Bio-Bootcamp

Part 1: Synthetic Biology and DNA Design

By Dominika Wawrzyniak and Michelle Yue

DNA Design - Overview

- Part 1 – Genome and gene biology
  - DNA at scale
  - The genome is organized
  - The genetic code – Codon Usage
  - The anatomy and function of DNA (eukaryotic vs prokaryotic)
  - Gene mutations
    - Point mutations
    - Frameshift mutationsDesigning DNA
- Part 2 – Working with DNA
  - DNA engineering basics: Molecular Cloning
    - The DNA template
    - The vector
    - The host cell
  - Modifying DNA – the enzyme toolkit
- Part 3 – Gene editing (Michelle Yue)
  - Artificial Enzymes
  - Gene editing approaches

Part 1 - Genome and gene biology

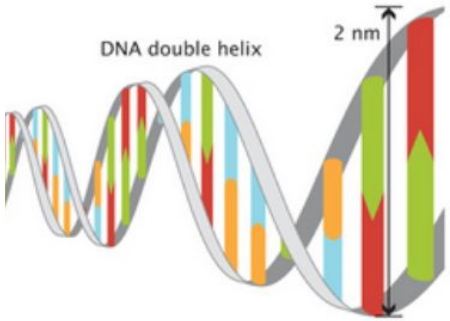
DNA at scale

1

A basepair:

- ~ 2 nm length
- ~ 0.34 nm bp to bp

Each helix turn (~10 bp)



2

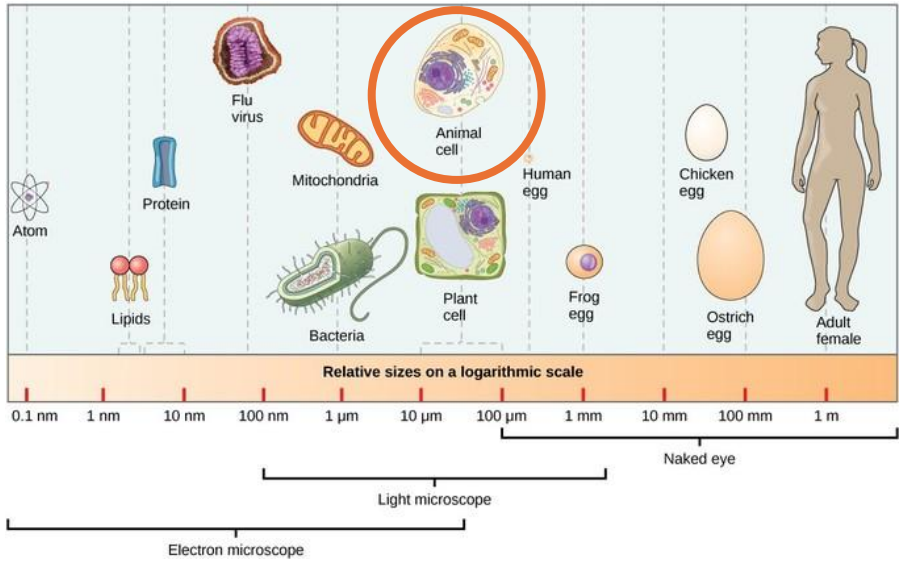
Table 4.3 | Genome size comparison

Genome	Size	RGS	Genes
HIV1 (human immunodeficiency virus; RNA)[1]	9.6 kb	0.002	9
Human mt DNA[2]	16.6 kbp	0.004	37
Bacteriophage lambda (λ)[1]	49.2 kbp	0.01	68
Escherichia coli (bacterium)[3]	4.6 Mbp	1	4 240
Saccharomyces cerevisiae (budding yeast)[3]	12.2 Mbp	2.7	6 600
Aspergillus nidulans (fungus)[3]	30.5 Mbp	6.6	10 534
Caenorhabditis elegans (nematode worm)[3]	100 Mbp	22	20 191
Arabidopsis thaliana (plant)[3]	135 Mbp	29	27 655
Drosophila melanogaster (fruit fly)[3]	144 Mbp	31	13 968
Mus musculus (mouse)[3]	2.7 Gbp	0.9	22 468
Homo sapiens (human)[3]	3.1 Gbp	1	20 442
Nicotiana tabacum (tobacco)[4]	4.6 Gbp	1.5	69 500
Triticum aestivum (wheat)[3]	17 Gbp	5.5	107 891
Neoceratodus forsteri (Australian lungfish)[5]	43 Gbp	13.9	31 120
Protopterus aethiopicus (marbled lungfish)[a,6]	130 Gbp	42	Uncertain
Paris japonica (Japanese canopy plant)[a,7]	149 Gbp	48	Uncertain

Diploid human genome

- Male: 6.27 Gbp
- Female: 6.37 Gbp

3

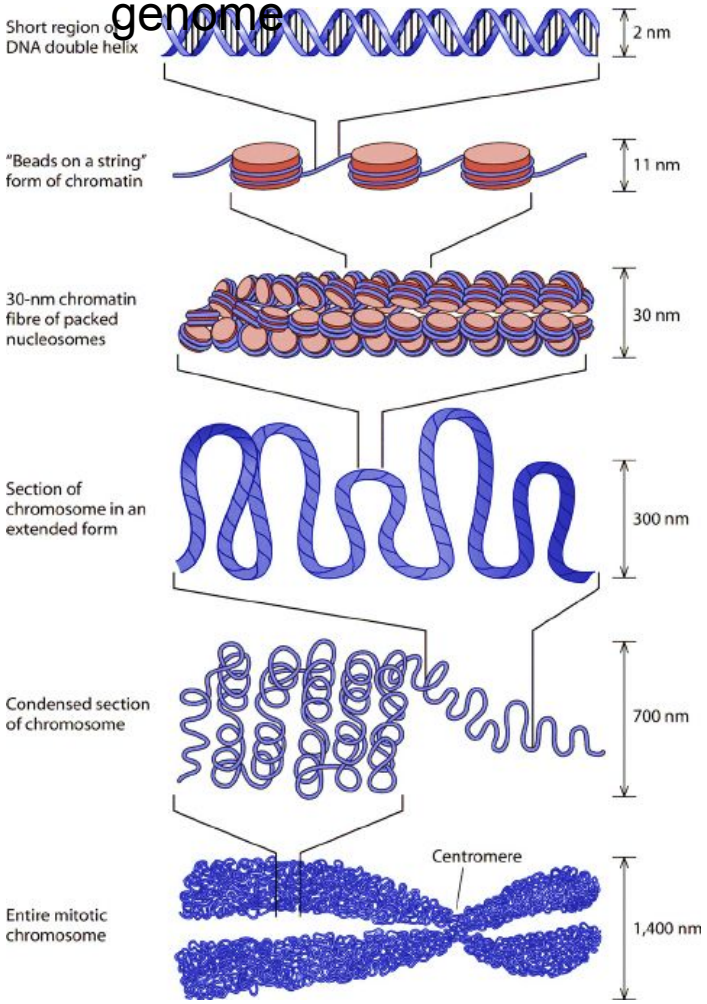


Size of an animal cell: ~10 μm – 100 μm

The human genome is about 2 m long – how does it fit?

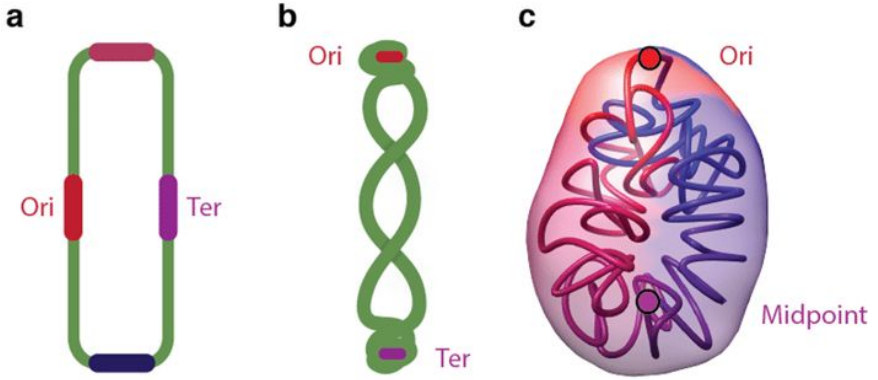
The genome is organized

The **eukaryotic genome**

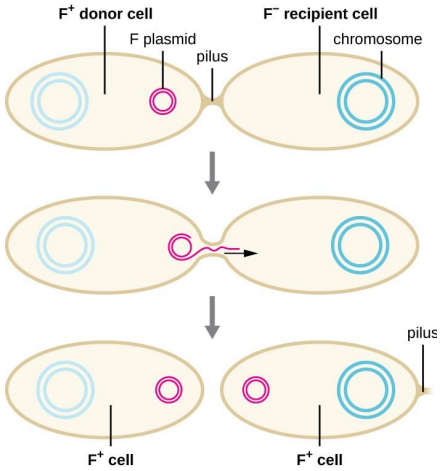


The **prokaryotic genome**

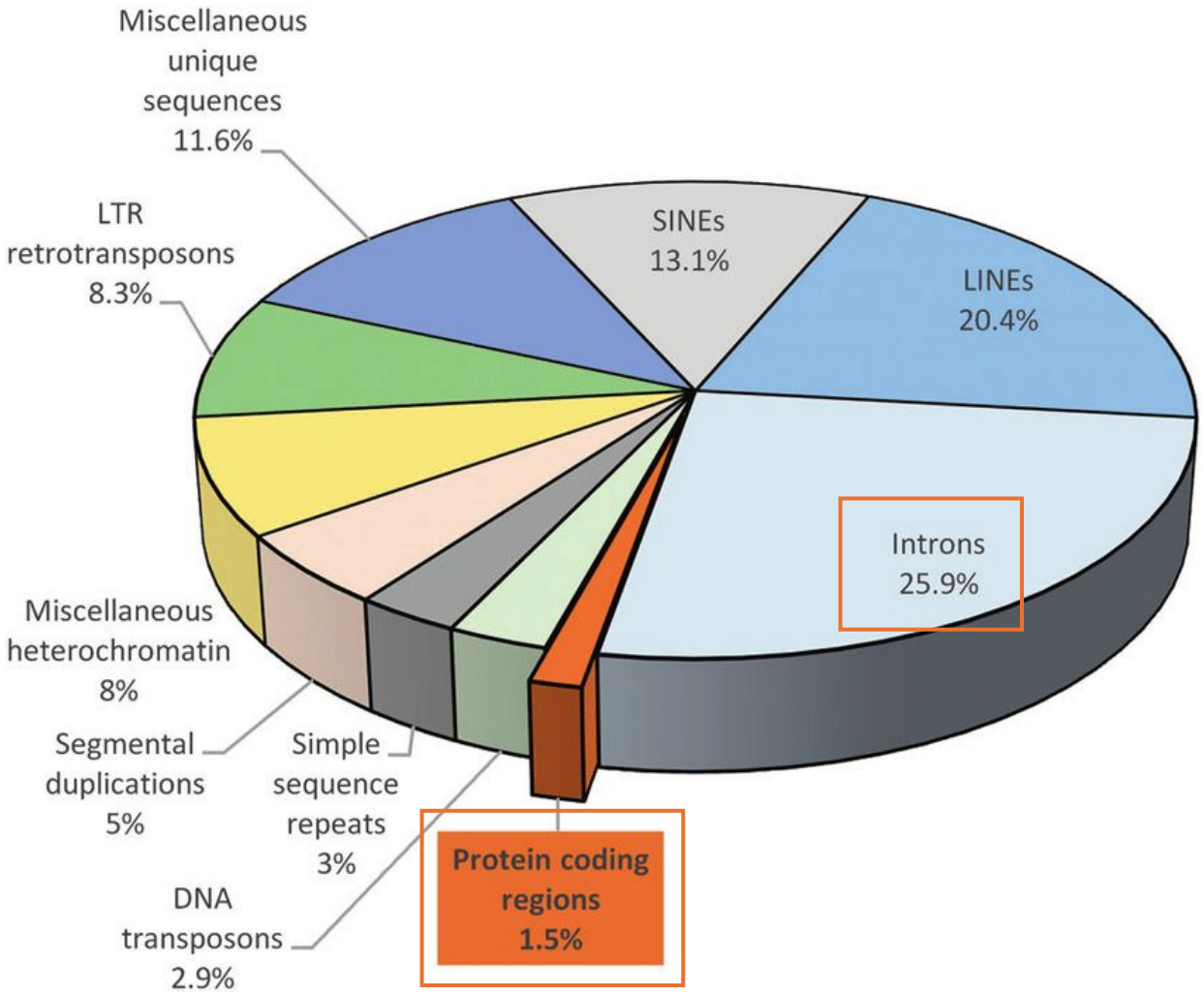
1 Bacterial chromosome



2 Bacterial plasmids

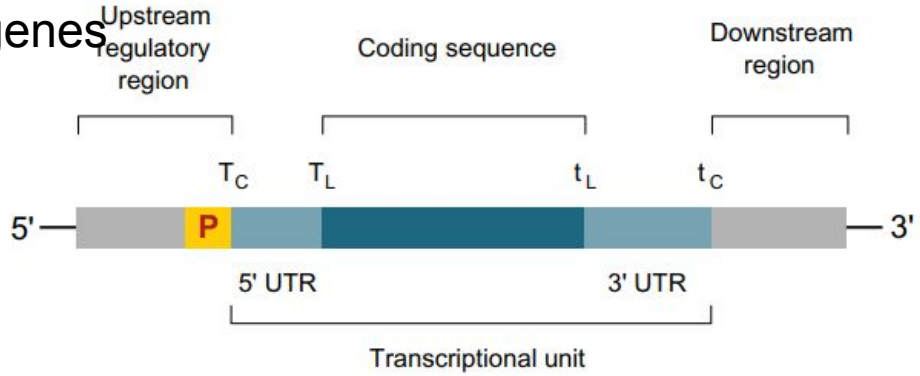


Only a small portion encodes proteins



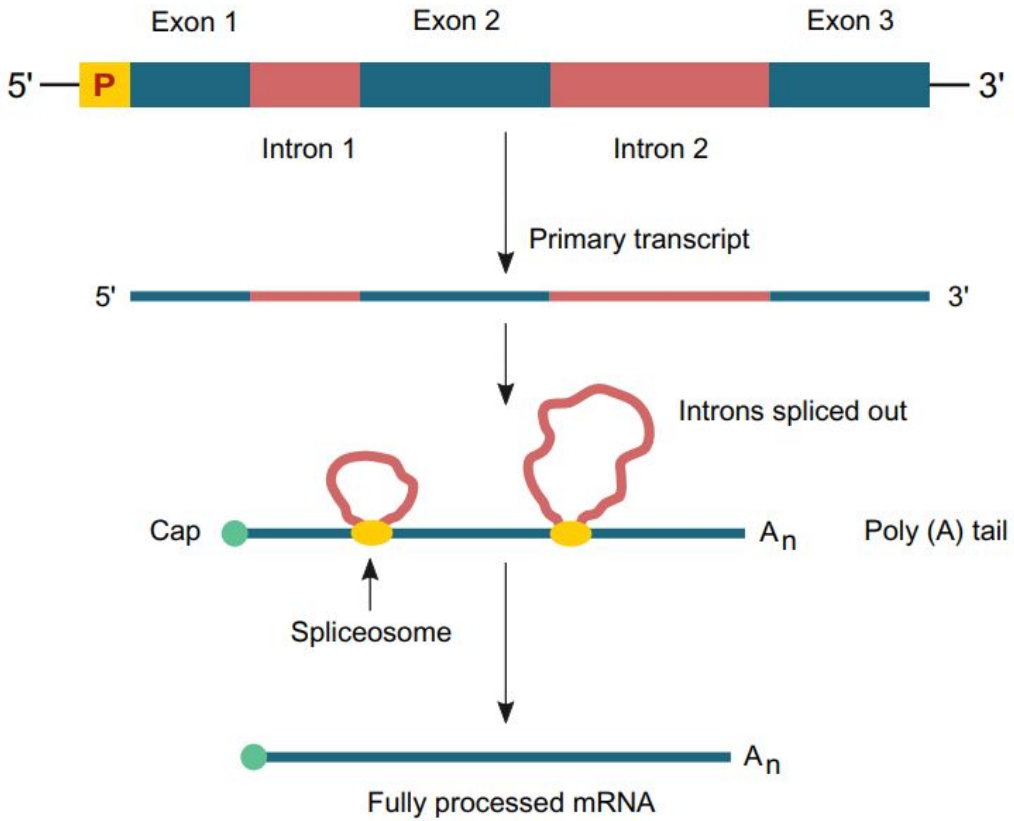
The anatomy and function of DNA – Eukaryotic genes

The **anatomy** of eukaryotic genes



- Elements of a eukaryotic gene:
  - **P**: Promoter (recruits RNA-Pol, transcription start)
  - **T_C**: transcription start, **t_C**: transcription end (both make one transcriptional unit)
  - **T_L**: translation start (on mRNA), **t_L**: translation end (both make **coding sequence, CDS**)
  - **5'UTR, 3' UTR**: untranslated regions (important for regulation)
  - Other elements (not pictured)
    - **Enhancers** (can be distal, thousands of nucleotides away) or **silencers**
    - **Terminator** (marks the end of a gene)

Processing of eukaryotic genes



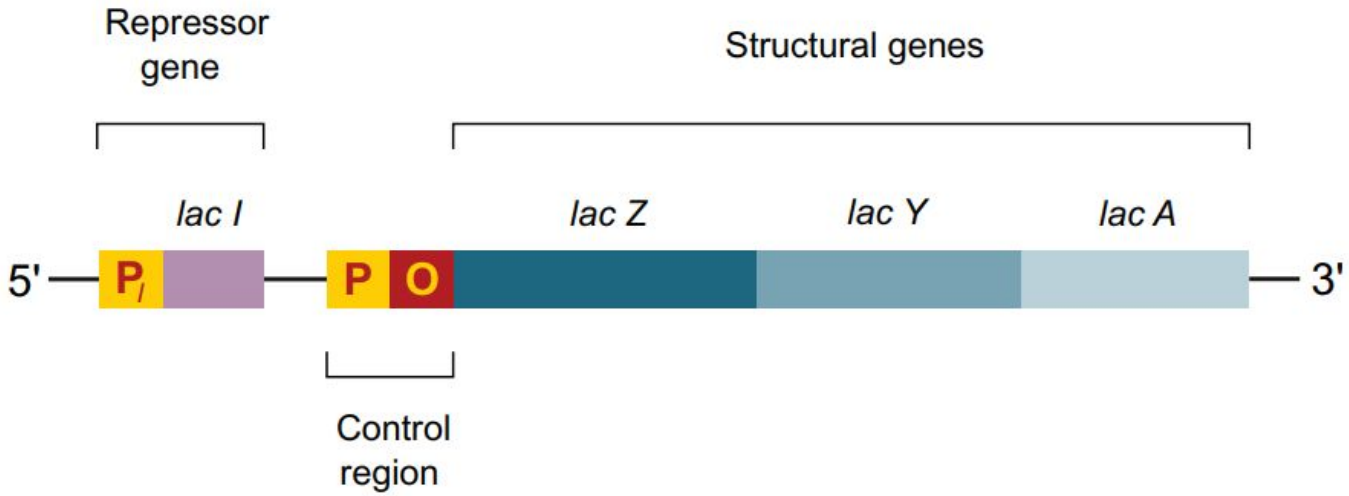
Eukaryotic genes are composed of **introns** and **exons**

- During transcript processing, **introns are**

The anatomy and function of DNA – Prokaryotic genes

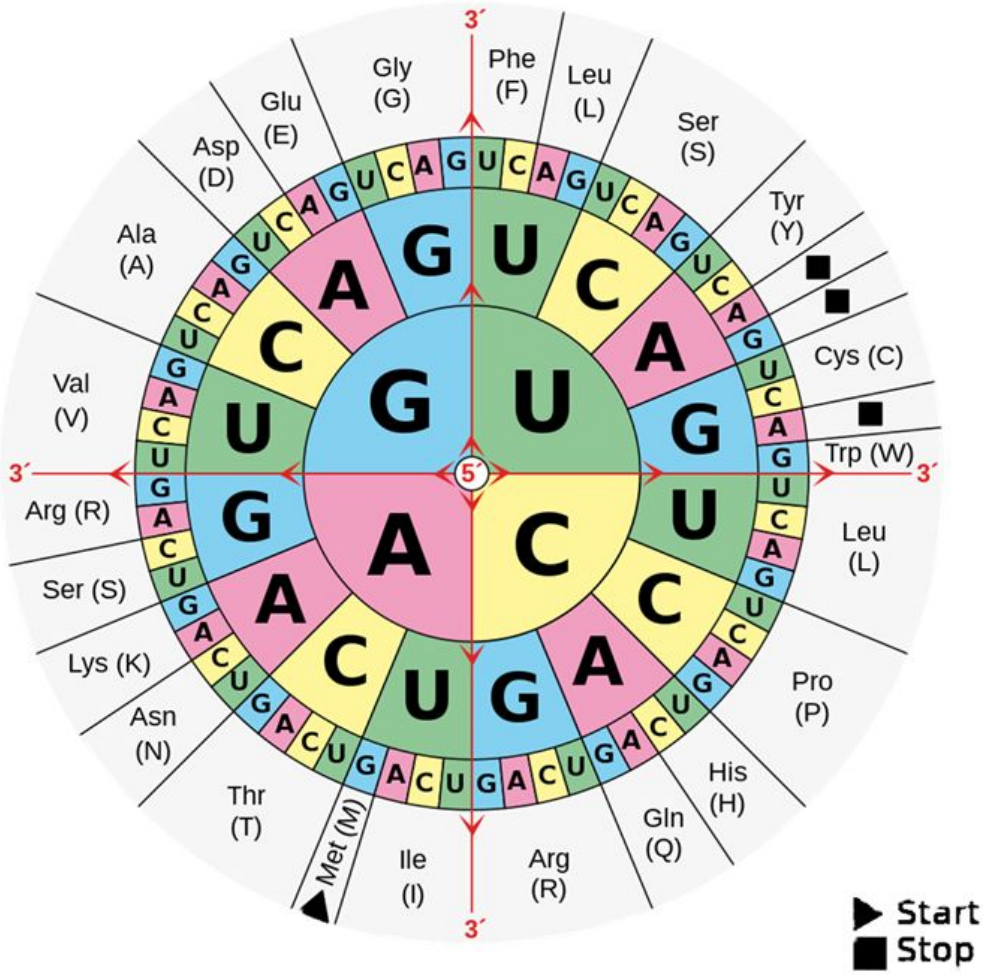
Prokaryotic genes are usually organized in operons

- Operons encode a cluster of genes that are related (e.g. encode genes for the same metabolic pathway)
  - Structural genes
  - The resulting mRNA encodes multiple genes = polycistronic mRNA
- Genes within an operon are controlled by the same promoter (P)/operator (O) region
- **There is no mRNA processing in prokaryotes**



 If you want to express an eukaryotic transgene in bacteria, you need to use the coding sequence for your designs (= exclude introns)

The genetic code – Codon usage



		E.coli high	E.coli low	B.subtilis high	B.subtilis low	S.cerevisiae high	S.cerevisiae low	S.pombe high	S.pombe low	Drosophila high	Drosophila low	Human G+C	Human A+T	
Phe	UUU	0.34	1.33	0.70	1.48	0.19	1.38	0.44	1.28	0.12	0.86	0.27	1.20	UUU
	UUC	1.66	0.67	1.30	0.52	1.81	0.62	1.56	0.72	1.88	1.14	1.73	0.80	UUC
Leu	UUA	0.06	1.24	2.71	0.66	0.49	1.49	0.28	1.79	0.03	0.62	0.05	0.99	UUA
	UUG	0.07	0.87	0.00	1.03	5.34	1.48	2.16	0.80	0.69	1.05	0.31	1.01	UUG
Leu	CUU	0.13	0.72	2.13	1.24	0.02	0.73	2.44	1.55	0.25	0.80	0.20	1.26	CUU
	CUC	0.17	0.65	0.00	0.93	0.00	0.51	1.13	0.31	0.72	0.90	1.42	0.80	CUC
	CUA	0.04	0.31	1.16	0.34	0.15	0.95	0.00	0.87	0.06	0.60	0.15	0.57	CUA
	CUG	5.54	2.20	0.00	1.80	0.02	0.84	0.00	0.68	4.25	2.04	3.88	1.38	CUG
Ile	AUU	0.48	1.38	0.91	1.38	1.26	1.29	1.53	1.77	0.74	1.27	0.45	1.60	AUU
	AUC	2.51	1.12	1.96	1.14	1.74	0.66	1.47	0.59	2.26	0.95	2.43	0.76	AUC
	AUA	0.01	0.50	0.13	0.48	0.00	1.05	0.00	0.64	0.00	0.78	0.12	0.64	AUA
Met	AUG	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	AUG
Val	GUU	2.41	1.09	1.88	0.83	2.07	1.13	1.61	2.04	0.56	0.74	0.09	1.32	GUU
	GUC	0.08	0.99	0.25	1.49	1.91	0.76	2.39	0.65	1.59	0.93	1.03	0.69	GUC
	GUA	1.12	0.63	1.38	0.76	0.00	1.18	0.00	1.06	0.06	0.53	0.11	0.80	GUA
	GUG	0.40	1.29	0.50	0.92	0.02	0.93	0.00	0.24	1.79	1.80	2.78	1.19	GUG
Ser	UCU	2.81	0.78	3.45	0.77	3.26	1.56	3.14	1.33	0.87	0.55	0.45	1.63	UCU
	UCC	2.07	0.60	0.00	0.81	2.42	0.81	2.57	0.52	2.74	1.41	2.09	0.80	UCC
	UCA	0.06	0.95	1.50	1.29	0.08	1.30	0.00	1.56	0.04	0.84	0.26	1.23	UCA
	UCG	0.00	1.04	0.00	0.94	0.02	0.66	0.00	0.67	1.17	1.30	0.68	0.13	UCG
Pro	CCU	0.15	0.75	2.29	0.99	0.21	1.17	2.00	1.21	0.42	0.43	0.58	1.50	CCU
	CCC	0.02	0.68	0.00	0.27	0.02	0.75	2.00	0.83	2.73	1.02	2.02	0.83	CCC
	CCA	0.42	1.03	1.14	1.08	3.77	1.38	0.00	1.51	0.62	1.04	0.36	1.57	CCA
	CCG	3.41	1.54	0.57	1.66	0.00	0.70	0.00	0.45	0.23	1.51	1.04	0.10	CCG
Thr	ACU	1.87	0.76	2.21	0.39	1.83	1.23	1.89	1.52	0.65	0.70	0.36	1.45	ACU
	ACC	1.91	1.29	0.00	0.98	2.15	0.78	2.11	1.04	3.04	1.58	2.37	0.92	ACC
	ACA	0.10	0.68	1.38	1.64	0.00	1.38	0.00	1.04	0.10	0.77	0.36	1.45	ACA
	ACG	0.12	1.28	0.41	0.98	0.01	0.60	0.00	0.40	0.21	0.95	0.92	0.18	ACG
Ala	GCU	2.02	0.61	2.94	0.78	3.09	1.07	2.30	1.79	0.95	0.91	0.45	1.59	GCU
	GCC	0.18	1.18	0.08	1.14	0.89	0.76	1.49	0.50	2.82	1.93	2.38	0.92	GCC
	GCA	1.09	0.79	0.60	1.19	0.03	1.49	0.21	1.14	0.09	0.59	0.36	1.38	GCA
	GCG	0.71	1.42	0.38	0.89	0.00	0.68	0.00	0.57	0.14	0.57	0.82	0.11	GCG

Redundant codons are not necessarily redundant:

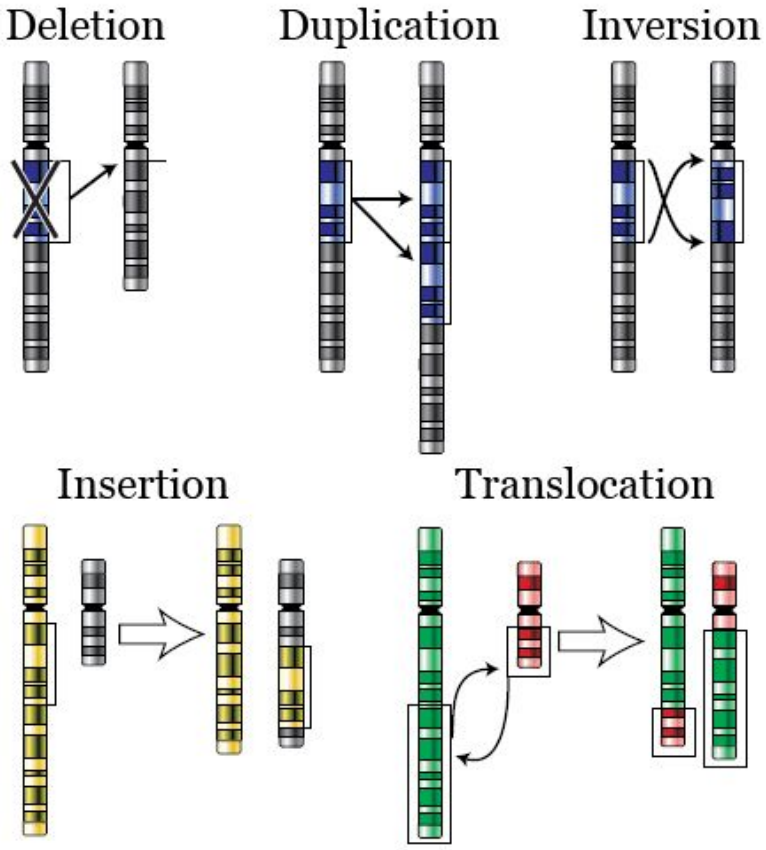
- Different organisms use different sets of degenerate codons more frequently than others
- Within species there are different codon usage frequencies between highly and low expressed genes

Genomic and gene mutations

- Any kind of change to the DNA sequence compared to standard is a mutation
- Occur mostly randomly (vs. induced by environment or inherited)
  - In humans, parents pass on an average of 60 new mutations onto their children

Some types of mutations

- Large-scale mutations
  - Chromosomal alteration
- Small-scale mutations
  - Point mutations

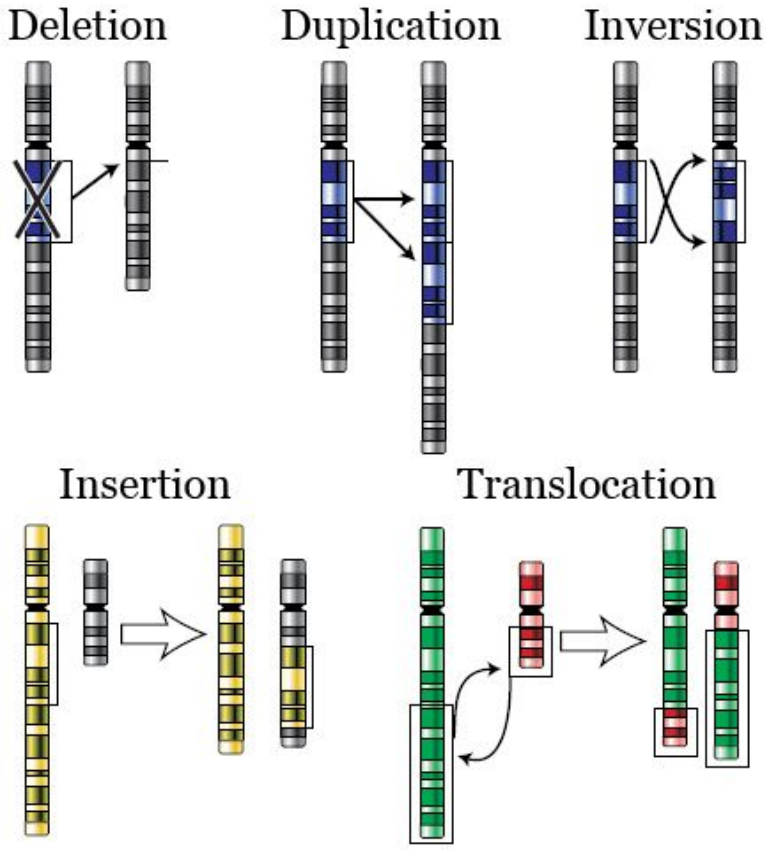


Genomic and gene mutations

- Any kind of change to the DNA sequence compared to standard is a mutation
- Occur mostly randomly (vs. induced by environment or inherited)
  - In humans, parents pass on an average of 60 new mutations onto their children

Some types of mutations

- Large-scale mutations
  - Chromosomal alteration
- Small-scale mutations
  - Point mutations

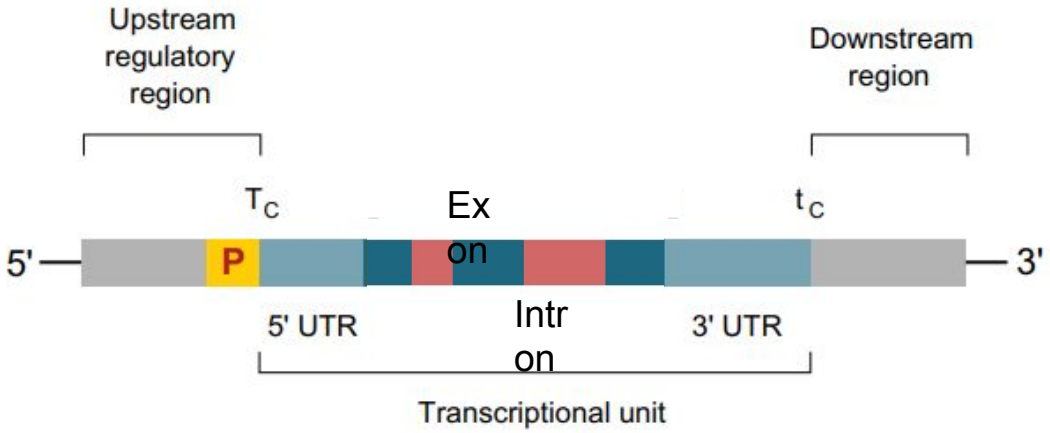


Genomic and gene mutations – Point mutations

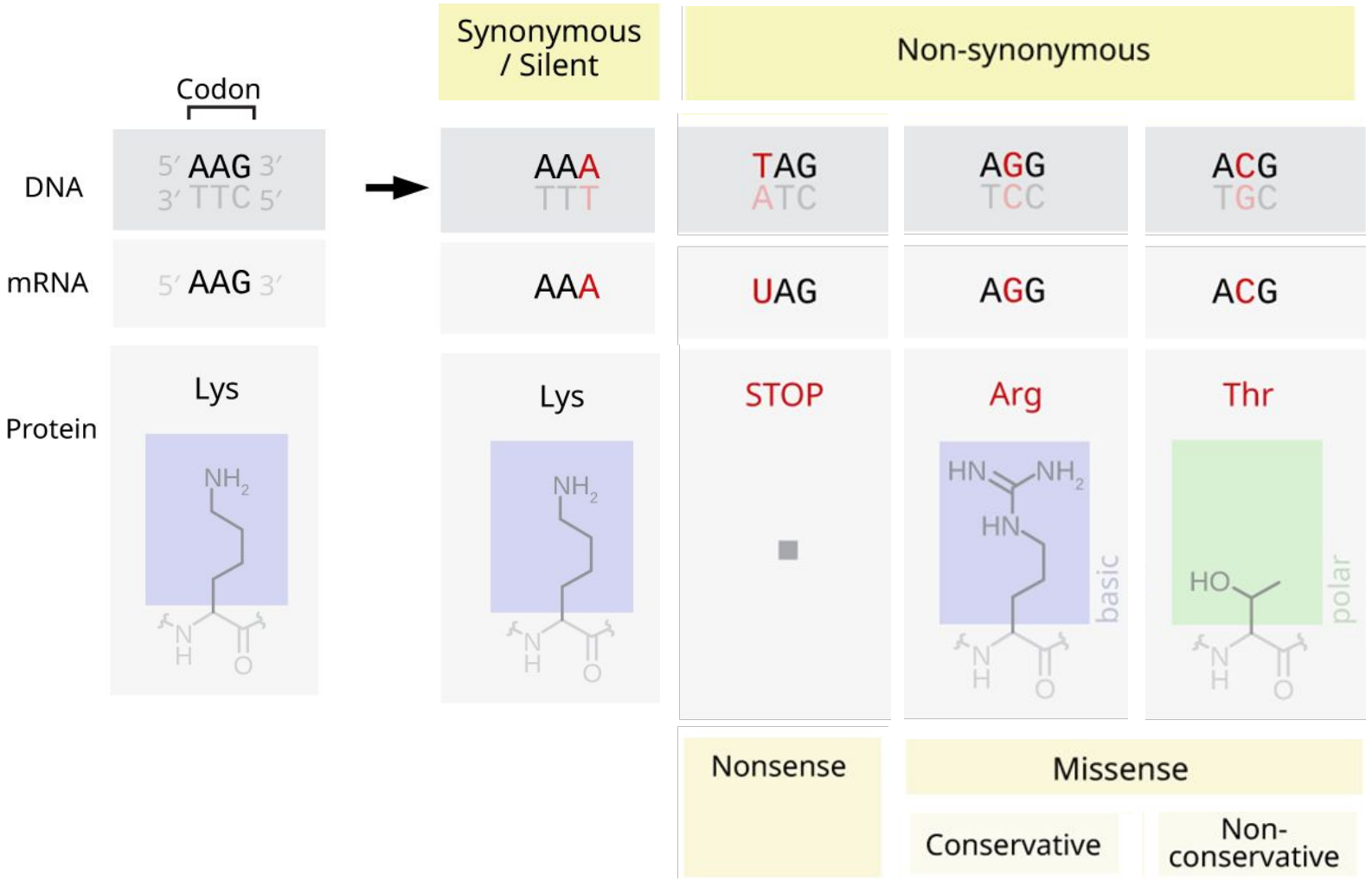
Location matters!

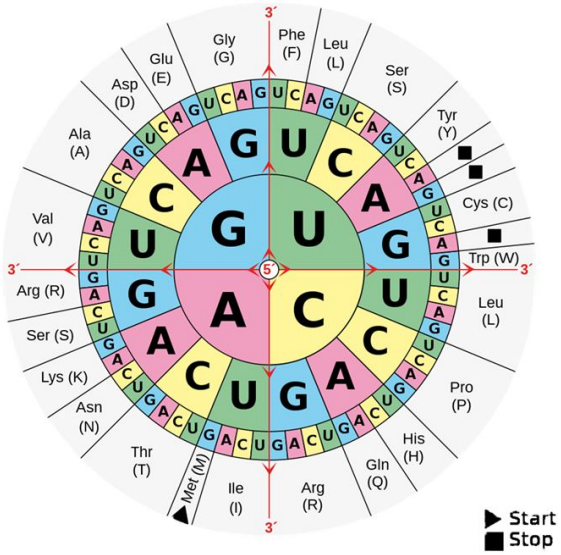
- Non-coding region?
  - Introns: Likely to be neutral (if it does not affect splicing)
  - Regulatory regions: Might affect protein expression

- Coding region?
  - Can affect the protein

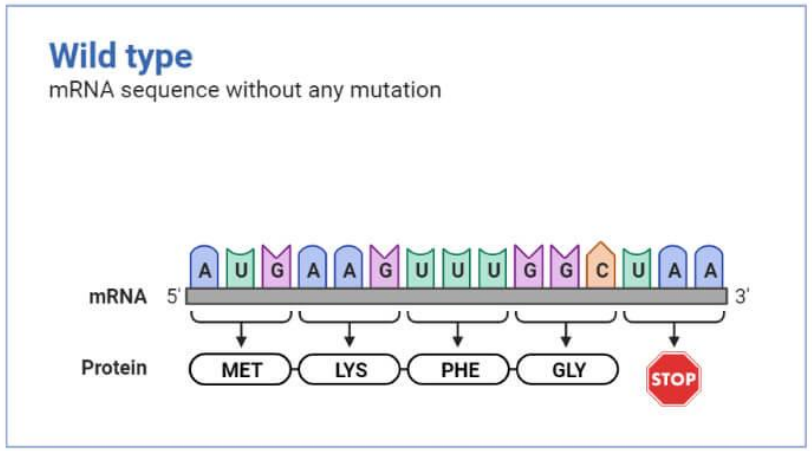


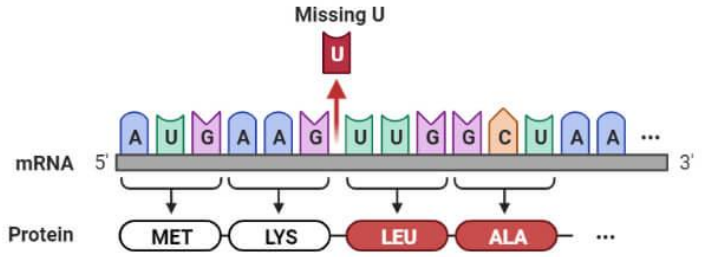
Genomic and gene mutations – Point mutations



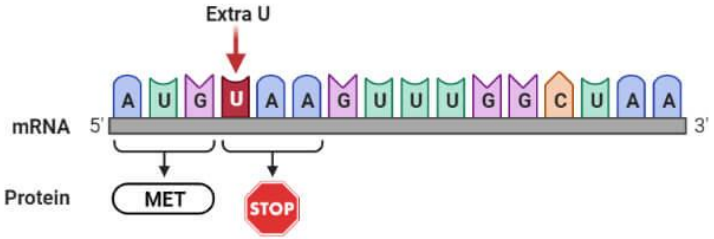


Genomic and gene mutations – Frameshift mutations



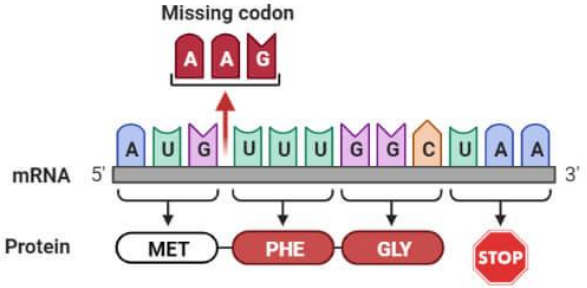


Base-pair insertion



Base-pair deletion

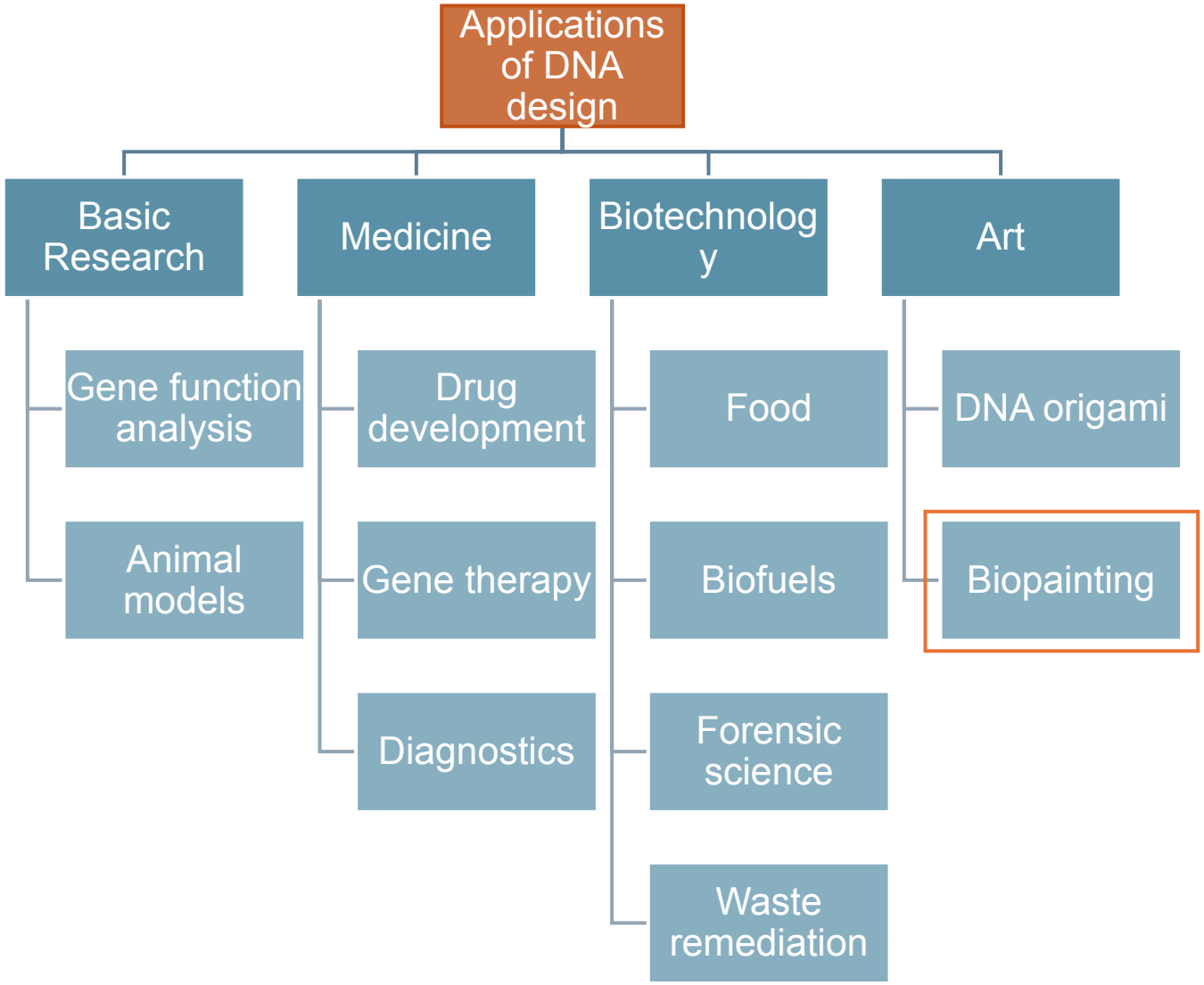
Frameshift = n(insertions or deletions) not divisible by 3



Not a frameshift: Protein with missing amino acid

Part 2 – Working with DNA

Why do we want to design DNA?



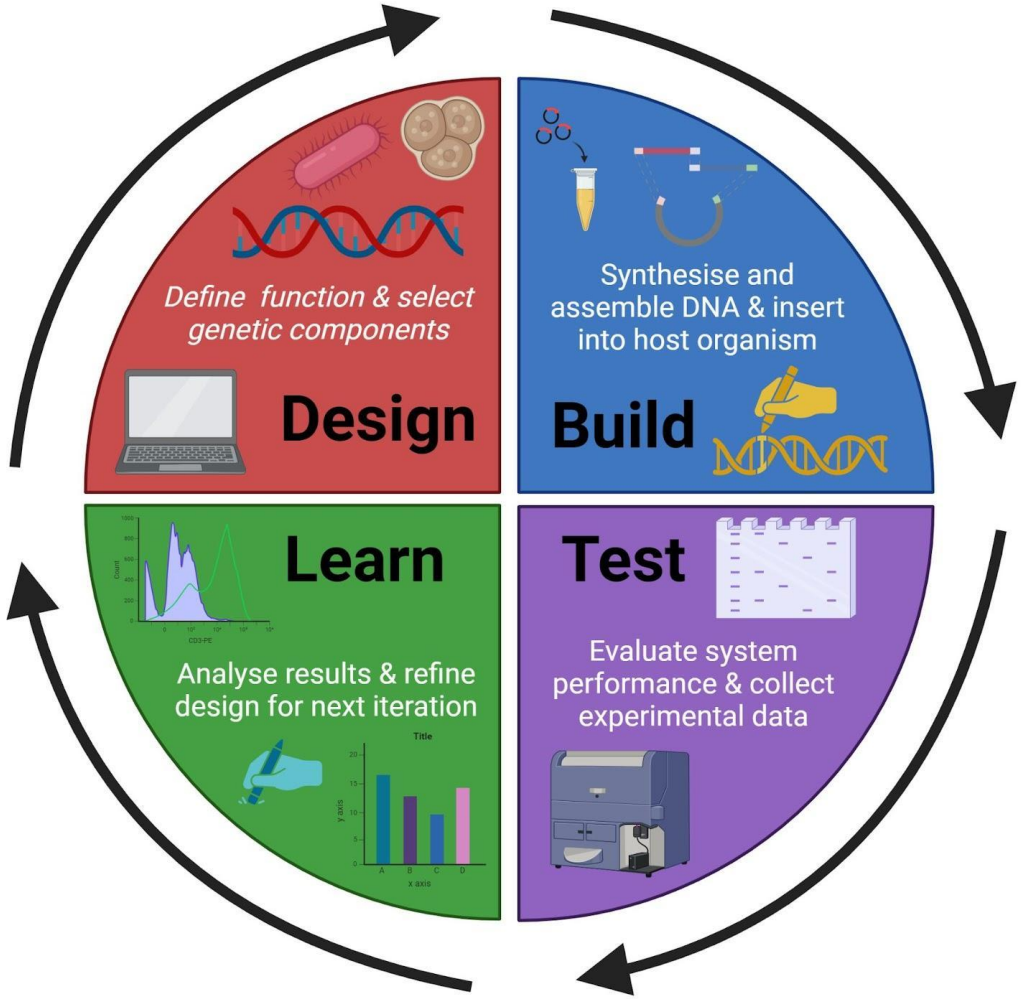
The synthetic biology design cycle

Design a DNA/genome

Design a pathway

Design experiments

Select a chassis



DNA assembly

Gene editing

Synthetic chromosomes

Introduce into chassis

Data analysis

Trouble Shooting

Modeling/Simulation

Re-Design

PCR/qPCR

Protein biochemistry

Next-Gen-Sequencing

Flow-cytrometry

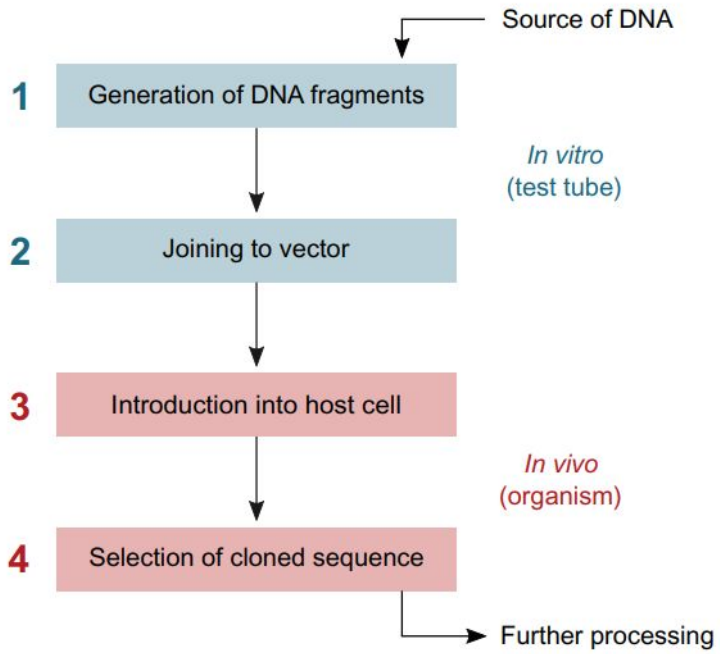
Mass spectrometry

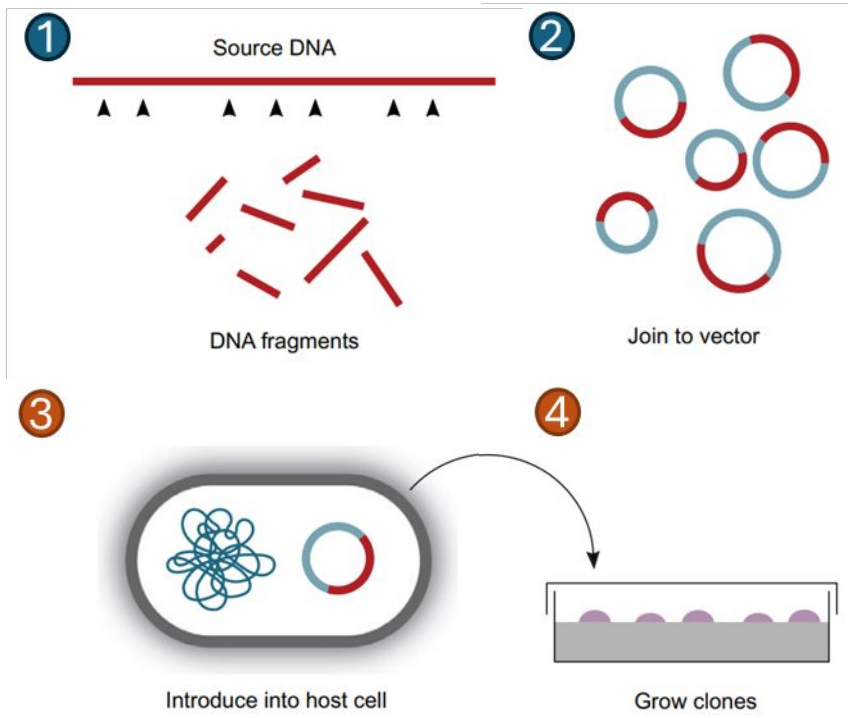
Microfluidics

How to engineer DNA? - The basics

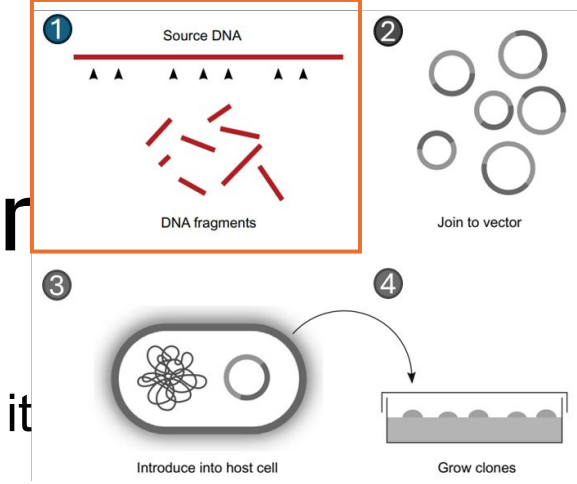
- Want to engineer DNA or analyse the output (transcripts, protein) of any DNA of interest (wild type, recombinant, engineered and otherwise)? Molecular cloning is necessary

Basic steps of Molecular cloning





1 The DNA template (gene/DNA of inter



You are interested in analysing, modifying or using a gene. Where do you get it from?

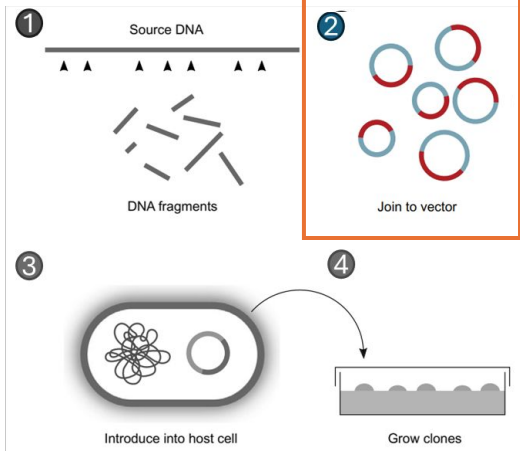
- From the genome.

  But which genome is your gene/DNA fragment from?
  - If from bacteria: Design primers for your gene, isolate the genome or perform colony PCR
  - If from eukaryotic cells: Complicated. You need to isolate mRNA, translate it into DNA to create cDNA first. Exception: Yeast, which barely has any introns.
- From an existing plasmid (ordered, or available as a prep)
  - Easy, just do a PCR
- DNA that you created/isolated using PCR methods/gene editing methods
- A de-novo synthetized DNA fragment
  - Easy, but expensive and limited by size.

- At all steps: Be mindful of codon usage if you want to express foreign or recombinant genes in a host cell!

2 The vector

- Allows to introduce into and selectively maintain recombinant DNA in host cells
- Various kinds of vectors
  - Bacterial plasmid
  - Viral vectors (e.g. Bacteriophages)
  - YAC/BAC
  - Etc.



Type of plasmid	Plasmid	Size (kb)	Vendor/Genbank accession no
Bacterial expression and cloning vectors	pUC19	2.686	M77789
	pBR327	3.274	L08856
	pBR322	4.361	J01749
	pET3a	4.64	New England Biolabs

Choose plasmid by application:

- Controlled expression?
- Protein purification?
- Visualization of cellular processes/Tagging?
- Measure/quantify?
- Genome/gene editing?

Table 7.5 Some vectors for plant and animal cells

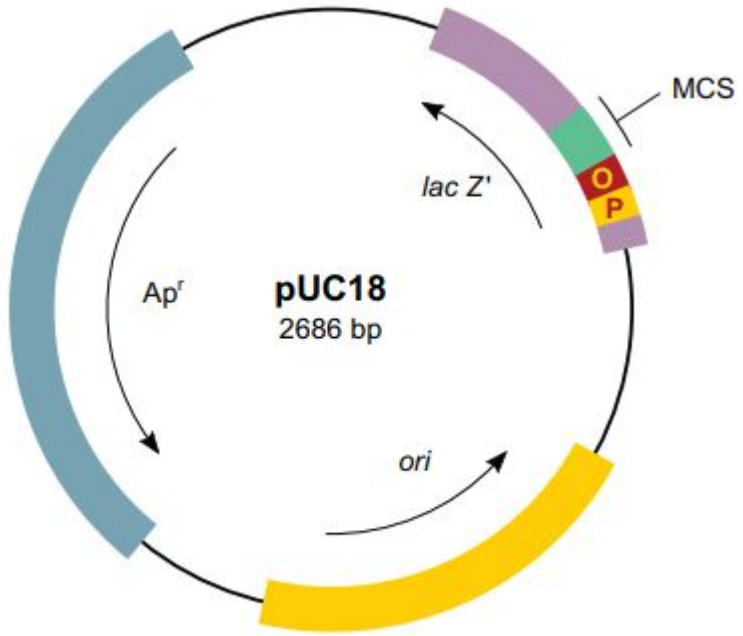
Cell type	Vector type	Genome	Examples
Plant cells	Plasmid	DNA	Ti plasmids of *Agrobacterium tumefaciens* are well-established vector systems
	Viral	DNA	Cauliflower mosaic virus Geminiviruses
		RNA	Tobacco rattle virus
Animal cells	Plasmid	DNA	Many vectors available commercially, often using sequences/promoters/origins of replication from Simian virus 40 (SV40) and/or cytomegalovirus (CMV)
	Viral	DNA	Baculoviruses for insect cells Papilloma viruses Adenovirus SV40 Vaccinia virus
	Viral	RNA	Retroviruses
	Transposon	DNA	P elements in *Drosophila melanogaster*

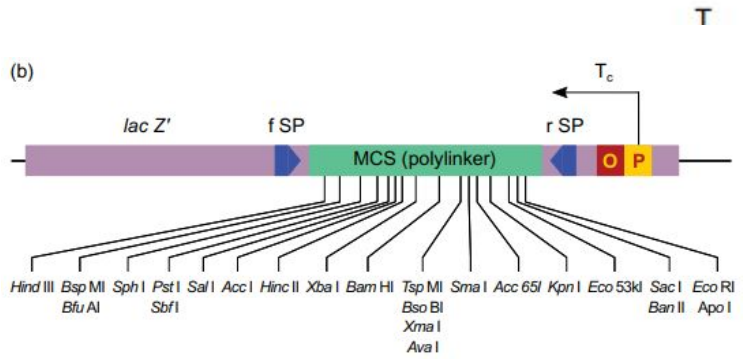
2 The bacterial plasmid

DNA is commonly introduced into bacterial hosts using bacterial plasmid vectors.

- Choose vector as one of the first steps of your experiment

- Anatomy of a bacterial vector
  - **Origin of replication (ori)**:
    - Determines whether plasmid is high- (<700 plasmids) or low (1-12 plasmids) copy
  - **Selectionmarker(s):**
    - Antibiotics: Ampicillin, Tetracycline, Kanamycin etc.
  - **Multiple cloning site (MCS):**
    - Accumulation of restriction sites for cloning
  - **Promoter**
    - Can be constitutive , inducible or repressible





3 Choosing the host cell

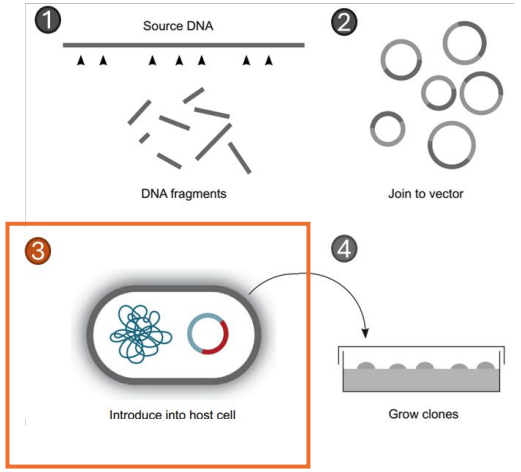


Table 7.1 Some host cells used for genetic engineering

Major group	Prokaryotic/eukaryotic	Type	Examples
Bacteria	Prokaryotic	Gram −	*Escherichia coli*
		Gram +	*Bacillus subtilis*
			Streptomyces spp.
Fungi	Eukaryotic	Microbial	*Saccharomyces cerevisiae*
		Filamentous	*Aspergillus nidulans*
Plants	Eukaryotic	Protoplasts	Various types
		Intact cells	Various types
		Whole organism	Various types
Animals	Eukaryotic	Insect cells	*Drosophila melanogaster*
		Mammalian cells	Various types
		Oocytes	Various types
		Whole organism	Various types

Note: Bacteria and fungi are generally cultured in liquid media and/or agar plates, using relatively simple growth media. Plant and animal cells may be subjected to manipulation either in tissue culture or as cells in the whole (or developing) organism. Growth requirements for these cells are often more exacting than for microbial host cells.

Fast, easy, high-throughput

Compromise between prokaryotic and eukaryotic properties

DNA modifications
RNA processing

How to modify DNA?

Modes of DNA modification

- Targeted modification
  - Requires knowledge of the sequence
- Random modification
  - No knowledge of the sequence necessary

DNA modifying/mutagenesis methods

- Physical methods: UV-radiation, heat, shear stress
- Chemical methods: intercalating agents, base analoga, reacting chemicals)
- Biological methods: enzymes

How to modify DNA?

Modes of DNA modification

- Targeted modification
  - Requires knowledge of the sequence
- Random modification
  - No knowledge of the sequence necessary

DNA modifying/mutagenesis methods

- Physical methods: UV-radiation, heat, shear stress
- Chemical methods: intercalating agents, base analoga, reacting chemicals)
- Biological methods: enzymes

The genetic toolkit: Enzymes

Nuclease
s

Degrade nucleic acids by breaking the phosphodiester bond between nucleotides

Types of Nucleases

- Restriction Endonucleases (= restriction enzymes)

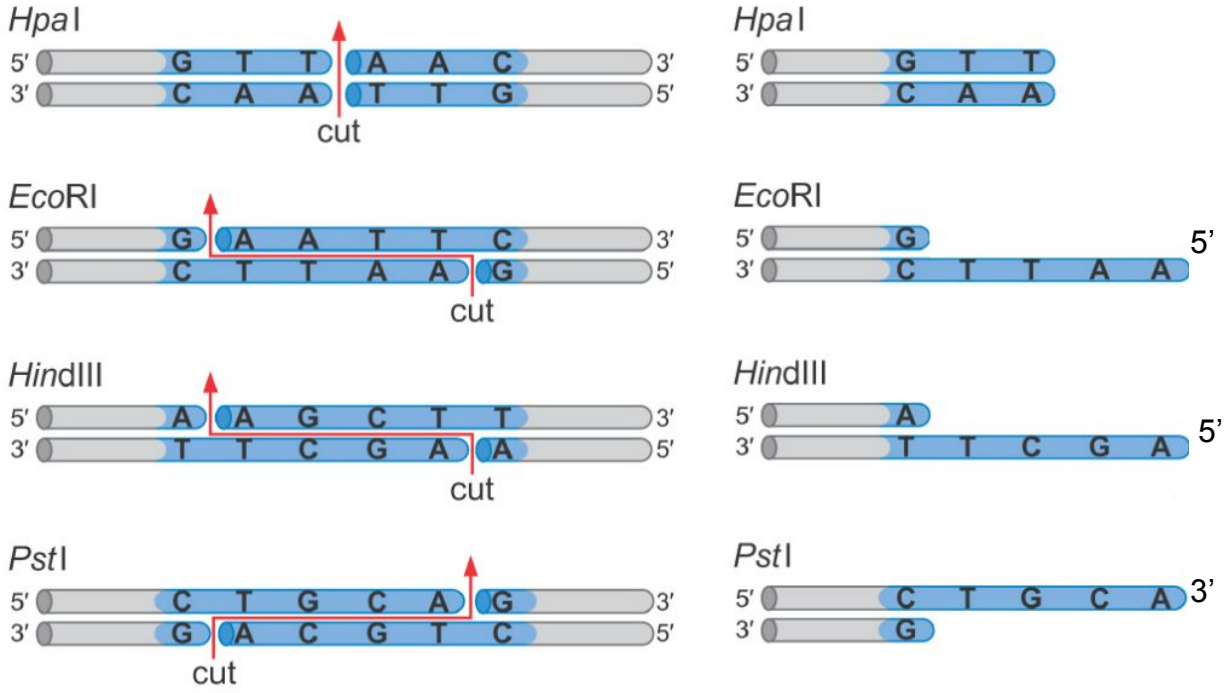
Table 5.1 General properties of restriction enzymes

Type	Characteristics
I	The first restriction enzymes to be discovered. These enzymes are large, have combined restriction/modification functions and recognise asymmetric recognition sequences (RS). Cleavage sites (CS) are random sites some 1 000 bp from the RS. Of limited use in gene manipulation.
II	The most useful type for routine use. Loosely grouped into several subtypes, depending on specific characteristics. Type IIP (often just called type II) are the most common. These recognise an inverted palindrome RS and cut within it. Other types recognise asymmetrical RS and/or may cut a few base-pairs from the site, or may cut both ends of the RS. One of the most useful subtypes is type IIS. These enzymes have two domains, one for recognition and one for cleavage. They recognise an asymmetric RS and cut a small number of base-pairs from this. They have become a key part of a cloning method called Golden Gate cloning (see Section 8.4.5).
III	Type III are large combined restriction–modification systems and cut about 20 bp from the RS. Require two inverted RS and often do not cleave DNA fully.
IV	Cleave modified regions such as methylated DNA.
V	The Cas (CRISPR-associated protein) enzyme system is considered a form of restriction enzyme that utilises guide RNAs to cleave DNA. Most well known as the CRISPR-Cas9 editing system (see Chapter 12).
Artificial	Enzymes can be engineered by combining a DNA recognition domain with a functional nuclease domain. This approach opens up the design of specific recognition sequences that can be used for very precise manipulation of gene sequences. Examples include zinc-finger nucleases, TALENs and the CRISPR-Cas9 system (see Section 5.2.1 and Chapter 12).

= commonly referred to as restriction enzymes in practice

Restriciton Endonucleases Type II (restriction enzymes)

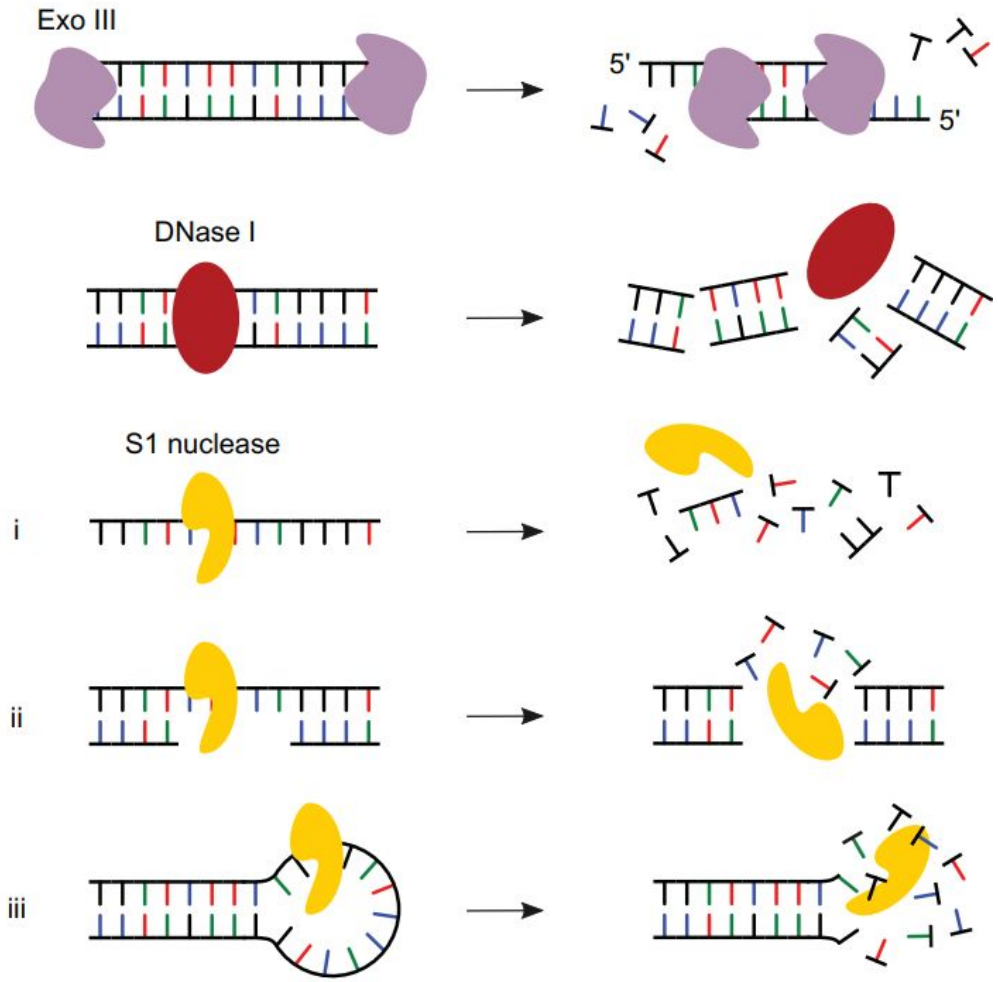
- Cut DNA site-specifically
- Used for restriction digest
  - Identify/confirm DNA sequence
  - Cloning
- Recognize specific DNA motifs (= recognition sequence)
  - Palindromic
  - Usually 4 bp – 8bp long
  - Can create sticky or blunt ends



The genetic toolkit: Enzymes

Nucleases
Degrade nucleic acids by breaking the phosphodiester bond between nucleotides

Types of Nucleases
- Restriction Endonucleases (= restriction enzymes)
- Other endonucleases:
  - Desoxyribonuclease I (DNAse I): Cut DNA nonspecifically
  - S1 endonuclease: Splits ssDNA
  - RNAses



The genetic toolkit: Enzymes

Nuclease

- Degrade nucleic acids by breaking the phosphodiester bond between nucleotides

Types of Nucleases

- Restriction Endonucleases (= restriction enzymes)
- Other endonucleases:
  - Desoxyribonuclease I (DNAse I): Cut DNA nonspecifically
  - S1 endonuclease: Splits ssDNA
  - RNAses
  - CRISPR/Cas9

DNA-Polymerases

- Amplifies DNA in 5'>3' direction
- Used in PCR, needs primers

Reverse Transcriptase

- RNA-dependent DNA-Polymerase
- Translates RNA into DNA

Alkaline Phosphatase

- Removes 5' phosphate group
  - Prevent unwanted ligation

Polynucleotide Kinase (PNK)

- Adds 5' phosphate group
  - Allows for ligation reaction

Ligases

- Makes phosphodiester bonds
- Needs 5' Phosphate, 3'OH group

Terminal Transferase

- Adds nucleotides to any available 3'OH terminus
- Practical example: Adding radioisotope labeled nucleotides

Sources

Content sources

- Pray, L. (2008) Eukaryotic genome complexity. Nature Education 1(1):96
- D.S.T. (2023) An introduction to genetic engineering, Fourth Edition, Cambridge University Press
- Reichetzeder, Christoph. (2021). Overweight and obesity in pregnancy: their impact on epigenetics. Figure 6. European Journal of Clinical Nutrition. 75. 1-13. 10.1038/s41430-021-00905-6.
- Jansen, An & Verstrepen, Kevin. (2011) Microbiology and molecular biology reviews. 75. 301-20. 10.1128/MMBR.00046-10.
- Trussart et al. (2017). Nature Communications. 8. 14665 EP. 10.1038/ncomms14665..
- Oliveira, Pedro H. & Prather, Kristala & Prazeres, Duarte & Monteiro, Gabriel. (2010). Analysis of DNA repeats in bacterial plasmids reveals the potential for recurrent instability events. Applied microbiology and biotechnology. 87. 2157-67. 10.1007/s00253-010-2671-7
- Sharp, P. M. et al; Nucleic acids research 16 17 (1988): 8207-11 .

Image Sources

- https://bio.libretexts.org/Bookshelves/Introductory_and_General_Biology/Book%3A_General_Biology_%28Boundless%29/04%3A_Cell_Structure/4.04%3A_Studying_Cells_-_Cell_Size, taken 01/02/2024, 8:00 pm.
- https://courses.lumenlearning.com/suny-microbiology/chapter/how-asexual-prokaryotes-achieve-genetic-diversity/, Taken 02/02/24, 6:34 pm
- https://commons.wikimedia.org/wiki/File:Aminoacids_table.svg#file, 02/01/24, 8:50 pm
- https://sivabio.50webs.com/dam.htm, DNA Damage and Repair, https://sivabio.50webs.com/dam.ht7.gif, taken 17/01/25 at 10:00 pm
- https://en.wikipedia.org/wiki/Point_mutation#/media/File:Point_mutations-en.png, Point mutations of a codon, classified by their impact on protein sequence, taken 17/01/25 17:35 pm
- microbenotes.com, Frameshift Mutation- Definition, Causes, Mechanism, Applications, Examples, https://microbenotes.com/wp-content/uploads/2021/03/Frameshift-Mutation-1.jpeg, 01/17/25, 7:22 pm
- Oliveira, Pedro H. & Prather, Kristala & Prazeres, Duarte & Monteiro, Gabriel. (2010). Analysis of DNA repeats in bacterial plasmids reveals the potential for recurrent instability events. Applied microbiology and biotechnology. 87. 2157-67. 10.1007/s00253-010-2671-7

Thank you for your attention!

Coming up next: Genome Engineering Methods