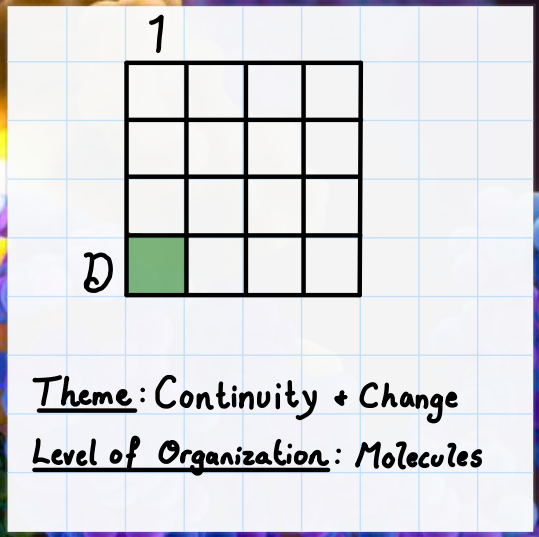
D1.3 MUTATIONS AND GENE EDITING

Ver. 2

Guiding Questions
How do gene mutations occur?
What are the consequences of gene mutation?

Linking Questions
How can natural selection lead to both a reduction in variation and an increase in biological diversity?
How does variation in subunit composition of polymers contribute to function?



Theme: Continuity + Change

Level of Organization: Molecules

Written and drawn by:

Peter Marier

NOTEWORTHY SCIENCE

MARIER

SL LEARNING OUTCOMES

D1.3.1	Gene mutations as structural changes to genes at the molecular level	Distinguish between substitutions, insertions and deletions.
D1.3.2	Consequences of base substitutions	Students should understand that single-nucleotide polymorphisms (SNPs) are the result of base substitution mutations and that because of the degeneracy of the genetic code they may or may not change a single amino acid in a polypeptide.
D1.3.3	Consequences of insertions and deletions	Include the likelihood of polypeptides ceasing to function, either through frameshift changes or through major insertions or deletions. Specific examples are not required.
D1.3.4	Causes of gene mutation	Students should understand that gene mutation can be caused by mutagens and by errors in DNA replication or repair. Include examples of chemical mutagens and mutagenic forms of radiation.
D1.3.5	Randomness in mutation	Students should understand that mutations can occur anywhere in the base sequences of a genome, although some bases have a higher probability of mutating than others. They should also understand that no natural mechanism is known for making a deliberate change to a particular base with the purpose of changing a trait.
D1.3.6	Consequences of mutation in germ cells and somatic cells	Include inheritance of mutated genes in germ cells and cancer in somatic cells.
D1.3.7	Mutation as a source of genetic variation	Students should appreciate that gene mutation is the original source of all genetic variation. Although most mutations are either harmful or neutral for an individual organism, in a species they are in the long term essential for evolution by natural selection. **NOS:** Commercial genetic tests can yield information about potential future health and disease risk. One possible impact is that, without expert interpretation, this information could be problematic.

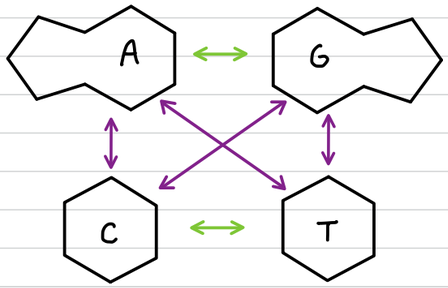
HL LEARNING OUTCOMES

D1.3.8	Gene knockout as a technique for investigating the function of a gene by changing it to make it inoperative	Students are not required to know details of techniques. Students should appreciate that a library of knockout organisms is available for some species used as models in research.
D1.3.9	Use of the CRISPR sequences and the enzyme Cas9 in gene editing	Students are not required to know the role of the CRISPR–Cas system in prokaryotes. However, students should be familiar with an example of the successful use of this technology. **NOS**: Certain potential uses of CRISPR raise ethical issues that must be addressed before implementation. Students should understand that scientists across the world are subject to different regulatory systems. For this reason, there is an international effort to harmonize regulation of the application of genome editing technologies such as CRISPR.
D1.3.10	Hypotheses to account for conserved or highly conserved sequences in genes	Conserved sequences are identical or similar across a species or a group of species; highly conserved sequences are identical or similar over long periods of evolution. One hypothesis for the mechanism is the functional requirements for the gene products and another hypothesis is slower rates of mutation.

D1.3.1—Gene mutations as structural changes to genes at the molecular level. D1.3.2—Consequences of base substitutions. D1.3.3—Consequences of insertions and deletions

gene mutation: permanent structural changes to genes at the molecular level (i.e. change in DNA nucleotide base sequence)

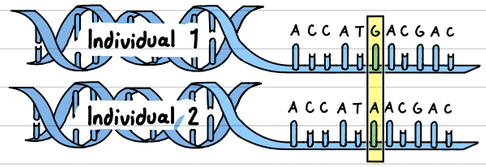
↳ **Base substitution mutation**: replacement of one base in the coding section of a gene with another —— can result in ——→ **single-nucleotide polymorphisms (SNPs)**: a variation at a single base in a DNA sequence present in more than 1% of the population. SNPs can occur in both coding and non-coding sequences. If it occurs in a gene it may or may not result in a new amino acid and polypeptide due to degeneracy of genetic code



Transversions: point mutation that changes a pyrimidine nucleotide (single ring) to a purine (double ring) or vice-versa

Transitions: point mutation that changes a purine nucleotide (double ring) to another purine or a pyrimidine (single ring) to another pyrimidine. More common than transversions (as same ring structure kept) and more likely to result in silent mutations and persist as single-nucleotide polymorphisms

↳ SNPs can be used as genetic markers in genome-wide association studies (GWAS) where SNP patterns are compared across large populations to identify associations between SNP and disease susceptibility or drug response

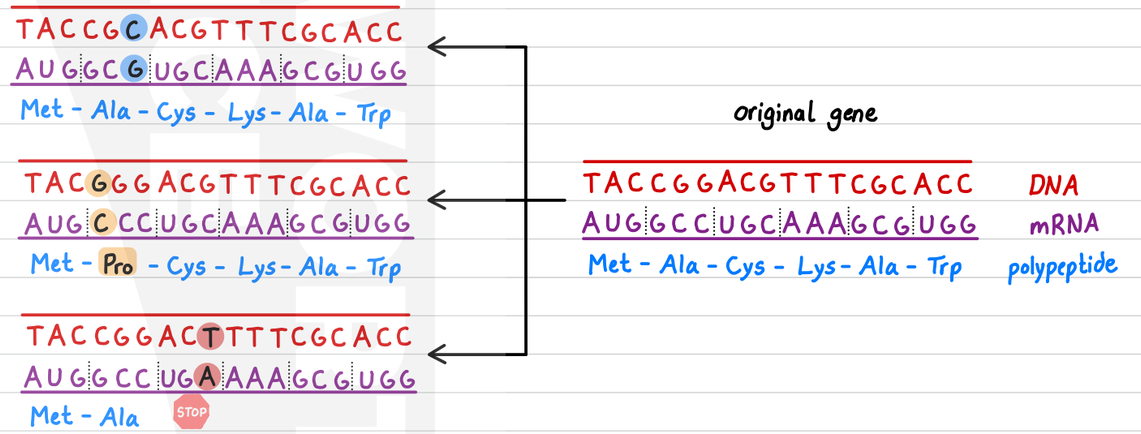


Base substitution mutations can be classified based on the resulting consequence:

- **same-sense (silent) mutation**: base substitution which alters the codon for an amino acid into another codon for the <u>same</u> amino acid due to degeneracy of genetic code. No effect on the polypeptide produced and the phenotype

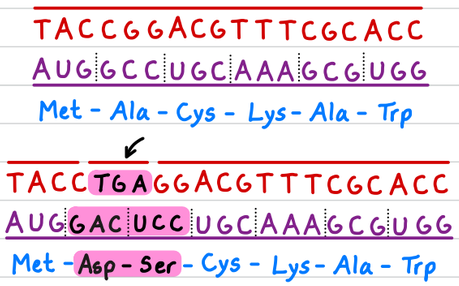
- **missense mutation**: base substitution which alters the codon for an amino acid into another codon for a <u>different</u> amino acid. Changes the polypeptide produced. If the new amino acid causes the protein to fold differently it may alter protein function and the resulting phenotype (ex: Sickle cell anemia)

- **nonsense mutation**: base substitution which alters the codon for an amino acid into a <u>STOP</u> codon, causing translation to end prematurely. Polypeptide produced is shortened which may alter its function or more likely make it non-functional, altering the phenotype (ex: β-thalassemia)

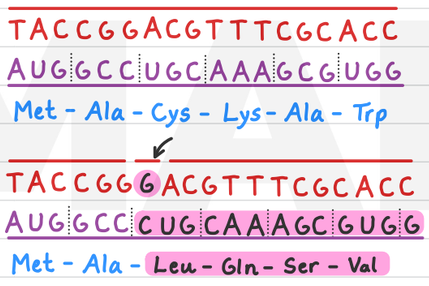


↳ **Insertion mutation**: <u>addition</u> of one or more nucleotides within the base sequence. Effect depends on the position and how many nucleotides are added (ex: HTT gene and Huntington's disease)

- <u>number inserted is multiple of 3</u>
  extra codon(s) added, resulting in a longer polypeptide. May alter protein function. Majority of protein unaltered

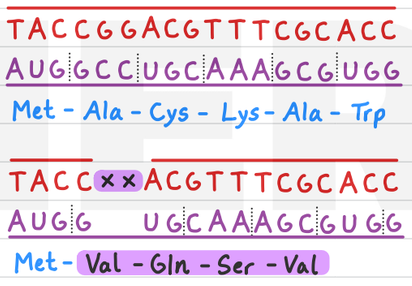


- <u>number inserted is not multiple of 3</u>
  Results in a **frameshift mutation** where the reading frame (grouping of codons) is shifted, altering many/all amino acids downstream of mutation. The earlier in the sequence it occurs the greater the impact. Protein produced is very different from original and is likely non-functional

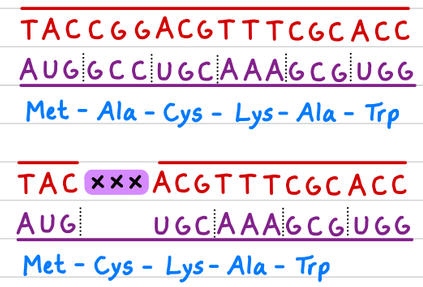


↳ **Deletion mutation**: <u>deletion</u> of one or more nucleotides within the base sequence. Effect depends on the position and how many nucleotides are removed (ex: CCR5 gene and HIV resistance)

- number deleted is not multiple of 3



- <u>number deleted is multiple of 3</u>
  codon(s) deleted, resulting in a shorter polypeptide. May alter protein function. Majority of protein unaltered



D1.3.4—Causes of gene mutation. D1.3.5—Randomness in mutation. D1.3.6—Consequences of mutation in germ cells and somatic cells. D1.3.7—Mutation as a source of genetic variation

Gene mutations can be caused by a variety of factors:

↳ Errors in DNA replication and/or DNA repair

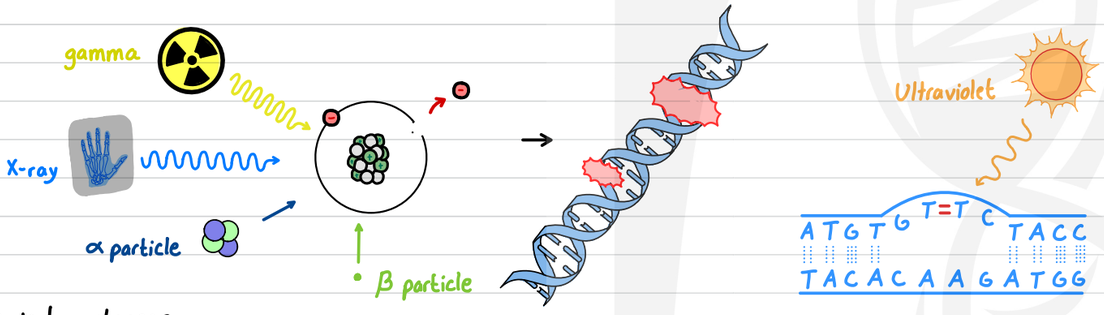
DNA polymerases may make errors, leading to mismatched nucleotides, insertions or deletions. Typically errors are identified and corrected via proofreading but on occasion it is not and persists. Thus, the next time the cell replicates, the mutated base sequence will be used as a template and be passed onto next generation of cells.
These spontaneous mutations are rare but normal and expected.

DNA replication D1.1

Mutagen: agent which can induce a change in DNA, increasing mutation frequency above background rate

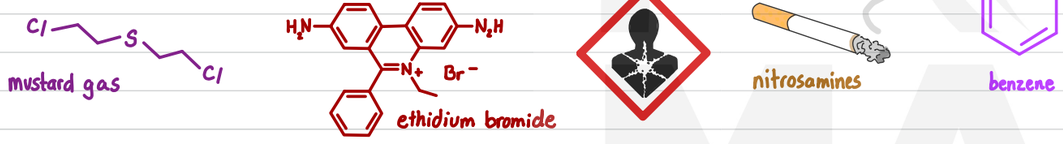
↳ Mutagenic radiation

Ionizing radiation are types of high energy radiation which can displace electrons from atoms, leading to DNA damage such as single-strand or double-strand breaks. Types include: high frequency EM radiation such as Gamma and X-ray and particulate radiation such as α particles and β particles
Ultraviolet (UV) radiation is non-ionizing but causes pyrimidine dimers, distorting DNA - leading to errors during replication



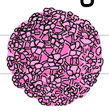
↳ Chemical mutagens

Mutagenic chemicals can induce mutations via a variety of mechanisms such as acting like nitrogenous bases but pairing incorrectly during DNA replication, inserting themselves into DNA causing a distortion in double helix, or by changing pairing properties.



↳ Infectious agents

Infection of organisms by some viruses and some bacteria can result in DNA damage and mutations

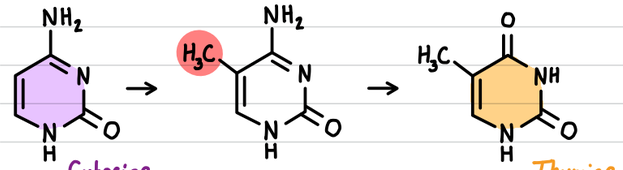


HPV infection related to cervical cancer

H. pylori infection related to gastric cancer

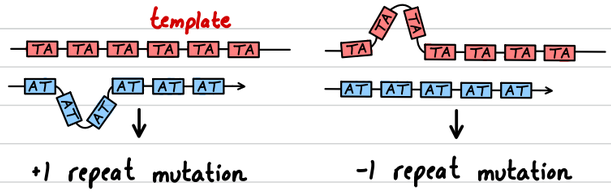
Mutations are random, meaning they are unpredictable.
↳ mutations can occur anywhere in a base sequence of a genome but some areas are more frequently impacted:

- **Cytosine-Guanine (CG) rich regions - CpG sites**
  In order to silence a gene, methylation is done to cause DNA to become more tightly packed When Cytosine is methylated it can spontaneously mutate to a Thymine. CG regions are often methylated and show higher mutation rates than any other base combination



Gene expression D2.2

- **Microsatellites / Short-Tandem Repeats** DNA replication D1.1
  STRs are repetitive regions in the DNA (ex: ATATAT) and have higher mutation rates than any other region. One reason could be due to slippage of DNA polymerase during replication where entire repeats can be gained or lost.

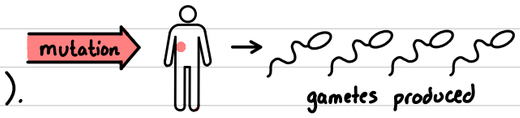


※ No natural mechanism is known for making a deliberate change to a particular base sequence with the purpose of changing a protein or trait. i.e an organism cannot decide or select sequences to mutate

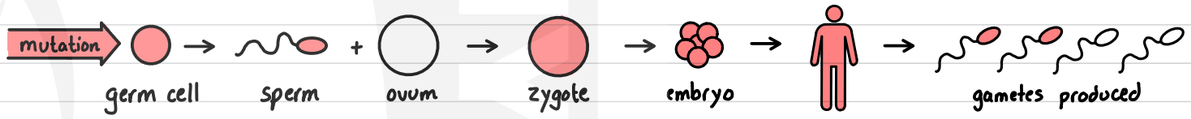
The consequence of a mutation in humans will depend on which type of cell mutates:

cell + nuclear division D2.1

↳ if **somatic cell** (body cell) DNA is mutated, only this cell and all resulting daughter cells will carry the mutation. Depending on the gene mutation it may result in cancer (uncontrolled cell division).
As mutation is not present in a gamete, it cannot be inherited by offspring



↳ if **germ cell** (cells that develop into sperm/ovum) DNA is mutated and it fertilizes an ovum, the mutation will be present in the zygote and all somatic cells of the resulting embryo and adult. Due to meiosis, half of all produced gametes will carry mutation and mutated DNA may be passed onto offspring and inherited



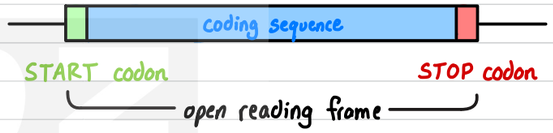
Gene mutations alter base sequence, potentially resulting in a new protein This forms a new allele (gene variant)
↳ members of a species may differ in which alleles they carry (genetic variation) - some being negative, neutral or positive
∴ mutations are the source of genetic variation and a prerequisite for evolution via natural selection, where alleles associated with advantageous traits are selected for and become more common in population (and vice-versa) → adaptation

NOS: Commercial genetic tests can provide consumers information regarding disease susceptibility and health risks
potential issues: ✗ some allele variants may be missing ✗ life insurance companies may use information to raise premiums
✗ expert interpretation/genetic counseling required so consumers understand and can make informed decisions

HL

D1.3.8—Gene knockout as a technique for investigating the function of a gene by changing it to make it inoperative.
D1.3.10—Hypotheses to account for conserved or highly conserved sequences in genes

When analyzing the DNA of an organism, determining which sequences are genes (coding) can be done by looking for **open reading frames** →



However, how can the function of the gene be determined?

Gene Knockout : technique that produces a genetically-modified organism with one specific non-functional (inoperative) gene

↳ by observing the phenotype and comparing the modified organism to a normal organism, it allows deduction of the function of inoperative gene

ex: leptin gene is made inoperative in mice

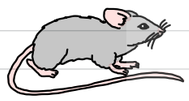
LEP Knockout mice eat excessively and become obese

∴ hormone leptin involved in appetite suppression and regulation of fat stores

↳ thousands of knockout strains in model organisms have been produced, acting as a library for genetic research

General process:
1. Observation of phenotype in humans leads to hypothesis - suggesting role of gene 'X'
2. test hypothesis experimentally using Knockout for gene 'X' in model organism
3. validate findings in human cells / clinical trials

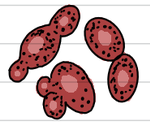
↳ many different model organisms can be used to experimentally test impact of genes, each offering advantages



mice share 99% of the same genes with humans making them good analogues

zebrafish embryos are transparent, allowing early development to be easily observed

Fruit flies reproduce quickly and in large numbers and have similar genes in few chromosomes - convenient



Yeast are used due to fast reproduction times and ease of genetic modification

---

While genetic variation is present within and across species, some sequences are identical or very similar - showing very little variation

evolution + speciation A4.1

↳ **Conserved sequences** : identical or similar across a species or a group of species

↳ **Highly conserved sequences** : identical or similar over long periods of evolution

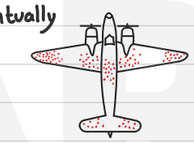
ex: cyt c gene codes for cytochrome c ; ~100 amino acid long protein used in aerobic cellular respiration
its sequence and polypeptide product is nearly identical in many organisms

※ by comparing gene or polypeptide sequences evolutionary relatedness can be estimated

2 main hypotheses (not mutually exclusive) for conserved and highly conserved sequences:

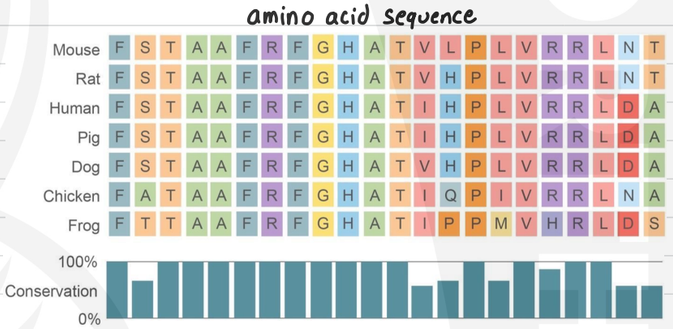
- <u>Functional requirements</u>

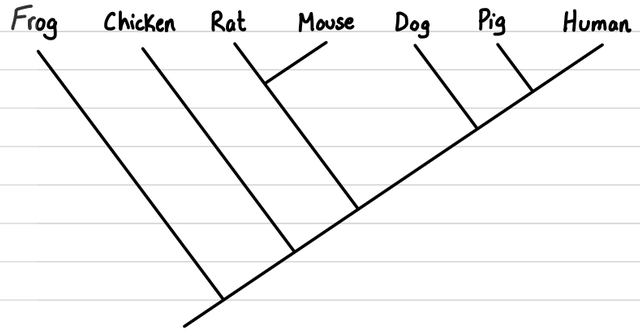
  gene sequences may code for proteins essential for basic cellular stability, function, or reproduction
  ∴ change in these sequences could alter the protein's functionality and lead to cell death or impede reproduction, preventing many changes to accumulate over evolutionary history as these are selected against and eventually eliminated by natural selection. Akin to survivorship bias as changes aren't seen as they don't persist

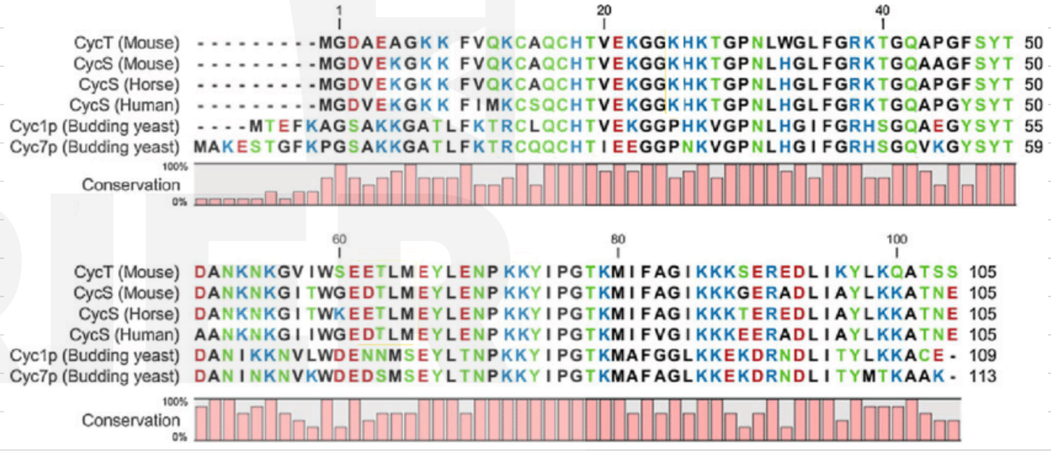


- <u>Slower mutation rates</u>

  gene sequences are in regions of the genome where mutation rates are low due to higher gene expression.
  mutation rate is linked to gene expression: highly transcribed genes have lower rates as these regions may have more enhanced proofreading and repair mechanisms







adapted from Kulikov et al 2011

HL D1.3.9 — Use of the CRISPR sequences and the enzyme Cas9 in gene editing.

Like all organisms, bacteria are susceptible to viral infection. When a virus infects a cell, it inserts its viral DNA into the bacterial genome and can hijack the cell's metabolism to make more copies. viruses A2.3

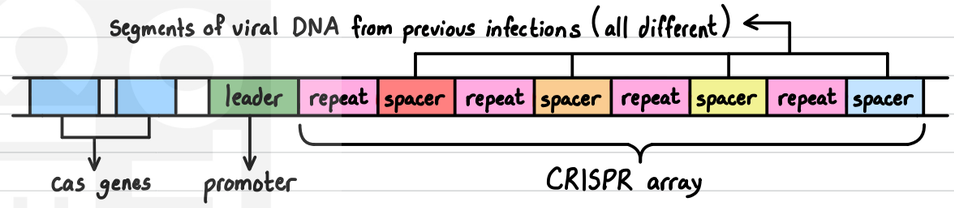
↳ bacteria have evolved a counter-measure (immune system) to identify and excise viral DNA using CRISPR-Cas9

Cas9 (CRISPR-associated protein 9): enzyme that uses CRISPR sequences as a guide to recognize and cut-open specific sections of DNA

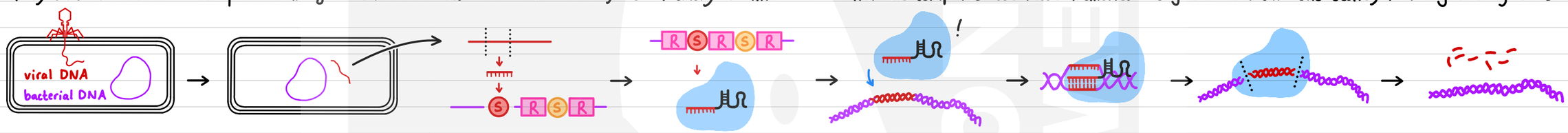
CRISPR - **C**lustered **R**egularly **I**nterspaced **S**hort **P**alindromic **R**epeats

- **C**lustered: grouped in one part of the genome
- **R**egularly: repeats separated by spacer sequences at regular intervals
- **I**nterspaced: small number of bases (23-47)
- **S**hort: section of sequence read the same forward and backwards
- **P**alindromic: same base sequence occurs several times in genome

ACGCGT
TGCGCA



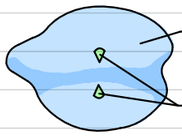
↳ CRISPR-Cas9 system in bacteria: Cas proteins integrate viral DNA into the CRISPR array as a 'memory' of infection. Other Cas complexes use these 'memories' as guides to locate and destroy matching viral genomes



Gene editing: changing the genetic material of an organism by deliberately removing, inserting or altering genes. CRISPR-Cas9 provides a mechanism to precisely cut and edit genomes

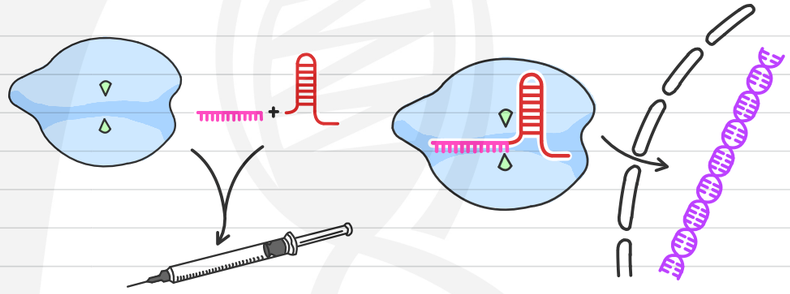
※ prime editing is a different technique where a modified cas9 is used (which cuts only 1 strand) along with a prime editing guide (peg) RNA, guide RNA and reverse transcriptase. This allows a precise 'search and replace' all-in-one!

Components:

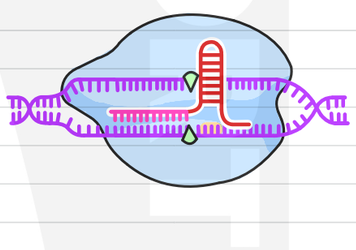


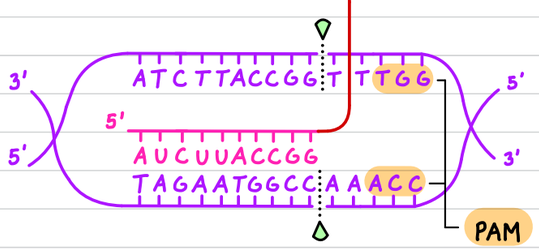
Cas9: enzyme which cleaves DNA at specific endonuclease active sites

guide RNA: RNA engineered to contain a complementary sequence to the target gene allowing it to 'guide' cas9 to desired location



1. guide RNA is created with a complementary sequence to target gene and mixed with cas9
2. cas9-guide RNA complex forms and is introduced into target cell, entering nucleus
3. cas9-guide RNA complex moves and scans DNA until a PAM (Protospacer Adjacent Motif) is located (next to target)
4. cas9-guide RNA complex binds to target sequence complementarily and both strands of DNA are cut by active sites





5. Gene Knockout: Cut DNA is repaired by the cell but incorrectly, resulting in mutations (additions/deletions) making gene non-functional

Gene Knock-in: New gene template is added and inserted during repair allowing cell to express gene of interest

NOS: potential use of CRISPR-Cas system in gene editing raises several ethical issues which need to be addressed and regulations agreed upon prior to implementation

↳ germ-line editing has severe ethical concerns:

- ♥ Safety: risk of off-target edits/mutations causing potential unforeseen consequences in other genes
- Slippery slope: if genetic diseases can be edited, why not edit for enhancement/cosmetic purposes?
- ⚖ justice: gene editing is costly and likely only available to the wealthy → inequity in health and genetic status

Example application - Sickle-cell anemia

protein synthesis D1.2

- Sickle-cell anemia have mutated β chains in the haemoglobin, causing erythrocyte sickling
- Fetal haemoglobin (HbF), unlike the adult form is composed of two α and two γ subunits but this stops being made in favor of adult haemoglobin after birth due to repression of the γ globin gene by a repressor protein, coded by BCL11A gene. CRISPR silences BCL11A to make HbF

Hematopoietic stem cell (HSC) harvested from bone marrow

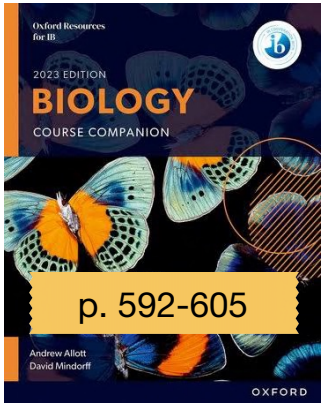
CRISPR-Cas9 causes disruption in BCL11A gene

modified HSC re-injected into patient. HbF expression increases, producing normal erythrocytes

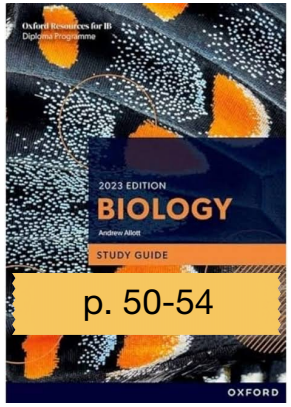
Resource Links

each resource is hyperlinked

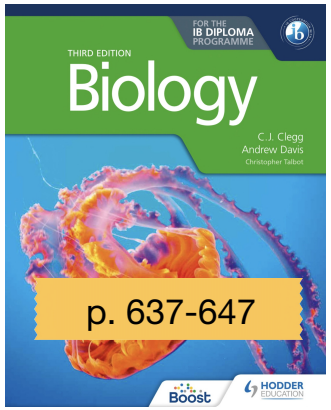
Textbooks



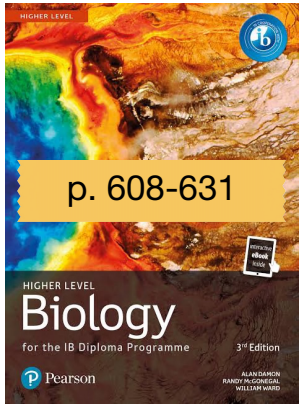
p. 592-605



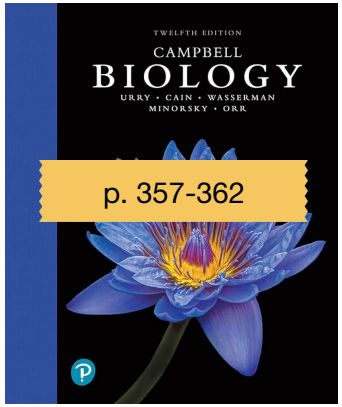
p. 50-54



p. 637-647

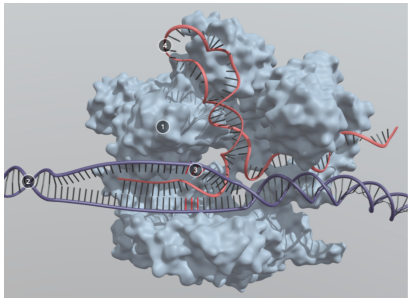


p. 608-631

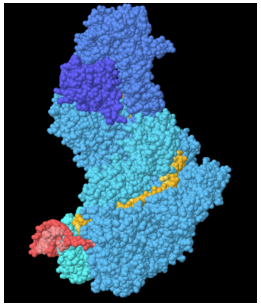


p. 357-362

3D models

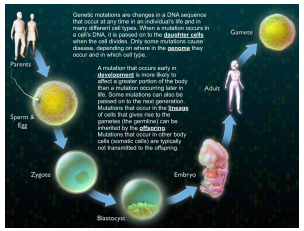


CRISPR/Cas9

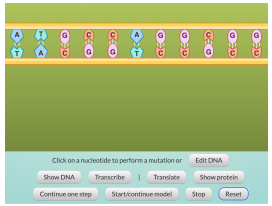


Cascade and CRISPR

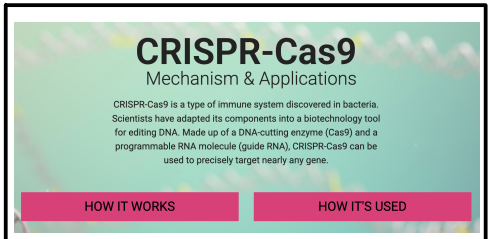
Simulators / Interactives



mutations and disease



Gene mutation consequences



CRISPR-Cas9

Articles

CRISPR in Nature - Innovative Genomics Institute (IGI). (2025, January 20). Innovative Genomics Institute (IGI). https://innovativegenomics.org/crisprpedia/crispr-in-nature/

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable Dual-RNA–Guided DNA endonuclease in adaptive bacterial immunity. Science, 337(6096), 816–821. https://doi.org/10.1126/science.1225829

Kulikov, A. V., Shilov, E. S., Mufazalov, I. A., Gogvadze, V., Nedospasov, S. A., & Zhivotovsky, B. (2011). Cytochrome c: the Achilles' heel in apoptosis. Cellular and Molecular Life Sciences, 69(11), 1787–1797. https://doi.org/10.1007/s00018-011-0895-z

Lino, C. A., Harper, J. C., Carney, J. P., & Timlin, J. A. (2018). Delivering CRISPR: a review of the challenges and approaches. Drug Delivery, 25(1), 1234–1257. https://doi.org/10.1080/10717544.2018.1474964

What Is CRISPR: The Ultimate Guide To CRISPR Mechanisms, Applications, Methods & More. (n.d.). Synthego. https://www.synthego.com/learn/crispr