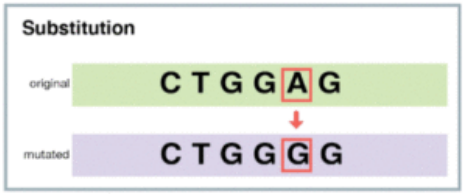
Phylogeny of COI exercise – Part 2

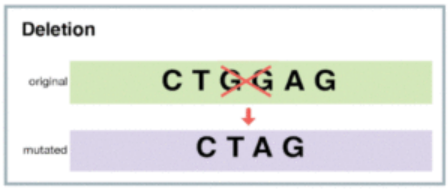
Mutations are changes in the genetic information of a cell – meaning changes in the DNA sequence. Usually occurred during DNA replication (either in mitosis or meiosis).

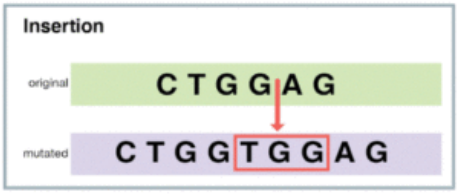
Types:

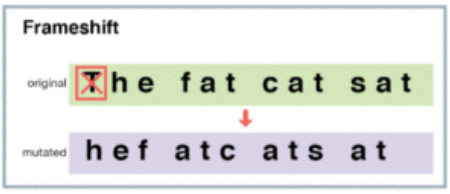
Point mutations: changes in one nucleotide base pair.

- **Substitutions**: replaces one base-pair with another
- **Insertions**: insertion of a base-pair
- **Deletions**: deletion of a base-pair









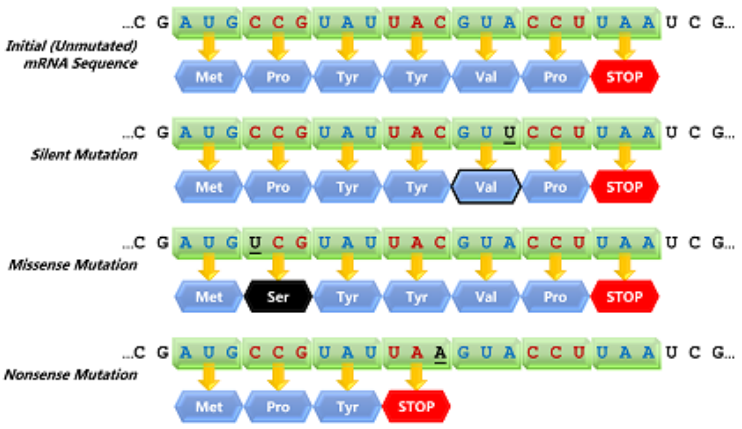
Frameshift mutations: When an insertion or deletion causes a change in the reading frame of the gene.

Silent mutations: a substitution mutation that has no effect on the amino acid produced by a codon because of redundancy in the genetic code

Missense mutations: a substitution mutation that results in the gene coding for another amino acid.

Nonsense mutations: a substitution mutation that results in a switch from a codon for the original amino acid to a stop codon. The resulting protein would not be complete.

You can also have **insertions** or **duplications** of multiple base pairs. If you insert a multiple of 3 base pairs you will keep the same reading frame.



Hypothetical cub exercise:

Note: this exercise is an overly simplified scenario (not realistic).

You are a zookeeper. One of your lions just gave birth to 6 cubs.

Two of the cubs seemed normal.
One cub is a bit lethargic (tired).
One cub is very lethargic (very tired).
Two cubs are deceased (dead).

You decide to sequence the genomes of the cubs and the lion mom and the lion father to see if you can figure out what happened. You swab the saliva of the lions, label the tubes, and send the samples for sequence. A few weeks later you receive the results.

You look at the sequences and you can tell that the COI gene does not appear the same across all sequences.

Unfortunately, some of the labels of the samples were mixed up and you are only confident in which sequences belong to the lion mother and father. You can't tell which sequence comes from which cub.

You decide to discard the lion father sequence and only focus on the mom and the cubs to figure out what went wrong.

1) Why did you decide to discard the lion father sequence from further analysis?

You need to figure out:

2) What types of mutations are present in the sequences.
3) Which sequence likely came from which cub.

You can do several things to achieve this:

- Use MEGA open the nucleotide fasta file of the sequences (Lions COI.fasta) and look if you can identify any changes to the genetic code between the mom and the cubs. You can look at the sequence with and without alignment.

- You can translate the sequences in MEGA to the appropriate protein to help you distinguish between different kinds of mutations. <u>You need use the mitochondrial vertebrate genetic code</u>. Do all mutations in the nucleotide sequence have an effect on the amino acid sequence.

There are two differences in the mitochondrial genetic code from the nuclear genetic code: in the nuclear genetic code, UGA is a stop codon, whereas in the mitochondrial genetic code UGA codes for tryptophan; AUA codes for isoleucine in the nuclear genetic code, whereas AUA codes for methionine in the mitochondrial genetic code. - *Gerald Litwack*

Mitochondrial genetic code

1st base	2nd base								3rd base
	U		C		A		G		
U	UUU	(Phe/F) Phenylalanine	UCU	(Ser/S) Serine	UAU	(Tyr/Y) Tyrosine	UGU	(Cys/C) Cysteine	U
	UUC		UCC		UAC		UGC		C
	UUA	(Leu/L) Leucine	UCA		UAA	Stop	UGA	(Trp/W) Tryptophan	A
	UUG		UCG		UAG		UGG		G
C	CUU		CCU	(Pro/P) Proline	CAU	(His/H) Histidine	CGU	(Arg/R) Arginine	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	(Gln/Q) Glutamine	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	(Ile/I) Isoleucine	ACU	(Thr/T) Threonine	AAU	(Asn/N) Asparagine	AGU	(Ser/S) Serine	U
	AUC		ACC		AAC		AGC		C
	AUA	(Met/M) Methionine	ACA		AAA	(Lys/K) Lysine	AGA	Stop	A
	AUG[A]		ACG		AAG		AGG		G
G	GUU	(Val/V) Valine	GCU	(Ala/A) Alanine	GAU	(Asp/D) Aspartic acid	GGU	(Gly/G) Glycine	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	(Glu/E) Glutamic acid	GGA		A
	GUG		GCG		GAG		GGG		G

The endosymbiotic origin of mitochondria would explain why these organelles have their own genome with a genetic code different from nuclear DNA.

4) Do any of the mutations have an affect on the 3D structure of the protein?
    a. To do this you need to export the amino acid fasta file or copy/paste the amino acid sequences
    b. SwissModel cannot have * in the sequence
    c. A * in the sequence means there is a STOP codon
    d. This means you have to remove all amino acids after the first * STOP codon
    e. You can not have gaps "-" in the sequence for SwissModel. These need to be deleted (but everything after can stay)