# Protein Translation

Shifra Ben-Dor

# Translation is relatively simple…

You need:

- A DNA sequence
- A table to convert 3mers to amino acids
- A program to put the two together

Position 2

Position 1

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| TTT | F | Phe | TCT | S | Ser | TAT | Y | Tyr | TGT | C | Cys |
| TTC | F | Phe | TCC | S | Ser | TAC | Y | Tyr | TGC | C | Cys |
| TTA | L | Leu | TCA | S | Ser | TAA | * | Ter | TGA | * | Ter |
| TTG | L | Leu | TCG | S | Ser | TAG | * | Ter | TGG | W | Trp |
| CTT | L | Leu | CCT | P | Pro | CAT | H | His | CGT | R | Arg |
| CTC | L | Leu | CCC | P | Pro | CAC | H | His | CGC | R | Arg |
| CTA | L | Leu | CCA | P | Pro | CAA | Q | Gln | CGA | R | Arg |
| CTG | L | Leu | CCG | P | Pro | CAG | Q | Gln | CGG | R | Arg |
| ATT | I | Ile | ACT | T | Thr | AAT | N | Asn | AGT | S | Ser |
| ATC | I | Ile | ACC | T | Thr | AAC | N | Asn | AGC | S | Ser |
| ATA | I | Ile | ACA | T | Thr | AAA | K | Lys | AGA | R | Arg |
| ATG | M | Met | ACG | T | Thr | AAG | K | Lys | AGG | R | Arg |
| GTT | V | Val | GCT | A | Ala | GAT | D | Asp | GGT | G | Gly |
| GTC | V | Val | GCC | A | Ala | GAC | D | Asp | GGC | G | Gly |
| GTA | V | Val | GCA | A | Ala | GAA | E | Glu | GGA | G | Gly |
| GTG | V | Val | GCG | A | Ala | GAG | E | Glu | GGG | G | Gly |

Position 3

# General rules:

- Changing the second position always changes the amino acid
- Changing the last position usually doesn't change the amino acid
- Changing the first amino acid is somewhere in between

```
AAs    = FFLLSSSSYY**CC*WLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts= ---M------**--*----M---------------M---------------------------
Base1  = TTTTTTTTTTTTTTTTCCCCCCCCCCCCCCCCAAAAAAAAAAAAAAAAGGGGGGGGGGGGGGGG
Base2  = TTTTCCCCAAAAGGGGTTTTCCCCAAAAGGGGTTTTCCCCAAAAGGGGTTTTCCCCAAAAGGGG
Base3  = TCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAG
```

# But there are several genetic codes

https://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi?mode=c

# And several definitions of ORF

- Start to Stop : ATG > *
- Stop to Stop : * > *
- Open ended (either requiring ATG or stop or neither)

M *

* M *

M M M

*

* *

# Defining an ORF

- For example: An ORF is defined as the coding sequence:
  - from the beginning of the transcript to the first stop
  - between two stops
  - from the last stop to the end of the transcript
- minimum length is also defined

# Codon Usage

- For amino acids with more than one codon, there is preferential usage of the different codons
- This usage differs by species (and by organelle, for example mitochondria)
- There are databases and programs to calculate codon usage

# Codon Optimization

- This is the process of optimizing the codons for maximal translation in a given species

- The same protein sequence can have very different DNA sequences in optimized sequences for different species