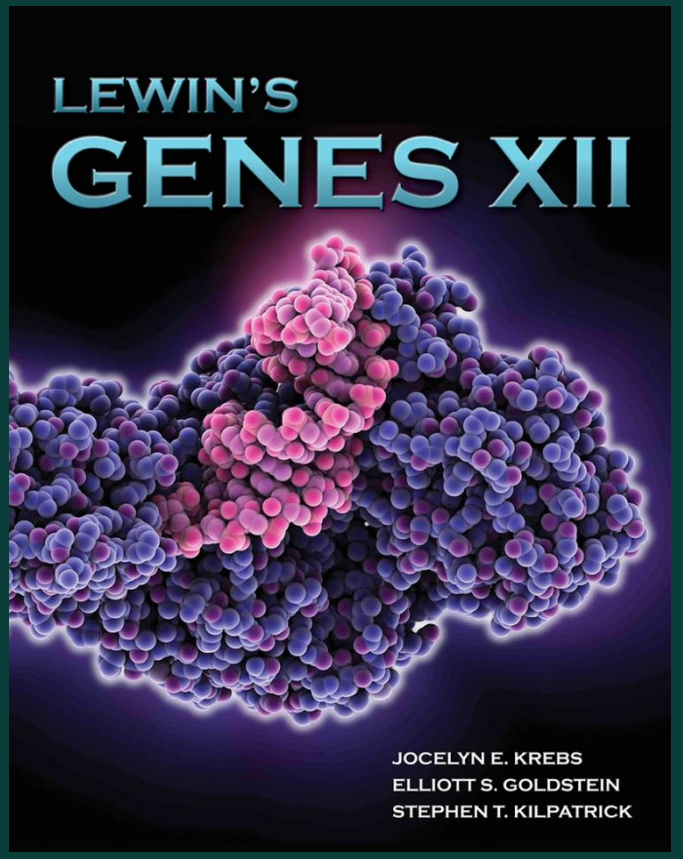
Genes & Genomes: Structure, Content, and Evolution

Genes, Genomas y Proteomas



Source: Lewin's Genes XII (Krebs, Goldstein & Kilpatrick, 2018) · Chapters 1, 3, 4, 5, 6

A Detective Story: How We Learned the Genome

- **Discovery of nucleic acids** — a chemical no one suspected of carrying heredity.
- **DNA is the hereditary molecule** — transformation, then DNase, then phage labeling.
- **Genes encode proteins** — one gene–one enzyme, colinearity, the cistron.
- **Gene architecture: pro- vs eukaryotes** — compact operons vs. interrupted, regulated genes.
- **The rest of the genome** — repeats, transposons, pseudogenes, organelles.
- **Genome evolution** — duplication, transposition, drift, the molecular clock.

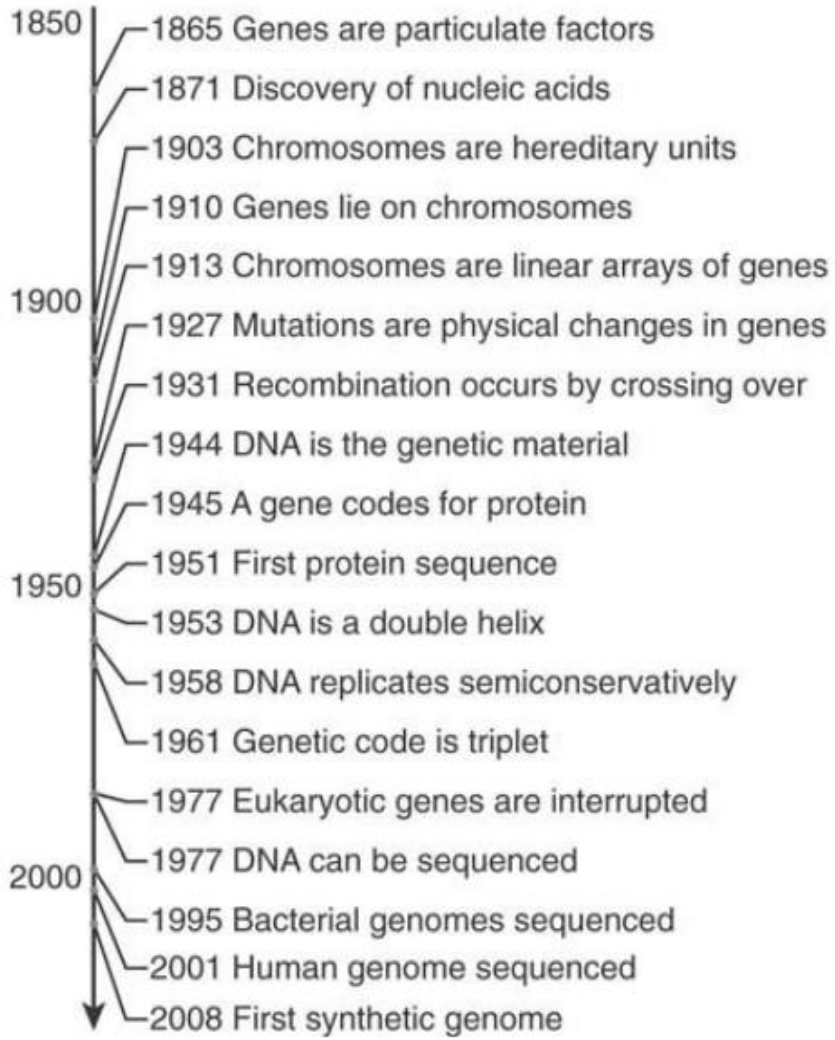


FIGURE 1.1 A brief history of genetics.

PART 1 · DISCOVERY

Is There a Molecule of Heredity?

From 'Nuclein' to a Chemical Suspect

- **Friedrich Miescher** isolated 'nuclein' from leukocyte nuclei (1869) — acidic, phosphorus-rich, clearly not a protein.
- Chemistry resolved: 4 bases (A, G, C, T) on a sugar–phosphate backbone — a 5′→3′ polynucleotide.
- **Levene's tetranucleotide hypothesis** — wrongly assumed equal, repeating bases. Too monotonous to be the gene; protein looked richer.
- **Chargaff's rules (1950):** the quantitative clue: %A = %T and %G = %C, but (A+T)/(G+C) varies by species — not monotonous after all.
  - Realization: composition is species-specific and constrained — DNA could carry information.

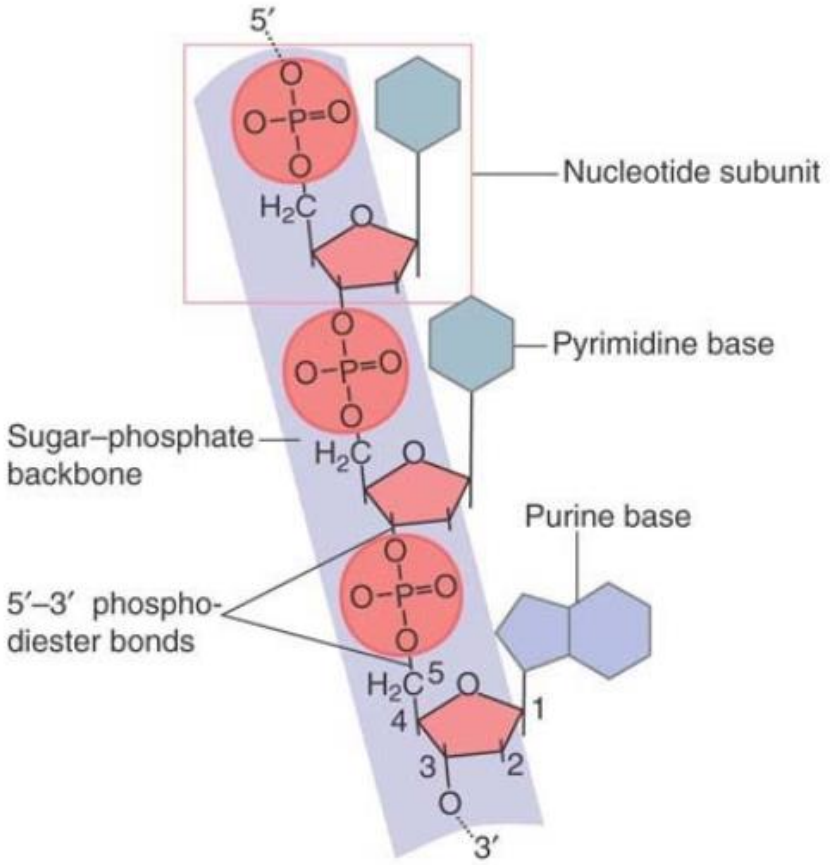


FIGURE 1.9 A polynucleotide chain consists of a series of 5′–3′ sugar–phosphate links that form a backbone from which the bases protrude.

Griffith 1928: An Invisible 'Transforming Principle'

- **Two strains:** smooth (S, virulent, capsule) vs rough (R, harmless) pneumococcus.
- Live R → mice live. Live S → die. Heat-killed S → live. But heat-killed S + live R → mice DIE.
- Live S-type bacteria are recovered from the dead mice — the R cells were permanently changed.
- **Inference:** Something released by dead cells reprogrammed living ones — heredity rides on a transferable molecule.
  - The identity of the 'principle' stayed unknown for 16 years.

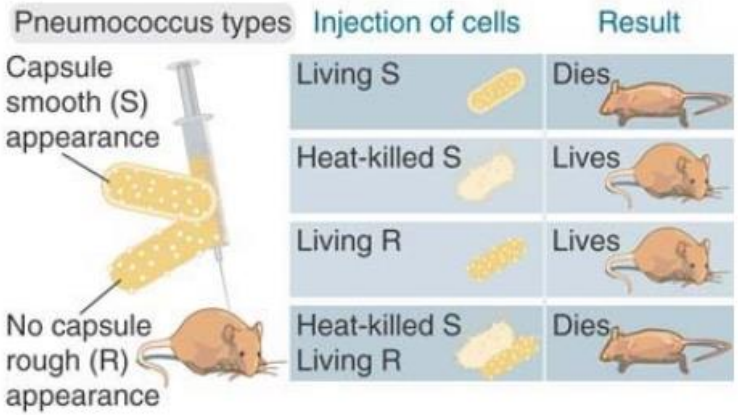


FIGURE 1.4 Neither heat-killed S-type nor live R-type bacteria can kill mice, but simultaneous injection of both can kill mice just as effectively as the live S type.

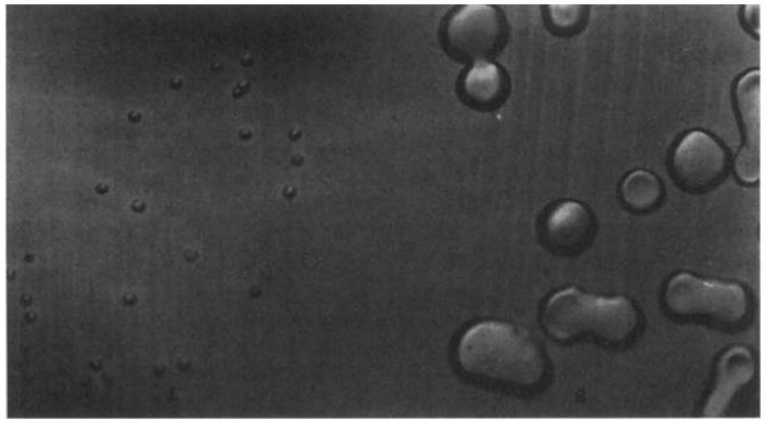


FIGURE 1.6 Rough (left) and smooth (right) colonies of *S. pneumoniae*.

Avery, MacLeod & McCarty 1944: The Principle Is DNA

- **Strategy:** purify Griffith's transforming activity, then destroy candidate molecules one at a time.
- Proteases and RNase → transformation still works. DNase → transformation is abolished.
- Purified DNA from S cells alone converts R → S, heritably.
- **Conclusion:** The transforming principle is DNA — the first direct evidence the gene is made of DNA.
  - Resistance was strong: many still believed 'simple' DNA couldn't encode genes.

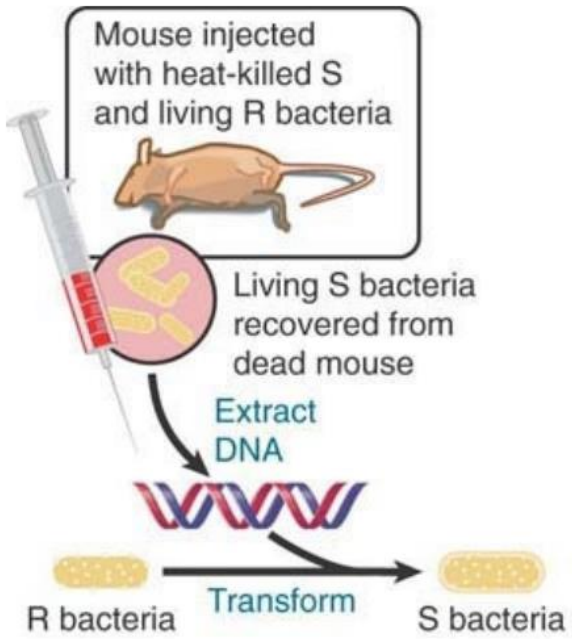
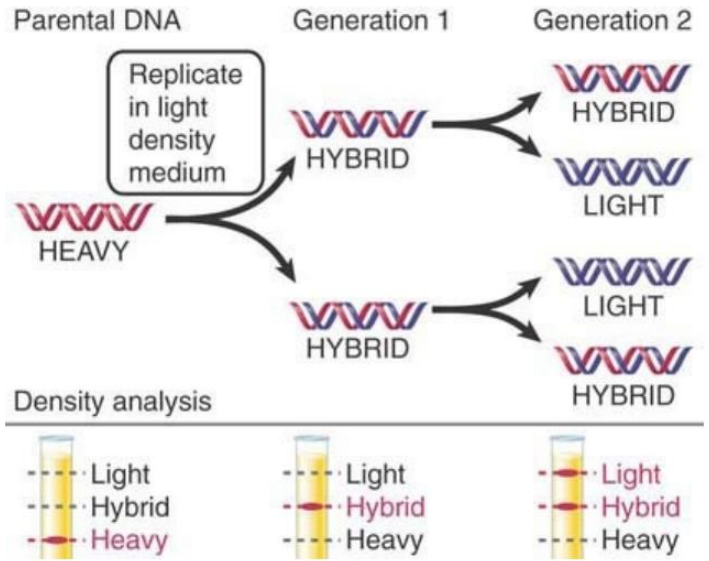


FIGURE 1.5 The DNA of S-type bacteria can transform R-type bacteria into the same S type.

Hershey–Chase 1952 & the Double Helix (1953)

- **Hershey–Chase:** phage protein with ^{35}S, DNA with ^{32}P; infect, then shear cells in a blender.
- Only ^{32}P (DNA) enters the bacterium and is inherited by progeny phage — DNA is the genetic material.
- **Watson & Crick:** antiparallel complementary strands; A·T and G·C pairing (explaining Chargaff).
- Complementarity immediately suggested a copying mechanism — 'it has not escaped our notice...' – Pattern → mechanism: structure told scientists how heredity could be replicated.



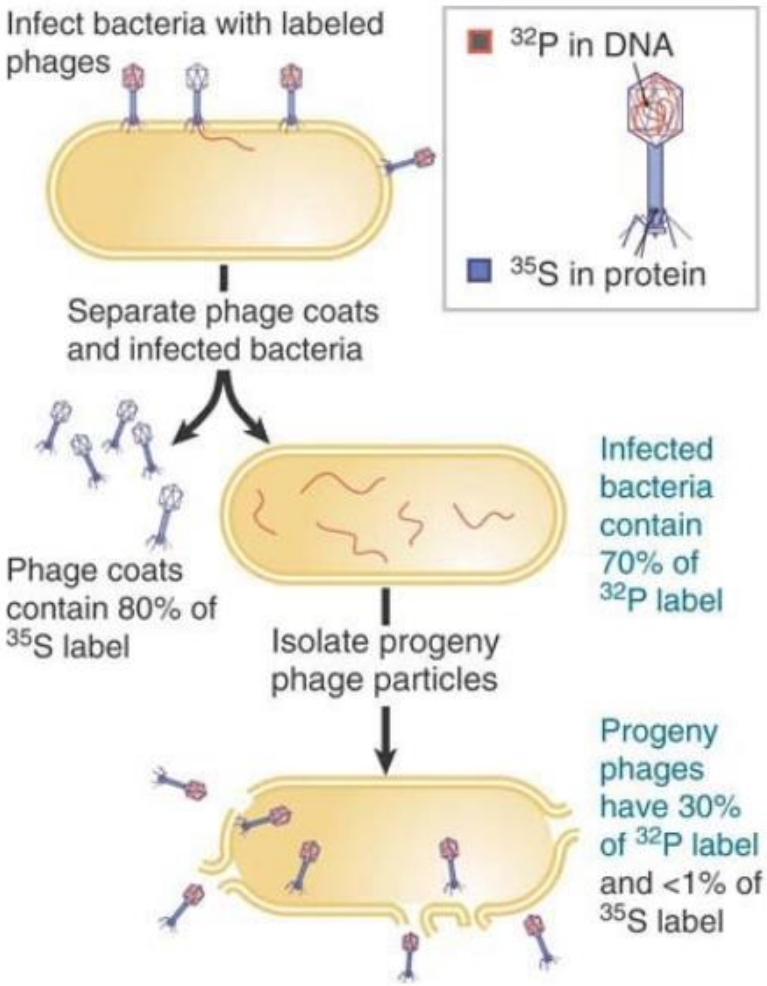


FIGURE 1.7 The genetic material of phage T2 is DNA.

PART 2 · GENE = INFORMATION

Linking Nucleic Acids to Genes

Genes Specify Proteins: One Gene–One Enzyme

- **Garrod:** 'inborn errors of metabolism' — a blocked step = a missing enzyme = a heritable defect (early 1900s).
- **Beadle & Tatum (1941):** X-ray–mutagenize Neurospora; each mutant fails at one biosynthetic step and is rescued by one nutrient.
- Each single-gene mutation knocks out one enzyme → one gene–one enzyme (later one gene–one polypeptide).
- **Take-home:** A gene is a unit of information read out as a specific protein.

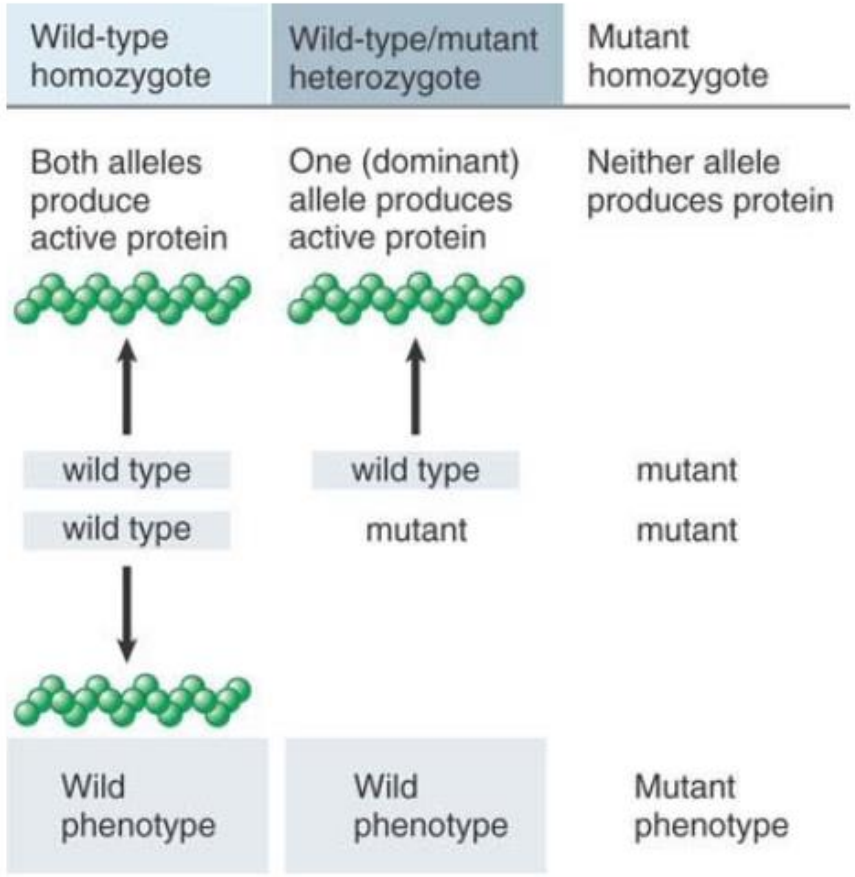


FIGURE 1.33 Genes encode proteins; dominance is explained by the properties of mutant proteins. A recessive allele does not contribute to the phenotype because it produces no protein (or protein that is nonfunctional).

The Gene Is Colinear with Its Protein

- **Yanofsky:** the order of mutations in trpA maps in the same order as the altered amino acids — DNA and protein are colinear.
- **Crick & Brenner:** frameshift experiments show the code is read in non-overlapping triplets from a fixed start.
- **The cistron (complementation test)** defines a gene functionally: two recessive mutations that fail to complement lie in the same unit.
- **cis vs trans:** — sites act only on their own DNA (promoters); trans-acting products diffuse (proteins).

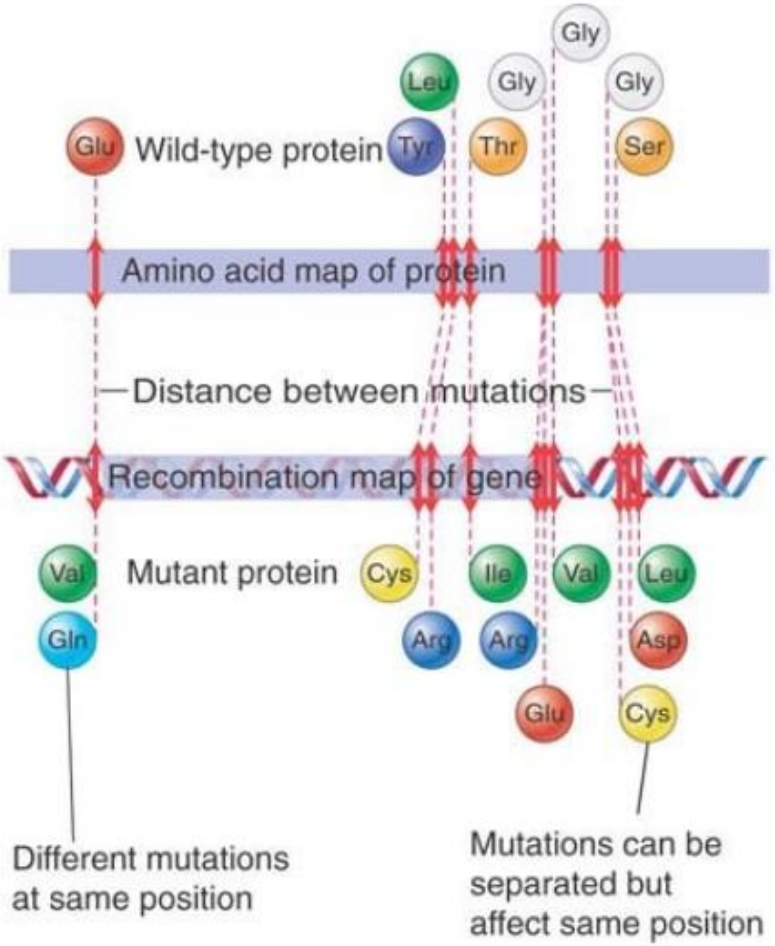


FIGURE 1.42 The recombination map of the tryptophan synthetase gene corresponds with the amino acid sequence of the polypeptide.

The Gene Is Colinear with Its Protein

- **Yanofsky:** the order of mutations in trpA maps in the same order as the altered amino acids — DNA and protein are colinear.
- **Crick & Brenner:** frameshift experiments show the code is read in non-overlapping triplets from a fixed start.
- **The cistron (complementation test)** defines a gene functionally: two recessive mutations that fail to complement lie in the same unit.
- **cis vs trans:** — sites act only on their own DNA (promoters); trans-acting products diffuse (proteins).

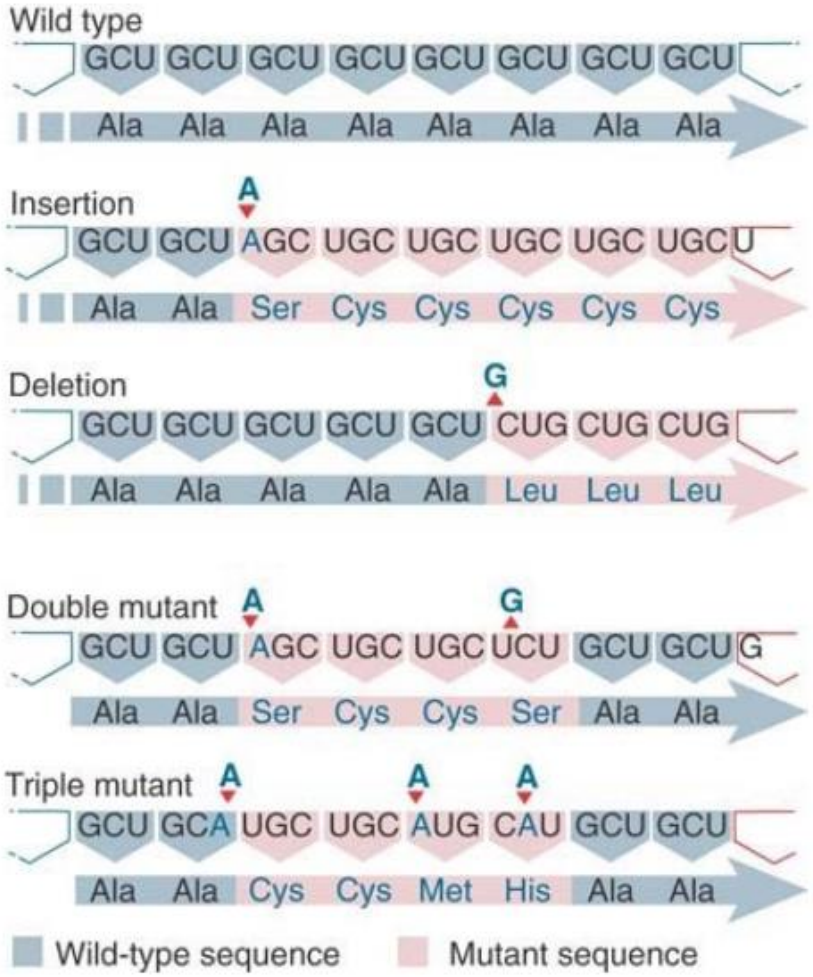


FIGURE 1.40 Frameshift mutations show that the genetic code is read in triplets from a fixed starting point.

The Gene Is Colinear with Its Protein

- **Yanofsky:** the order of mutations in trpA maps in the same order as the altered amino acids — DNA and protein are colinear.
- **Crick & Brenner:** frameshift experiments show the code is read in non-overlapping triplets from a fixed start.
- **The cistron (complementation test)** defines a gene functionally: two recessive mutations that fail to complement lie in the same unit.
- **cis vs trans:** — sites act only on their own DNA (promoters); trans-acting products diffuse (proteins).

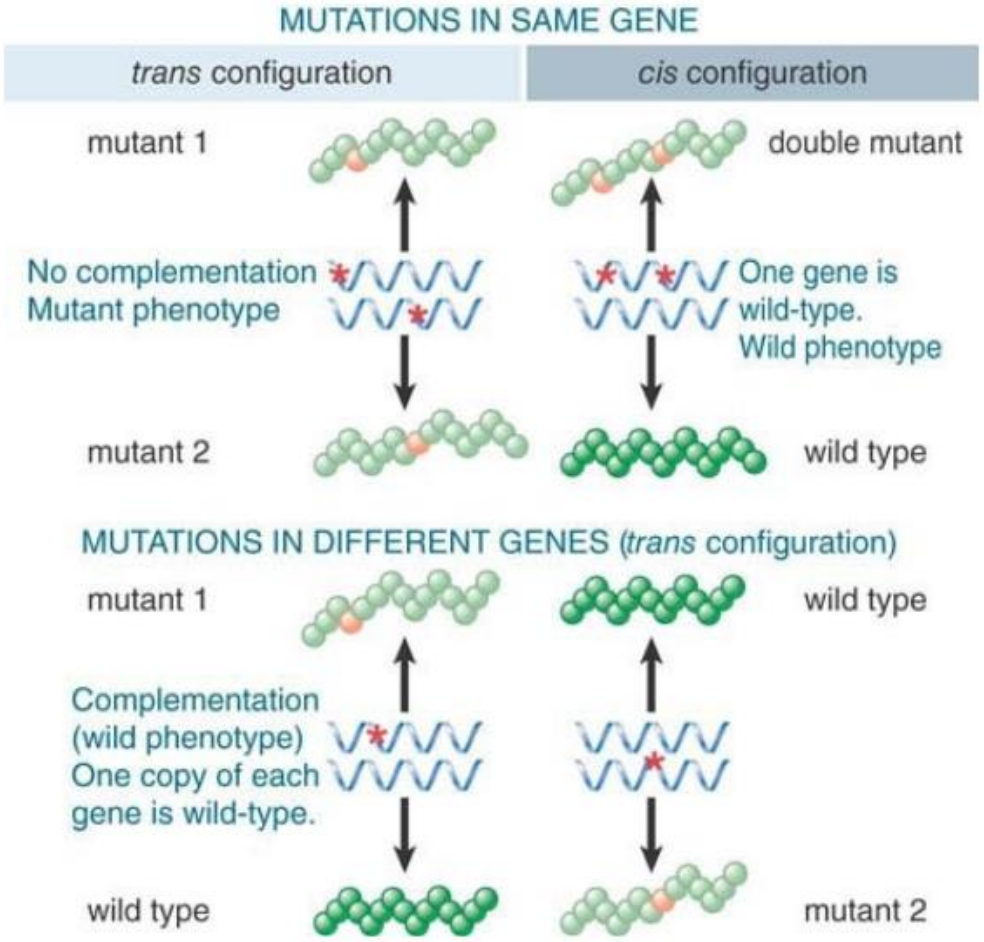


FIGURE 1.34 The cistron is defined by the complementation test. Genes are represented by DNA helices; red stars identify sites of mutation.

PART 3 · GENE ARCHITECTURE

How Genes Are Built — Prokaryotes vs Eukaryotes

Prokaryotic Genes: Compact and Colinear

- An uninterrupted ORF — the coding sequence is continuous and colinear with the mRNA and protein.
- **Genes are clustered by function** — one promoter drives several genes as a single polycistronic mRNA (operon).
- No nucleus → transcription and translation are coupled; ribosomes load on nascent mRNA.
- Little spacer DNA; regulation via promoters/operators sitting right next to the genes.
- **Net:** Genome is dense — most of it codes or directly controls coding.

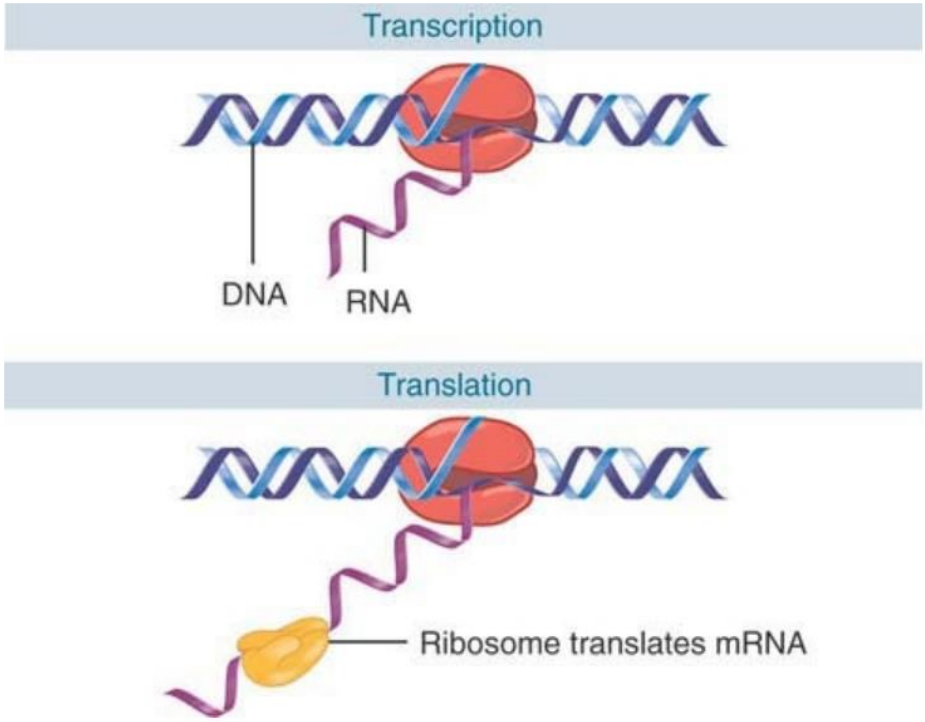


FIGURE 1.45 Transcription and translation take place in the same

1977: A Shock — Genes Are Interrupted

- **The experiment:** hybridize mature mRNA to its genomic DNA and view by electron microscopy (R-loop mapping).
- The DNA loops out single-stranded — stretches present in the gene are ABSENT from the mRNA: introns.
- Restriction maps of cDNA (mRNA copy) vs genomic DNA differ: the gene is longer than its message.
- Exons stay in the same order, but DNA distances ≠ mRNA/protein distances — splicing removes introns.
  - Sharp & Roberts shared the 1993 Nobel; 'the gene' had to be redefined.

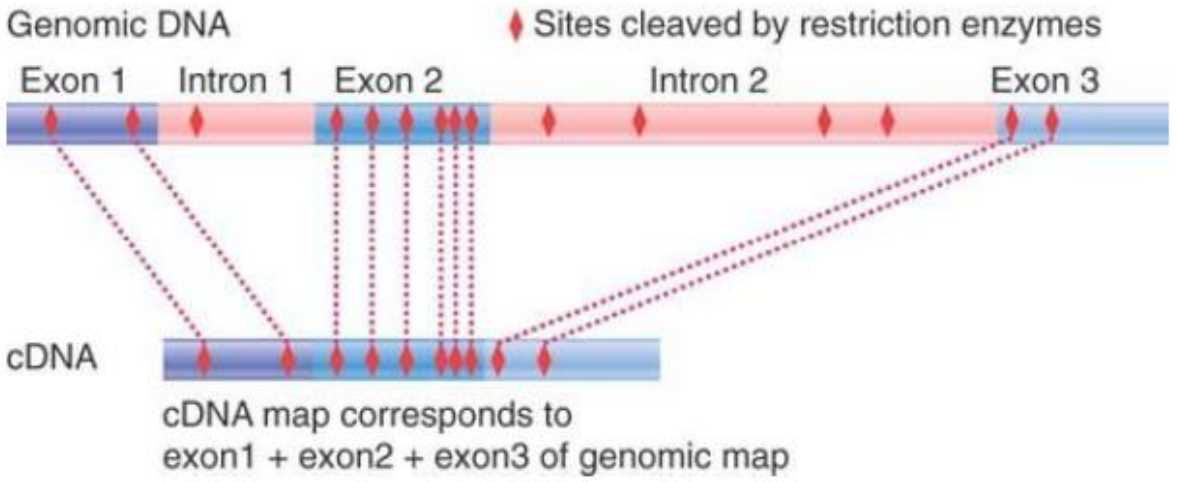


FIGURE 3.3 Comparison of the restriction maps of cDNA and genomic DNA for mouse βb-globin shows that the gene has two introns that are not present in the cDNA. The exons can be aligned exactly between cDNA and the gene.

1977: A Shock — Genes Are Interrupted

- **The experiment:** hybridize mature mRNA to its genomic DNA and view by electron microscopy (R-loop mapping).
- The DNA loops out single-stranded — stretches present in the gene are ABSENT from the mRNA: introns.
- Restriction maps of cDNA (mRNA copy) vs genomic DNA differ: the gene is longer than its message.
- Exons stay in the same order, but DNA distances ≠ mRNA/protein distances — splicing removes introns.
  - Sharp & Roberts shared the 1993 Nobel; 'the gene' had to be redefined.

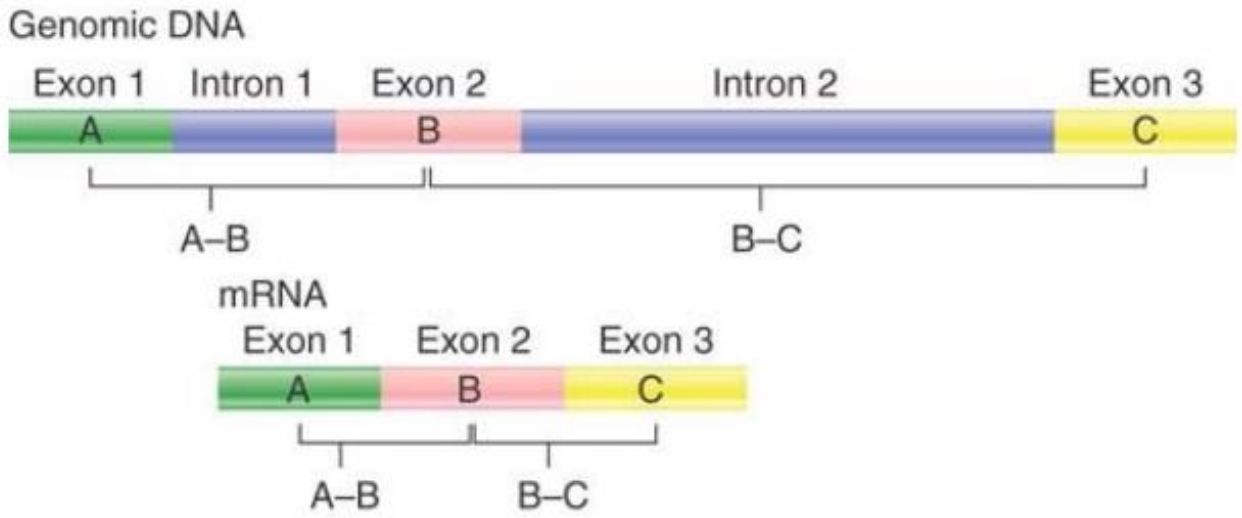


FIGURE 3.2 Exons remain in the same order in mRNA as in DNA, but distances along the gene do not correspond to distances along the mRNA or polypeptide products. The distance from A–B in the gene is smaller than the distance from B–C, but the distance from A–B in the mRNA (and polypeptide) is greater than the distance from B–C.

Eukaryotic Genes: Exons, Introns & Regulation

- **Split structure:** exons (kept) and introns (spliced out); mature mRNA = exons joined in order.
- Monocistronic; a promoter plus distant enhancers/silencers integrate many inputs.
- **Exons as modules** — exons often map to protein domains; shuffling builds new proteins (e.g. LDL receptor).
- **Alternative splicing** — one pre-mRNA → many proteins, multiplying proteome diversity from one gene.
  - Cost: most of the gene (and genome) is non-coding — setting up the next question.

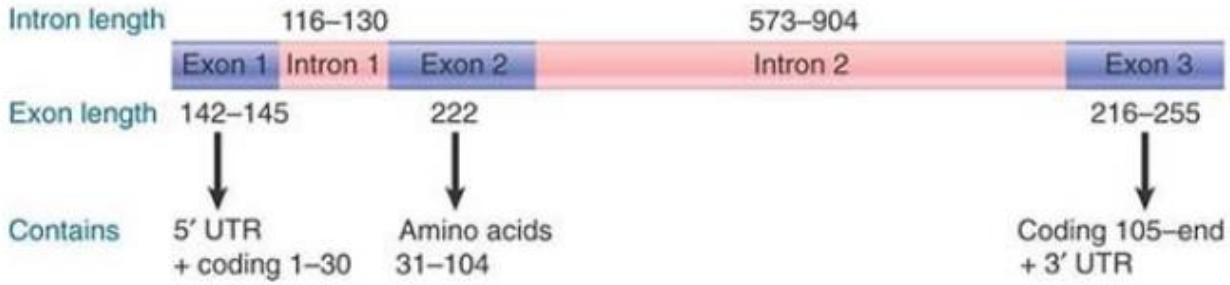


FIGURE 3.4 All functional globin genes have an interrupted structure with three exons. The lengths indicated in the figure apply to the mammalian βb-globin genes.

Eukaryotic Genes: Exons, Introns & Regulation

- **Split structure:** exons (kept) and introns (spliced out); mature mRNA = exons joined in order.
- Monocistronic; a promoter plus distant enhancers/silencers integrate many inputs.
- **Exons as modules** — exons often map to protein domains; shuffling builds new proteins (e.g. LDL receptor).
- **Alternative splicing** — one pre-mRNA → many proteins, multiplying proteome diversity from one gene.
  - Cost: most of the gene (and genome) is non-coding — setting up the next question.

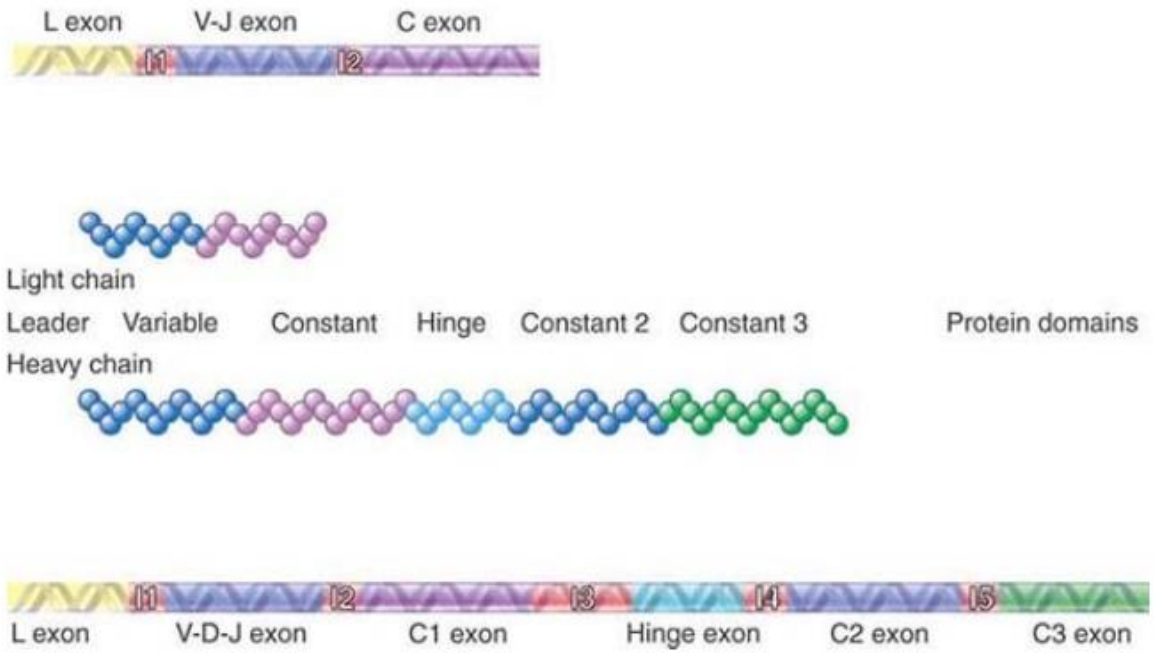


FIGURE 3.15 Immunoglobulin light chains and heavy chains are encoded by genes whose structures (in their expressed forms) correspond to the distinct domains in the protein. Each protein domain corresponds to an exon; introns are numbered I1 to I5.

PART 4 · THE GENOME

Beyond Genes — and How Genomes Evolve

What's Actually in a Genome?

- In humans, genes span ~25% of the DNA, but protein-coding sequence is only ~1–2%.
- **The rest:** introns, regulatory DNA, pseudogenes, RNA genes, organelle genomes — and lots of repeats.
- Gene number does NOT track complexity (human ≈ worm ≈ ~20,000) — regulation does the work.
- **Endosymbiosis** — mitochondria/chloroplasts are ex-bacteria, with their own small circular genomes.
- **Surprise:** The biggest single component of the human genome is transposable elements.

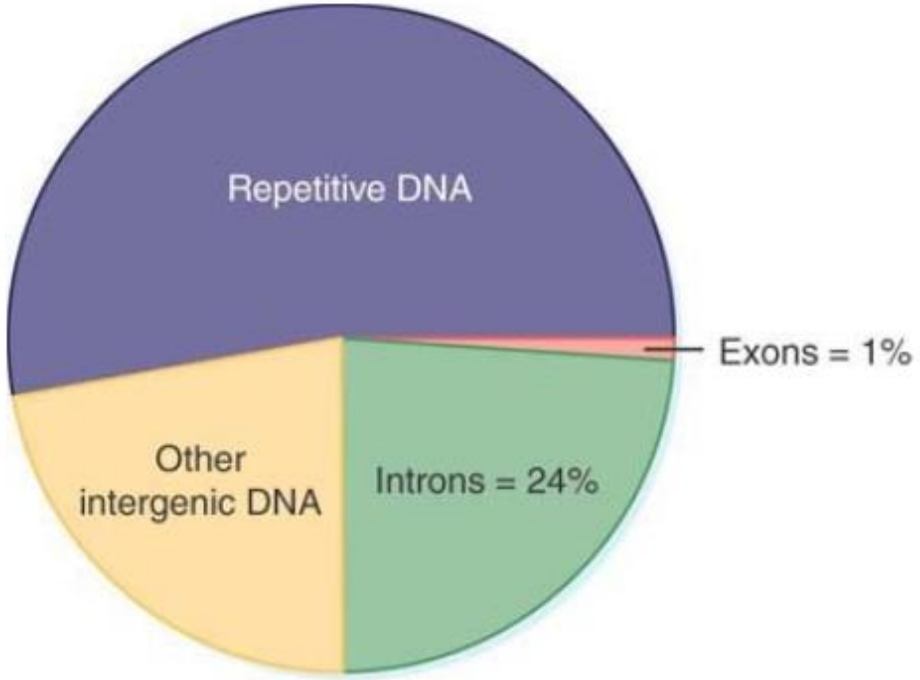


FIGURE 5.9 Genes occupy 25% of the human genome, but protein-coding sequences are only a small part of this fraction.

Repetitive DNA: Revealed by How Fast DNA Re-anneals

- **Britten–Kohne (C_0t) curves:** denature genomic DNA, let it reassociate; abundant sequences find partners FAST.
- Three kinetic classes appear → highly-repetitive, moderately-repetitive, and single-copy DNA.
- **Satellite DNA** — tandem repeats that form a separate CsCl density band; sit at centromeres.
- **Mini-/microsatellites** — variable tandem repeats; allele length varies → basis of DNA fingerprinting (VNTRs).
  - Short repeats expand/contract by replication slippage and unequal crossing-over.

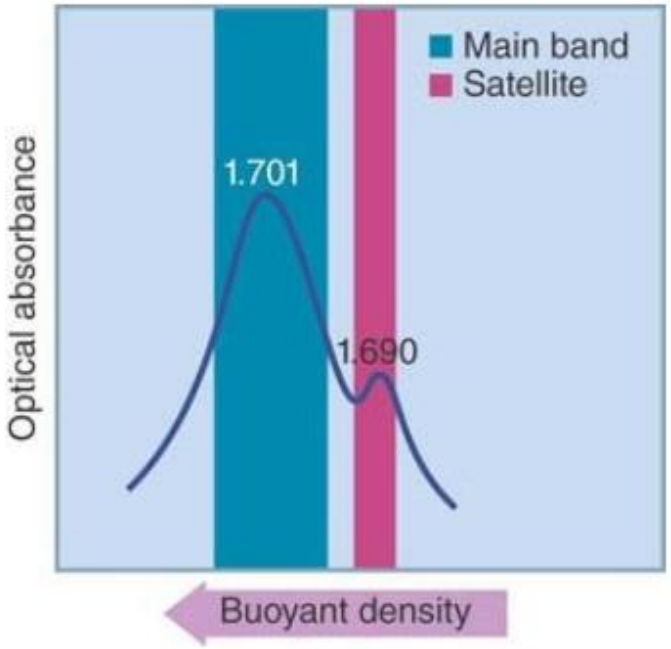


FIGURE 6.11 Mouse DNA is separated into a main band and a satellite band by centrifugation through a density gradient of CsCl.

Mobile DNA: McClintock's 'Controlling Elements'

- **McClintock:** inferred mobile elements in maize from sectored kernels (1940s–50s); vindicated molecularly (1983 Nobel).
- **DNA transposons** — cut-and-paste by a transposase; leave target-site duplications.
- **Retroelements** — move via an RNA + reverse-transcriptase copy (LINEs, SINEs/Alu); one-way expander.
- **Key contrast:** the two classes differ in HOW they move (Ch 15, Fig 15.2).
- **Scale:** ~45% of the human genome — mobile DNA, not genes, dominates.

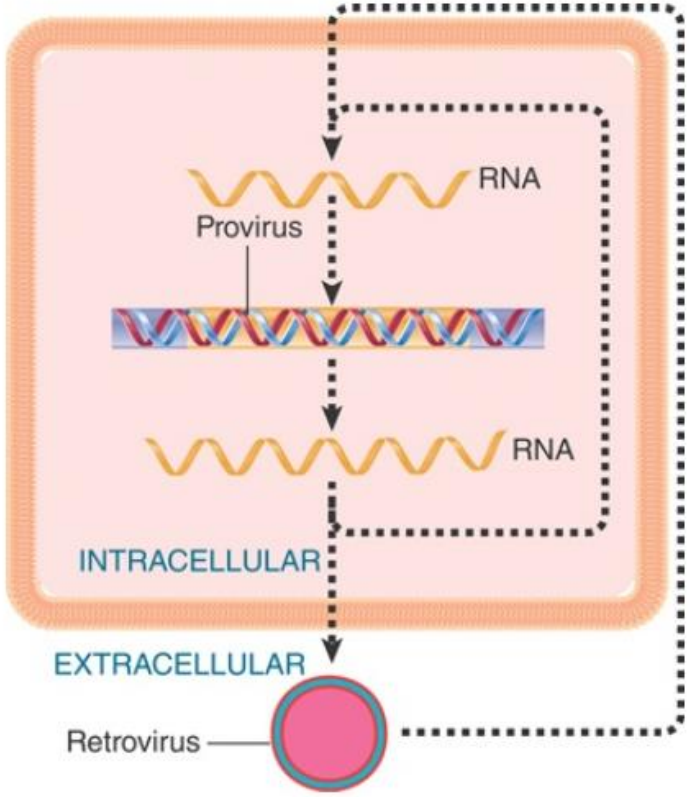


FIGURE 15.2 The reproductive cycles of retroviruses and retrotransposons alternate reverse transcription from RNA to DNA with transcription from DNA to RNA. Only retroviruses can generate infectious particles. Retrotransposons are confined to an intracellular cycle.

How Genomes Evolve: Duplication, Drift & Time

- **C-value paradox:** genome size varies ~100,000× and does not match organismal complexity (driven by repeats/introns).
- **Gene & genome duplication** — the raw material of innovation: one copy keeps the job, the other is free to diverge or decay.
- Build gene families and pseudogenes (the globin clusters are the textbook case).
- **Concerted evolution** — unequal crossing-over homogenizes tandem repeats so family members evolve together.
- **Molecular clock** — silent sites tick ~neutrally; read selection from K_a/K_s (dN/dS): <1 purifying, >1 positive.

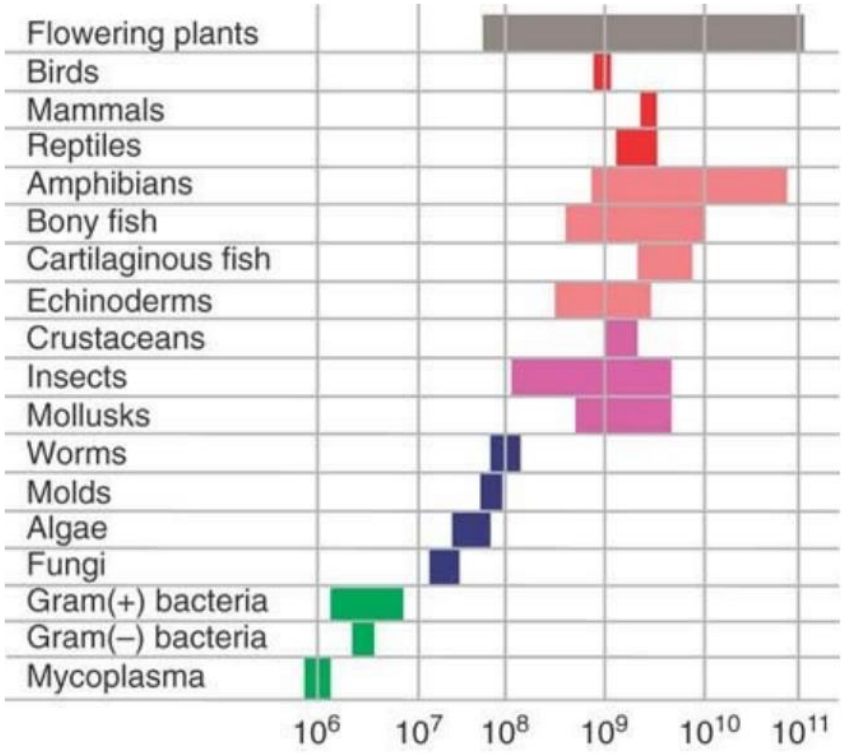


FIGURE 5.29 DNA content of the haploid genome increases with morphological complexity of lower eukaryotes, but varies

The Through-Line: How We Learned the Genome

- Heredity is molecular — transformation → DNase → phage labeling → the double helix.
- A gene is colinear information for a protein — but interrupted and heavily regulated in eukaryotes.
- The genome is mostly non-coding and dynamic — repeats and transposons dominate it.
- It evolves by duplication, transposition, and the sorting of variants by drift and selection.
- **Meta-point:** Recurring lesson: each conceptual leap followed a method that made the invisible visible.