

GÉNÉTIQUE

Structure du génome & variabilité génétique

I. ORIGINE ET NATURE DE LA GÉNÉTIQUE

C'est la science de la genèse des êtres vivants. Plus modestement : l'**étude des relations génotype/phénotype**. Elle a plusieurs branches :

- génétique du développement
- génétique de l'évolution
- génétique médicale = génétique des maladies

1- Bref historique

Depuis environ 20.000 ans, les hommes font de la génétique sans le savoir : élevage, agriculture. Quelques dates marquantes :

- 1866 : découverte des lois de Mendel
- 1900-1910 : théorie chromosomique de l'hérédité : chromz. qui se séparent à la mitose
- 1944 : identification de l'ADN comme vecteur d'hérédité : observations sur les bactéries : expériences sur les pneumocoques pathogènes inactivés
- 1953 : la double hélice : CRICK & WATSON : observations structurales (rayons X sur cristaux d'ADN : déductions : mécanisme de réplication semi-réplivative, une hypothèse plus importante que l'observation elle-même, vérifiée par la suite
- 1970 : le clonage de l'ADN : permet d'en étudier des fragments spécifiques
- 2002 : séquence du génome humain

2- La génétique permet d'affirmer l'unité du monde vivant

a. Mécanisme schématique de la reproduction :

Grandes **similarités de constitution et d'organisation** :

- les acides nucléiques sont le vecteur de l'information,
- un même lecture de l'information : code génétique.

In vivo : ADN – transcription → ARN – traduction → protéine

L'ADN permet l'**auto-renouvellement du vivant dans un contexte cellulaire** : il est difficile de parvenir à une cellule fonctionnelle à partir du seul ADN, hors de son contexte cellulaire.

La cellule est un **système complexe intégré capable de se reproduire** (à l'identique), mais aussi d'**évoluer** : ceci vient à l'appui de l'hypothèse d'une ascendance commune pour tous les êtres vivants (DARWIN).

b. Conséquences de la reproduction et diversité des organismes vivants :

- une grande variété de « formes »,
- des millions d'organismes vivants (espèces).

L'**apparition de la nouveauté** provient de la **reproduction « presque » à l'identique** ; à chaque

division cellulaire, il apparaît des différences, simples ou plus complexes.
L'évolution est possible grâce à la nouveauté, les possibilités de changements :

- mutation ponctuelle (dans la réplication de l'ADN),
- duplication, délétion,
- **recombinaison** (rôle important) :
 - entre séquences homologues = méiose (reproduction sexuée),
 - mais plus large, peut aussi se produire entre segments non homologues, à partir de segments existants => toute modification dans l'**ordre** des séquences d'ADN, échange de brins.

II. PRINCIPES DE L'ÉVOLUTION

1- Mutation et évolution

Les mécanismes de renouvellement et de transmission du matériel génétique impliquent le phénomène de **mutation** : qui permet l'apparition de la **nouveauté**, qui peut être suivie :

- de son **élimination** : 2 causes
 - principe de cohérence interne : certaines mutations ne sont pas compatibles avec l'organisme, son développement...(souvent du seul fait d'un dysfonctionnement créé)
 - le nouveau phénotype est pénalisé dans un environnement donné
- de son **maintien**, ou de son **expansion** (effet + : avantage sélectif, survie, reproduction, face à un environnement donné)

La **cohérence interne** de l'organisme muté, et son **insertion** dans son **environnement** sont donc les deux facteurs décisifs qui font qu'une mutation restera isolée et disparaîtra avec l'organisme qui le porte, ou bien contribuera à l'évolution et à la divergence des espèces.

Le plus fréquemment, la mutation est **neutre**, car elle n'a aucune conséquence sur le phénotype.

2- Les mutations expliquent le polymorphisme des individus au sein des espèces

2 types :

- polymorphisme des phénotypes : caractères qui distinguent les individus (résulte de l'interaction GxE, part environnementale importante)
- polymorphisme des génotypes (ex: 2 séquences d'ADN ≠ par 1 nt sur 2 K homologues), souvent n'a aucune conséquence sur le phénotype

> **Polymorphismes et évolution** : valeur adaptative et dérive génétique

Le polymorphisme des individus :

- peut aboutir à des différences **génotypiques** et **phénotypiques** (des caractères observables),
- Ces caractères peuvent être transmis : hérédité du polymorphisme des individus
- peut aboutir à la sélection adaptative : cette pression s'exerce d'abord sur les phénotypes, mais a ensuite une incidence sur le génotype ; elle dépend de la **valeur adaptative** ; (= le caractère est ± transmis)
- provient de ce que l'on appelle la **dérive génétique** : chaque individu ne transmet qu'une partie de son patrimoine génétique (en moyenne 75%), et le nombre d'allèles varie, surtout dans les populations peu nombreuses. Les fréquences d'allèle sont modifiées de façon aléatoire de manière d'autant plus importante que la population est petite
- L'**isolement reproductif** est souvent impliqué dans la naissance d'une espèce ≠

Le polymorphisme est donc la première étape et le premier signe visible d'un processus qui aboutit à l'évolution des espèces, ou **spéciation**.

On parle d'**espèces différentes** lorsque des êtres à l'origine génétique commune **ne peuvent plus se reproduire ensemble**.

3- Grandes lignes d'une classification du vivant

On peut opérer un regroupement par :

a. Similitude : caractères voisins

- procaryotes vs. Eucaryotes (compartimentation intra ϕ)
- unicellulaires/multicellulaires

b. Ascendance commune = phylogénie :

Il s'agit de regroupement par **ascendance commune supposée**.

On peut comparer les eubactéries (bactéries classiques, procaryotes uni ϕ) aux archéobactéries (idem, mais caractéristiques très différentes) qui ont des ascendances communes, mais dont la divergence est très ancienne.

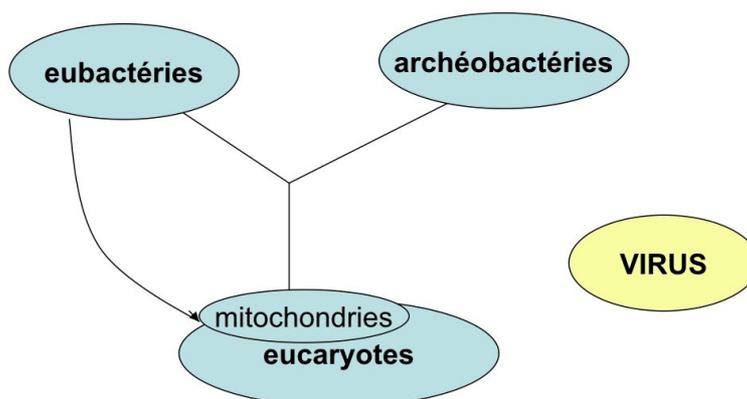
Si on compare leur génome, la distance phylogénétique est équivalente entre eubactéries, archéobactéries et eucaryotes.

Les virus constituent un groupe à part, qui ont besoin de détourner la machinerie génétique des cellules qu'ils infectent pour se reproduire.

Il y a aussi des cas plus complexes, dont un appelé **endosymbiose** :

Certains procaryotes (eubactéries, archéobactéries), qui ont été absorbés (phagocytés) mais n'ont pu être digérés, ont lié des **relations symbiotiques** avec leur cellule **eucaryote** hôte. Le fait de posséder en son sein des organismes capables de lutter contre l'oxygène, surtout il y a 1,5 milliards d'années, quand la teneur atmosphérique en oxygène atteignit son maximum, a probablement sauvé la cellule eucaryote, en plus du fait que la chaîne respiratoire permet la régénération d'ATP. Peu à peu ces organismes se sont intégrés à la cellule (leur ADN s'est combiné à l'ADN de la cellule) et ont formé les organites cellulaires (mitochondries, chloroplastes, et peut être les peroxysomes). L'**origine** vraisemblable de **ces organites** des procaryotes est donc une **bactérie** (comparaison des gènes).

On peut en partie résumer par ce schéma :



III. ORGANISATION DU GÉNOME ET MUTATIONS

A. TAILLE ET ORGANISATION DU GÉNOME

1- Taille comparée de différentes espèces

Génomes de	Nombre de nt	gènes	% ADN codant
E. coli (eubactérie)	$\approx 4,2 \cdot 10^6$	≈ 4.000	≈ 90
Levure (eucaryote uniç)	$1,3 \cdot 10^7$	≈ 7.000	≈ 70
Drosophile*	$1,8 \cdot 10^8$	≈ 20.000	≈ 33
Homo sapiens*	$3,0 \cdot 10^9$	≈ 30.000	≈ 10
Xénope*	$2,0 \cdot 10^{10}$	≈ 25.000	$\approx 1,5$

*taille du génome haploïde.

La grande disparité dans la taille est expliquée par le pourcentage d'ADN codant : ARNm, ARNr, ARNt, ARN de régulation.

Qu'en est-il du lien entre complexité de l'organisme et :

- taille du génome (nombre de nucléotides) : hypothèse assez probante, mais pas systématique : le cas du xénope, dont le génome est 7 fois plus grand que celui de l'humain, dément tout lien simple ;
- nombre de gènes codant pour des protéines : le saut entre drosophile et humain semble faible, encore plus pour le xénope : peu concluant ;
- % de gène codant : (= ADN qui est transcrit en ARNm ou en ARN, machinerie cellulaire) on ne peut conclure ; par contre il y a clairement décorrélation avec la taille du génome.

2- Cas de l'organisation (structure) du génome humain

	Génome humain 3.200 MB	c'est à dire 3.200 millions de bases ; (1) = (2) + (3)	
Gènes (1/3) 1.200 MB			ADN intergénique 2/3 2.000 MB
		Séquences répétées (3/4) 1.400 MB	Autres (1/4) 600 MB

L'ADN codant ne représente qu'un peu plus du tiers du génome humain.

L'ADN répété non codant représente environ 40% du génome, et 70% de l'ADN intergénique

3- ADN répété non codant

Il est composé de différents types de répétitions :

a. Séquences répétées en tandem :

Ce sont des séquences identiques, en plusieurs exemplaires les unes à côté des autres au sein du génome :

type	motif nt	nbre. répét.	localisation	commentaires
DNA satellite α	171	$8 \cdot 10^5$	centromères	jouerait un rôle dans la division ϕ dans la mitose ou la méiose
minisatellites télomériques	6	10^4	télomères	Maintien de la structure du K
microsatellites	1-5	$5 \cdot 10^5$	tous les chromosomes sauf télom. & centrom.	y compris dans les introns des gènes

Minisatellites télomériques : du type (TTAGGG) $_n$

Ils jouent un rôle dans la stabilité des chromosomes, en luttant contre le raccourcissement du génome (permettent une réplication jusqu'à l'extrémité).

Microsatellites : (GT) $_n$ ou (ATT) $_n$

Ils sont caractérisés par un **très grand polymorphisme** de ces séquences : le nombre de répétitions **n** est **variable** et distingue les allèles. Ce motif répété peut être trouvé dans de nombreux locus, mais constitue avec les séquences adjacentes un **marqueur génétique** d'un locus particulier.

b. Séquences répétées dispersées (réparties en \neq endroits du génome) :

type	nt	répét.	localisation	commentaires
LINE	6.000	10^6	tous les chromosomes	également dans les introns (entre les gènes & à l'intérieur des gènes)
ALU	300	10^6		

Varie de façon importante en fonction des espèces : dans le génome humain, un grand nombre de séquences ALU à des endroits où il n'y en a pas chez le chimpanzé, qui est pourtant proche, dans les espaces intergéniques, mais aussi à l'intérieur des gènes.

N.B. : bien qu'elles soient en quasi-totalité non codantes, les séquences LINE et ALU sont impliquées dans la **rétrotransposition** ; il s'agit d'exceptions.

4- ADN répété codant

Il peut coder pour :

- Des **ARN**
 - ADN codant pour les ARNs ribosomiques (10.000-50.000 gènes),
 - ADN codant pour les ARNt,
- Des **protéines**
 - **familles de gènes**, par exemple, les gènes de **globines**, des **histones**, les récepteurs olfactifs ; elles sont constituées :
 - o d'un nombre de quelques **gènes** (globines) à quelques **centaines** de gènes (histones),
 - o dont la séquence est proche.

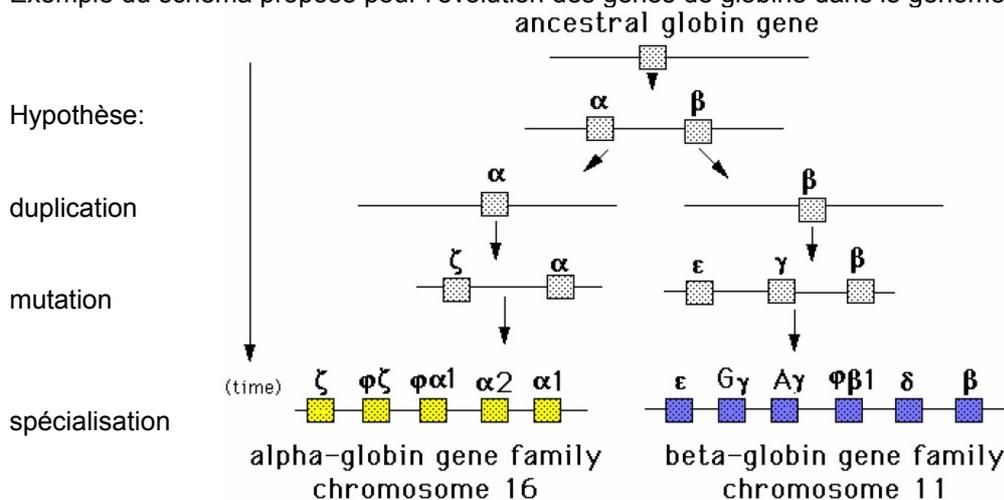
L'ADN répété codant est l'une des manifestations et le siège de mécanismes de diversité et d'évolution les plus importants.

B. EXEMPLES DE MÉCANISMES À L'ORIGINE DE LA DIVERSITÉ GÉNÉTIQUE :

1- Origines des séquences d'ARN répété codant

Chaque famille de gène dérive d'un gène unique (gain et perte de gènes, évolution par duplication). A la suite de sa **duplication**, il subit différents gains ou pertes de gènes, et des changements de séquences. L'évolution séparée par **mutation** amène une **différenciation** de chaque copie qui peut correspondre à une **spécialisation**.

Exemple du schéma proposé pour l'évolution des gènes de globine dans le génome humain :



On connaît actuellement 2 sous-familles **α** et **β** (séparation chromosomique à la 2^{ème} étape), chacune ayant évolué jusqu'à la situation actuelle, sur les chromosomes 16 (**α**) et 11 (**β**). Chacune a des rôles différents, il y a eu une spécialisation des 2 familles, **α** et **β**, et les deux sont indispensables pour faire une Hb fonctionnelle.

La **famille β** possède 2 globines particulières dont l'expression est différente au cours du développement : **ε** **spécifique de l'embryon**, et **γ** conservée jusque dans la **petite enfance** (15 mois environ).

L'affinité de chacune de ces globines pour l'oxygène n'est pas la même. Chez les mammifères, il y a un nombre variable de gènes de globines.

2- Mécanismes de duplication des séquences

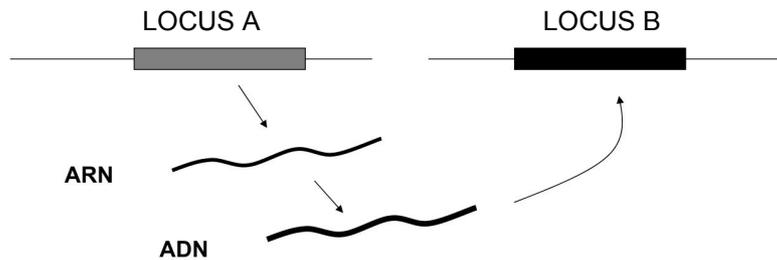
Ils sont au nombre de 3 :

- Duplication de l'ensemble du génome,
- Transposition (cas le plus fréquent: rétrotransposon)
- Recombinaison inégale.

a. Duplication suivie d'une perte partielle

La duplication se produit plusieurs fois, toujours suivie de la perte d'une partie du patrimoine dupliqué. Ce mécanisme expliquerait l'augmentation de la taille du génome, accompagnée de répétitions en proportion croissante.

b. Rétrotransposition (rétrotransposon)



C'est le type de transposition le plus rencontré dans le génome humain.

Un élément mobile est présent dans le locus A, pas le locus B. Un locus est normalement transcrit en ARN, mais ensuite, une **rétrotranscriptase inverse**, à partir de cet ARN simple brin, produit de l'ADN simple brin, puis il se duplique et devient double brin, et s'intègre à l'ADN, mais à un autre endroit du génome.

Résultat : une **2ème copie du locus, à un endroit différent** (locus B); on appelle cette copie un **rétrotransposon**.

La comparaison de génomes de mammifères en donne des preuves, comme celle du chimpanzé et de l'humain, qui en possède un nombre beaucoup plus grand que le chimpanzé, mais aussi certaines maladies génétiques, comme dans l'**hémophilie**, un gène devient dysfonctionnel à cause d'un rétrotransposon intercalé dans une séquence codante.

> Les rétrotransposons les plus abondants :

Sur le million de séquences **LINE** et le million de séquences **ALU** présentes dans le génome humain, seul un **très petit nombre sont actives**, les autres étant non codantes : on peut les compter sur les doigts de la main.

Certaines des séquences **LINE** (capable de coder pour une transcriptase inverse), qui font 6.000 nt lorsqu'elles sont complètes, possèdent :

- une séquence **polyA**,
- une séquence qui ressemble à des gènes que l'on trouve dans des rétrovirus, codant pour une **transcriptase inverse**.

Les séquences ALU n'ont pas cette capacité. On suppose donc que les séquences ALU actives utilisent une transcriptase inverse codée par un autre gène pour se rétrotransposer (ex: LINE).

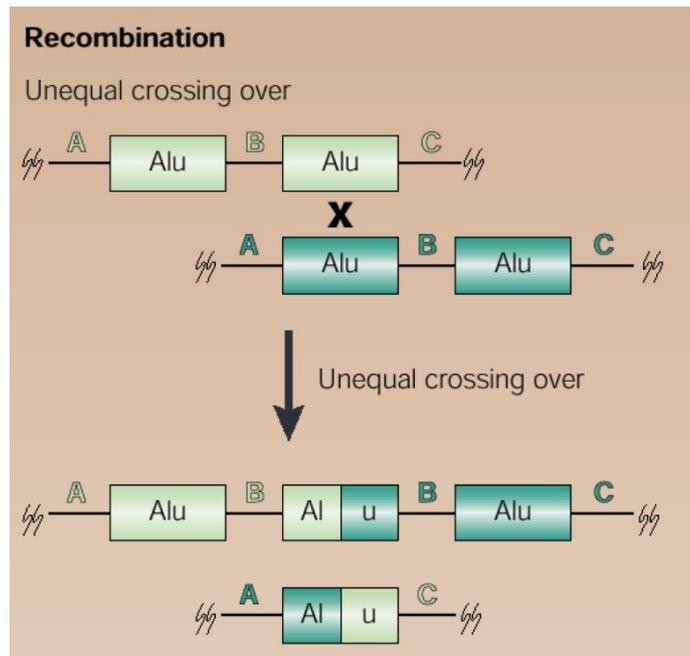
Ce mécanisme de rétrotransposition aboutit à une **modification du nombre de copies du gène** concerné, mais génère aussi des anomalies génétiques, par exemple si le rétrotransposon s'insère à l'intérieur d'un gène qu'elle rend non fonctionnel.

c. Recombinaison illégitime

Rq : « illégitime » car non-méiotique

La fréquence des recombinaisons illégitimes est une conséquence de la **fréquence de séquences répétées**, d'autant plus si leur proximité est grande.

La recombinaison se produit en général à la suite de l'appariement de segments de brins complémentaires, c'est à dire présentant des **homologies de séquences importantes**, mais dont les séquences complètes ne sont malgré tout **pas homologues**, donc pas en position homologue d'une même paire de chromosomes.



L'exemple donné présente 2 séquences ALU (appariement décalé), mais il pourrait s'agir d'autres séquences, y compris codante. Ici les séquences codantes sont A, B et C.

L'exemple montre les 2 types de conséquences possibles sur les séquences issues de la recombinaison illégitime :

- la 1ère produit une insertion du gène B, qui est donc en double = **duplication** sur le K1
- la deuxième produit une **délétion** du même gène B sur le K2

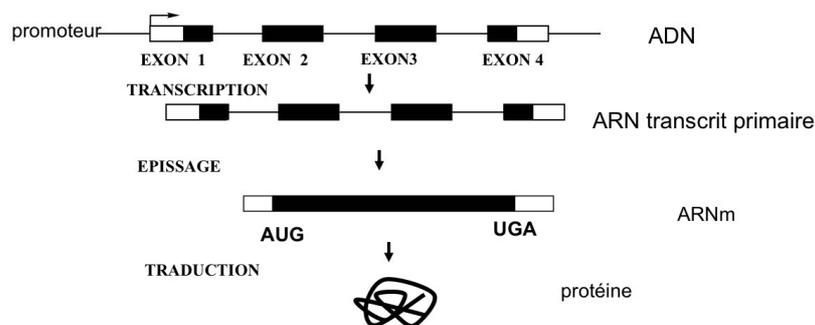
Comme pour la rétrotransposition, on trouve des exemples de ce phénomène à la fois par comparaison du génome de mammifères aux origines phylogénétiques identiques, et dans certaines maladies génétiques.

IV. ORGANISATION DES GÈNES EUCARYOTES

1- Structure du gène - rappels

Un gène (d'un eucaryote supérieur) est constitué de :

- un promoteur,
- plusieurs **exons*** (présents au niveau de l'**ARNm** : traduit),
- séparés par des introns



*exon / exon = partie de l'ADN qui sort du noyau (ARNm mature) pour être traduit en général, mais pas toujours (codon-stop...) => !!! exon ≠ codant.

Le gène est transcrit à partir du promoteur (qui permet l'assemblage des protéines : complexe de transcription), ce qui forme un transcrit primaire, qui est ensuite modifié, notamment par l'**épissage qui élimine les introns** et produit l'ARNm, qui porte une phase de lecture ouverte, qui est traduite (figurée en noir), et des extrémités non traduites (figurées en blanc : avant AUG=signal de début de traduction et après UGA=signal de fin de traduction).

Les exons comportent de **150 à 200 nt** en moyenne.

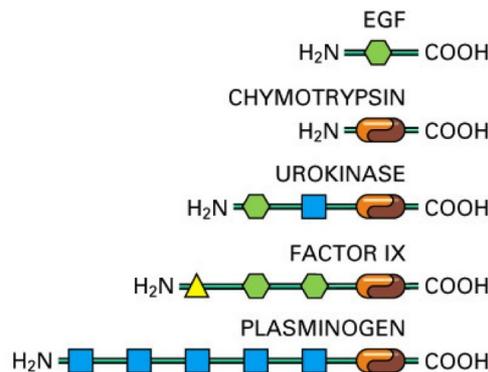
Nombre d'exons par gène = très variable : de 1 (exceptions) à plusieurs centaines : certains gènes sont très morcelés ; le plus souvent : entre **10 et 20 exons** pour la majorité des gènes.

Les introns : de 60 à 10⁶ nt, de **taille très variable** selon les espèces et les gènes au sein d'une espèce ; leur contenu est également très variable, y compris de nombreuses séquences répétées, comme vu précédemment.

2- Structure modulaire et combinatoire des protéines

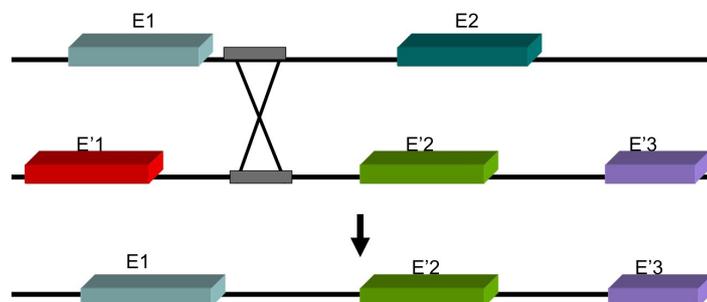
Un grand nombre de protéines ont des fonctions et séquences globales très différentes (conformation spatiale propre, fonction particulière), mais sont composées de certaines parties identiques:

- soit des modules différents
- soit des modules répétés



Considérations sur la structure modulaire des gènes : dans le génome humain, 1/3 des gènes codent pour plusieurs protéines. On peut expliquer cette « combinatoire » des protéines par différents mécanismes dont :

a. Recombinaison non allélique du génome



On peut défendre l'hypothèse qu'au cours de l'évolution, il y a eu recombinaison **illégitime** entre **exons**, entre 2 molécules d'ADN, entre gènes différents

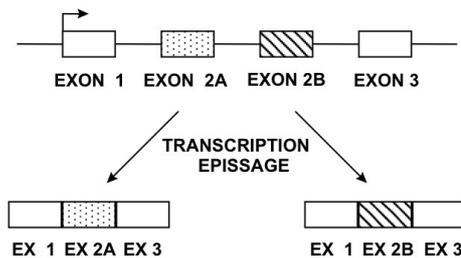
Dans l'exemple, il y a recombinaison illégitime entre introns entre les gènes A et B. Un produit possible de cette recombinaison est le gène C qui combine 1 exon du gène A et 2 exons du gène B. Résultat => nouvelle combinatoire d'exons.

Ceci peut expliquer la structure modulaire et combinatoire des protéines : 1 ou plusieurs exons => 1 module particulier qui se trouve dans plusieurs protéines.

Si les recombinaisons peuvent avoir un effet pathologiques, elles sont également à l'origine de la diversité des protéines et enzymes, en expliquant la présence de séquences identiques mais à des locus différents, ce qui produit finalement des protéines « modulaires », aux fonctions très différentes, à partir d'un matériau génétique d'origine plus réduit.

b. La combinatoire des exons

i. Épissage alternatif :



Dans un exemple simple, 1 gène de 4 exons est transcrit en ARNm primaire.

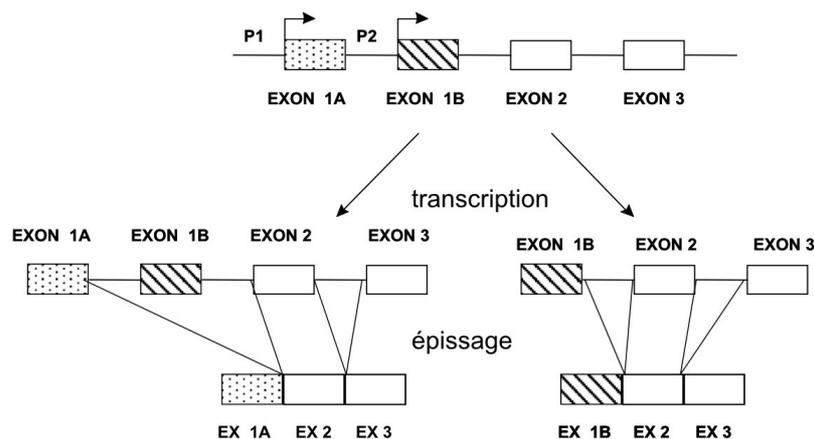
Deux épissages (alternatifs) produisent des ARNm matures différents, dont la traduction peut donner naissance à des protéines différentes.

Éléments qui déterminent le choix d'un épissage :

- il est général déterminé par le type cellu dans lequel s'exprime le gène considéré : les facteurs d'épissage sont différents d'un type à l'autre ; ils sont donc fonction de la différenciation cellulaire ;
- parfois des épissages différents sont possibles dans le même type cellulaire.

ii. Promoteurs alternatifs :

L'utilisation de promoteurs alternatifs permet à 1 seul gène de donner plusieurs transcrits primaires : la transcription peut commencer à des endroits différents s'il y a plusieurs promoteurs. C'est un autre mécanisme à l'origine de la combinatoire dans l'expression génétique :



Il peut exister, au sein d'un gène, plusieurs promoteurs (ici, 2 : P1 et P2) : la transcription peut commencer par l'un ou par l'autre , ce qui produit 2 ARN transcrits primaires différents.

Le choix du promoteur peut dépendre de la **différenciation cellulaire** mais il est possible que les 2 promoteurs soient utilisés dans le même type cellulaire.

iii. Combinaison des deux :

L'**épissage alternatif** peut suivre l'utilisation de promoteurs différents d'un même gène.

À partir des transcrits primaires différents résultant de la transcription à partir de promoteurs différents, l'épissage alternatif ajoute encore des combinaisons supplémentaires, aboutissant à plusieurs ARNm différents à partir d'un même gène, et donc à plusieurs protéines différentes.

Dans l'exemple ci-dessus, on note que, dans l'épissage du transcrit de gauche, l'exon 1B n'est pas reconnu comme un exon.

3- Fonctions des introns

a. micro-ARN :

Certains introns sont source de petits **ARN non codants, mais régulateurs** de l'expression des gènes, que ce soit de la transcription ou de la traduction. Ils influent sur :

- la stabilité des ARNm
- leur traduction.

b. Régulation de la transcription :

Dans l'ADN, il y a au niveau des introns des domaines qui régulent **la transcription** : ils ont un rôle de fixation de protéines nucléaires, qui rentrent alors en interaction avec les facteurs de transcription, ayant potentiellement un effet d'activation ou d'inhibition.

c. Épissage alternatif :

Ils peuvent générer la **diversité de production de protéines** différentes à partir d'un même gène par épissage alternatif (traité ci-dessus IV.2.b.i.).

d. Au cours de l'évolution :

Ils permettent le **partage de domaines** entre \neq gènes par **recombinaison** au niveau des mlc d'ADN (voir III.B.2.c.).

e. ADN égoïste :

On désigne sous ce vocable beaucoup de séquences introniques qui n'ont **pas de fonction connue**.