

Comment l'information de séquençage numérique est-elle produite ?

Dr Wael Houssen

Membre du Conseil EPSRC

Institut des sciences médicales

Université d'Aberdeen

<https://www.abdn.ac.uk/people/w.houssen>





CBD



**Convention on
Biological Diversity**

Distr.
GENERAL

CBD/DSI/AHTEG/2020/1/3
29 January 2020

ENGLISH ONLY

AD HOC TECHNICAL EXPERT GROUP ON
DIGITAL SEQUENCE INFORMATION ON
GENETIC RESOURCES

Montreal, Canada, 17-20 March 2020

Annex

**Digital Sequence Information on Genetic Resources: Concept, Scope and
Current Use**

Wael Houssen, Rodrigo Sara, Marcel Jaspars

Liens :

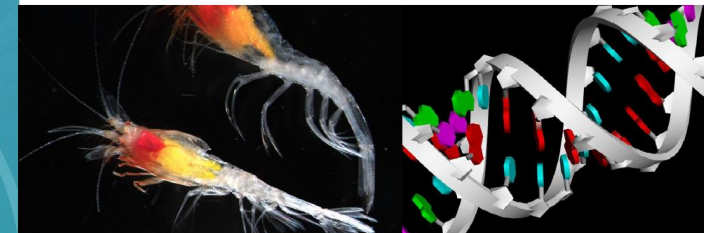
- <https://www.cbd.int/doc/c/fef9/2f90/70f037ccc5da885dfb293e88/dsi-ahteg-2020-01-03-en.pdf>
- <https://www.dosi-project.org/wp-content/uploads/070-DSI-Policy-brief-V4-WEB.pdf>

POLICY BRIEF

MARCH 2020



Digital Sequence Information – Clarifying Concepts

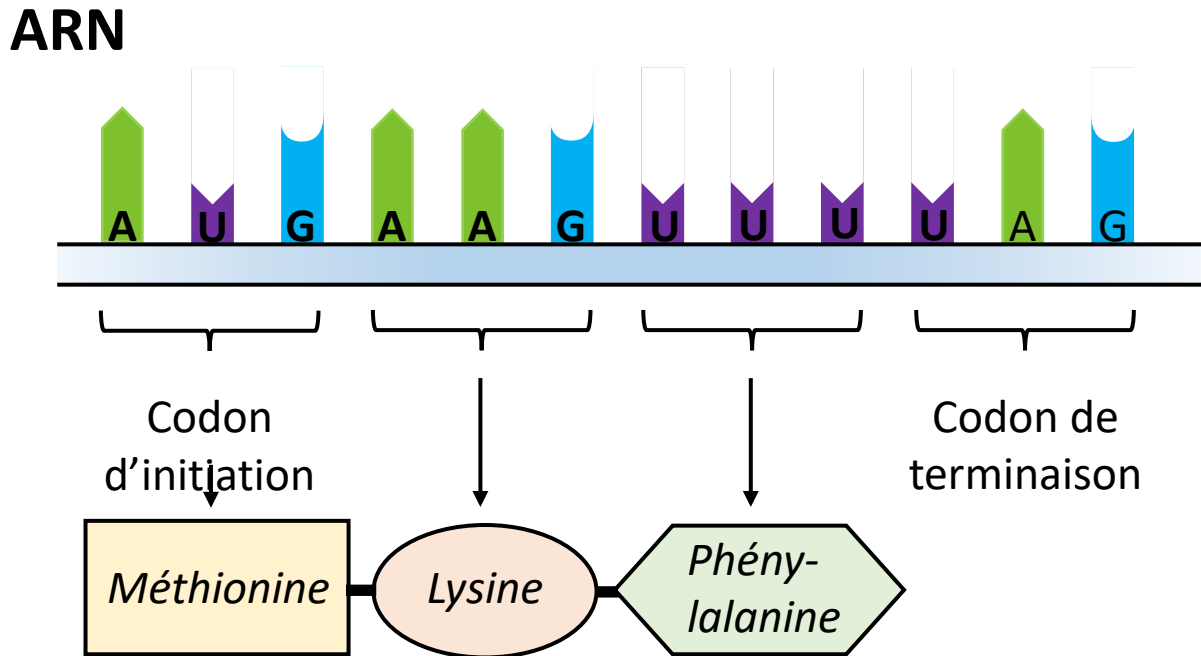
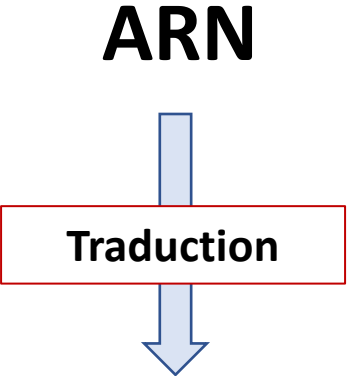
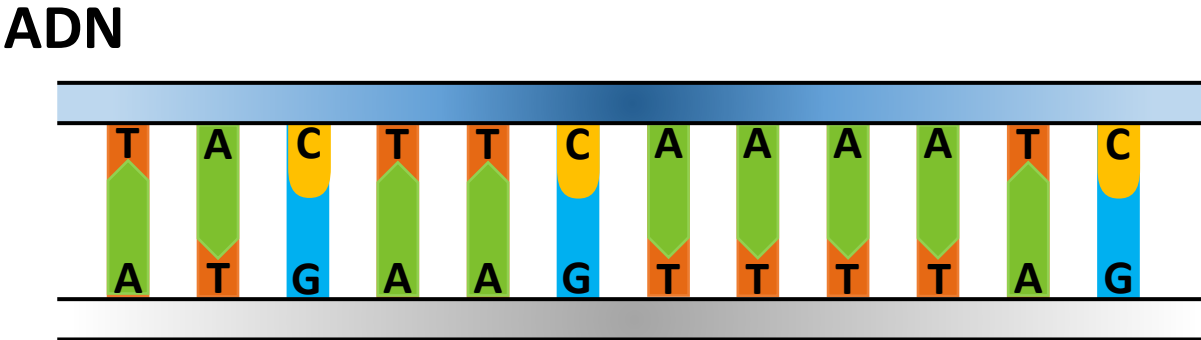
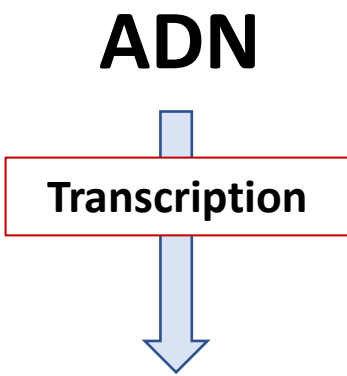


This Policy Brief is Based on:

“Digital Sequence Information on Genetic Resources: Concept, Scope and Current Use”. Authors:

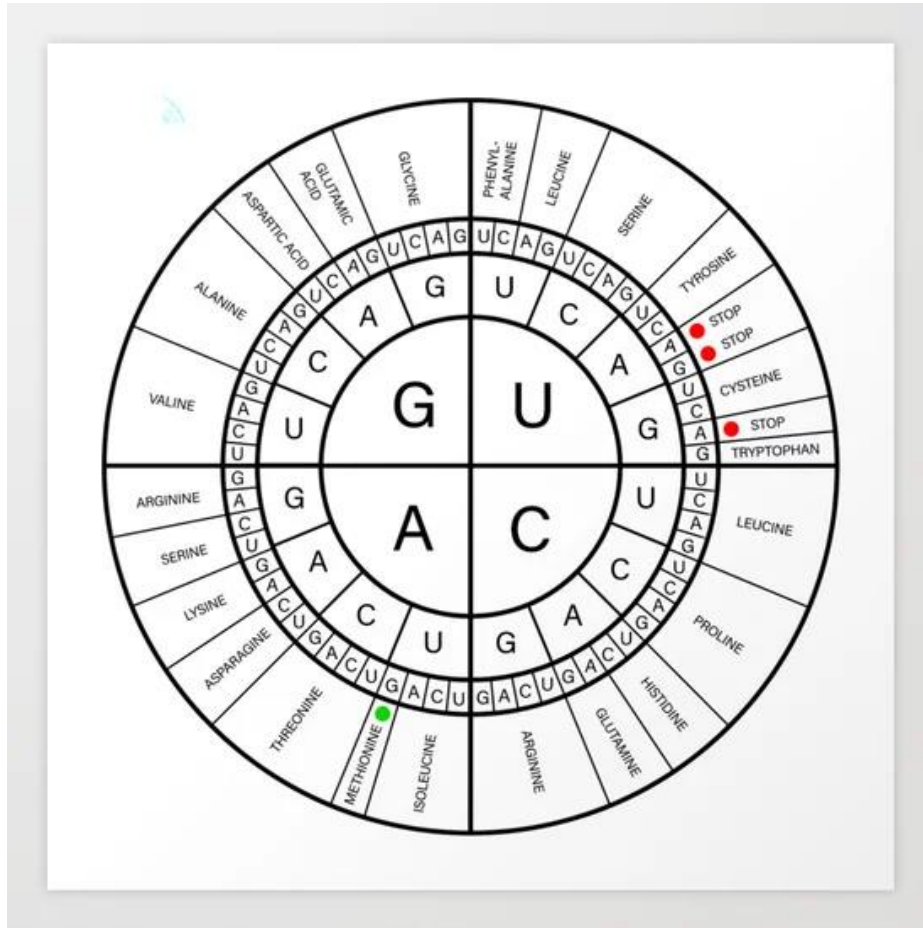
Wael Houssen, Rodrigo Sara, Marcel Jaspars

Le dogme central de la biologie moléculaire



Métabolites

Roue de codon



- Les vingt acides aminés naturels sont encodés par soixante-quatre codons.
- Le codon AUG encode la méthionine et active le ribosome pour démarrer la traduction d'une protéine. On parle de « **codon d'initiation** ».
- Les « **codons de terminaison** » quant à eux marquent la fin de la traduction en protéines.
- On est en présence de dégénérescence du code génétique lorsque plusieurs acides aminés sont codés par différents codons.
- Chaque organisme a sa « préférence » quant à l'utilisation d'un codon particulier pour un acide aminé spécifique. C'est ce qu'on appelle un « biais de codon ».

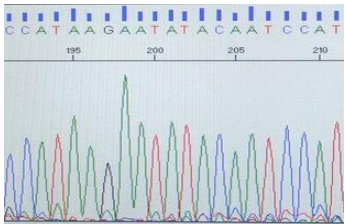
Origine de l'information de séquençage numérique

Applications

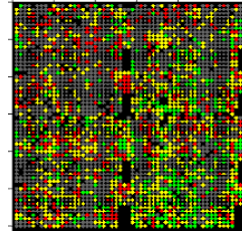
Métadonnées



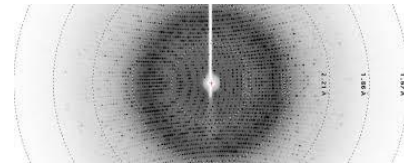
Annotations fonctionnelles



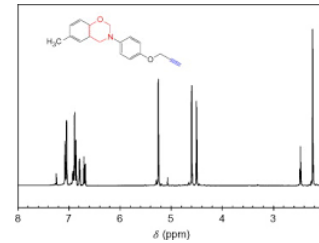
Profilage d'expression génétique



Structure protéique et caractérisation fonctionnelle



Structure des métabolites et bioactivité



Génomique

Transcriptomique

Protéomique

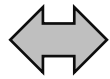
Métabolomique



Ressource génétique



ADN



ARN



Protéine



Métabolite

Séquence brute par opposition à ISN à valeur ajoutée

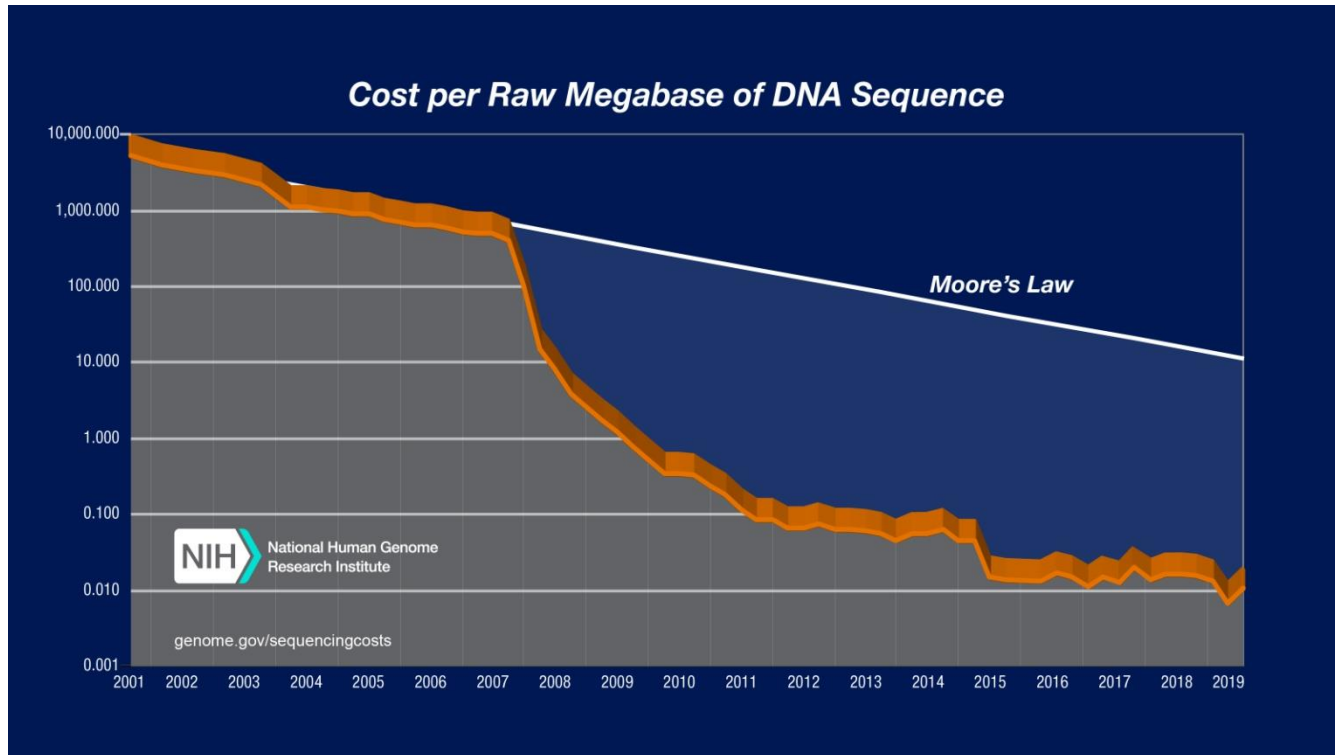
La séquence de l'ADN en elle-même ne fournit aucune information quant

- à la fonction de chaque gène,
- au niveau d'expression de chaque gène,
- au mode de repliement d'une séquence de protéine codée. Les protéines mal pliées sont connues pour être à l'origine de maladies graves.
- aux modifications post-traductionnelles des protéines, p. ex. la glycosylation des protéines.

Technologies de séquençage de la prochaine génération

Plateforme / instrument	Plage de débit (Gb)	Longueur de lecture (bp)	Force	Faiblesse
Séquençage Sanger				
ABI 3500/3730	0,0003	jusqu'à 1kb	Longueur et exactitude de la lecture	Coût et débit
Illumina				
MiniSeq	de 1,7 à 7,5	de 1×75 à 2×150	Investissement initial minimal	Longueur de lecture et temps de passage
MiSeq	de 0,3 à 15	de 1×36 à 2×300	Longueur de lecture, extensibilité	Temps de passage
NextSeq	de 10 à 120	de 1×75 à 2×150	Débit	Longueur de lecture et temps de passage
HiSeq (2500)	de 10 à 1 000	de 1×50 à 2×250	Exactitude de la lecture, débit, coût minimal par échantillon	Investissement initial important, temps de passage
NovaSeq 5000/6000	de 2 000 à 6 000	de 2×50 à 2×150	Exactitude de la lecture, débit, coût minimal par échantillon	Investissement initial important, longueur de lecture et temps de passage
Ion Torrent				
PGM	de 0,08 à 2	jusqu'à 400	Longueur de lecture, vitesse	Débit, homopolymères
S5	de 0,6 à 15	jusqu'à 400	Longueur de lecture, vitesse, extensibilité	Homopolymères
Proton	de 10 à 15	jusqu'à 200	Vitesse, débit	Homopolymères
Pacific BioSciences				
PacBio RSII	de 0,5 à 1	jusqu'à 60 kb (moyenne 10 kb, N50 20 kb)	Longueur de lecture, vitesse	Investissement initial important et taux d'erreur élevé, faible débit
Sequel	de 5 à 10	jusqu'à 60 kb (moyenne 10 kb, N50 20 kb)	Longueur de lecture, vitesse	Taux d'erreur élevé
Oxford Nanopore				
MinION	de 0,1 à 1	jusqu'à 100 kb	Longueur de lecture, portabilité	Taux d'erreur élevé, temps de passage, faible débit

Coûts de séquençage



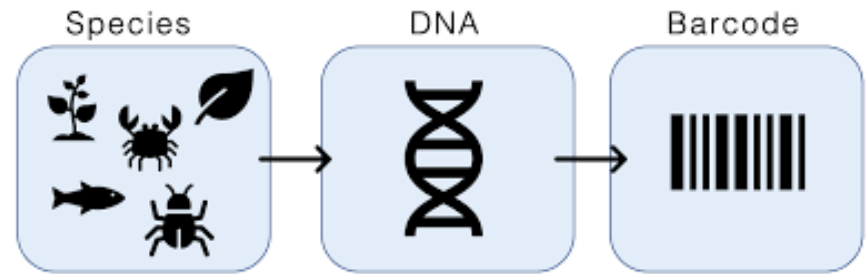
- Le séquençage se fait de **plus en plus vite** et à **des coûts de plus en plus bas**.
- Le nombre de séquences ne cesse d'augmenter.

Qu'est-ce qui fait l'objet de séquençage ?

- génome entier
- métagénome
- 16S/18S/ITS/CO1 amplicon pour l'identification moléculaire des microorganismes
- ADN environnemental (ADNe)

Exemples d'application d'ISN

- Identification taxonomique des organismes : Codage à barres de l'ADN



- Conservation de la diversité biologique
- Agriculture et sécurité alimentaire : Marqueurs génétiques pour élevage de sélection ; Développement d'organismes génétiquement modifiés (OGM)
- Recherche de nouveaux médicaments : identification de nouvelles cibles pour les médicaments
- Médicaments : détection d'agents pathogènes infectieux
- SynBio and biotechnologie : production recombinante de protéines, par exemple d'insuline et d'enzymes pour les détergents à lessive