. Each pair should describe observable, contrasting states of the same character. “Large” versus “green” is a poor pair: a leaf can be both. Avoid using habitat or an invisible feature when the key is intended for a detached leaf. State the set of organisms and life stages for which the key was built.

For a local practical, select four permitted plant species or non-destructively observed animal species from the school grounds or a nearby accessible habitat. Confirm their names with a suitable field guide or authoritative collection. Photograph the actual specimens, record several individuals where possible, and make a character table. Do not claim an unverified species name merely because a phone application proposes it. Test the draft key with another student’s specimens and revise ambiguous couplets.

The following is an invented leaf dataset, not a claim about species observed at this school:

Sample	Observable state
A	Needle-like leaf
B	Broad fan-shaped blade with veins repeatedly dividing into two
C	Broad blade, not fan-shaped, with no lobes
D	Broad blade, not fan-shaped, with lobes



Invented character set A–D.
Test a local-species key against observed, verified specimens.

The key identifies A–D within this declared dataset. Replace labels with verified local species only after observing and testing the relevant characters.

One valid key is: 1a needle-like → A; 1b broad → 2. Then 2a fan-shaped with repeatedly dividing veins → B; 2b not fan-shaped → 3. Finally 3a lobed → D; 3b unlobed → C. Every sample should reach one endpoint. An identification key is a decision tool, not automatically a phylogenetic tree: an efficient first split need not represent the earliest ancestral divergence.

AHL: DNA barcodes and environmental DNA

A DNA barcode is a standardized, relatively short sequence used for identification by comparison with reference sequences. A reference must be linked to a reliably identified organism. Different taxonomic groups may require different suitable markers; no single barcode automatically separates every species. A close database match is stronger when the reference coverage, sequence quality and discriminatory power of the marker are known.

Environmental DNA (eDNA) is DNA recovered from environmental samples such as water or soil rather than necessarily from a captured whole organism. Organisms shed cells and other material; sequencing suitable barcode regions can reveal evidence of taxa in a habitat. This can investigate biodiversity with less disturbance and can detect organisms difficult to observe directly.

Use collection blanks, extraction controls and comparison samples to assess contamination and method performance. DNA can move with water and persist after an organism has left. A detection does not automatically locate a living individual at that exact point and time. Failure to detect DNA does not prove absence if sampling was limited, the DNA degraded or the chosen marker amplified poorly.

Sequence-read abundance is not automatically the number or biomass of organisms: shedding, transport, extraction and amplification differ. Use eDNA results alongside repeated sampling and other ecological observations. Barcoding supports identification; whole-genome sequencing examines much more information and answers additional questions.

Practice

Use stated species concepts and units. All unreferenced numerical datasets are invented.

1. Explain how inherited and environmental differences can both contribute to variation. Give one strength and two limitations of grouping species by morphology alone.
2. Write the scientific name *Pan troglodytes* with the correct capitalization and explain both parts. Compare it with *Pan paniscus*. Why is an epithet alone an incomplete name?
3. Explain the biological species concept and why two colour forms producing fertile offspring are not automatically separate species. Describe one difficulty when two populations are only partly reproductively isolated.
4. State typical human and chimpanzee diploid and gamete chromosome numbers. For a human cell after DNA replication but before chromatid separation, state chromosome and chromatid numbers and explain the difference.

5. The invented karyogram has four homologous pairs. Find $2n$, n , and the number of chromatids after replication but before separation. Give three features used to pair the chromosomes.
6. Explain two predictions or observations relevant to the human chromosome 2 fusion hypothesis. Why are chromosome count alone and the claim “humans came from modern chimpanzees” inadequate?
7. Define genome and SNP. Explain why genome size, gene count and chromosome count are different measurements, and why not every SNP produces a visible trait.
8. An alignment has 8 differences among 50 sites; another has 7 among 25. Calculate both percentages. Explain why these are neither universal species thresholds nor calibrated dates. What does rotating branches at an unscaled node change?
9. Use the three NCBI assembly rows to identify the largest and smallest reported assembled sequence totals. Explain why the reference’s chromosome sequence-type count must not be substituted for human diploid chromosome number.
10. An invented plant and animal have respectively 150 and 3,000 million bp per haploid nuclear complement. Calculate the ratio. If the plant measurement instead represents a diploid cell containing 300 million bp, show the correction needed for a fair comparison.
11. Give two research uses of whole-genome sequencing and one potential use in personalized medicine. Explain two reasons a sequence does not provide a complete prediction of phenotype.
12. **AHL:** Explain why asexual reproduction and horizontal gene transfer challenge a fertile-offspring species test. Describe how chromosome-number differences can reduce hybrid fertility, without treating count as a perfect species boundary.
13. **AHL:** Write a key for A–D in the leaf table, trace sample D through it, and explain how to replace the invented labels with a tested local-species key. Why is this key not automatically an evolutionary tree?
14. **AHL:** Explain how barcodes and eDNA can identify taxa. Give two controls or validation steps, two reasons a positive detection need not locate a living organism exactly, and one reason read number is not organism number.

■ Worked answers

1. Different inherited alleles and different conditions such as light or water can affect a trait. Morphology enables practical identification, including fossils. Limits include variation with age/sex/environment and similar-looking distinct species or convergent structures. Use multiple diagnostic characters and suitable comparison samples.
2. *Pan* has an initial capital; *troglodytes* is lowercase; italicize both. *Pan* is the genus and the whole binomial names the species. *P. paniscus* shares the genus but is a different species. The epithet alone omits the genus needed for the complete species name.

3. Populations belong to a potentially interbreeding group producing viable, fertile offspring and reproductively isolated from other groups. Fertile interbreeding colour forms can be variation within one species. Partial isolation can develop gradually; different concepts or evidence may give different boundary decisions.
4. Humans: 46 diploid, 23 in a normal gamete. Chimpanzees: 48 and 24. A replicated human cell has 46 chromosomes and 92 chromatids before separation. DNA copying makes sister chromatids without yet separating their centromeres into different chromosomes.
5. $2n = 8$, $n = 4$ and 16 chromatids after replication before separation. Match relative length, centromere position and banding pattern. The four displayed pairs are eight chromosomes, not four.
6. Corresponding sequence/banding blocks should align with two separate ancestral chromosome types; internal oppositely oriented telomere-like arrays and an additional ancestral centromere region fit fusion. Count alone cannot specify the structural event. Modern chimpanzees and humans share ancestors; one modern species is not automatically the other's direct ancestor.
7. A genome contains complete genetic information, including coding and non-coding regions; specify nuclear versus organellar DNA when needed. A SNP is nucleotide-position variation in a population. DNA amount differs from how DNA is divided among chromosomes or how many genes it contains. A variant can be non-coding, have a silent coding effect or depend on biological context.
8. $8/50 \times 100 = 16\%$ and $7/25 \times 100 = 28\%$. These describe the selected alignments, without a universal species cutoff or a calibrated change-rate model. Rotating branches changes page order, not branching relationships; arbitrary branch lengths do not measure elapsed time.
9. The human reference row has the largest total, about 3,100 million bp; the bacterial row the smallest, about 4.64 million bp. An assembly is a sequence representation with a declared component set, not an ordinary diploid cell. A reference can represent both X and Y types without containing a real 46-chromosome body-cell complement.
10. $3000/150 = 20$. For the alternative plant measurement, a comparable haploid amount is $300/2 = 150$ million bp, again giving 20. Dividing $3000/300 = 10$ would mix haploid and diploid quantities. No ratio establishes a complexity or gene-count ratio.
11. Examples include evolutionary comparisons, variant discovery, inherited-condition research, pathogen studies and crop research. A potential medical use is considering relevant variants with other evidence when choosing prevention or treatment. Unknown variant effects and environmental/developmental influences limit prediction; sequence quality and interpretation also matter.
12. **AHL:** Asexual lineages cannot be grouped through normal interbreeding tests, while horizontal transfer can move genes outside parent-offspring descent. Different chromosome numbers/arrangements can disrupt meiotic alignment and pro-

duce unbalanced gametes, reducing hybrid fertility. Equal counts do not ensure compatibility, and count differences require case-specific interpretation.

13. **AHL:** 1a needle-like $\rightarrow$ A; 1b broad $\rightarrow$ 2. 2a fan-shaped/dividing veins $\rightarrow$ B; 2b not fan-shaped $\rightarrow$ 3. 3a lobed $\rightarrow$ D; 3b unlobed $\rightarrow$ C. D follows 1b, 2b, 3a. Observe and verify local species, record characters across samples, then test and revise the key. Its splits are identification decisions, not necessarily ancestry claims.
14. **AHL:** Compare a suitable marker from a sample with reliably identified reference sequences. Use collection/extraction blanks and validated references or positive controls, for example. DNA transport and persistence can separate detection from a current living individual. Unequal shedding and amplification can make read counts differ from organism counts; non-detection also needs cautious interpretation.

Exam tips

Name the species concept and the evidence it needs. Keep chromosome, chromatid, gene and genome-size counts separate. Match karyogram chromosomes by several features, and evaluate fusion using several predictions. Record database versions and compare the same DNA quantity. For a key or eDNA result, state the organisms and conditions for which identification is supported.

References: IB Biology guide, A3.1; chromosome 2 fusion study; chimpanzee genome comparison; NCBI bacterial assembly, yeast assembly, human assembly; fern DNA-content study.