

2. Gènes et protéines

- La séquence est-elle un bon modèle de l'ADN ?
- Les gènes, de Mendel à la biologie moléculaire
- **Le code génétique**
- Un algorithme de traduction
- Implémenter le code génétique
- Algorithmes + structures de données = programmes
- Les compromis de la conception d'algorithmes
- Les technologies de séquençage de l'ADN
- Le séquençage de génomes complets
- Comment trouver les gènes ?

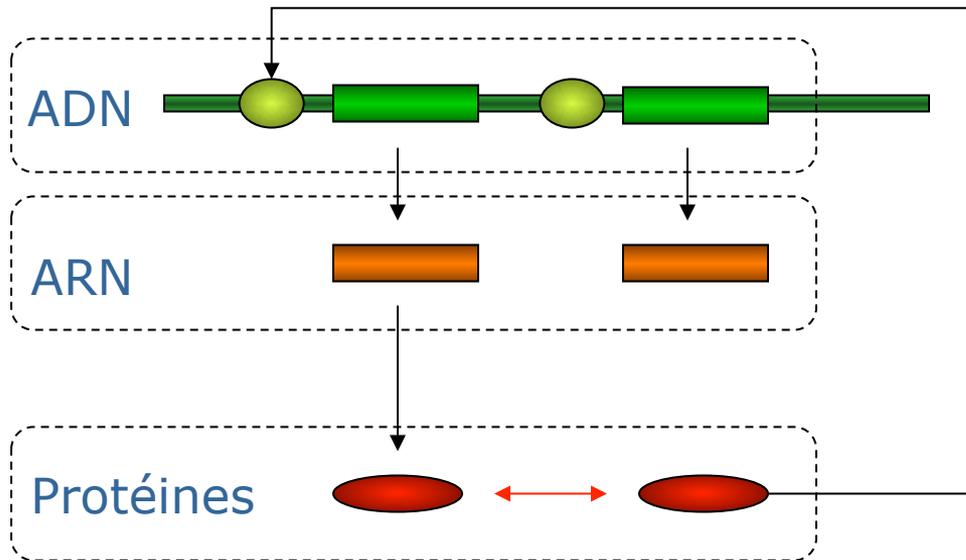
Protéines

- Les **gènes** sont des **régions de l'ADN** qui codent les **protéines**
- Une **protéine** est constituée d'une **succession d'acides aminés (AA)**
- Il existe **20 acides aminés** distincts
- Désigné chacun par un **nom complet**, un **nom à 3 lettres** ou par **une seule lettre**

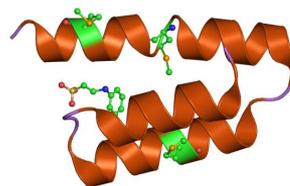
Acides aminés

Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Transcription



Traduction



Traduction

- L'**ARN** des gènes est traduit en **séquences d'acides aminés**
- **Quelle est le code ?**
c'est-à-dire la correspondance entre
une séquence d'**ARN**, écrite dans un **alphabet de 4 lettres**
et la séquence d'**AA**, écrite dans un **alphabet de 20 lettres ?**
- 1 (ribo)nucléotide / lettre → seuls 4 AA possibles codables
- 2 (ribo)nucléotides / lettres → seuls 16 AA possibles
AA AC AG AT CA CG CC CT GA GC GG GT TA TC TG TT
- 3 (ribo)nucléotides / lettres → 64 ($4 \times 4 \times 4$) AA

Le code génétique, une table à 3 entrées

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		UCA		UAA	Stop (Ochre)	UGA	Stop (Opal)	A
	UUG		UCG		UAG	Stop (Amber)	UGG	(Trp/W) Tryptophan	G
C	CUU	(Leu/L) Leucine	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		ACA		AAA	(Lys/K) Lysine	AGA	(Arg/R) Arginine	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

[A] : codon start

Le code génétique est redondant

- **Plusieurs triplets différents codent le même acide aminé**
 - Exemple: CCU, CCC, CCA et CCG codent tous pour la proline / P
- Les triplets de nucléotides au sein des gènes sont appelés « **codons** »

- Le triplet **ATG** (or AUG) code la **méthyonine**
mais il peut jouer également le rôle de **codon initiateur** de la traduction (ou **codon start**)
- **UAA, UAG** et **UGA** sont les triplets de **terminaison** de la traduction (ou **codons stop**)

Le code génétique dans un tableau

- Le code génétique peut être représenté par un tableau de **64 lignes** et **2 colonnes**
- L'ordre des triplets est destiné à faciliter l'utilisation du tableau
- Ici, seules les 12 premières lignes ont été représentées

TTT	F	
TTC	F	
TTA	L	
TTG	L	
TCT	S	
TCC	S	
TCA	S	
TCG	S	
TAT	Y	
TAC	Y	
TAA	*	stop
TAG	*	stop

Illustrations & photos : crédits

p. 4 : Domaine public, <http://www.ebi.ac.uk/>