

Université Frères MENTOURI Constantine 1

Faculté des Sciences de la Nature et de la Vie
Département de Biologie Appliquée

Master en : Bioinformatique

Cours de : **Génomique Fonctionnelle**

**INTRODUCTION À LA
GÉNOMIQUE FONCTIONNELLE**

Notion de Génétique Inverse

La génétique : est, la Science de l'hérédité.

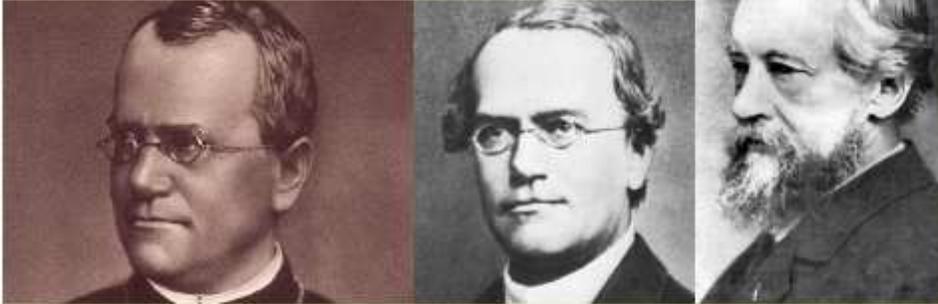
La génétique s'intéresse aux caractères transmissibles de génération en génération.

Le génome : est l'ensemble des gènes portés par les chromosomes.

La génomique : est donc la Science qui étudie l'ensemble des gènes.

Notion de Génétique Inverse

Rappels



1822 - 1884 au république Tchèque.
Il a publié son livre en 1865

Le père de la génétique (pour avoir décrit le mode d'action et d'héritage des gènes parentaux).

Les gènes déterminent les traits, (les caractères visibles que nous voyons autour de nous).

Il existe de nombreuses exceptions aux lois de Mendel :

Notion de Génétique Inverse

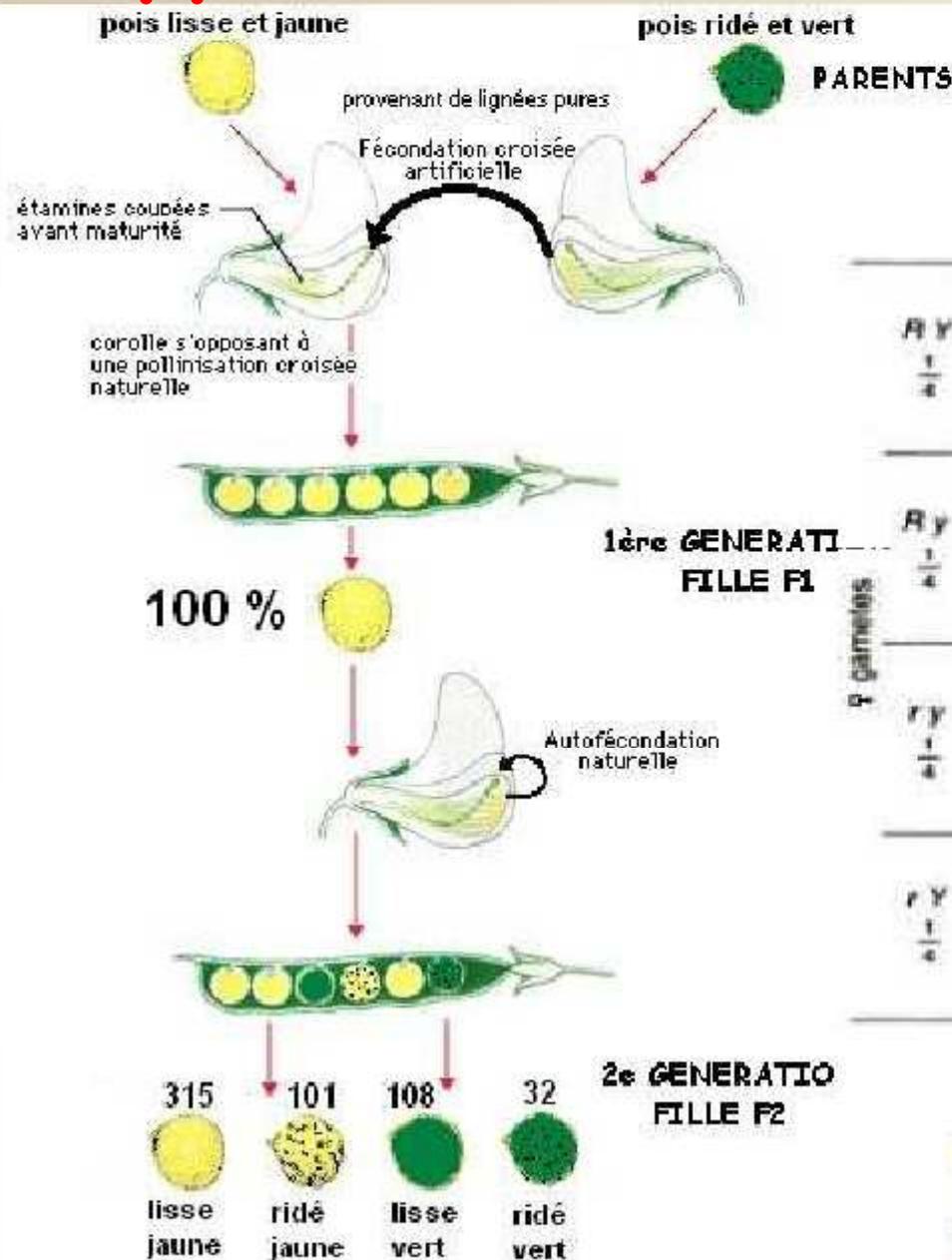
Rappels

le crossing-over (la recombinaison par enjambement),
le pluri-allélisme, la codominance, la dominance incomplète,
l'épistasie, la complémentation, la pléiotropie et les gènes
létaux.

Différents types de transformations génomiques (délétion,
duplication, translocation et inversion qui sont souvent le
résultat du "linkage").

Notion de Génétique Inverse

Rappels



♂ gametes

	RY $\frac{1}{4}$	Ry $\frac{1}{4}$	ry $\frac{1}{4}$	rY $\frac{1}{4}$
RY $\frac{1}{4}$	$RRYY$ $\frac{1}{16}$ 	$RRYy$ $\frac{1}{16}$ 	$RrYy$ $\frac{1}{16}$ 	$RrYY$ $\frac{1}{16}$
Ry $\frac{1}{4}$	$RRYy$ $\frac{1}{16}$ 	$RRyy$ $\frac{1}{16}$ 	$Rryy$ $\frac{1}{16}$ 	$RrYy$ $\frac{1}{16}$
ry $\frac{1}{4}$	$RrYy$ $\frac{1}{16}$ 	$Rryy$ $\frac{1}{16}$ 	$rryy$ $\frac{1}{16}$ 	$rrYy$ $\frac{1}{16}$
rY $\frac{1}{4}$	$RrYY$ $\frac{1}{16}$ 	$RrYy$ $\frac{1}{16}$ 	$rrYy$ $\frac{1}{16}$ 	$rrYY$ $\frac{1}{16}$

9 : 3 : 3 : 1

Round, yellow

Winkled, yellow

Round, green

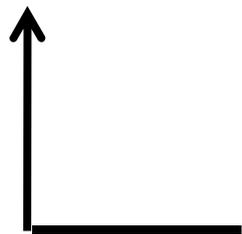
Winkled, green

Notion de Génétique Inverse



Identification des différences de séquences d'ADN entre variants génétiques distincts.

Identification du ou des gènes



Identification des différences moléculaires et développementales entre des individus de génotypes distincts

Notion de Génétique Inverse

Avec le développement de notre connaissance de l'ADN et de la façon dont il code l'information de la séquence d'acides aminés des protéines ainsi que le contrôle cellulaire de la production de celles-ci, une autre forme d'étude génétique devient possible :

L'Etude basée sur l'ADN : LA GÉNÉTIQUE INVERSE

Notion de Génétique Inverse

L'Etude basée sur l'ADN : LA GÉNÉTIQUE INVERSE

Au contraire, le programme de génétique inverse débute par des changements génétique connus et par la recherche des modifications qui en résultent dans l'organisme.

Notion de Génétique Inverse

L'Etude basée sur l'ADN : LA GÉNÉTIQUE INVERSE

Une analyse par génétique inverse est basée sur l'utilisation d'une séquence normale d'ADN. En insérant une mutation dans l'ADN (ou en le comparant avec l'ADN d'autres génomes), on peut analyser la fonction de la séquence d'ADN.

Notion de Génétique Inverse

L'Etude basée sur l'ADN : LA GÉNÉTIQUE INVERSE

À l'aide d'une séquence normale d'ADN, on peut lire la séquence d'acides aminés de la protéine synthétisée (correspondance code d'ADN/ aa codés).

L'ADN peut ensuite être modifié de façon spécifique,

Aboutit à un blocage total de la production de la protéine ou à un changement de la protéine,

(modification ou suppression de son activité métabolique).

Notion de Génétique Inverse

L'Etude basée sur l'ADN : LA GÉNÉTIQUE INVERSE

Les conséquences d'une telle mutation complète (knockout) sur le développement du phénotype peuvent ensuite être suivies de la même façon que les effets mutationnels dans le programme de génétique directe.

Notion de Génétique Inverse

L'Etude basée sur l'ADN : LA GÉNÉTIQUE INVERSE

L'un des avantages du programme inverse est qu'il permet la création sur mesure d'un grand nombre de mutations spécifiques.

Au lieu de cela, on utilise l'information déjà disponible sur les divergences des génomes de différents organismes modèles .

Notion de Génétique Inverse

L'Etude basée sur l'ADN : LA GÉNÉTIQUE INVERSE

Sachant que les génomes des deux espèces ont été séquencés et que l'on a observé une très faible différence entre les deux.

De nombreuses méthodologies différentes sont utilisées pour étudier les gènes et leurs activités :

Notion de Génétique Inverse

L'Etude basée sur l'ADN : LA GÉNÉTIQUE INVERSE

- * L'isolement des mutations affectant le processus biologique en cours d'étude ;
- * L'analyse des descendants d'unions contrôlées (croisements) entre des mutants et des individus de type sauvage ou d'autres variants discontinus ;
- * L'analyse génétique des processus biochimiques de la cellule ;
- * L'analyse microscopique ;
- * L'analyse directe de l'ADN (Clonage).

Notion de Génétique Inverse

Les génomes complets de nombreux organismes ont été séquencés grâce aux applications des techniques citées ci-dessus, ce qui a donné lieu à une nouvelle discipline au sein de la génétique, appelée **GÉNOMIQUE**.

Il s'agit de l'étude de **la structure**, de **la fonction** et de l'évolution des génomes considérés dans leur intégralité.

La génomique comprend également **la bioinformatique**, qui est l'analyse mathématique des informations contenues dans les génomes.

Génomique

La **génomique** se définit comme l'étude exhaustive des gènes :

localisation au sein du génome (établissement de cartes génétiques et physiques),

Séquençage,

identification de leur fonction biologique (en particulier par le biais de l'étude physiologique de mutants),

et variabilité au sein des individus d'une même espèce.

Génomique

1. La génomique structurale

La **génomique structurale** se concentre sur l'organisation physique du génome sous forme de chromosomes portant des séquences de types divers, parmi lesquelles les gènes constituent la cible principale des recherches.

2. La génomique fonctionnelle

La **génomique fonctionnelle** vise à caractériser l'expression du génome et son intégration dans l'élaboration des grandes fonctions métaboliques.

1. La génomique structurale

- la cartographie physique du génome
- le séquençage systématique du génome

2. La génomique fonctionnelle

- la mutagenèse insertionnelle
- le séquençage partiel de gènes exprimés
(Expressed Sequence Tags ou EST)
- * la transformation génétique
- * l'évaluation fine du matériel transformé

2. La génomique fonctionnelle

- la mutagénèse insertionnelle

Permet de créer des mutants en insérant au hasard, dans le génome, un segment d'ADN repérable.

Lorsque celui-ci s'intègre dans un gène, il en altère la fonction et provoque la modification du caractère correspondant donc une mutation.

2. La génomique fonctionnelle

- la mutagénèse insertionnelle

Le gène muté est localisé grâce à l'élément inséré, et sa fonction est identifiée grâce au caractère affecté.

Cette technique est développée sur les différents Organismes modèles (*animaux ou végétaux*).

2. La génomique fonctionnelle

- le séquençage partiel de gènes exprimés

Permet la caractérisation de "patrons" d'expression du génome.

L'analyse des ARN extraits sur certains organes de l'organisme étudié, dans des conditions données et à un moment donné, permet d'obtenir une image de l'ensemble des gènes qui s'expriment dans ces conditions.

2. La génomique fonctionnelle

- le séquençage partiel de gènes exprimés

En fonction de caractéristiques de séquence ou de patrons d'expression, on pourra attribuer à certains de ces EST une fonction hypothétique qui en fera des "gènes candidats".

Ceci est actuellement développé sur les organismes modèles et commence à s'étendre aux principales espèces d'intérêt.

2. La génomique fonctionnelle

- le séquençage partiel de gènes exprimés

En fonction de caractéristiques de séquence ou de patrons d'expression, on pourra attribuer à certains de ces EST une fonction hypothétique qui en fera des "gènes candidats".

L'analyse des séquences d'ADN et de la structure des génomes **ne permet pas d'associer directement une fonction à un gène.**

2. La génomique fonctionnelle

- le séquençage partiel de gènes exprimés

En d'autres termes, on ne peut pas inférer ce gène à :

=> Un ARN qui en est le produit d'expression,

=> Dans le cas d'un ARN messenger, à la fonction de la (ou des) protéine(s) pour laquelle (lesquelles) cet ARN messenger code.

Par ailleurs, tous les gènes ne s'expriment pas simultanément, ni au même taux.

2. La génomique fonctionnelle

- le séquençage partiel de gènes exprimés

=> Chez les Procaryotes, le niveau de transcription des gènes est contrôlé par la phase du cycle de division cellulaire et par l'incidence de l'environnement sur la cellule.

=> Chez les Eucaryotes, chaque type de cellule est caractérisé par la transcription de certains gènes qui lui sont propres et qui lui confèrent ses propriétés biologiques.

2. La génomique fonctionnelle

- le séquençage partiel de gènes exprimés

⇒ La **compartimentation** de la cellule Eucaryote est un élément capital en génomique :

Les produits de certains gènes, les protéines, sont adressés spécifiquement à tel ou tel compartiment.

Dans le cas d'un ARN messenger, déterminer quand et comment (régulation de la traduction) celui-ci est traduit en protéine est un point capital pour associer un gène à une **fonction biologique**. => C'est la finalité de la **protéomique**.

2. La génomique fonctionnelle

- le séquençage partiel de gènes exprimés

La caractérisation des fonctions précises des gènes candidats, identifiés grâce aux approches de la génomique, fait intervenir plusieurs étapes :

* la transformation génétique

Permet de tester l'effet d'un gène ou d'une séquence candidate sur le phénotype de l'organisme.

2. La génomique fonctionnelle

- le séquençage partiel de gènes exprimés

* la transformation génétique

Grâce à son insertion au sein du génome et/ou à la modification de son expression dans l'organisme ;

* l'évaluation fine du matériel transformé est nécessaire pour apprécier toutes les facettes de la fonction d'un gène.

2. La génomique fonctionnelle

- le séquençage partiel de gènes exprimés

* l'évaluation fine du matériel transformé

Car l'analyse des séquences d'ADN et de la structure des génomes ne permet pas d'associer directement une fonction à un gène.

2. La génomique fonctionnelle

- le séquençage partiel de gènes exprimés

* l'évaluation fine du matériel transformé

Elle met en jeu des outils de biologie cellulaire et moléculaire et de physiologie.

Ces technologies conduisent à : développer des ponts entre les génomes.

La conservation au fil de l'évolution de certains éléments de l'organisation générale des génomes, (conservation de la "synténie"),

2. La génomique fonctionnelle

- le séquençage partiel de gènes exprimés

* l'évaluation fine du matériel transformé

La colinéarité entre les chromosomes homologues, permet des transferts d'information entre plusieurs espèces.

C'est toute une dynamique collective que la génomique produit, rapprochant les équipes auparavant spécialisées sur différents organismes.

Objectifs de la *Génomique fonctionnelle*

En conséquence, afin de déterminer la fonction des ARN et des protéines associés à ces gènes, la génomique fonctionnelle analyse :

⇒ le **transcriptome** : les produits de la transcription des gènes, les ARN ;

⇒ le **protéome** : l'ensemble des protéines synthétisées (traduites) à partir des ARN messagers.

Objectifs de la *Génomique fonctionnelle*

La génomique vise à caractériser la fonction des milliers de gènes et à comprendre, grâce à des approches globales (**transcriptome** : étude globale de l'expression du génome ; **protéome** : complément **PROTE**ique du gén**OME** ; **métabolome** : complément en **MÉTABOL**ites du gén**OME**), les réseaux d'interaction entre eux en réponse à divers stress abiotiques ou biotiques.

Objectifs de la Génomique fonctionnelle

La génomique fonctionnelle a pour principaux buts de déterminer :

- le moment dans le cycle cellulaire où un gène est transcrit (appelé expression d'un gène) ;
- les conditions environnementales liées à la transcription ou la non-transcription d'un gène ;
- l'intensité (nombre de copies du ou des transcrits) avec laquelle ce gène est transcrit ;

Objectifs de la Génomique fonctionnelle

- le compartiment cellulaire où un gène est transcrit (ADN nucléaire, mitochondrial, chloroplastique) ;
- le compartiment cellulaire où est adressé le (ou les) produit(s) de la transcription d'un gène ;
- les interactions que le produit d'un gène peut établir avec d'autres produits de gènes et/ou d'autres types de molécules (*interactomique*). Ce type d'analyse débouche sur la construction de réseaux d'interactions ("*interaction networks*") ;

Objectifs de la *Génomique fonctionnelle*

- si le produit est un ARN, le rôle qu'il peut avoir dans la régulation post-transcriptionnelle (interférence ARN) ;
- si le produit intermédiaire est un ARN messenger et le produit final est donc une protéine, le rôle que cette protéine peut avoir dans une voie métabolique et/ou dans la régulation de cette voie métabolique.

...

Exemple de la *Génomique Végétale*

a pour but d'identifier des gènes d'importance agronomique majeure détectés chez diverses espèces (Cibles importantes dans les génomes des plantes cultivées), travaillent sur quatre axes de recherche :

- l'analyse structurale du génome des plantes,
- la recherche des gènes impliqués dans la résistance aux maladies,

Exemple de la Génomique Végétale

- les caractéristiques agronomiques (e.g., tolérance au stress hydrique, composants moléculaires du rendement), et
- les caractéristiques de qualité du produit fini (e.g., gènes du blé qui influent positivement sur la qualité du pain).

Vers l'élaboration de nouvelles ressources biologiques et moléculaires et d'outils de génomique.

Champs d'application de la *Génomique Fonctionnelle*

La *génomique fonctionnelle* permet :

- d'identifier les éléments constitutifs d'un gène (introns, exons, séquences de régulation de la transcription, ...);
- d'identifier les régions des génomes dont on ignore encore le rôle et élucider ce rôle ;
- d'étudier les différences de transcription des gènes dans le temps et pour chaque type de tissus et de cellules ;

Champs d'application de la *Génomique Fonctionnelle*

- d'étudier les différences d'activité biologique des produits des gènes dans le temps et pour chaque type de tissus, de cellules, de compartiments sub-cellulaires ;
- d'apporter des éléments qui contribuent à déterminer la fonction des ARN et des protéines pour lesquelles les gènes codent ;

Champs d'application de la *Génomique Fonctionnelle*

- d'intégrer toutes ces informations dans un ensemble plus vaste, celui du métabolisme (**métabolome**) en décrivant les interactions entre tous ces types de macromolécules biologiques (**interactome**);
- d'obtenir ces données pour le plus grand nombre d'organismes possibles. Aller à "**GOLD : Genomes OnLine Database**".

Champs d'application de la Génomique Fonctionnelle

La bioinformatique est une science nouvelle, se fondant sur le séquençage massif, récent, du génome de plusieurs organismes procaryotes et eucaryotes.

Sa mission est double :

1/ Annoter le génome des organismes vivants.

2/ Mettre en place des bases de données regroupant des informations structurales sur les gènes et les génomes et leur expression dans des contextes physiologiques.

Champs d'application de la Génomique Fonctionnelle

1/ Annoter le génome des organismes vivants.

L'un des enjeux majeur est de comprendre la structure des gènes (détermination des séquences promotrices gouvernant l'expression et la régulation des gènes, des séquences codantes dictant la synthèse des protéines impliquées dans l'activité cellulaire).

Champs d'application de la *Génomique Fonctionnelle*

1/ Annoter le génome des organismes vivants.

Un autre enjeu de cette annotation est de comprendre la fonction des gènes.

Il ressort en effet que près de 50 % des gènes prédits par le séquençage du génome d'*Arabidopsis* ont une fonction inconnue.

Champs d'application de la Génomique Fonctionnelle

2/ Mettre en place des bases de données regroupant des informations structurales sur les gènes et les génomes et leur expression dans des contextes physiologiques.

Ces bases de données constituent l'un des outils majeurs, via la génomique comparative, pour repérer des gènes jouant un rôle dans le développement ou en réponse aux stress biotiques et abiotiques.

Impact des nouvelles techniques de séquençage en masse

Les nouvelles techniques de séquençage en masse (ou massivement parallèles ou à très haut débit) ont encore élargi le champs d'investigation de la génomique fonctionnelle. On peut citer :

- le séquençage de *novo* ou le reséquençage d'un génome connu ;
- l'étude de la variabilité génétique et du polymorphisme de nucléotide simple (*SNP*) ;
- le séquençage d'haplotypes particuliers lors du clonage positionnel d'un gène d'intérêt ;

Impact des nouvelles techniques de séquençage en masse

- l'étude de plus en plus fine du transcriptome :

1. identification de transcrits rares, étude des phénomènes d'épissage alternatif, identification des séquences frontières intron/exon, analyse quantitative du niveau de transcription des gènes ;

2. étude du profil en petits ARN non codants ("*small ncRNAs*"), découverte de gènes codant ces types d'ARN ;

- l'étude des interactions ADN / protéines (régulation de la transcription, facteurs de transcription, ...) ;

- la génomique médicale ;

Impact des nouvelles techniques de séquençage en masse

- la génomique comparative qui compare la structure et les fonctions des génomes de différentes espèces (organisation et évolution des génomes) ;
- la métagénomique : étude du génome d'un organisme prélevé directement dans un environnement complexe (intestin, océan, sols, ...), à l'inverse d'un organisme de laboratoire. Le but est d'obtenir des informations sur l'incidence de cet environnement. Le préfixe "méta" signifie "après, au-delà de, avec, ...". Aller à : "*Metagenomics at EBI*".

Impact des nouvelles techniques de séquençage en masse

- l'épigénétique et l'épigénomique : étude de l'influence de l'environnement et de l'histoire individuelle sur les modifications de l'expression des gènes d'une génération à l'autre. Le préfixe "épi" signifie "sur, au-dessus, ..." ;
- l'étude du profil de méthylation (processus épigénétique).

Conclusion

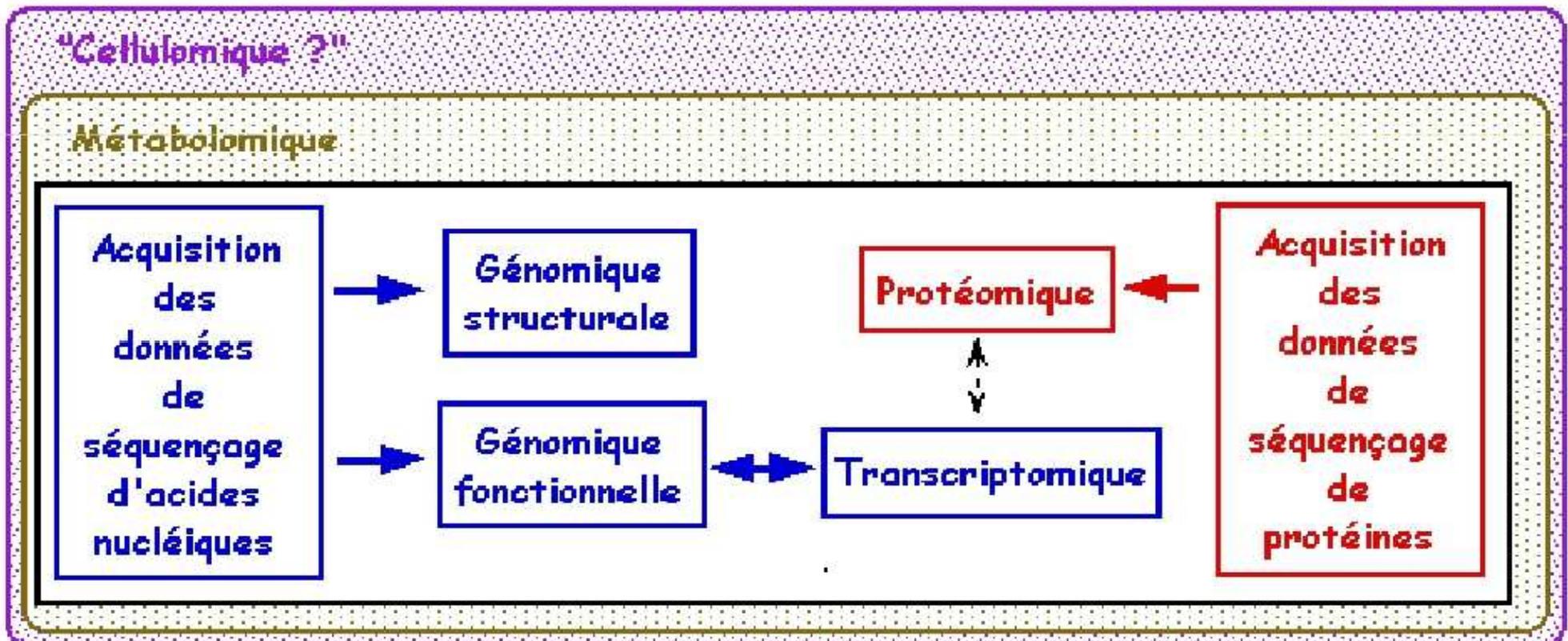
Le "matériau" de base de la génomique est l'ensemble des séquences d'acides nucléiques et des séquences polypeptidiques obtenues par différentes méthodes de séquençage.

Ces séquences et d'autres types d'informations qui découlent de leur analyse sont stockées dans des bases de données. L'accès aux bases de données s'effectue via le Web et Internet.

L'analyse de l'ensemble de ces données nécessitent des méthodes bioinformatiques.

Conclusion

La génomique structurale, la génomique fonctionnelle, la transcriptomique, la protéomique et la bioinformatique sont **des approches complémentaires.**



Université Frères MENTOURI Constantine 1

Faculté des Sciences de la Nature et de la Vie
Département de Biologie Appliquée

Master en : Bioinformatique

Cours de : *Génomique Fonctionnelle*

Génomique fonctionnelle et
Recherche de fonction

Comment trouver un gène ?

Les travaux de génétique directe ont toujours pour but le clonage du gène sauvage correspondant.

La recherche d'un gène se fait en plusieurs étapes :

Stratégie 1 (peu utilisée)

- recherche d'un gène chez une espèce par **synténie**

Stratégie 2 (la plus courante)

- attribution du gène à un chromosome
- cartographie précise du gène sur le chromosome
- clonage du gène

La tâche est plus ou moins difficile en fonction de l'espèce étudiée :

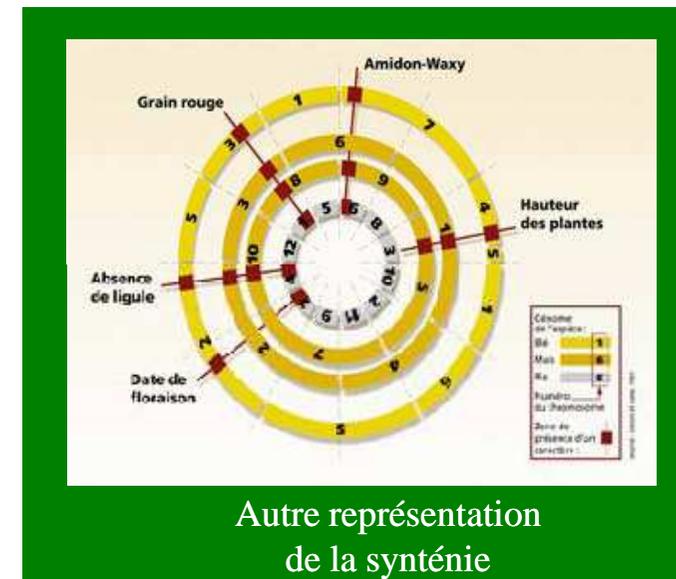
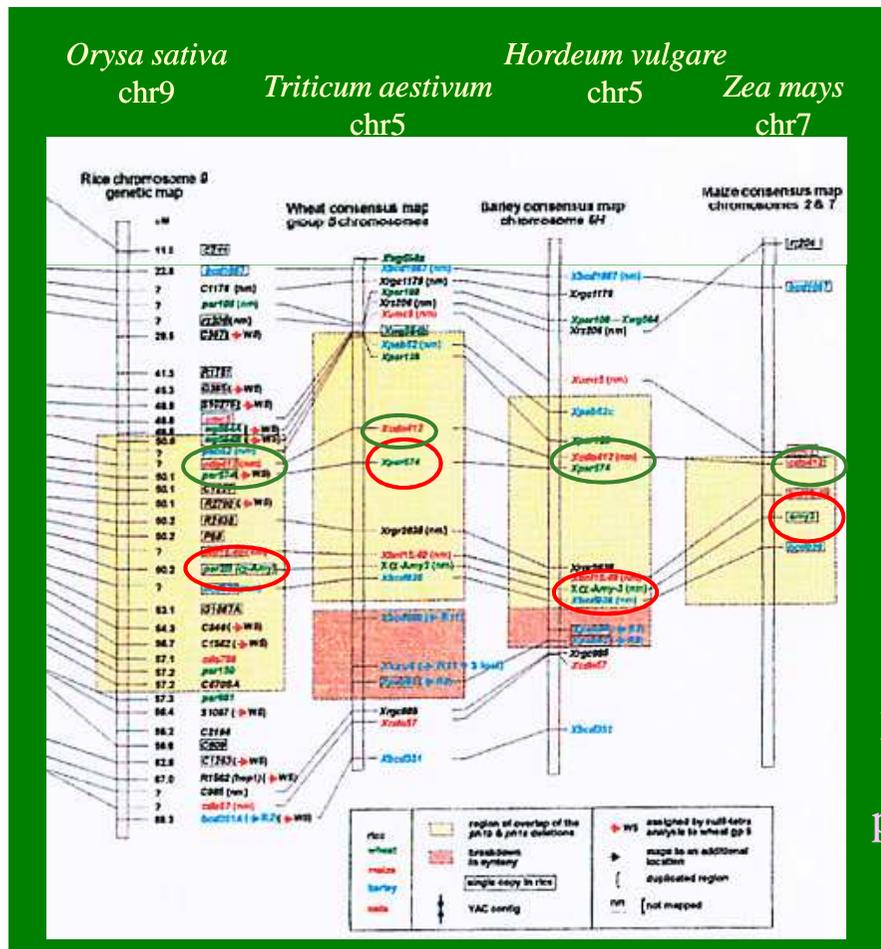
- "relativement facile" chez les organismes de laboratoire modèles dont le génome est entièrement séquencé.

- encore difficile chez des autres espèces dont le génome est en cours de séquençage (riz, maïs, blé, etc.)

- extrêmement difficile chez les espèces peu étudiée dans les laboratoires (peu ou pas de ressources disponibles).

Synténie

On retrouve chez des espèces différentes (mais proches) des blocs de gènes étroitement liés. Ces blocs sont le résultat d'inversions ou de translocations (cassure de chromosomes) qui ont été conservées durant l'évolution : c'est la **synténie**.



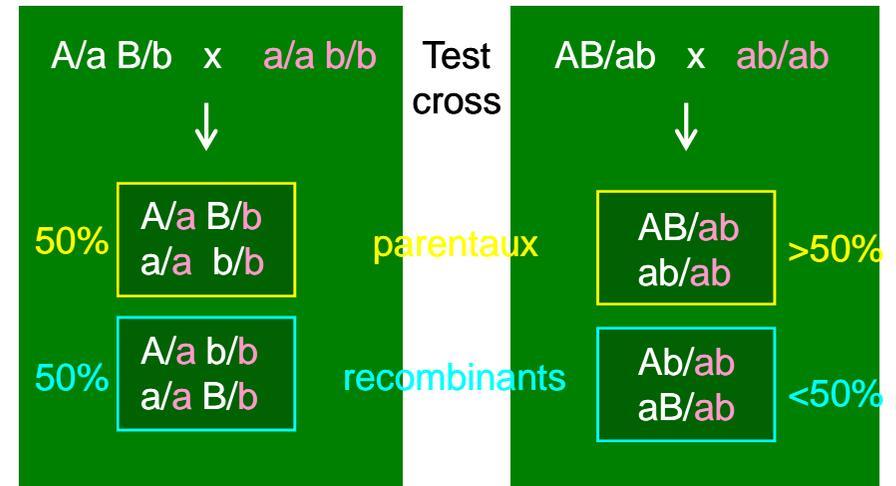
La synténie peut permettre de positionner rapidement un gène (Amy) chez le maïs, si sa position est connue chez le riz et qu'il fait partie d'un bloc synténique.

Attribution d'un gène à un chromosome

Rappel sur la liaison génétique :

Lorsque les 2 gènes sont indépendants (non liés)
50% phénotypes parentaux
50% phénotypes recombinants

Lorsque les 2 gènes ne sont indépendants (liés)
>50% phénotypes parentaux
<50% phénotypes recombinants



Pour attribuer un nouveau gène à un chromosome, on recherche une liaison génétique (traduisant une liaison physique) entre le gène et des marqueurs déjà positionnés sur la carte génomique.

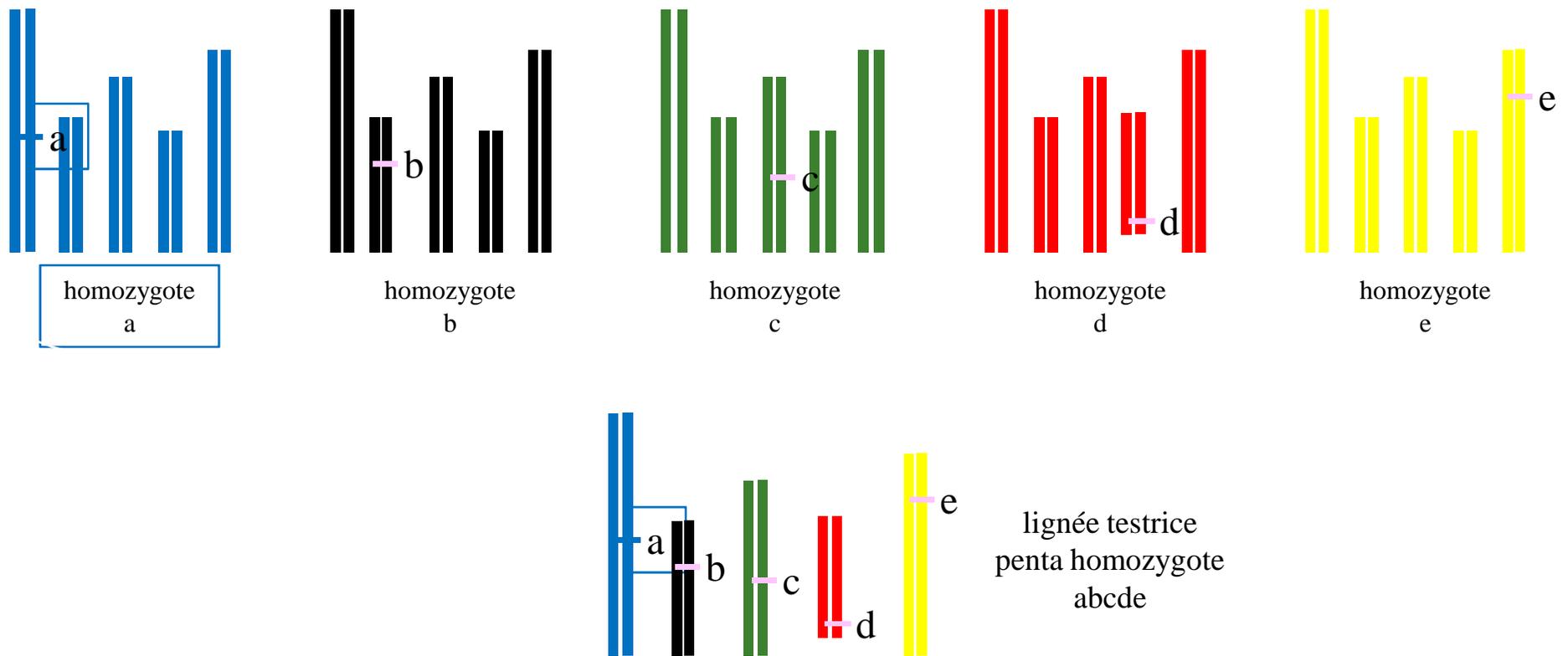
Quels marqueurs ?

Les 1^{er} marqueurs utilisés étaient des gènes mutés apportant un phénotype facilement observable.

Les lignées testrices (tester lines)

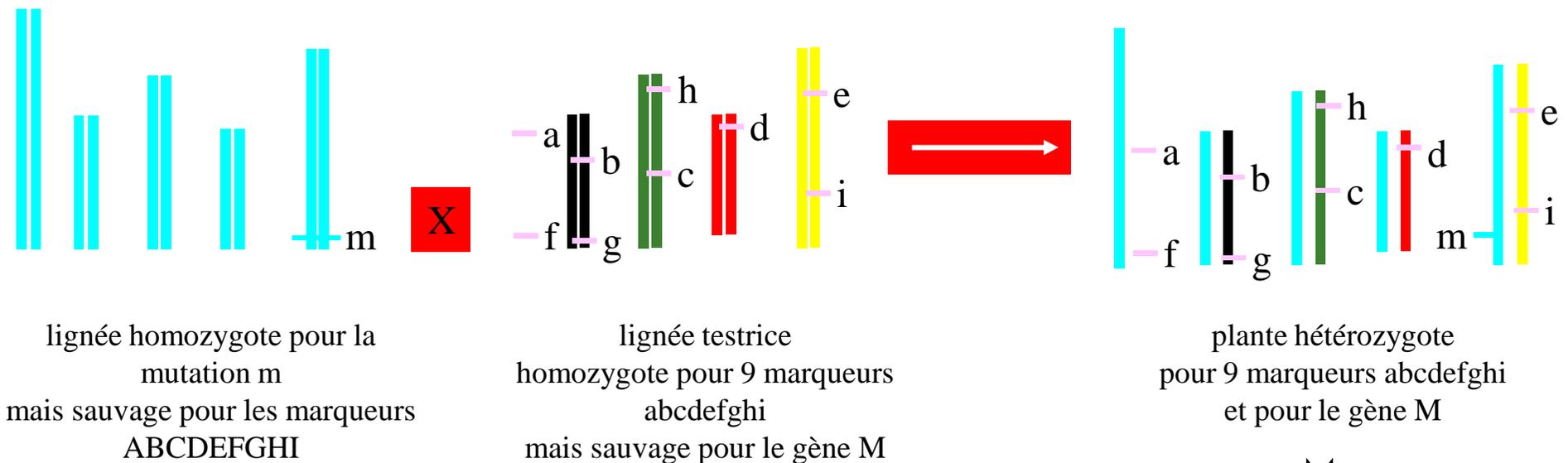
Création de lignées testrices (marqueurs phénotypiques) :

Croisement de plusieurs simples mutants pour rassembler le plus de mutations possibles dans un seul individu.



Utilisation des lignées testrices

Croisement de la plante dont on veut localiser la mutation avec plusieurs testrices comportants le plus de mutations possibles.



autofécondation

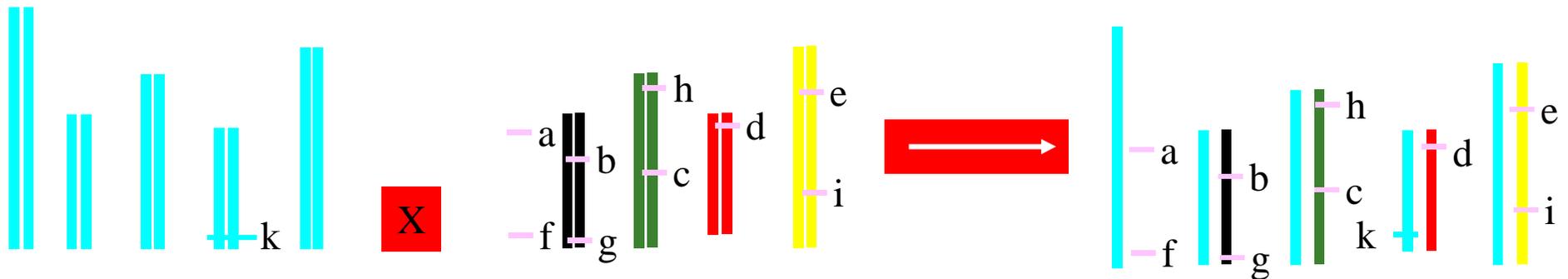
50% de parentaux (AA mm) et (aa MM)
50% de recombinants (AA MM) et (aa mm)
la mutation n'est pas liée au marqueur "a"

>50% de parentaux (II mm) et (ii MM)
<50% de recombinants (II MM) et (ii mm)
la mutation est liée au marqueur "i"

Observation des ségrégations entre le phénotype mutant (mm) et les phénotypes des différents marqueurs

Limite des lignées testrices

La densité de marqueurs présents sur les lignées testrices est déterminant :



autofécondation

50% de parentaux (DD kk) et (dd KK)
 50% de recombinants (DD KK) et (dd kk)
 la mutation n'est pas liée au marqueur "d"

la mutation n'est pas liée à AUCUN au marqueur

Observation des
 ségrégations entre le
 phénotype mutant (kk) et
 les phénotypes des
 différents marqueurs

Développement des marqueurs moléculaires

Les marqueurs moléculaires sont des variations neutres (d'un point de vue phénotypique) de la séquence d'ADN entre 2 lignées.

La découverte et le développement des marqueurs moléculaire a :

- augmenté considérablement le nombre de marqueurs sur les cartes génétiques
- facilité l'analyse des recombinants
- réduit le temps nécessaire à la cartographie d'un gène

On distingue différents types de marqueurs :

- RFLP (Restriction Fragment Length Polyporphism)
 - SSLP (Simple Sequences Length Polyporphism)
 - RAPD (Randomly Amplified Polymorphic DNA)
-

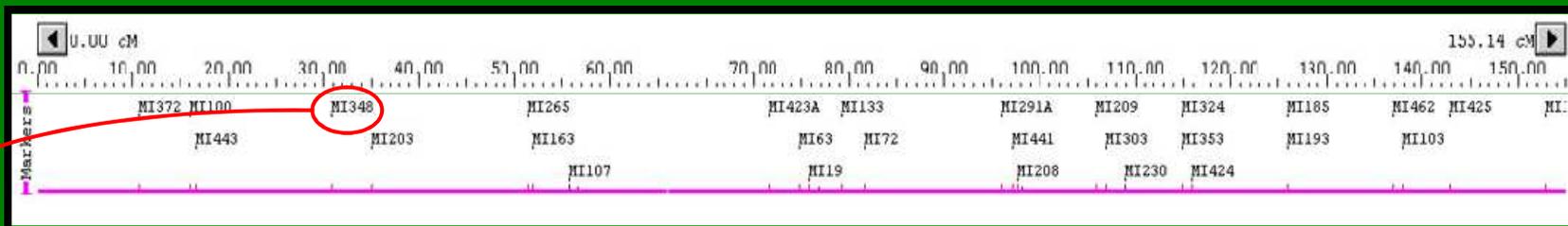
Utilisation des marqueurs moléculaires

1- Etablir une carte génétique de l'espèce :

On cherche de grands nombres de marqueurs différents. Lorsqu'on en a suffisamment, on peut les cartographier les uns par rapport aux autres, ou par rapports a des gènes déjà connus (carte RFLP par exemple).

2- Cartographier une mutation :

Le plus simple est d'isoler une mutation chez une variété puis de la cartographier en croisant la plante mutante avec une autre variété, fortement polymorphe.



Genetic Marker:MI348

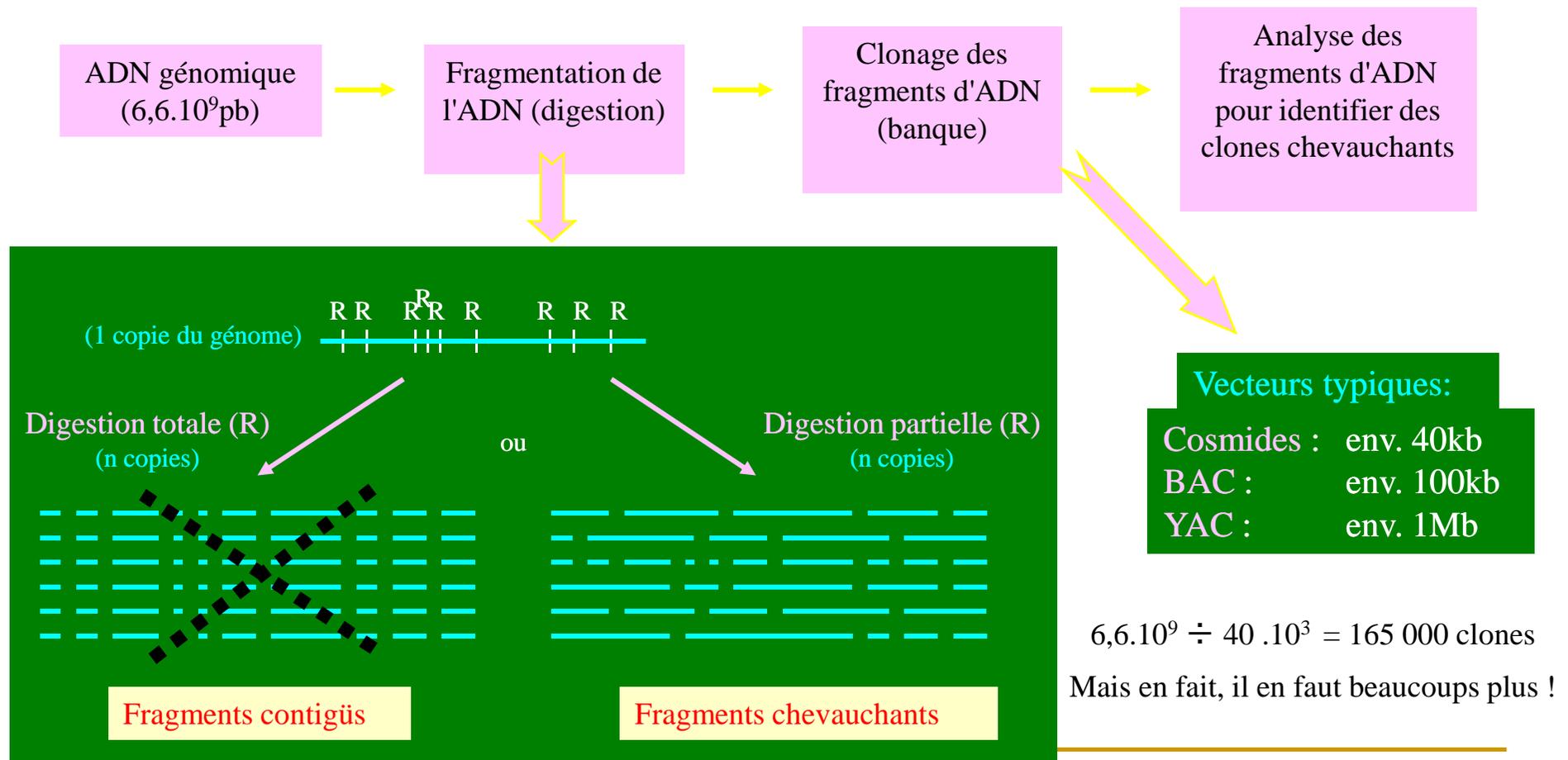
Date last modified	2001-02-18				
Name	MI348				
Tair Accession	GeneticMarker:1946014				
Type	RFLP				
Is PCR Marker	false				
Chromosome	1				
Made From	name	map element type			
	mi348	clone			
Associated Polymorphisms					
Digest Patterns	Polymorphism	species variant	restrict enzyme	fragment length	
	MI348	COLUMBIA	BAMHI	9.5	
	MI348	LANDSBERG ERECTA	BAMHI	14	
Map Locations	chrom	map	map type ?	coordinates	orientation
	1	RI	genetic	23.61 - 23.61 cM	unknown
	1	MIRFLP	genetic	30.8 - 30.8 cM	unknown

Carte RFLP du chromosome 1 d'*A. thaliana*

<http://www.arabidopsis.org/servlets/mapper?action=zoomto&band=3&zoomto=1>

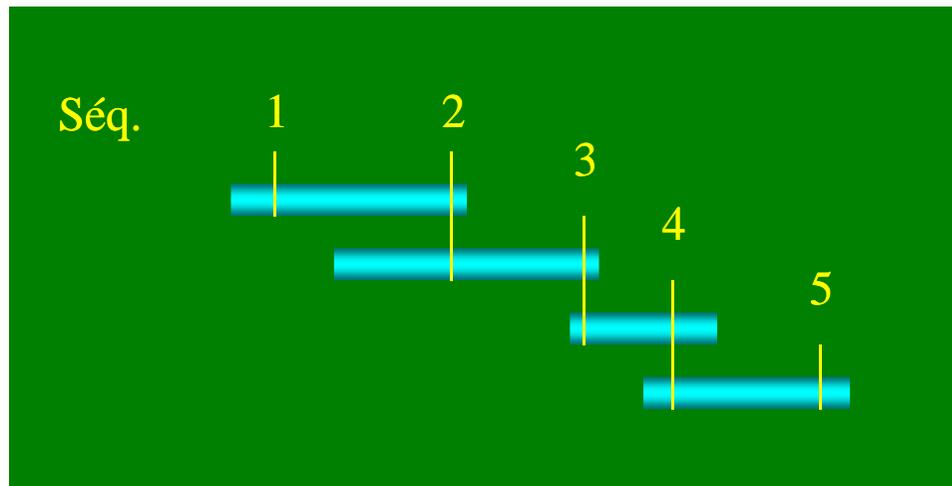
La cartographie physique

Le but est de produire des groupes de fragments d'ADN clonés chevauchants (contigs), dont l'ensemble couvre le génome complet.



La cartographie physique

Détermination de l'ordre des clones (séquençage de courtes régions des clones) pour former des contigs.



Au final, on obtient autant de contigs que de chromosomes de l'espèce

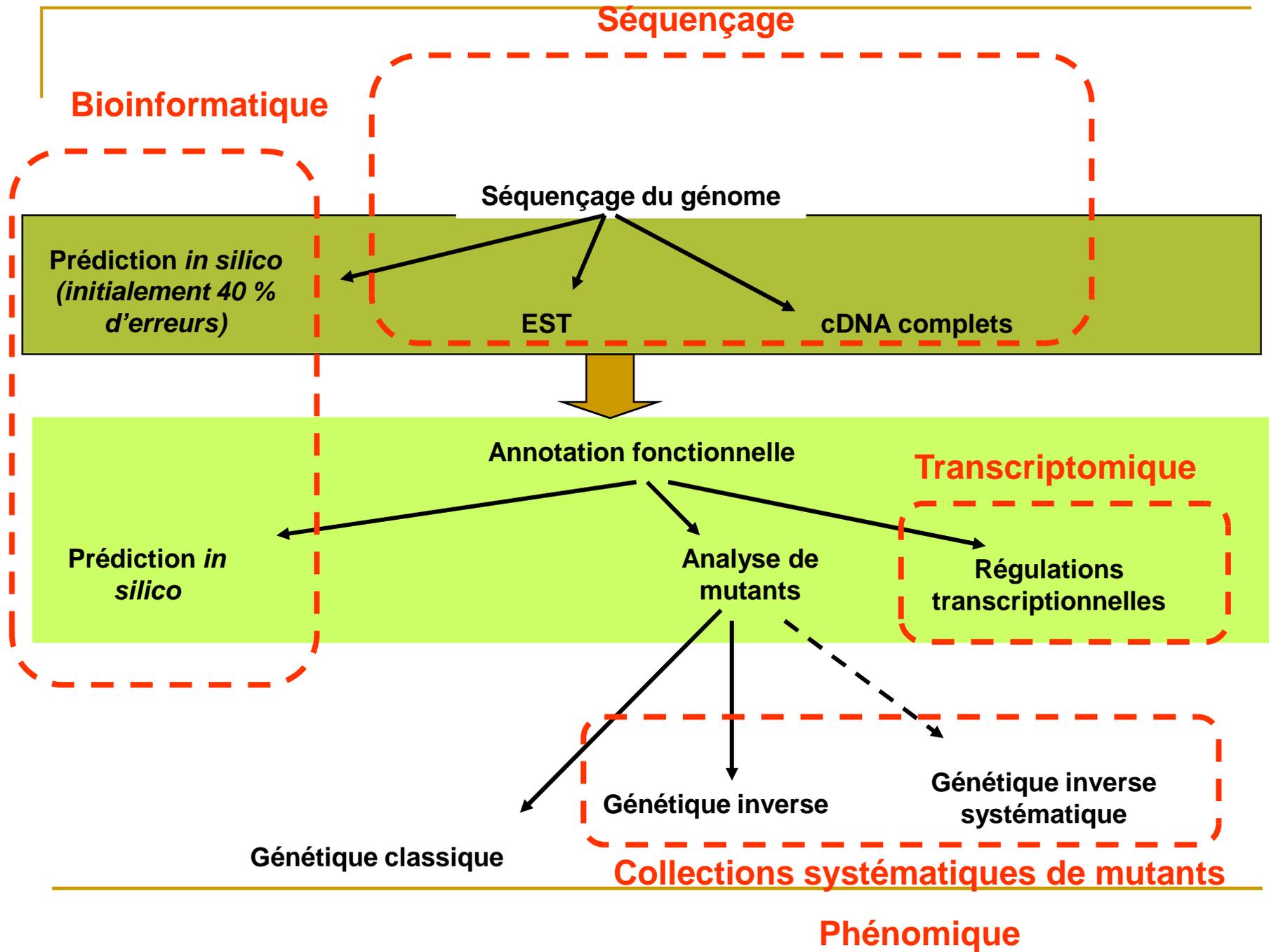


On Dec 14th, 2000, the Arabidopsis genome sequencing project is published in Nature.

Les cartes génétiques et physiques sont combinées (positionnement de marqueurs génétiques sur la carte physique).

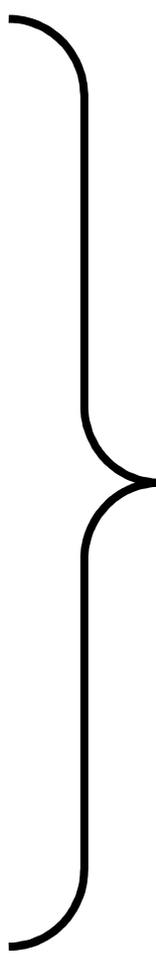
La carte physique sera utilisée :

- comme base pour le clonage de nouveaux gènes d'intérêts
- pour les projets de séquençage des génomes



Les outils de la génomique fonctionnelle

- Transcriptomique
- Bioinformatique
- Collections de mutants
- Phénomique



Investissement importants en parallèle des programmes de séquençage

Transcriptomique

- Objectifs :
 - Comprendre globalement les phénomènes de régulation de l'expression
 - Disposer d'une banque de données d'informations sur l'expression de chaque gène → fonctions potentielles
-

Collections de mutants

Quels mutants ?

- Mutagenèse chimique



Nombre important de mutations, mutations ponctuelles

- Mutagenèse par irradiation



Nombre important de mutations, mutations ponctuelles + délétions plus larges

- Mutagenèse insertionnelle

- Perte de fonction

- Gain de fonction



Faible nombre de mutations

Insertions

- RNAi



Extinction de familles de gènes

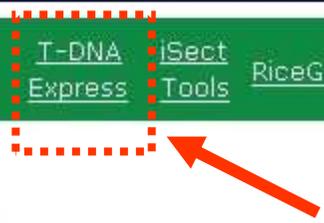
Simple pour les plantes non-modèles

T-DNA express

Un outil pour identifier l'existence de mutants T-DNA dans les différentes collections mondiales de graines d'*Arabidopsis thaliana*



- Data Release Policy
- cDNA Status
- T-DNA Genotyping
- T-DNA Express
- iSect Tools
- RiceGE
- Transcriptome
- ATGC Data
- Natural Variation
- Site Map



Arabidopsis Gene ORFeome Collection

Salk Insertion Sequence Database

Arabidopsis Tiling Array Transcriptome

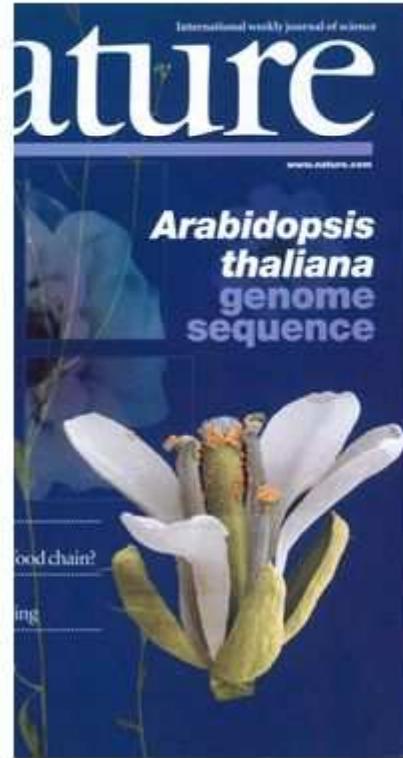
Phenome-Ready Unimutant Collection

High Resolution Arabidopsis Methylome

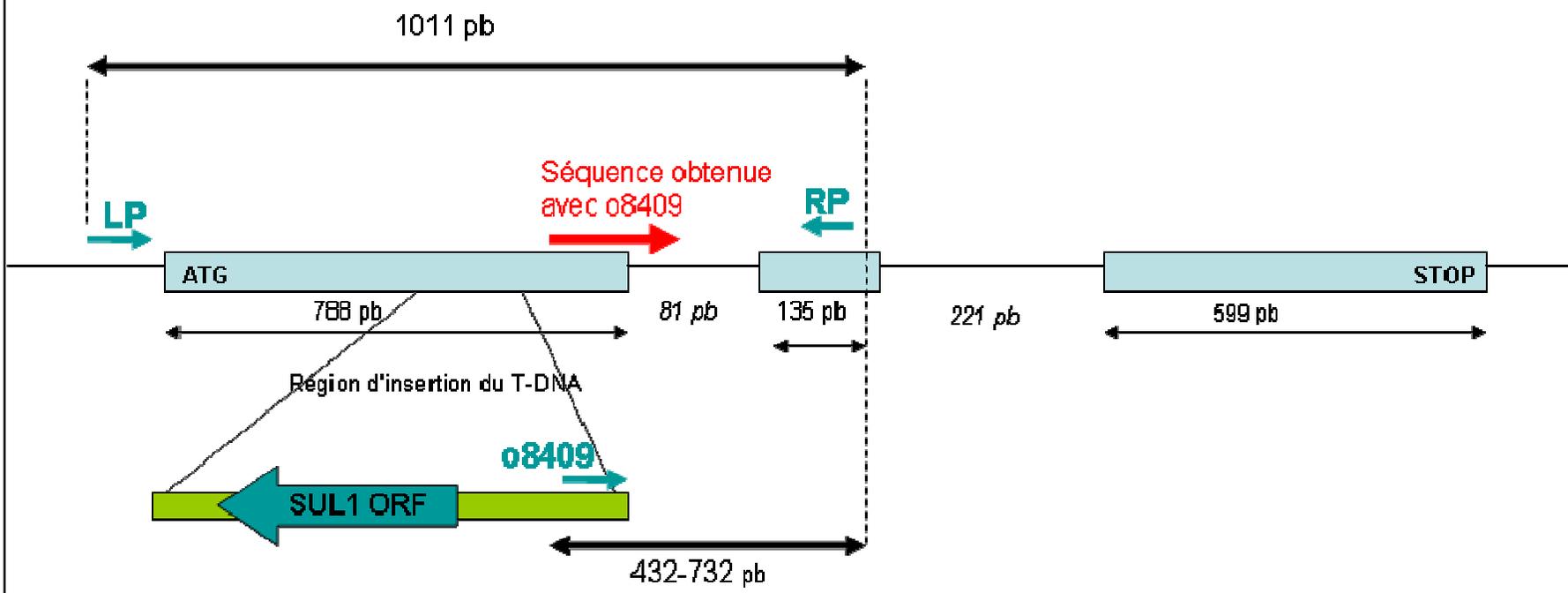
NEW

Single Feature Polymorphism Database

Rice Functional Genomics Database



Insertion du T-DNA de pAC161 dans la lignée knock out GABI-KAT GK_753B06 (=N472210)



Identification d'un gène

Deux méthodes principales pour identifier un gène chez une plante :

- Homologie de séquence

Dans le cas où le gène recherché a déjà été cloné chez une espèce voisine.

- Recherche et analyse de mutant(s)

- Interrogation de banques de données (recherche *in silico*)

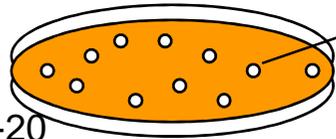
Pour les espèces dont le génome est entièrement séquencé (*A. thaliana*, le riz, et d'autres à venir).

Identification d'un gène par homologie de séquence

Recherche du gène X chez la tomate

Étalement sur boîtes de milieu gélosé

x 10-20

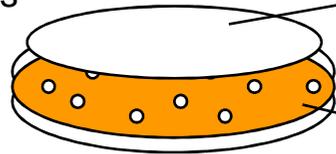


Étalement d'une banque d'ADN génomique ou d'ADNc de tomate

Plasmide contenant le gène X du tabac



Répliques sur filtres de nitrocellulose



Disque de nitrocellulose adsorbant partiellement les phages à partir des plages de lyse
Boîte mère conservée à 4°C

Marquage chaud (radioactif) ou froid (enzyme) de la sonde

Préparation des répliques pour l'hybridation

- Dénaturation
- Neutralisation
- Fixation sur le filtre



nitrocellulose solutions

Dénaturation de la sonde

Isolement des clones positifs sur la boîte mère
!!BINGO!!

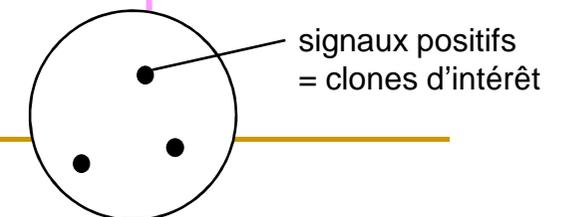
Hybridation moléculaire de la membrane avec sonde marquée

(type Southern)

- Pré-hybridation : blocage des sites non spécifiques (BSA, ADN thymus de veau)
- Hybridation : sonde hétérologue marquée
- Lavages : dissociation des hybrides non spécifiques (pas trop stringents)

Détection des signaux

d'hybridation (autoradiographie)



signaux positifs = clones d'intérêt

Autoradiogramme

Identification d'un gène par recherche de variants (mutants)

Les mutations peuvent être naturelles ou provoquées

- ponctuelles (erreurs durant la réplication, lésions spontanées, UV, EMS, etc.)

Grande variété de mutations possibles : allèle mutant faible, fort ou même gain de fonction.

- délétions (pendant les réarrangements chromosomiques, rayons gamma)

Mutations perte de fonction en général.

- insertions (saut de transposon, insertion d'ADN-T chez les dicots)

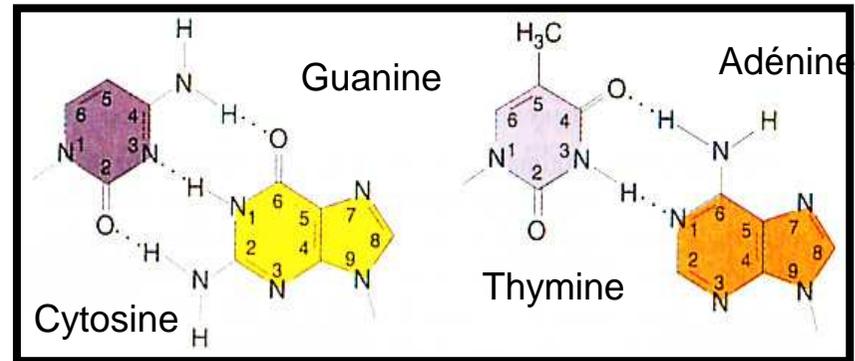
Mutations perte de fonction en général.

Mutations ponctuelles

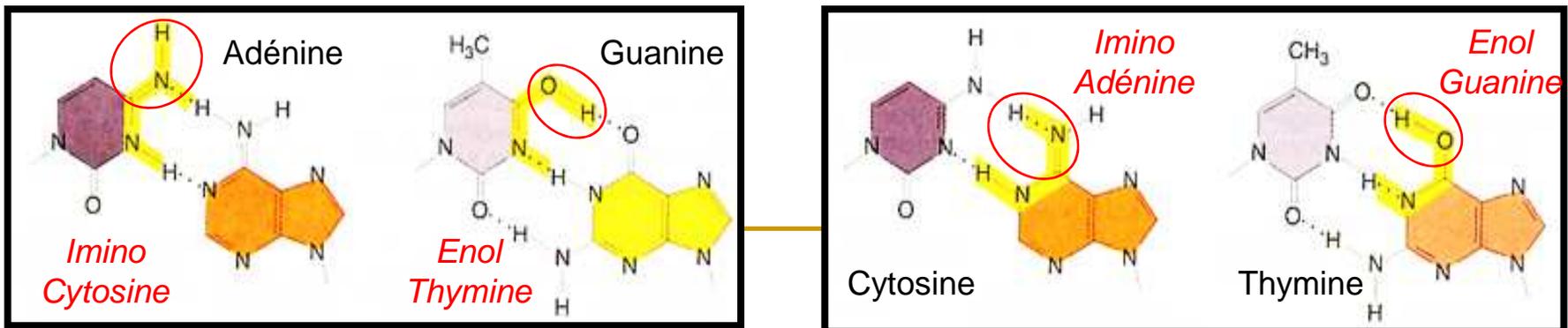
Silencieuses : AGG (Arg) CGG (Arg)
 Neutres : AAA (Lys) AGA (Arg)
 Faux-sens : GGC (Gly) GAC (Asp)
 Non-sens : UGG (Trp) UAG (STOP)

Addition décalage du
 Déletion cadre de lecture

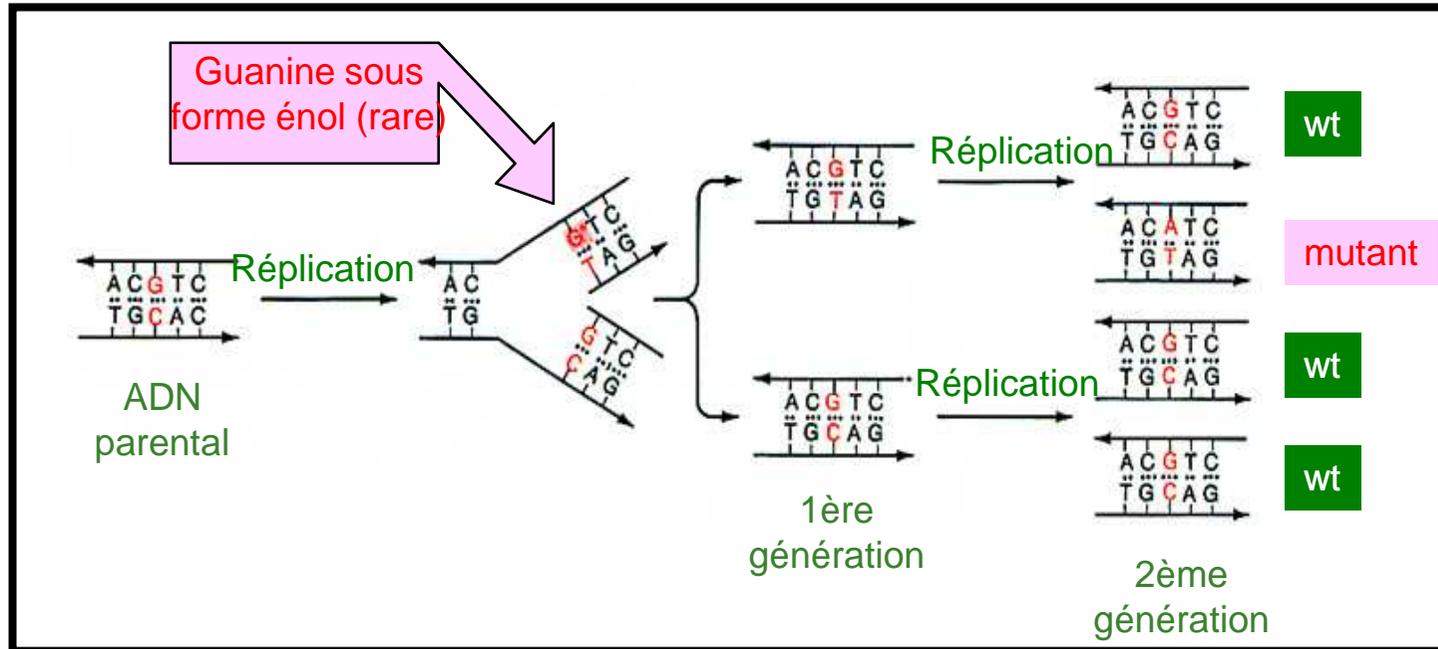
Appariements de Watson & Crick



Appariements illégitimes rares (commutations tautomériques)



Mutations ponctuelles pendant la réplication



Les mutations engendrées par ces mauvais appariements se classent en 2 catégories :

transitions

Mutations substituant une purine en une autre purine ou une pyrimidine en une autre pyrimidine.

A:T ↔ G C
T:A ↔ C G

transversions

Mutations substituant une purine en une pyrimidine ou une pyrimidine en une purine.

A:T ↔ C G
T:A ↔ G C
A:T ↔ T A
T:A ↔ A T
G:C ↔ C G
C:G ↔ G C

Mutations ponctuelles pendant la réplication

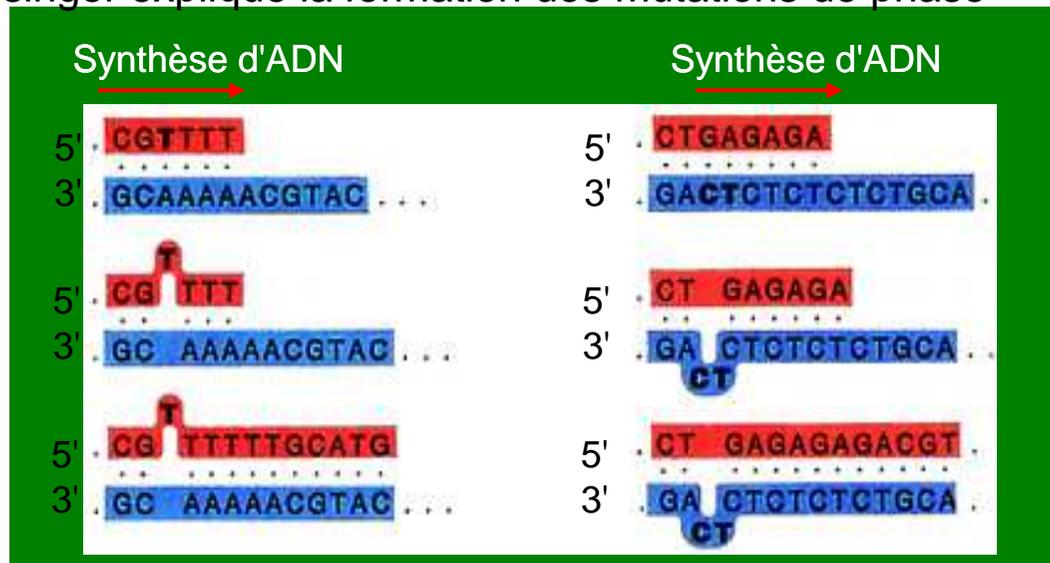
Les mutations de phase :

L'addition ou la délétion d'une ou plusieurs paires de bases (non multiple de 3) provoque des mutations sévères par décalage du cadre de lecture.

codons sauvages :	ILS	ONT	MIS	UNE	DES	SIX	VIS
délétion d'un nucléotide (M) :	ILS	ONT	*ISU	NED	ESS	IXV	IS
addition d'un nucléotide (G) :	ILS	GON	TMI	SUN	EDE	SSI	XVI S

Le modèle de Streisinger explique la formation des mutations de phase

Le glissement du brin néosynthétisé provoque une addition de nucléotide(s).

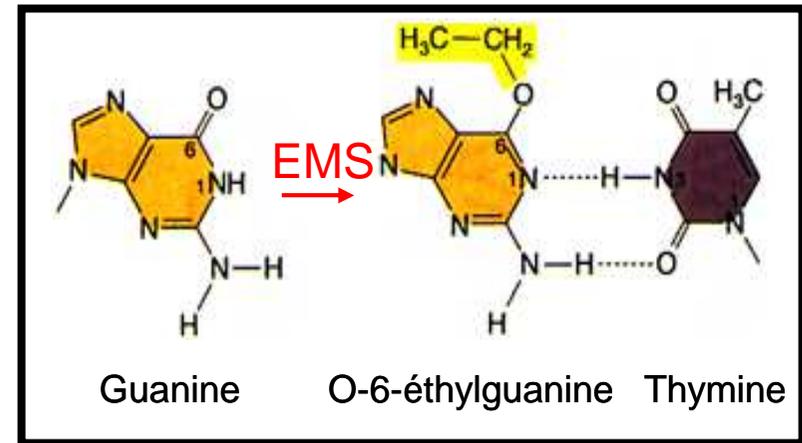


Le glissement du brin matrice provoque une délétion de nucléotide(s).

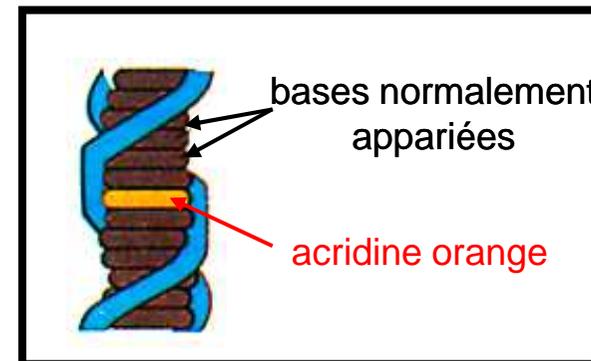
Fréquence d'apparition de ces mutations privilégiées aux points chauds : répétitions d'une base ou de courtes séquences.

Mutations ponctuelles provoquées

L'EMS (éthyl-méthane-sulfonate) :
L'EMS alkyle les guanines. La guanine alkylée en O-6-éthylguanine s'apparie illégitimement avec une thymine.
L'EMS produit des transitions G → A



Les agents intercalants :
sont des mutagènes à structure plane, analogue à une paire de bases. Ils peuvent s'intercaler dans la double hélice d'ADN et engendrer des mutations par addition ou délétion d'une paire de bases.



Mutations insertionnelles par ADN-T

Agrobacterium tumefaciens est une bactérie du sol (rhizobiacée, GRAM-) capable d'infecter les dicotylédones (rarement les dicots) et induisant la gale du collet (tumeur).

La maladie est due à un transfert d'ADN (ADN-T) de la bactérie dans le génome des cellules végétales : transformation.

Ce mécanisme est le seul exemple naturel de transport d'ADN entre règnes.

L'ADN-T :

Fragment d'ADN du plasmide Ti (Tumour inducing, 200kpb) délimité par 2 séquences répétées de 25pb (en orientation directes). Il code pour :

- 1 enzyme de la voie de biosynthèse des cytokinines
 - 2 enzymes de la voie de biosynthèse des auxines
 - enzyme de biosynthèses des opines (source de C et N pour la bactérie)
- } Oncogénèse

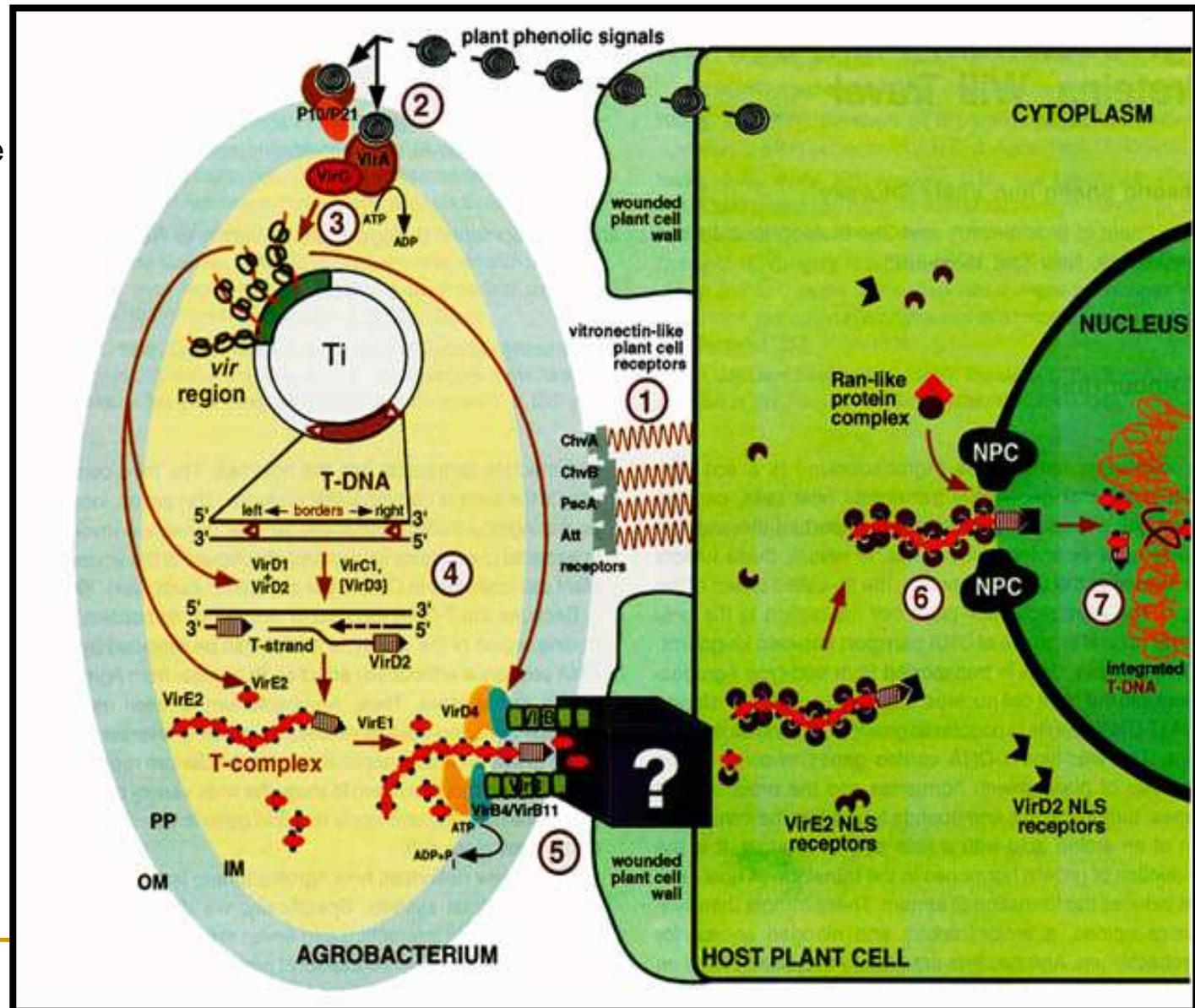
Les cellules infectées se divisent activement (tumeur) et synthétisent des opines utilisées par *A. tumefaciens*.

3 composantes génétiques nécessaires pour la transformation de la cellule végétale:

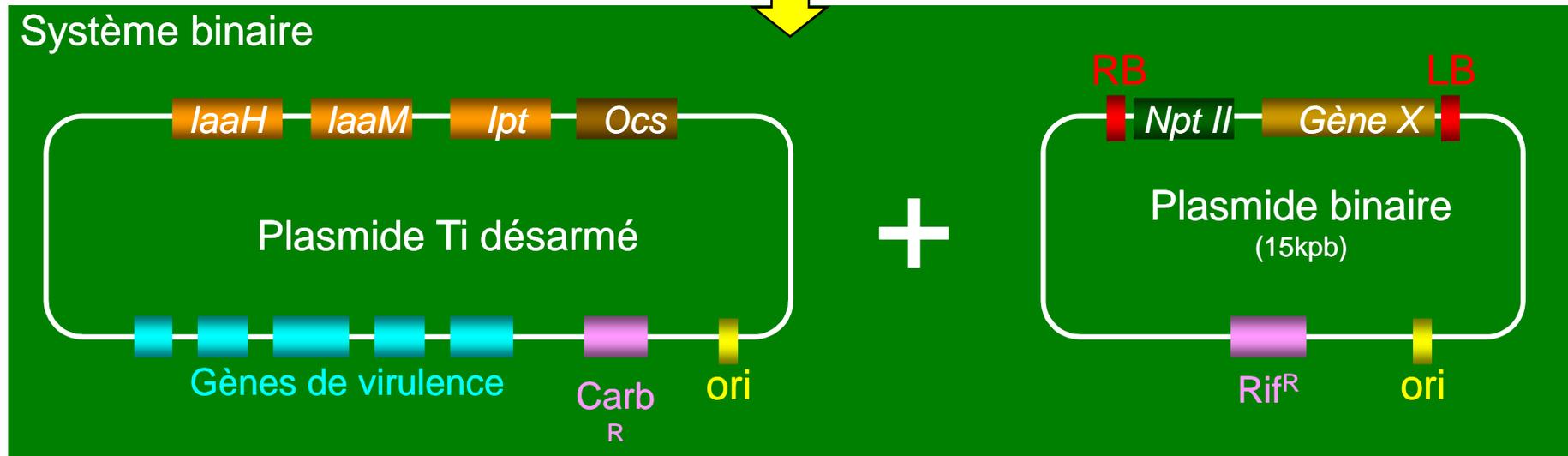
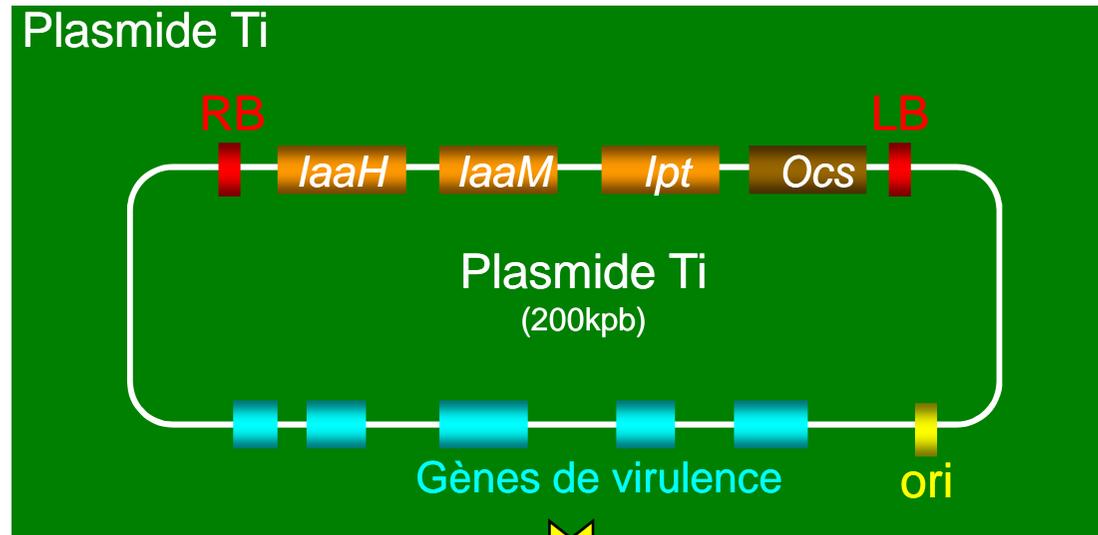
- l'ADN-T : seules les bordures gauches et droites sont indispensables (sur le Ti).
- les gènes *vir.* perception d'une blessure, synthèse transfert et intégration d'ADN-T (Ti).
- les gènes *chv.* chémotaxie et attachement à la cellule végétale (chromosome bactérien)

Mécanisme de transfert de l'ADN-T

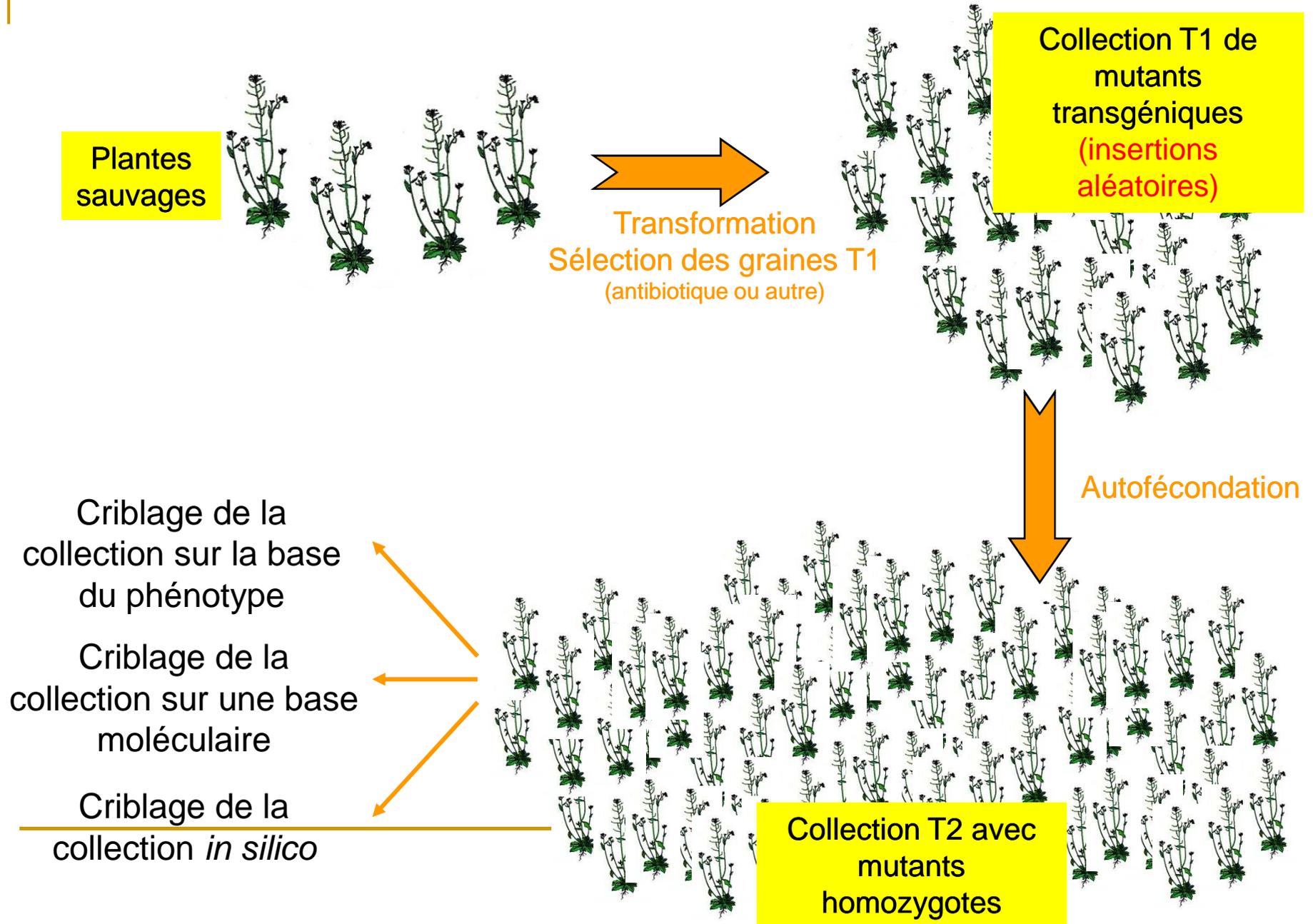
- ① Attachement d'*A. tumefaciens* à la cellule végétale (*chvB*, *pscA*, *Att*)
- ② Perception et transduction du signal de virulence (*virA*, *virG*)
- ③ Activation transcriptionnelle des gènes de virulence (*virG*)
- ④ Production du *brin-T* (*virD1*, *virD2*, *virE1*, *virE2*)
- ⑤ Transport intercellulaire (*virB*, *virD4*)
- ⑥ Importation nucléaire (*virD2*, *virE2*)
- ⑦ Intégration au génome végétal (*virD2*, *virE2*)



Le système binaire de transfert d'ADN-T



Collection de mutants ADN-T

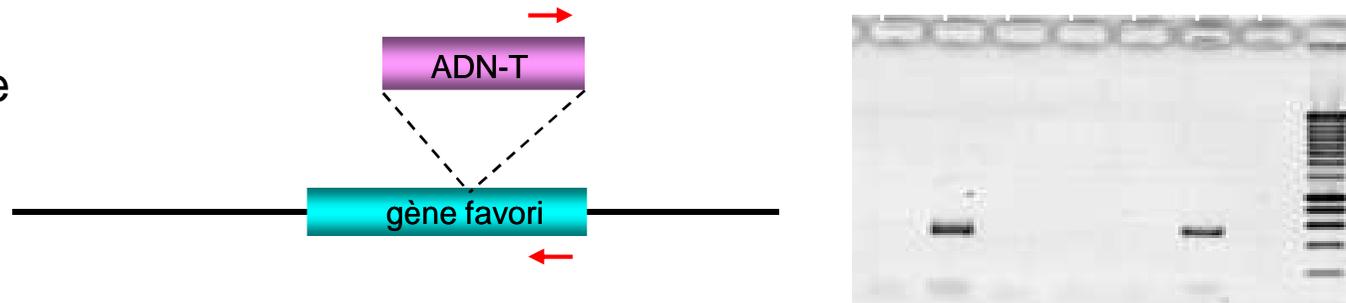


Criblage d'une collection de mutants ADN-T

Criblage Phénotypique
(gène inconnu)



Criblage Moléculaire
(gène connu)



Criblage *in silico*

(gène connu et
collections étiquetées
disponibles
sur internet)

```

ataaactttaagggttcattgtcttagagagtaagaacaaaaaccattttctttaaactcttctacattgttaaatatcttggataatattacaagacaca
====SALK_103658.30.55.X====>
=====SALK_031530=====
117201 cacacaactttaacttataacttaactcaactacactcttaggcacacacacaaagcatgccttaagcacattattaaccctagactctagactcttgcacaag
=====SALK_031530=====
ccacttaactacatggctcttggataacttgagtaaaaagtggaaatcttacaacacacttgcccaagtagaagaagacaggatccaaagataaata
=====SALK_031530=====
117401 cactgaacagactagagcttatgacttaactaagaacctaaacaacaccaaatgctacaaggcatattggtttcatCTAactatcattagctgctctgct
=====SALK_031530=====
AAGCTTGGCGGAAAATCTAAGCTTCTTGATATGGAATCTGTGTTCTCTCAACACAGATAAGGCTTTCATTCATCAGCTCAACTCTCTCTTTTACCATTTCGA
117601 AGGTTTCATCAGCTTCTCAAGCTGCAACTTCTTTTGATAACCACTTAAACCTTGGCACATCACGGCTTCAAAATGCATCTTCAACCGCCATAACTCATCAT
CATCATCATTTCTCATCATGCTCTCTTTTCCACTACTAGTCCCACTCTCTGTTGTTTGGCTGCCCACTTCTCATTAAAGTTCAGATATTAGCCAC
117801 TTCTCTCTGTTGCTCTCTCAACACTCTCTCCCGCTCTGTTAGTCCGGTTTTCCGCTGGCAACCACTTCCGTTTACCTCGTGCTCTTATCCATACCATCA
<====
TCATCACCTTGTCCCAATCATCGAAGAAAACCCAAAGGCTTCAGAATCACTAGACCTAGTAAATTTGGGATCATTCCTTGATTGATTTCTCCGAGT
=====SALK_115723.17.45.X=====|
<-----
118001 GAACATGAAACCTCCTTTTCACTTCTGATTTTACTCAAAAACCTGATCTTGGAGAATCTCTCAACGATAGAACTCCGAGGAAACTACAAAACGCATC
-----SALK_013473-----
CCAATCAATTTGGGATTAATCTGCTTAGCTCTGTAATCAACTAACCctgagcaatttgganaatctggttttagaaaaatcacaatgcaaaaccaa
-----SALK_013473-----
118201 atcaaaactataaatggagaagatcaatacaataacCTTAAGATAGAGAGCTCATCTCTCCGTTCCAGATCCGGACGATCAACGGCGGATCCAATTTCTT
-----SALK_013473-----
CTTCTCTTCTCATCTTCTCTCTCTCTCATAGCTCTCTCTTCTGGCAGCGAGCAGCAGCCTTGGAGAAAGAAATATCCGTGGAGATCTTCGATGG
-----SALK_013473-----
118401 GAGAGGATACCGTGGAGTTATCAGCACCAACACAGAAAGAAATCGATCTCTCCGGAATCAAGgcttattgctttagacttgcatttca
-----SALK_013473-----
ggggcgcagcagagcagtgattcccaatctcagcagcttataaaagtgaacccgagagaaaaagaaaaagcaaaataggatgagtaaacacttggccata
/
118601 cttttctcaacatcaataaaaaaacacactttcatcttttttctactgtttaatattcatcaatagtaacaatttgccttgcataaatctatcatt

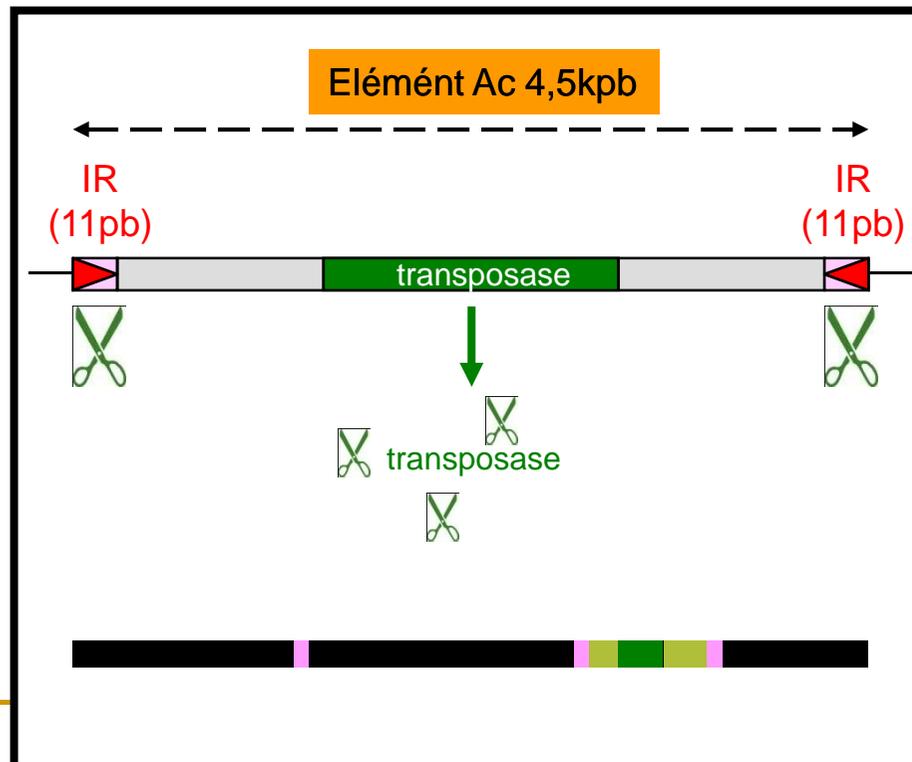
```

Mutations insertonnelles par transposon

Les transposons :

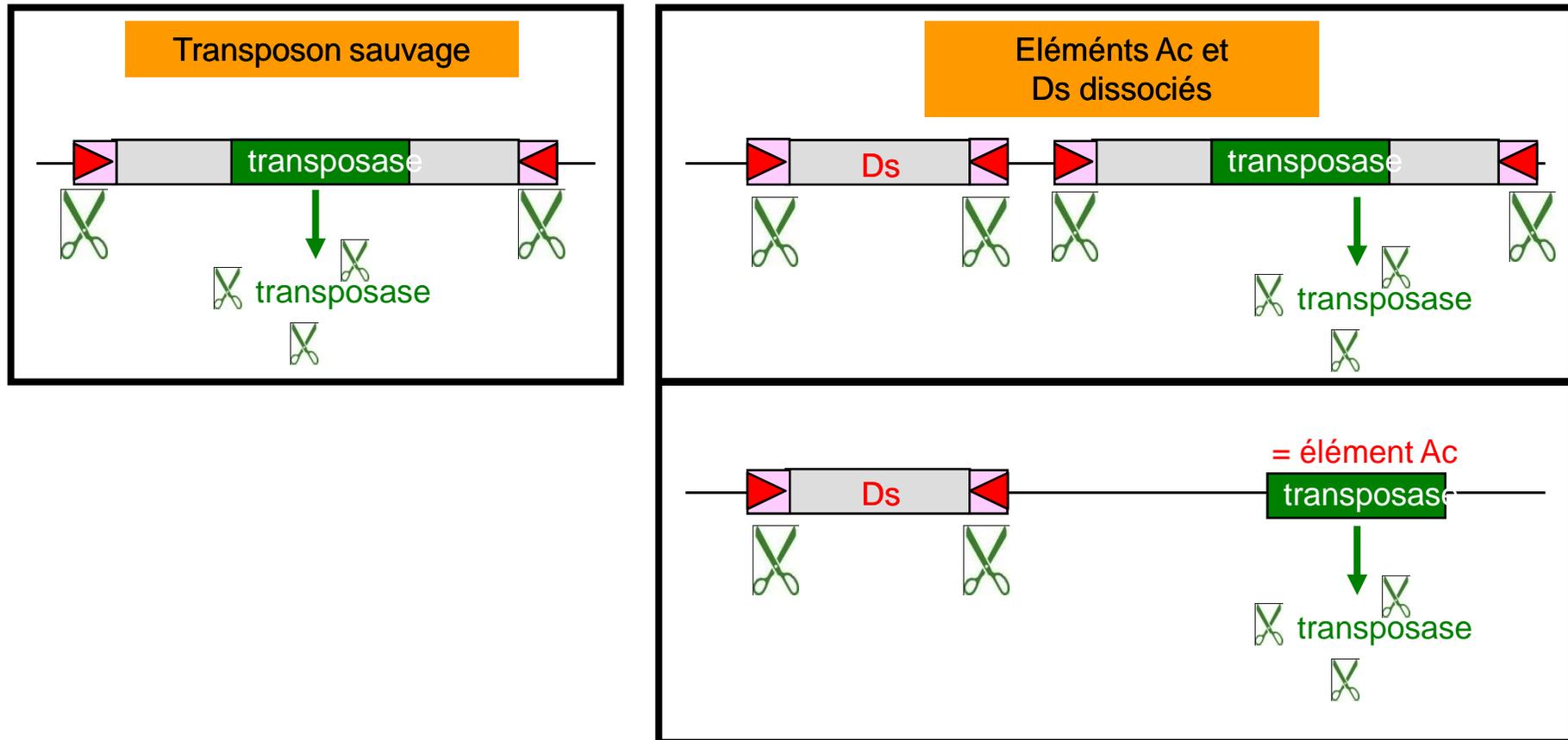
Éléments Ac/Ds découverts chez le maïs dans les années 1950
par

Barbara McClintock (prix Nobel 1983)



Découplage Ac et Ds

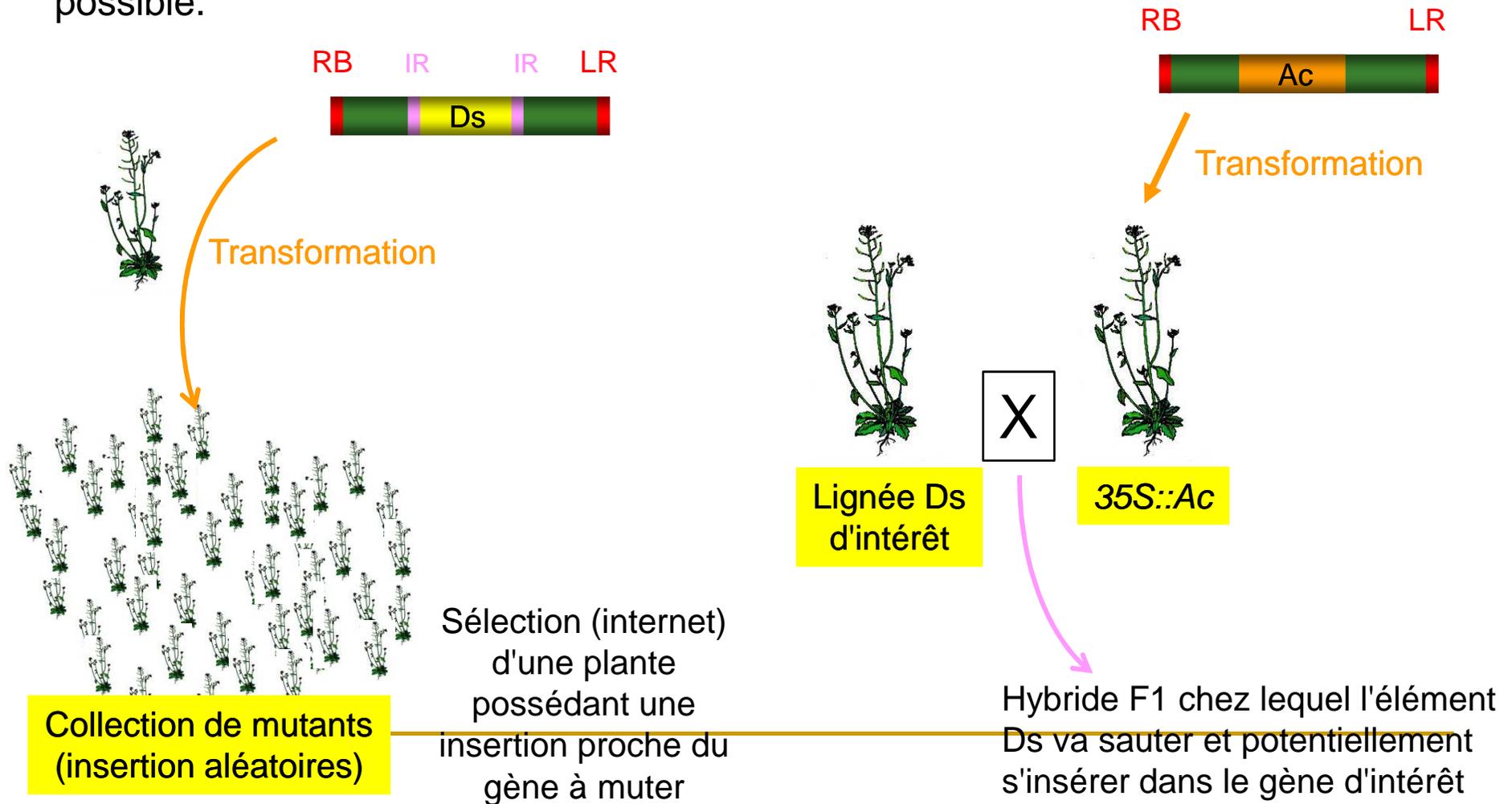
La transposase va mobiliser n'importe quelles séquences IR. Elle agit en *trans*.



Les transposons Ac/Ds existent aussi chez d'autres plantes que le maïs.
 On peut également les transférer chez des lignées qui en sont dépourvues.
 Si l'élément Ds s'insère dans un gène, il crée une mutation (forte).

Collection de mutants Ds

Lorsqu'un élément Ds "saute", il se réinsère (généralement) dans une séquence proche de son site d'origine. -> Nouveau type de collection possible.



Clonage d'un gène muté par mutation ponctuelle

1- Cartographie de la mutation :

- attribution à un chromosome
- localisation plus précise grâce aux marqueurs moléculaires

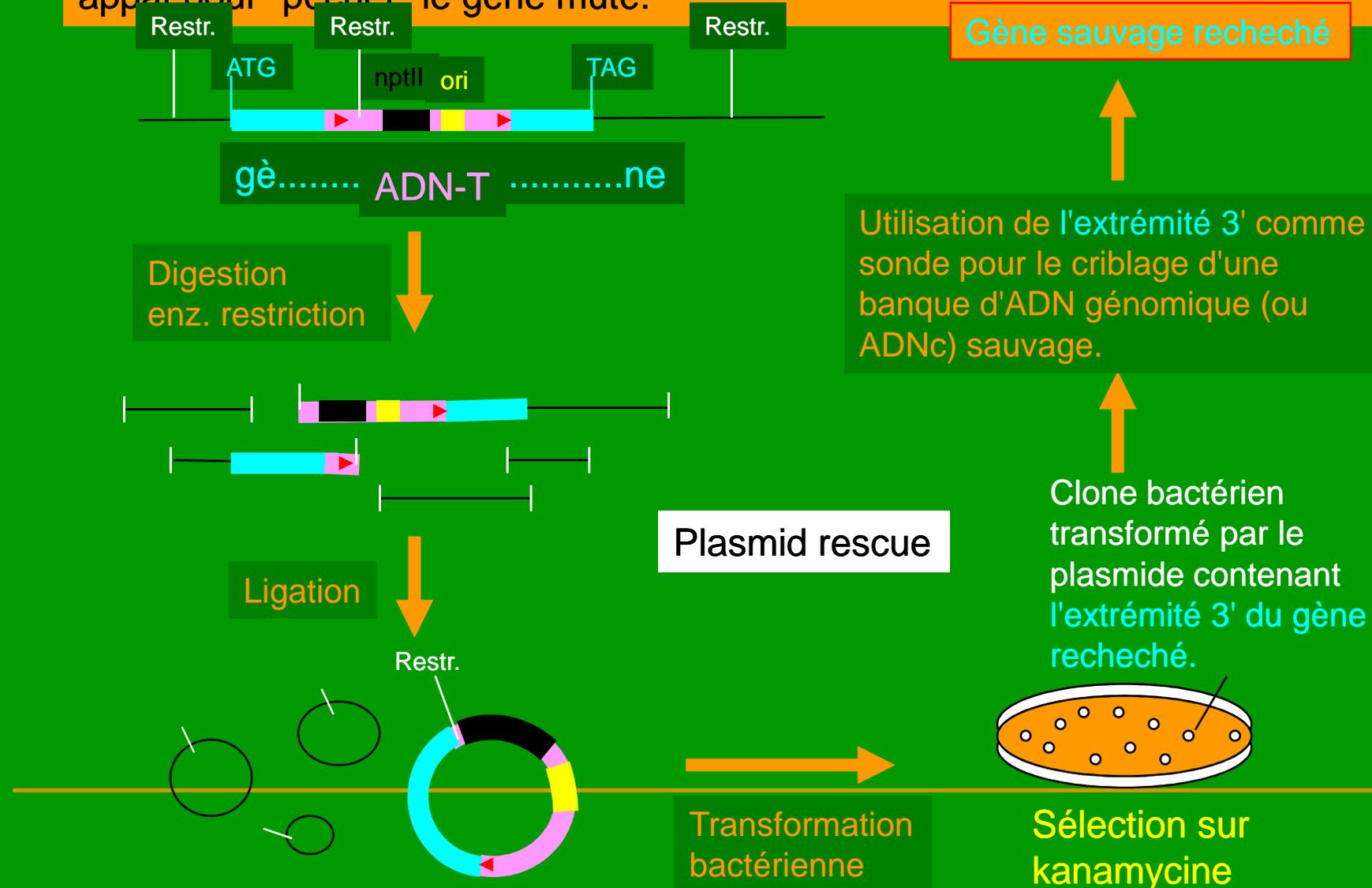
2- Cartographie physique de la mutation :

- utilisation des contigs (marche sur le chromosome)

3- Séquençage et recherche de gène candidat

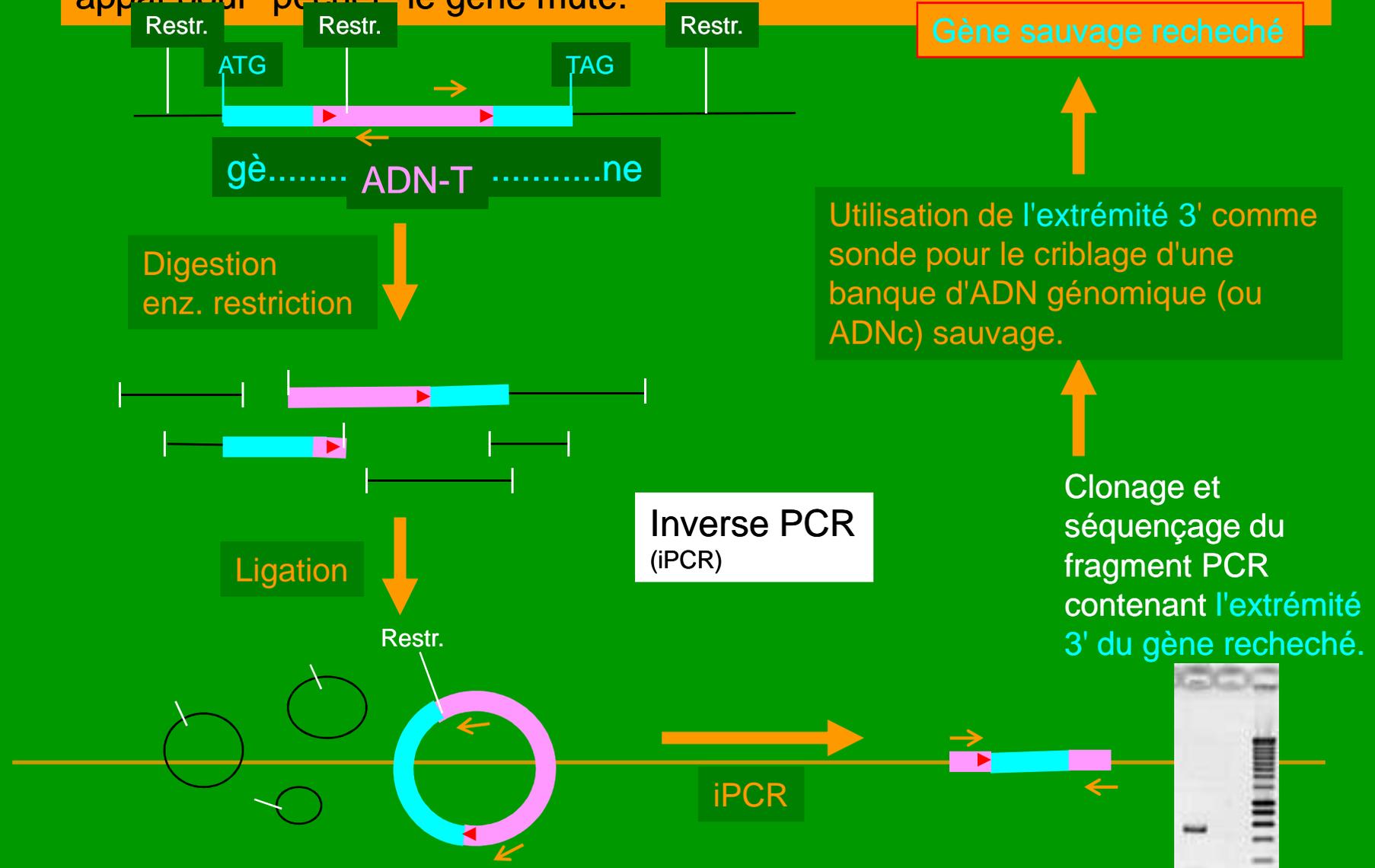
Clonage d'un gène étiqueté par l'insertion d'un ADN-T ou d'un transposon

La séquence de l'insertion étant connue (ADN-T ou Ds) on s'en sert comme appât pour "pêcher" le gène muté.

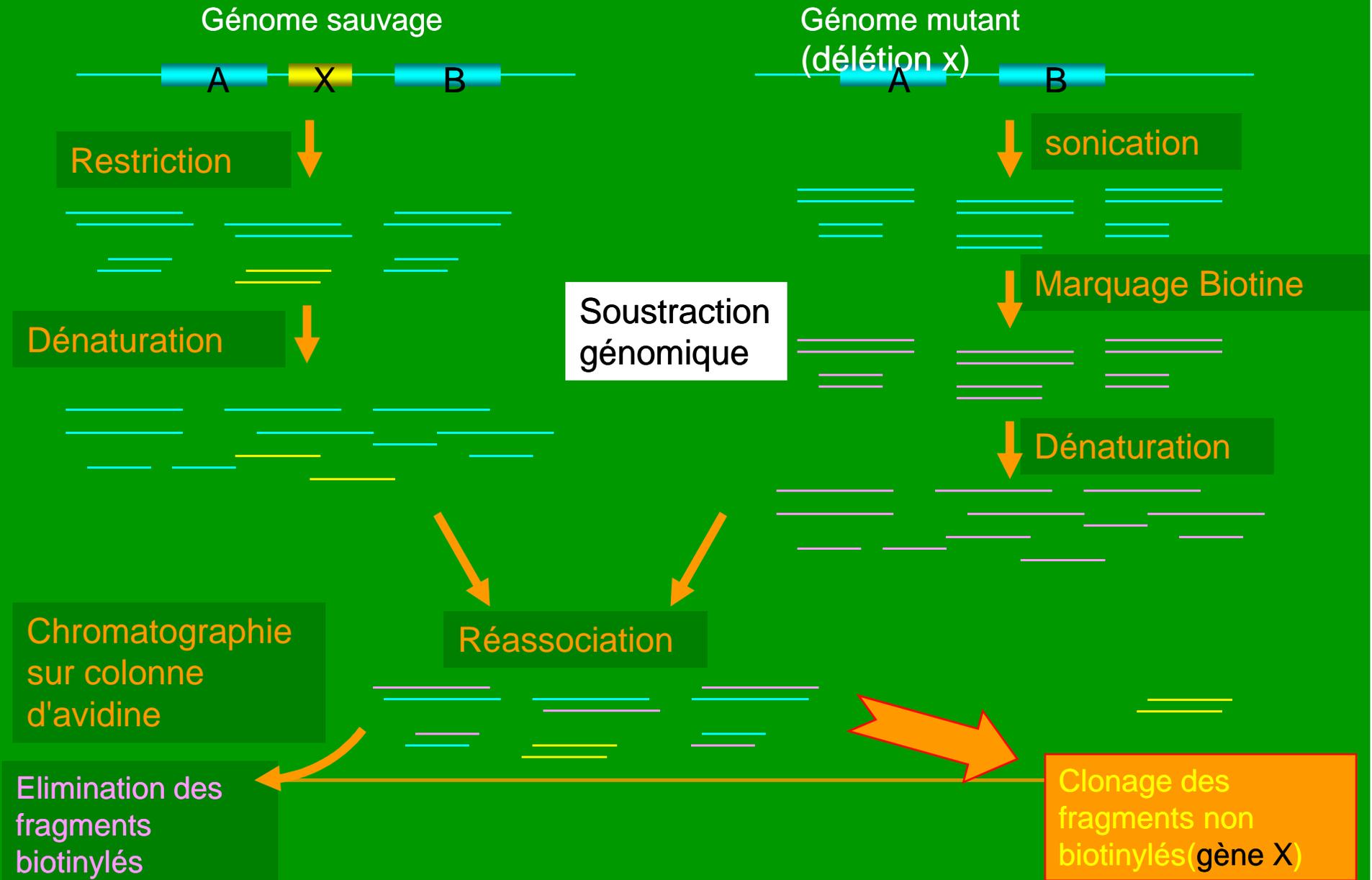


Clonage d'un gène étiqueté par l'insertion d'un ADN-T ou d'un transposon

La séquence de l'insertion étant connue (ADN-T ou Ds) on s'en sert comme appât pour "pêcher" le gène muté.



Clonage d'un gène à partir d'un mutant de délétion

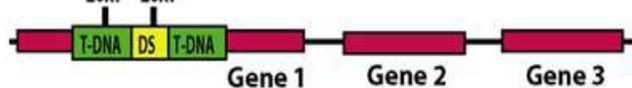


Induction de délétion par le système Cre-Lox

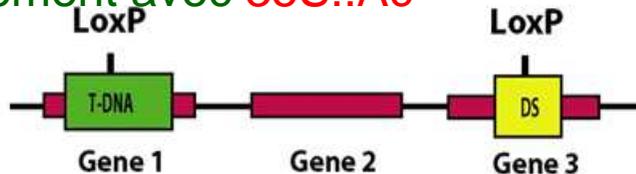
Stratégie générale

Insertion d'ADN-T contenant

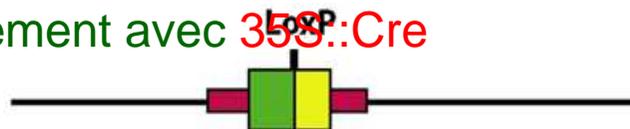
- 1 élément Ds
- les sites LoxP reconnu par la recombinaison



Mobilisation de l'élément Ds par croisement avec **35S::Ac**

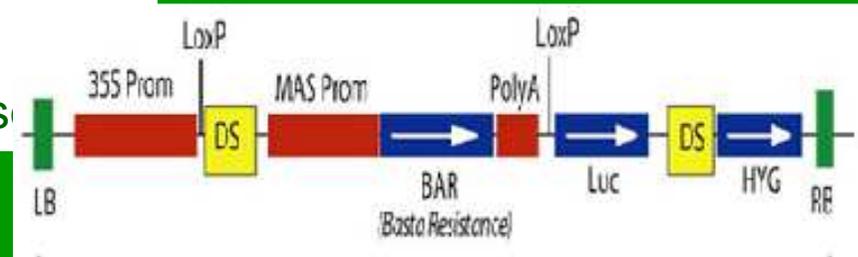


Excision de l'ADN situé entre les 2 sites LoxP par croisement avec **35S::Cre**



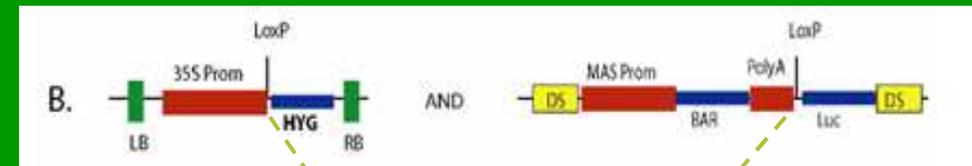
Lignées transgéniques WiscDsLox

<http://www.hort.wisc.edu/krysan/2010/default.htm>



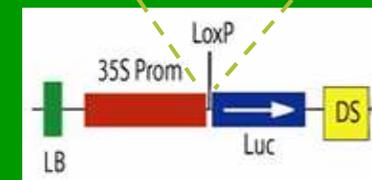
X **35S::Ac**

Sélection Basta



X **35S::Cre**

Sélection Hyg et Basta



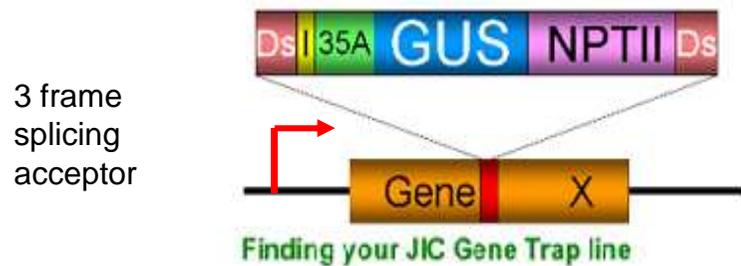
Sélection Luciférase

Clonage

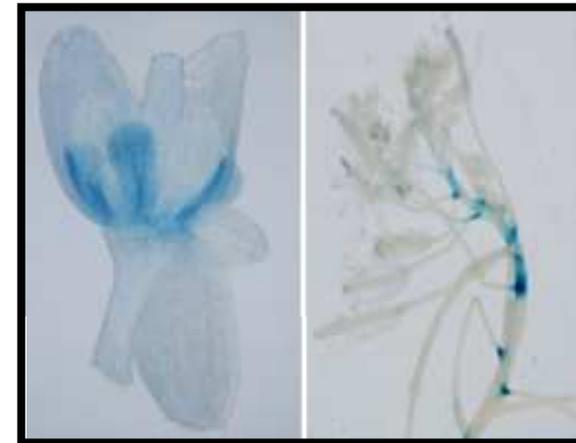
Piège à gènes (gene trap)

<http://www.jic.bbsrc.ac.uk/science/cdb/exotic/index.htm>

GUS et Gène X dans la même orientation



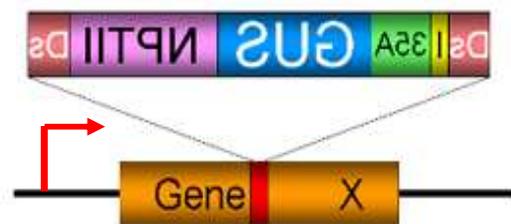
Les gènes piégés s'expriment dans



les pétales

les
jonctions
tige-
pédicelle

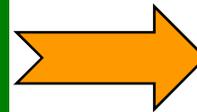
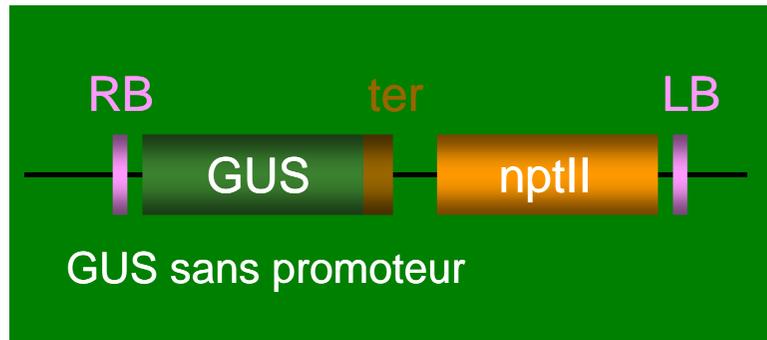
GUS et Gène X en orientation inverse



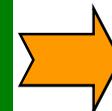
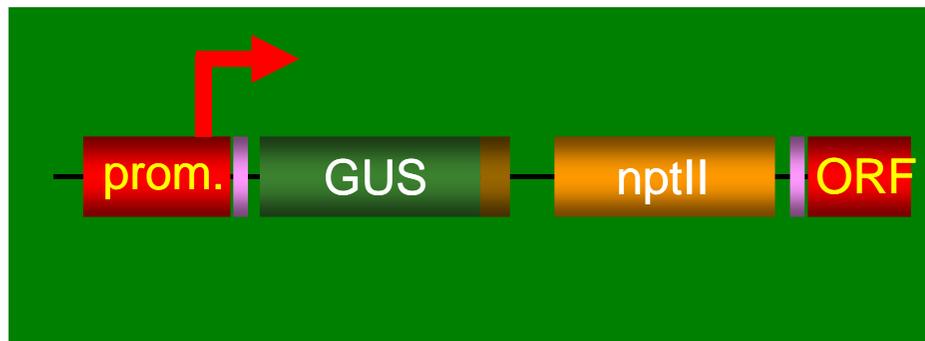
(simple knock out)

Piège à promoteurs (promoter trap)

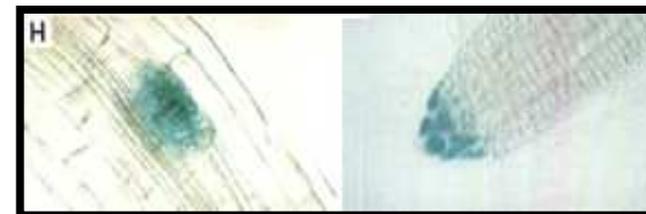
Identification de séquences régulatrices de la transcription



(pas d'expression)



Le promoteur piégé est celui d'un gène s'exprimant dans les méristèmes racinaires d'*A. thaliana*.

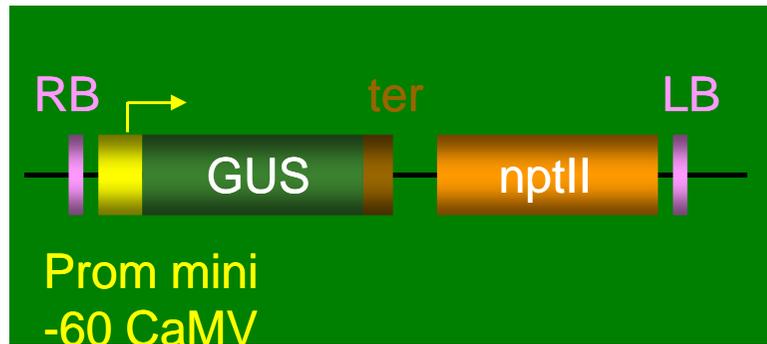


Ces lignées permettent l'étude de la régulation transcriptionnelle des gènes ciblés.

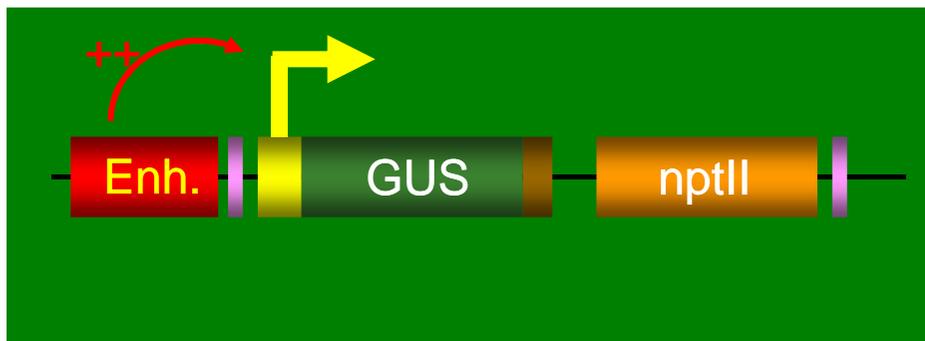
Piège à enhanceurs (enhancer trap)

Identification de séquences régulatrices de la transcription

<http://www.dartmouth.edu/~tjack/>



(ou extrêmement faible)

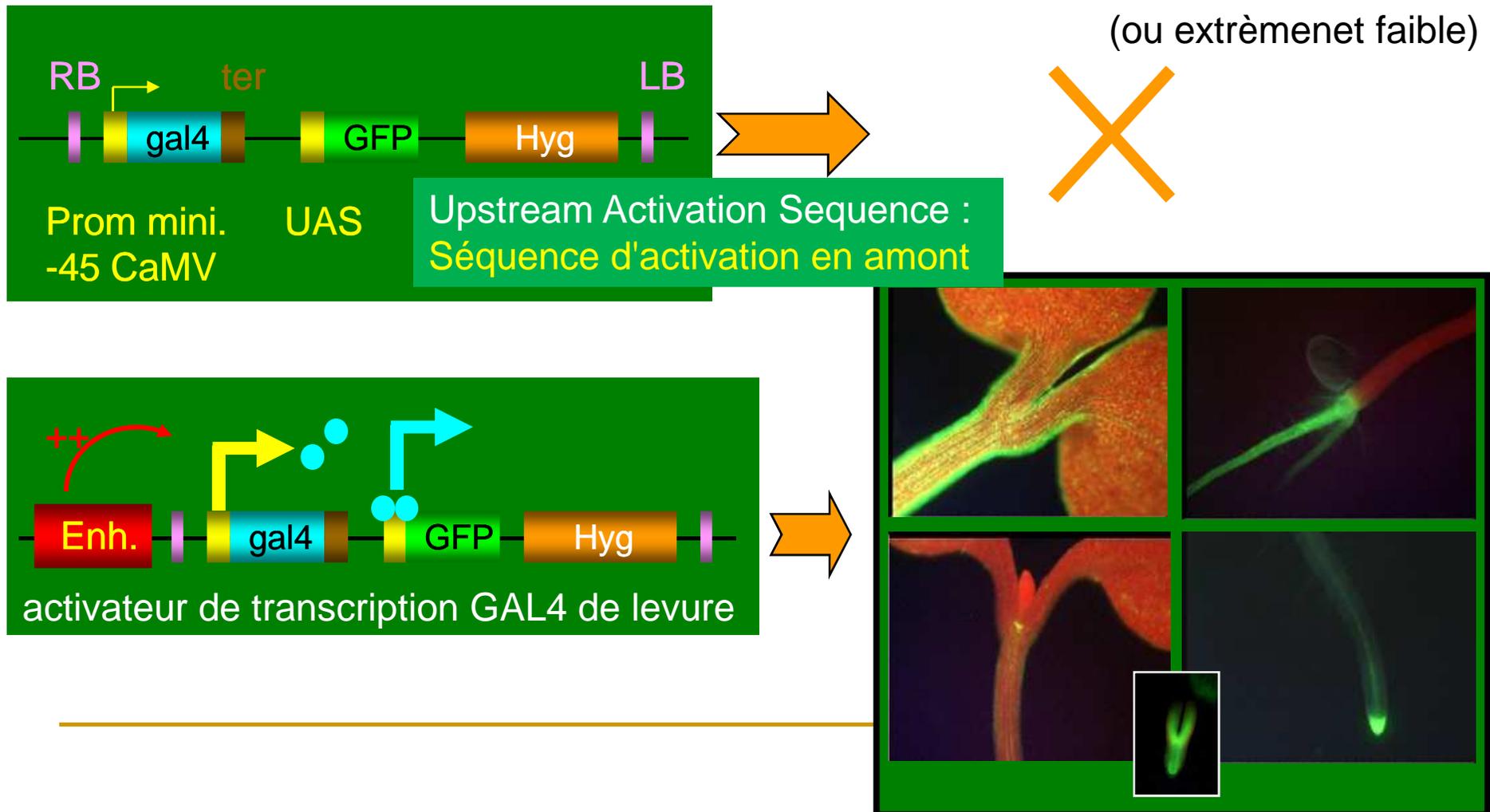


Un enhancer est fonctionnel en amont/aval, orientation direct ou inverse par rapport à la séq. codante.

Piège à enhanceurs (enhancer trap)

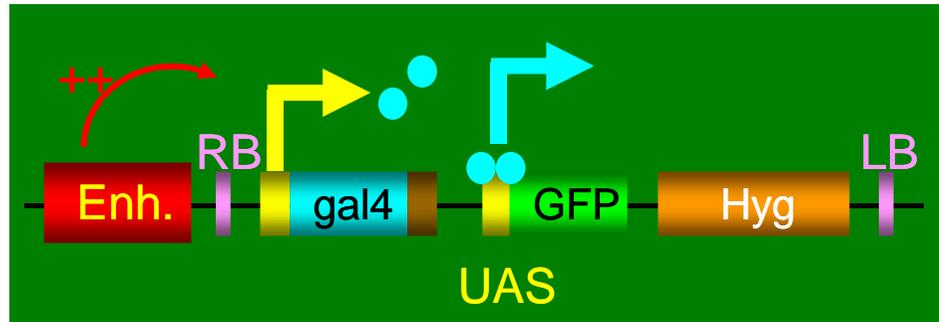
Identification de séquences régulatrices de la transcription

<http://www.plantsci.cam.ac.uk/Haseloff/construction/>

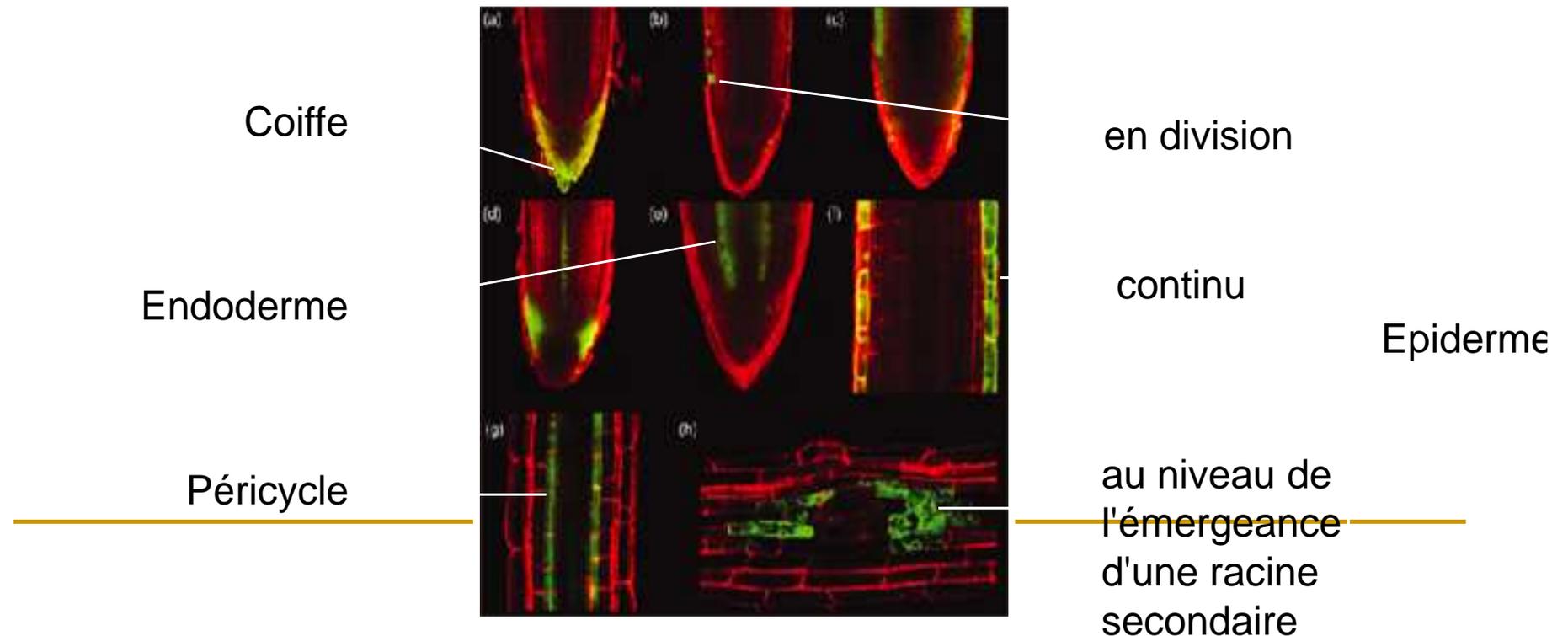


Piège à enhanceurs (enhancer trap)

Identification de séquences régulatrices de la transcription



La même construction a été transférée dans le génome d'*O. sativa* (riz), par transformation et régénération de plantes à partir de cals en culture.



Comprendre la fonction des gènes

Etude ou construction de lignées gain de fonction :

- Surexpression simple
- Surexpression conditionnelle (pour éviter une léthalité par exemple)
 - > la nouvelle fonction constatée est contrôlée par le gène sauvage.

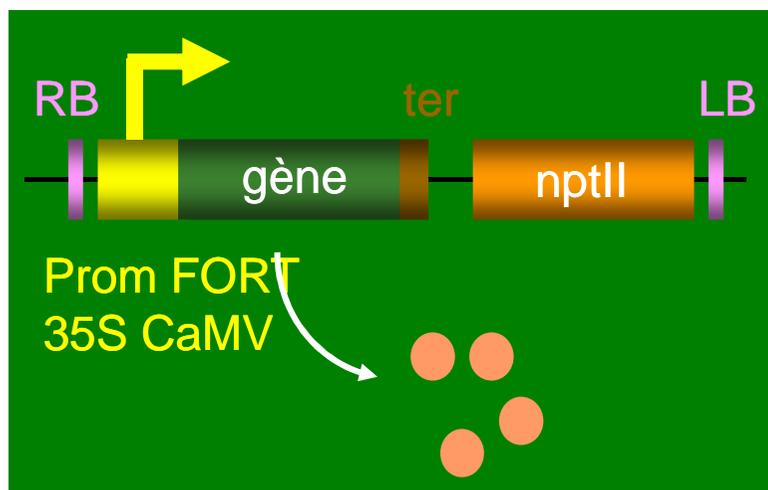
Etude ou construction de lignées perte de fonction :

- mutants naturels, isolés à partir de collections, ou construit spécifiquement
 - > la fonction perdue chez les mutants était contrôlée par le gène sauvage.

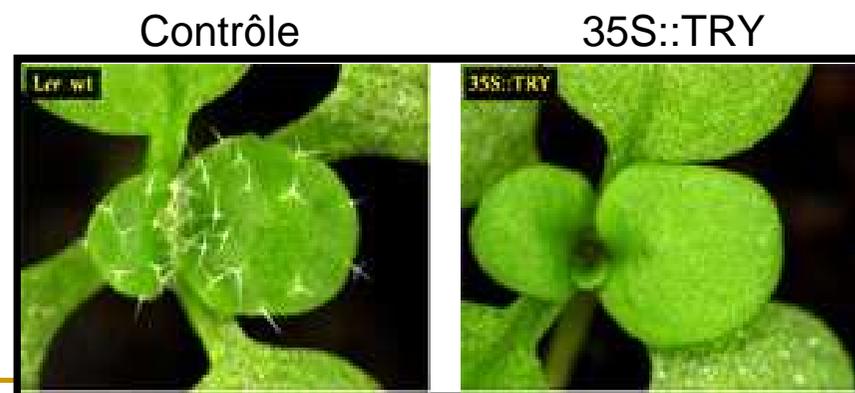
Surexpression d'un gène

La surexpression simple

Fusion transcriptionnelle :



BP stimule la production de méristèmes ectopiques
(Facteur de transcription homéodomaine)

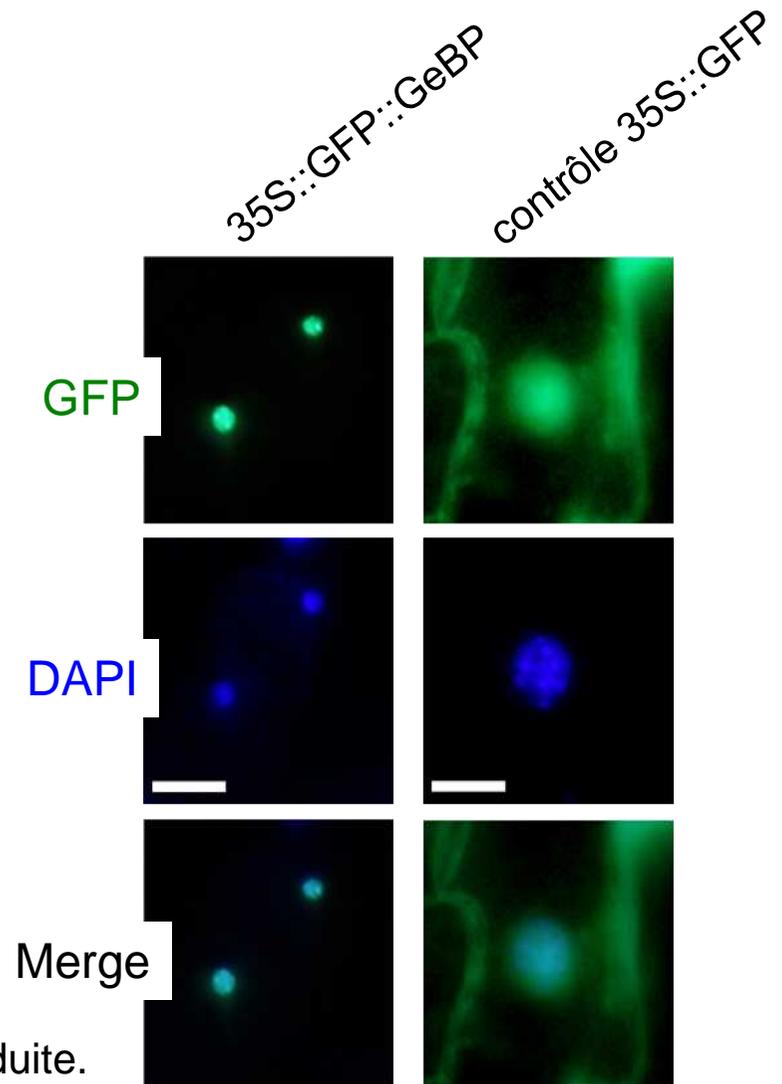
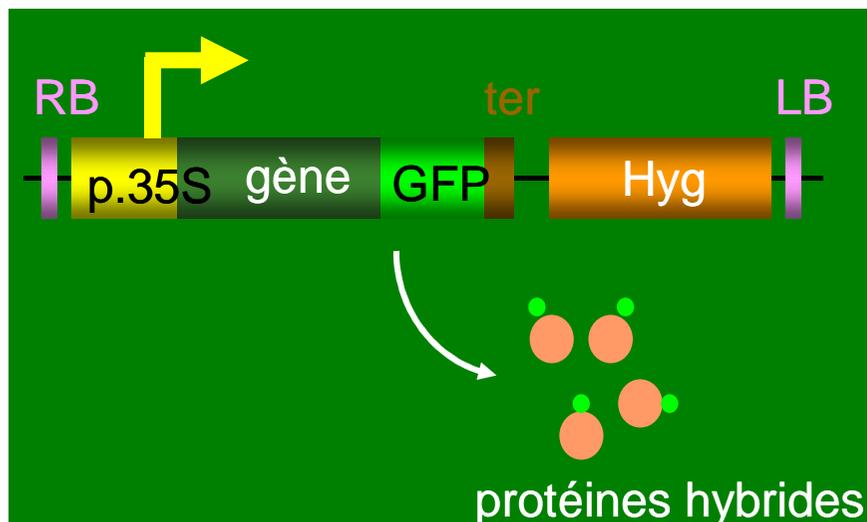


TRY réprime la production de trichomes
(Facteur de transcription Myb)

Surexpression d'un gène avec un gène rapporteur

La surexpression avec rapporteur

Fusion transcriptionnelle et traductionnelle :



La présence de GFP certifie que la protéine GeBP est produite.

La localisation intra-cellulaire de la GFP montre que GeBP est une protéine nucléaire
(Facteur de transcription Leu-Zip)

Systèmes inductibles d'expression de gènes

Il se peut que la surexpression d'un gène à un certain stade du développement soit létale.

On a alors besoin d'un système inductible.

Exemple : promoteur induit par l'éthanol

En absence d'autres sources de carbone, le champignon *Aspergillus nidulans* est capable de métaboliser l'éthanol. En présence d'éthanol, le facteur de transcription ALC R se fixe sur les séq. régulatrices pAlcA et stimule la transcription des gènes du métabolisme de l'alcool (opéron). Par ce mécanisme, *A. nidulans* s'assure de ne produire ces enzymes que si le substrat est présent.



Le gène ALC R codant pour le facteur de transcription est exprimé de manière constitutive.

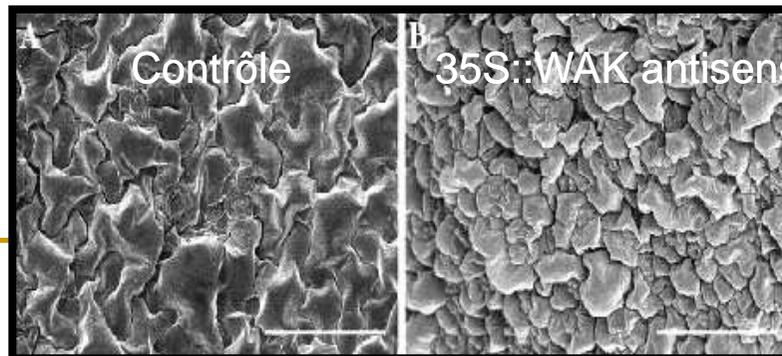
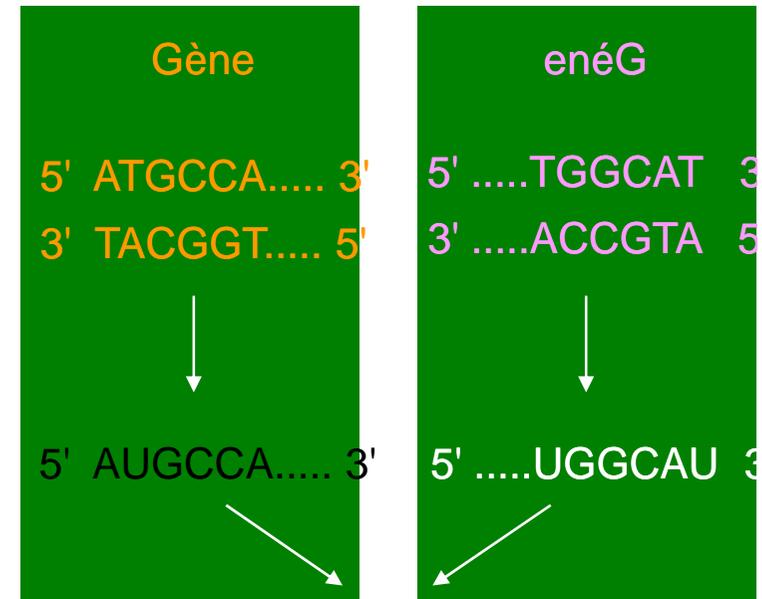
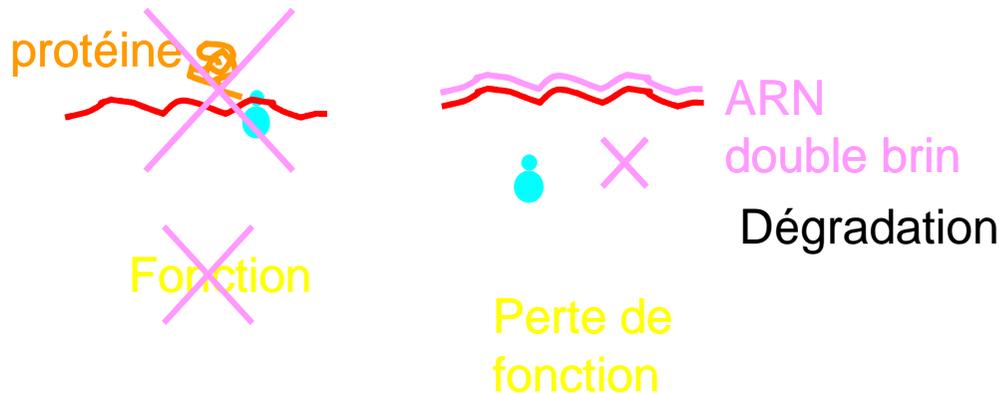
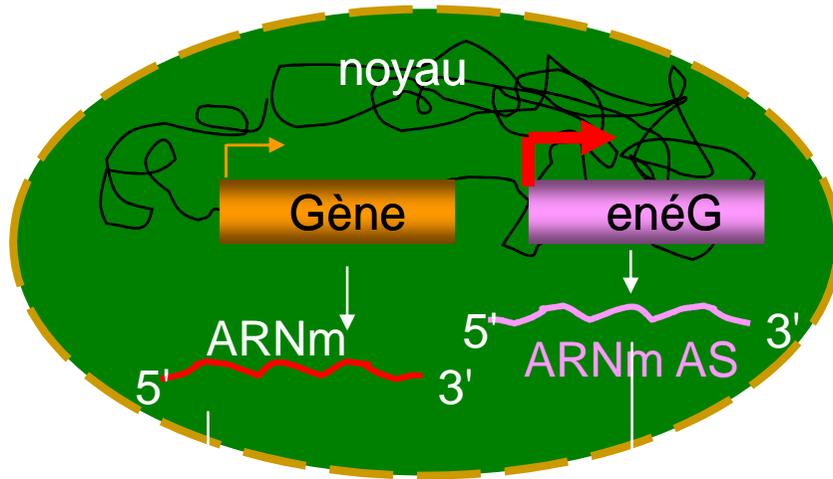
En absence d'alcool exogène, ALC R n'est pas actif.

L'apport d'éthanol exogène permet à ALC R d'activer la transcription à partir du promoteur pAlcA cloné en amont d'un gène d'intérêt.

La protéine à étudier n'est produite qu'en présence d'alcool.

D'autres systèmes existent : choc thermique, métaux, etc.

Inactivation de gènes par ARN antisens



Antisens pour le gène WAK (cell Wall Associated Kinase) impliqué dans les mécanismes d'élongation cellulaire chez *N. Benthamiana* (tabac)

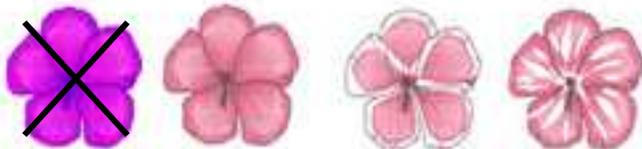
Inactivation de gènes par co-suppression

Mécanisme mis en évidence chez *Petunia hybrida* (1990)

Surexpression de la chalcone synthase

Transformation

F1



fleurs
désirées

wt

co-suppression

transgène et
gène
endogène
exprimés

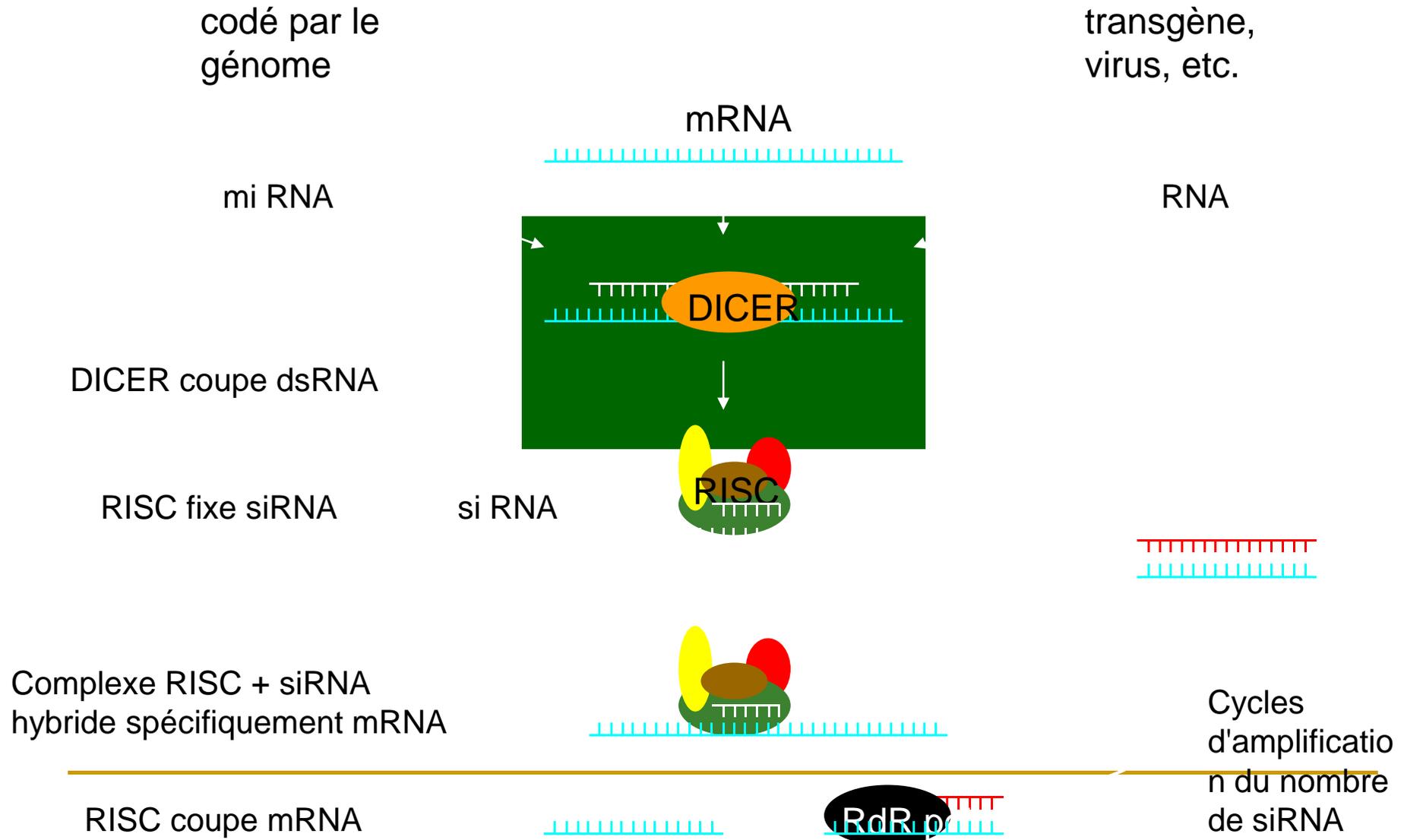
transgène et
gène
endogène
non exprimés



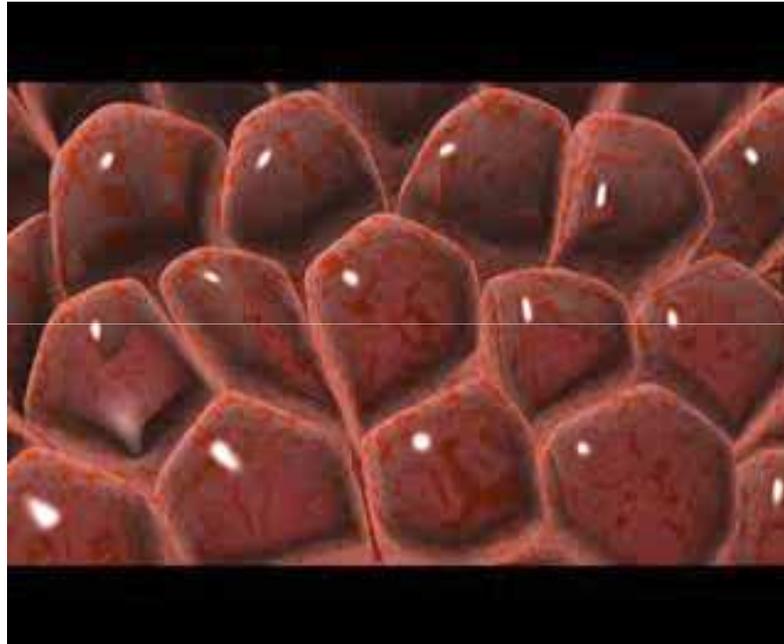
La tentative de surexpression d'un gène aboutit parfois à l'effet inverse : l'extinction du transgène et de la copie endogène.

Mécanisme d'inactivation de gènes par ARN interférence (RNAi)

Co-suppression = Post Transcriptional Gene Silencing (PTGS) = RNAi



L'activité RNAi en vidéo

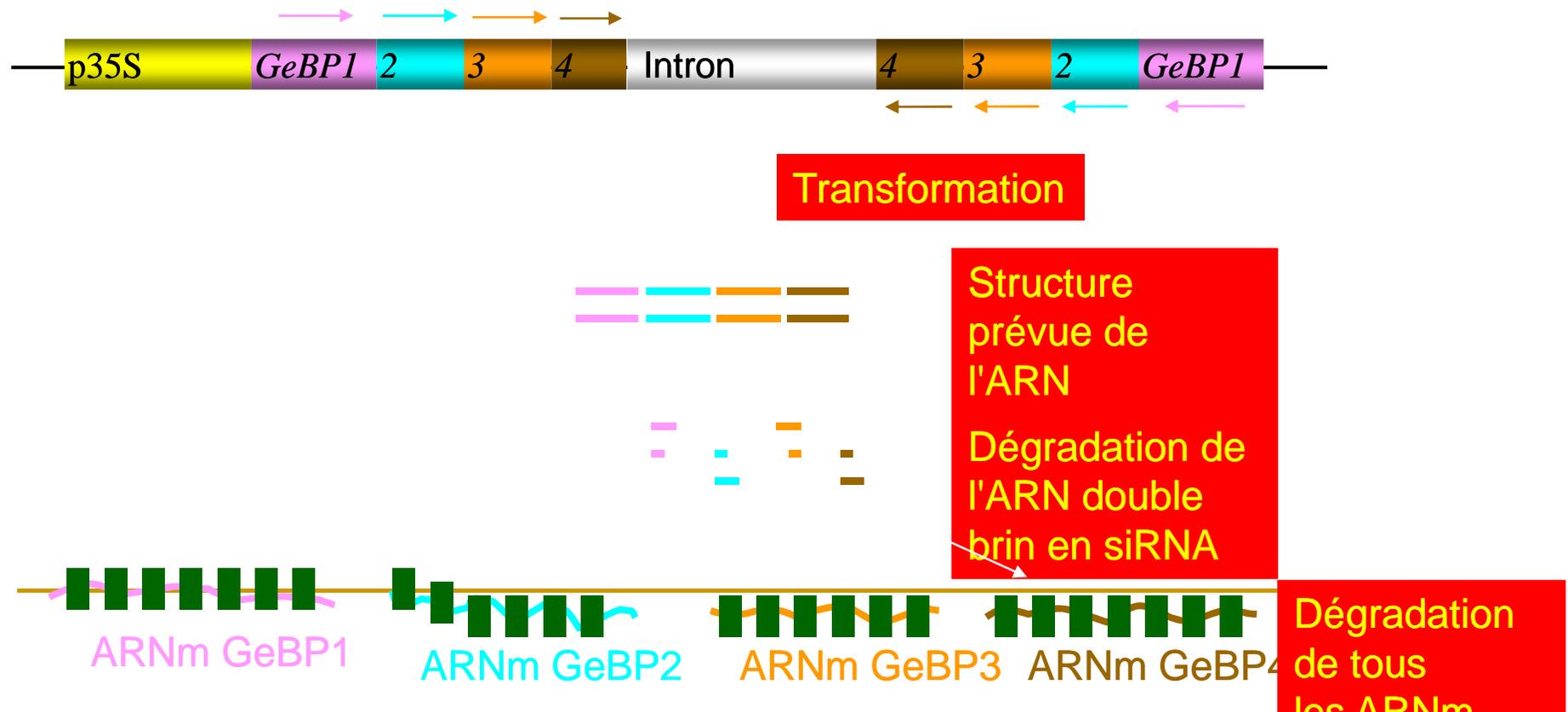


Inactivation de familles multigéniques par RNAi

Chez *A. thaliana*, la famille GeBP compte 4 membres (GeBP 1 à 4). Ils codent tous pour des facteurs de transcription et sont tous exprimés dans les mêmes organes. Leur fonction, encore inconnue est certainement redondante.

Les 4 simples mutants ne montrent aucun phénotype visible.

Construction RNAi visant à sous exprimer les 4 gènes *GeBP* simultanément :

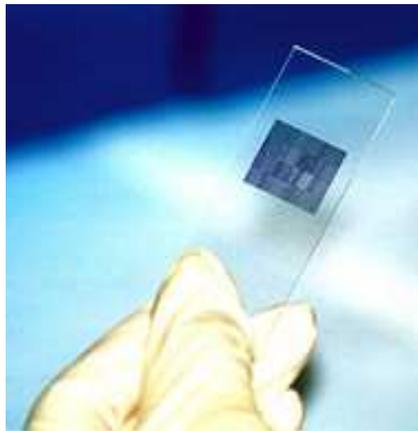


Etude de la transcription à l'échelle du génome

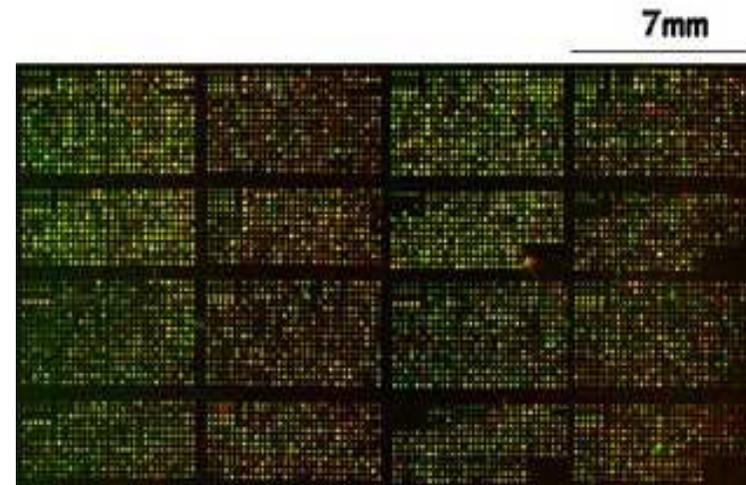
Puces à ADN (micro-array, DNA chips)

Les avancées dans les domaines de la génomique (prog. de séquençage des génomes, etc.) et des nanotechnologies (miniaturisation, automatisation, etc.) rendent possible l'analyse des niveaux de transcription de nombreux gènes simultanément: la transcriptomique.

Chez *A. thaliana*, l'ensemble des 25 000 gènes est représenté sur une puce.



Puces à ADN



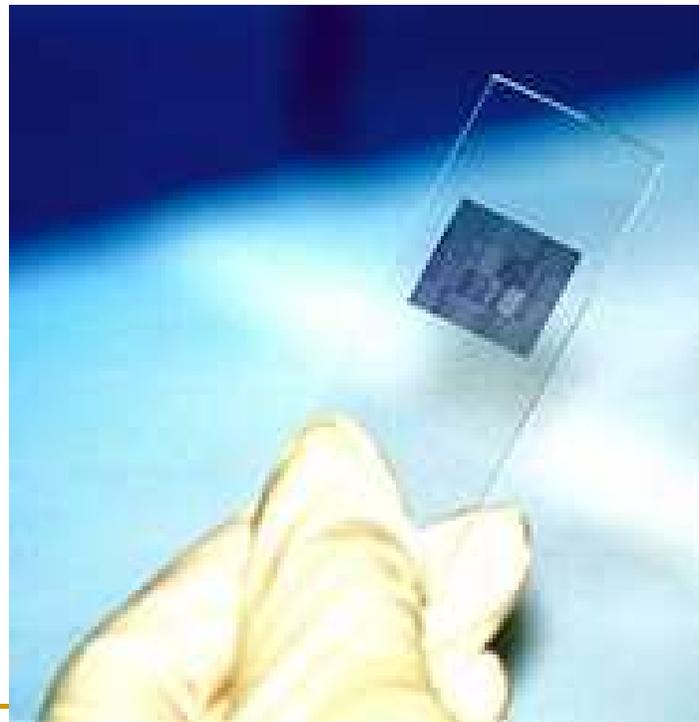
Résultat d'hybridation
d'une puces à ADN

Etude de la transcription à l'échelle du génome

Puces à ADN (micro-array, DNA chips)

Les avancées dans les domaines de la génomique (prog. de séquençage des génomes, etc.) et des nanotechnologies (miniaturisation, automatisation, etc.) rendent possible l'analyse des niveaux de transcription de nombreux gènes simultanément: la transcriptomique.

Chez *A. thaliana*, l'ensemble des 25 000 gènes est représenté sur une puce.

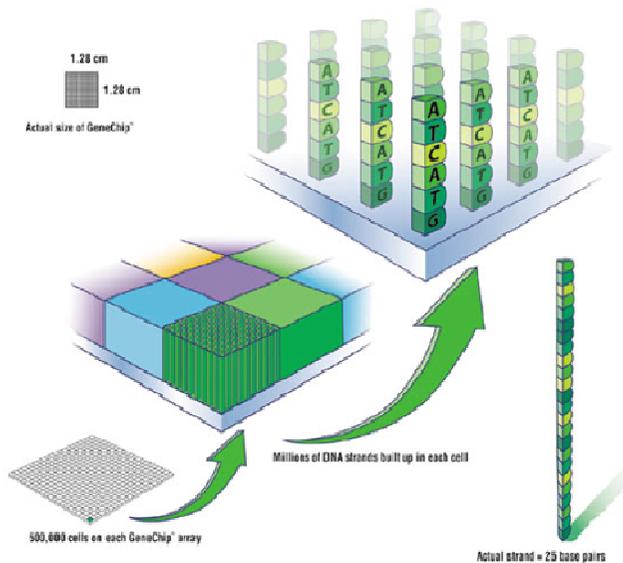


Puces à ADN

Elaboration et hybridation d'une puce

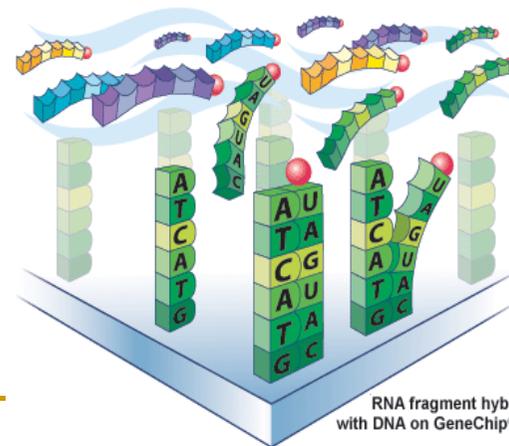


Les puces commerciales



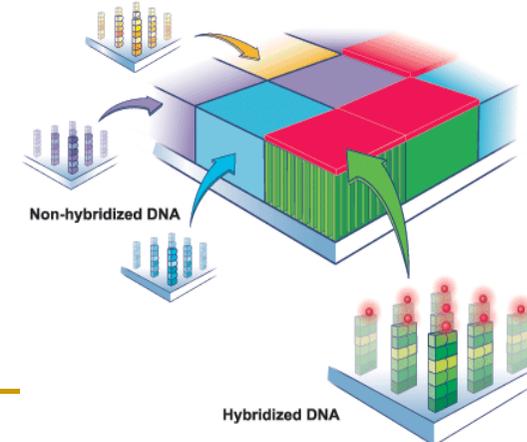
Fabrication de la puce

RNA fragments with fluorescent tags from sample to be tested



Hybridation de la puce

Shining a laser light at GeneChip® array causes tagged DNA fragments that hybridized to glow



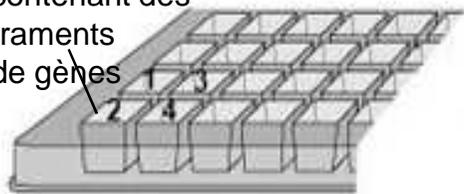
Détection des signaux

From Computer Desktop Encyclopedia
Reproduced with permission.
© 2007 Affymetrix

From Computer Desktop Encyclopedia
Reproduced with permission.
© 2007 Affymetrix

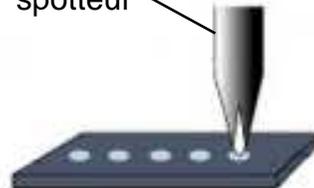
Etapes de l'analyse transcriptomique par micro-array

Plaque contenant des fragments de gènes



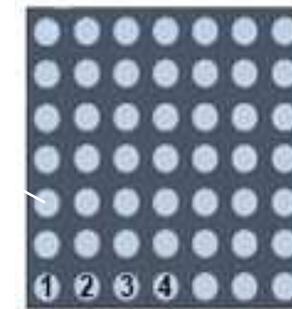
Fragments PCR spécifiques de chaque gène (1, 2, 3, ...)

robot spotteur



Fragments PCR sont déposés sur une lame de microscopie

ADN



Puce ADN (micro array)

hybridation avec un mélange des 2 sondes

sonde ADNc



synthèse d'ADNc avec nucléotides fluorescents

sonde ADNc



ARNm spécifiques des fleurs

ARNm spécifiques des feuilles



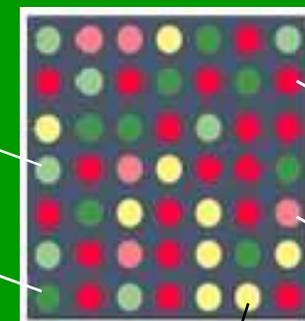
signal faible

signal fort

signal équivalent

signal fort

signal faible



Université Frères MENTOURI Constantine 1

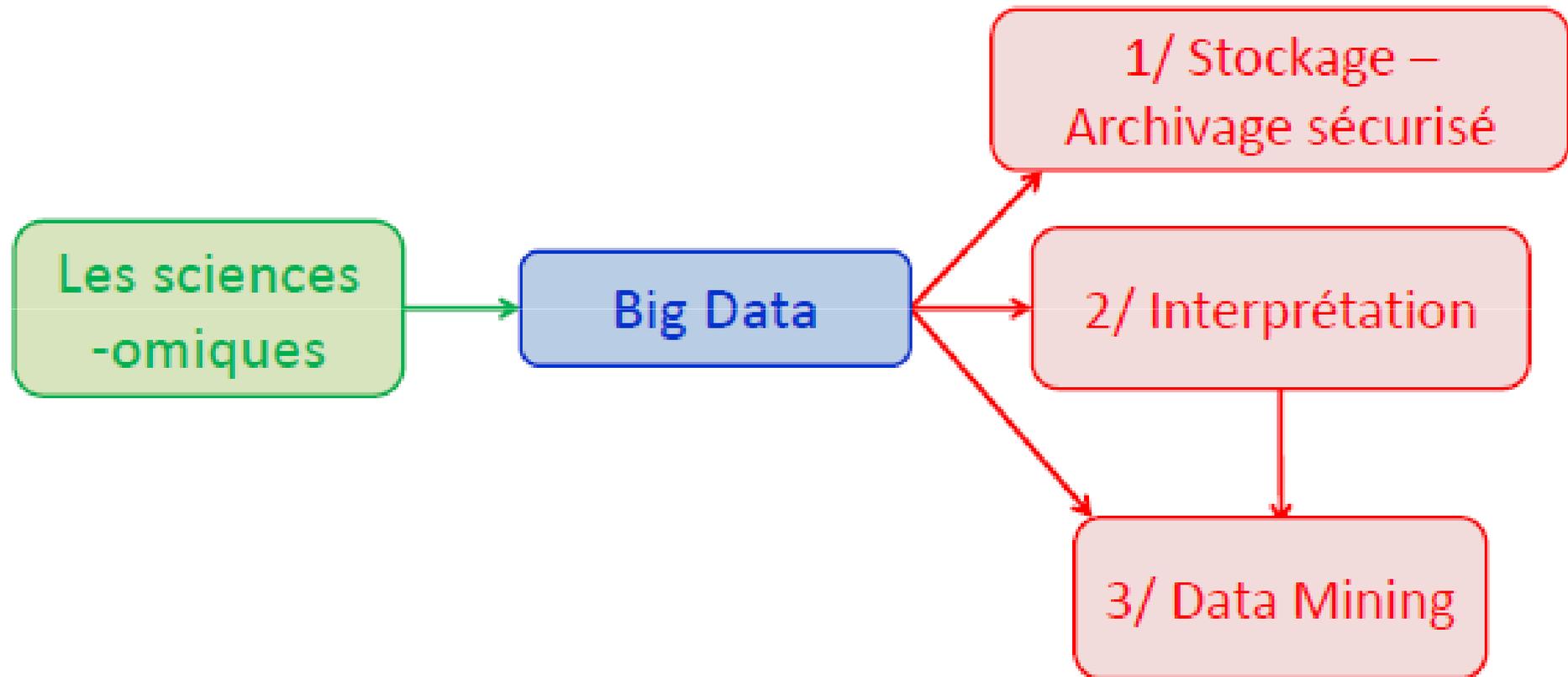
Faculté des Sciences de la Nature et de la Vie
Département de Biologie Appliquée

Master en : Bioinformatique

Cours de : Génomique Fonctionnelle

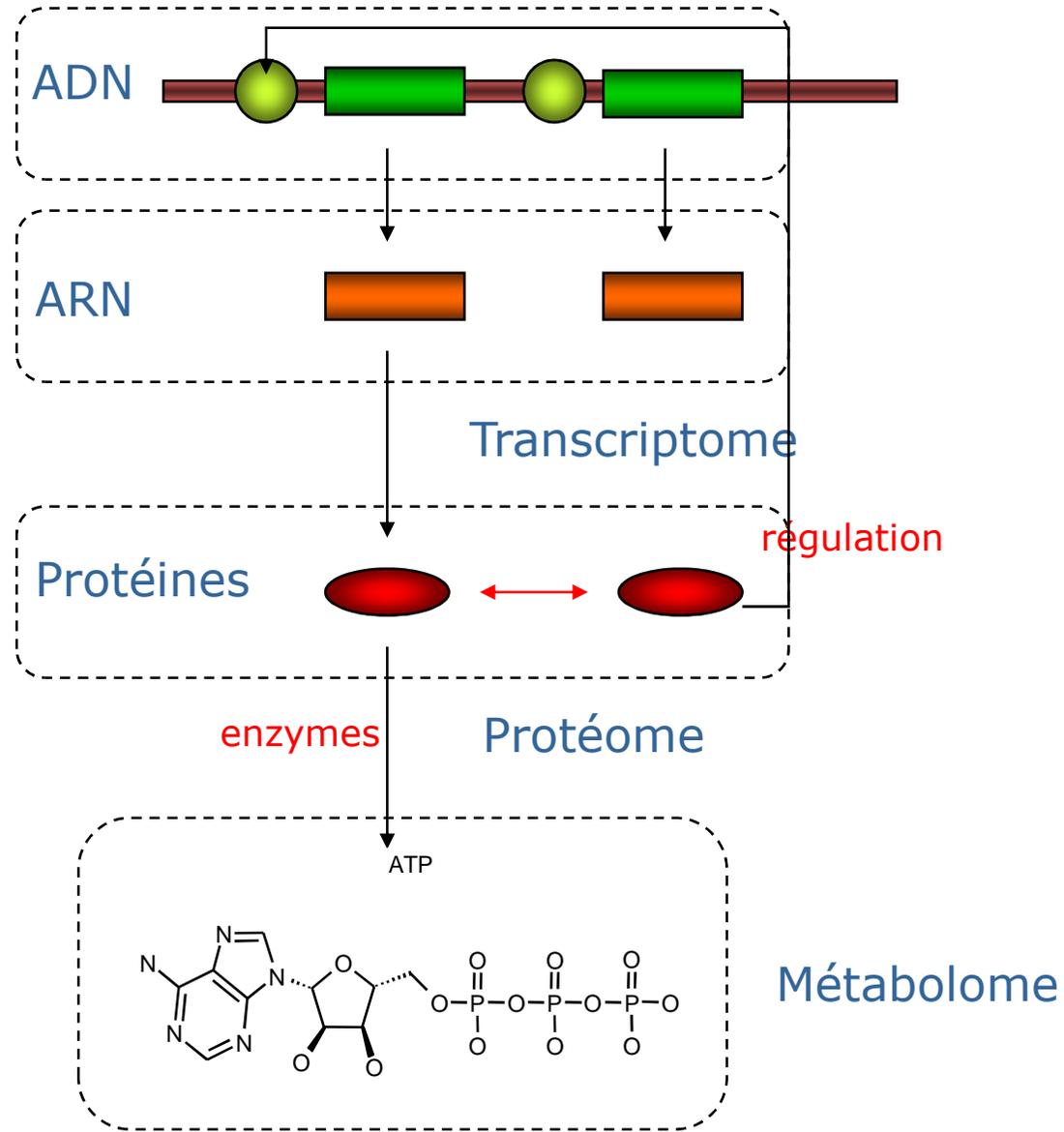
ANALYSE DES GÉNOMES

La problématique des -omiques



La problématique des -omiques

Génome



ADN double brin



transcription



messenger



traduction



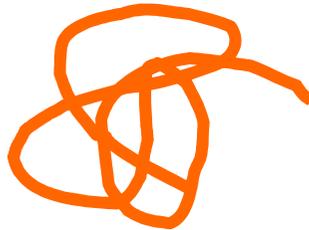
chaîne d'acides aminés



repliement



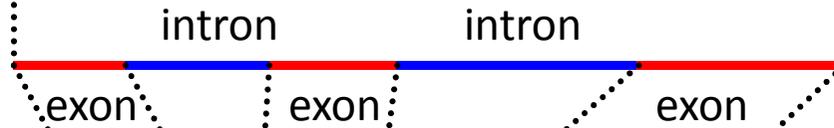
Cellules procaryotes



ADN double brin



ARN pré-messager



transcription

messager



*maturation
(excision -
épissage)*



traduction

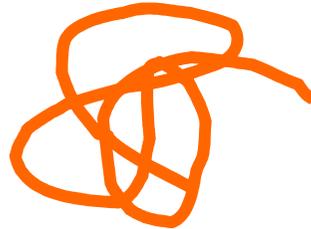
chaîne d'acides aminés



repliement



Cellules eucaryotes

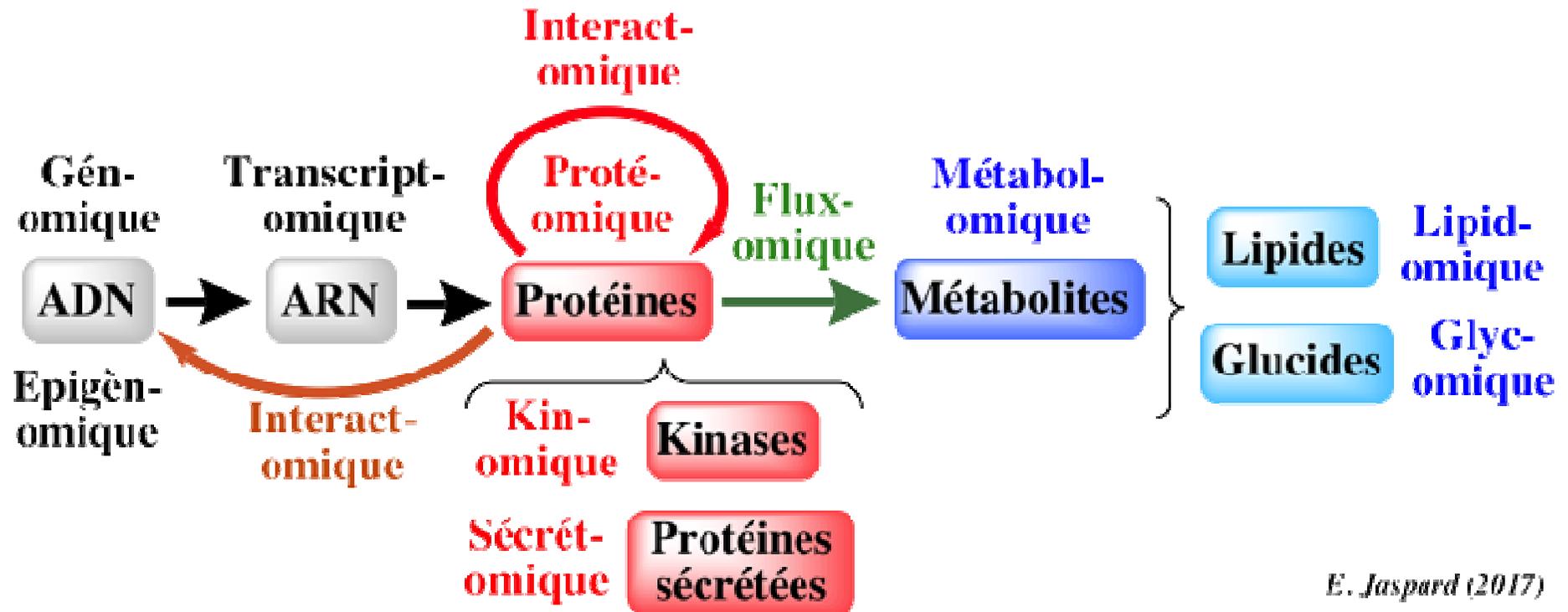


La problématique des -omiques



Métabolomique – Complexomique – Interactomique –
Métagénomique – Métaprotéomique – Protéogénomique, ...

La problématique des -omiques



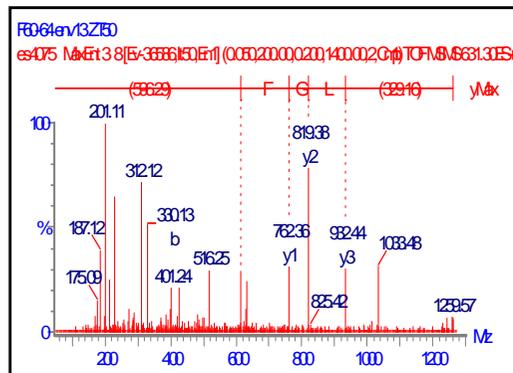
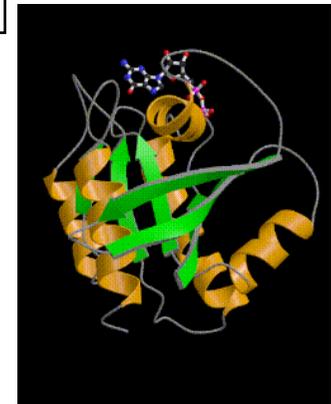
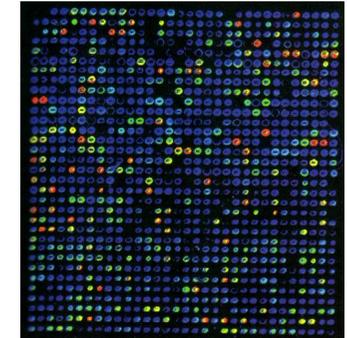
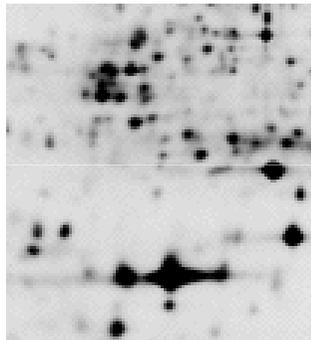
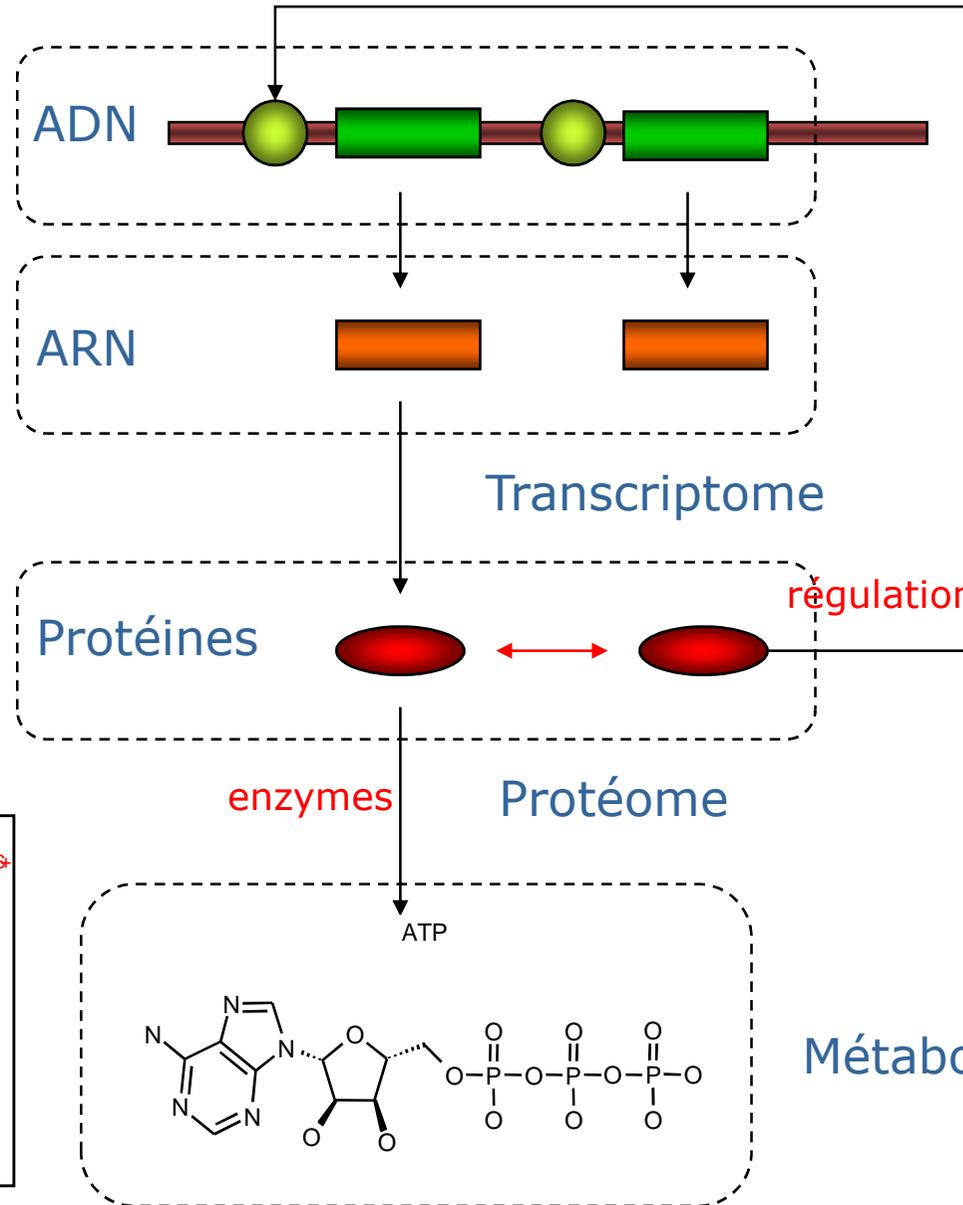
Deux types de molécules support de la bioinformation : les **acides nucléiques et les protéines**

Le "matériau de base" de la génomique, de la transcriptomique et de la protéomique est **la séquence** : l'enchaînement ordonné et orienté de **nucléotides** (acides nucléiques) ou d'**acides aminés** (protéines).

La problématique des -omiques

GATCACCTCACTACGG
 GTCAGGGGAAGGAAA
 GGGGAAGTGAAGATT
 TGTCAGTGTGAGAAGC
 AGTCCCAGGAGTTAGA
 AGTAGTGGCTCCATGA
 CTCACAAATTAAGTTC
 CCTTTCAGGCAGGGCT
 TCTTATTTTCCTTAGCA
 TCCCTGTCTTGATCCCA
 GCCTGCTCAGACCCCT
 GCCTCTCACTGCAAGA
 TGTGCTT

Génome



Métabolome

La problématique des -omiques



Code à 4 lettres
A, T, G, C



Code à 20 lettres
20 acides aminés

Le code génétique

		2 ^e nucléotide							
		T		C		A		G	
1 ^{er} nucléotide	T	TTT	phénylalanine	TGT	cytosine	TAT	tyrosine	TGT	lysine
		TTC		TGC		TAC		TGC	
		TTA	leucine	TCA	alanine	TAA	codon-stop	TGA	codon-stop
	C	CTT		CGT		CAT	histidine	CGT	
		CTC		CCG	proline	CAC		CCG	arginine
		CTA	isoleucine	CCA		CAA	glutamine	CGA	
	A	ATT		ACT		AAT	asparagine	AAT	alanine
		ATC	thréonine	ACC	thréonine	AAC		AGC	
		ATA		ACA		AAA	lysine	AGA	arginine
	G	GTT		GCT		GAT	acide aspartique	GAT	
		GTC	valine	GCC	valine	GAC		GCC	glycine
		GTA		GCA		GAA	acide glutamique	GGA	
	GTC		GCG		GAG		GGG		

La problématique des -omiques



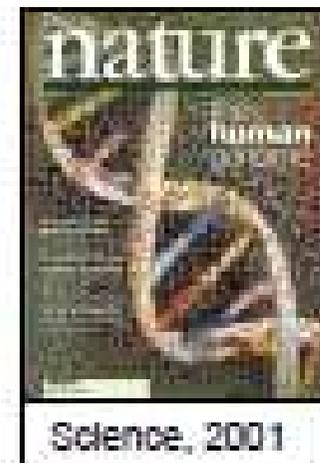
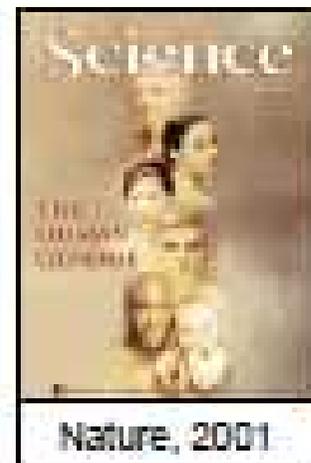
Premier génome séquencé en 1995:
Haemophilus influenzae (Taille $1,8 \cdot 10^6$ bps)



Génome de la levure en 1996:
Saccharomyces cerevisiae (Taille $14 \cdot 10^6$ bps)



Premier draft du génome humain 2001:
Homo sapiens (Taille $3,2 \cdot 10^9$ bps)

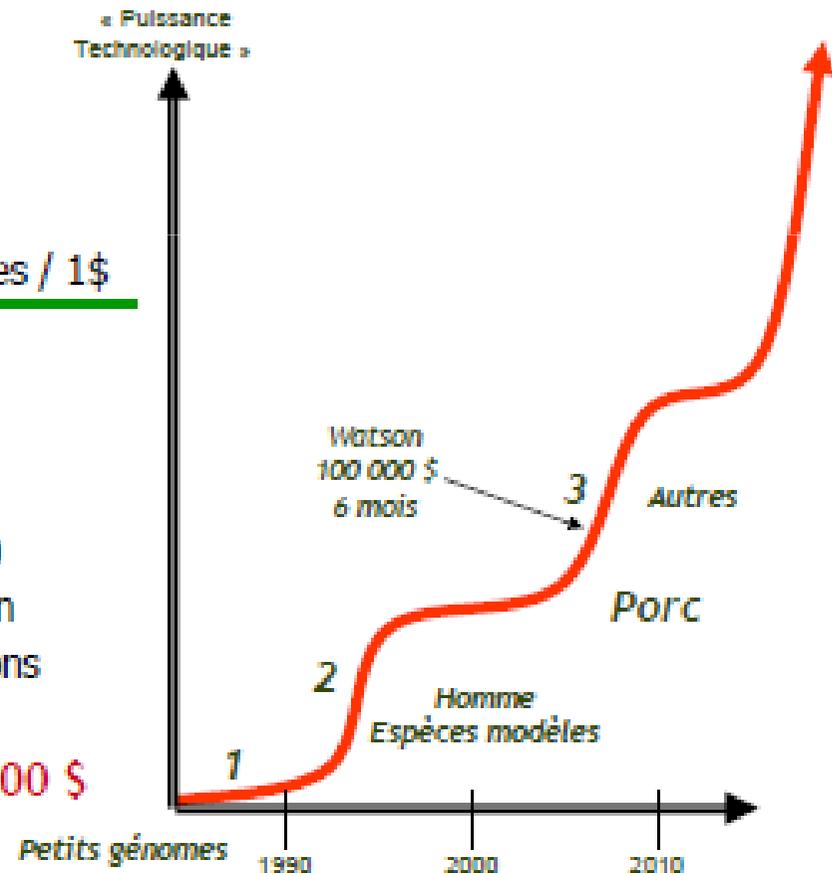


La problématique des -omiques

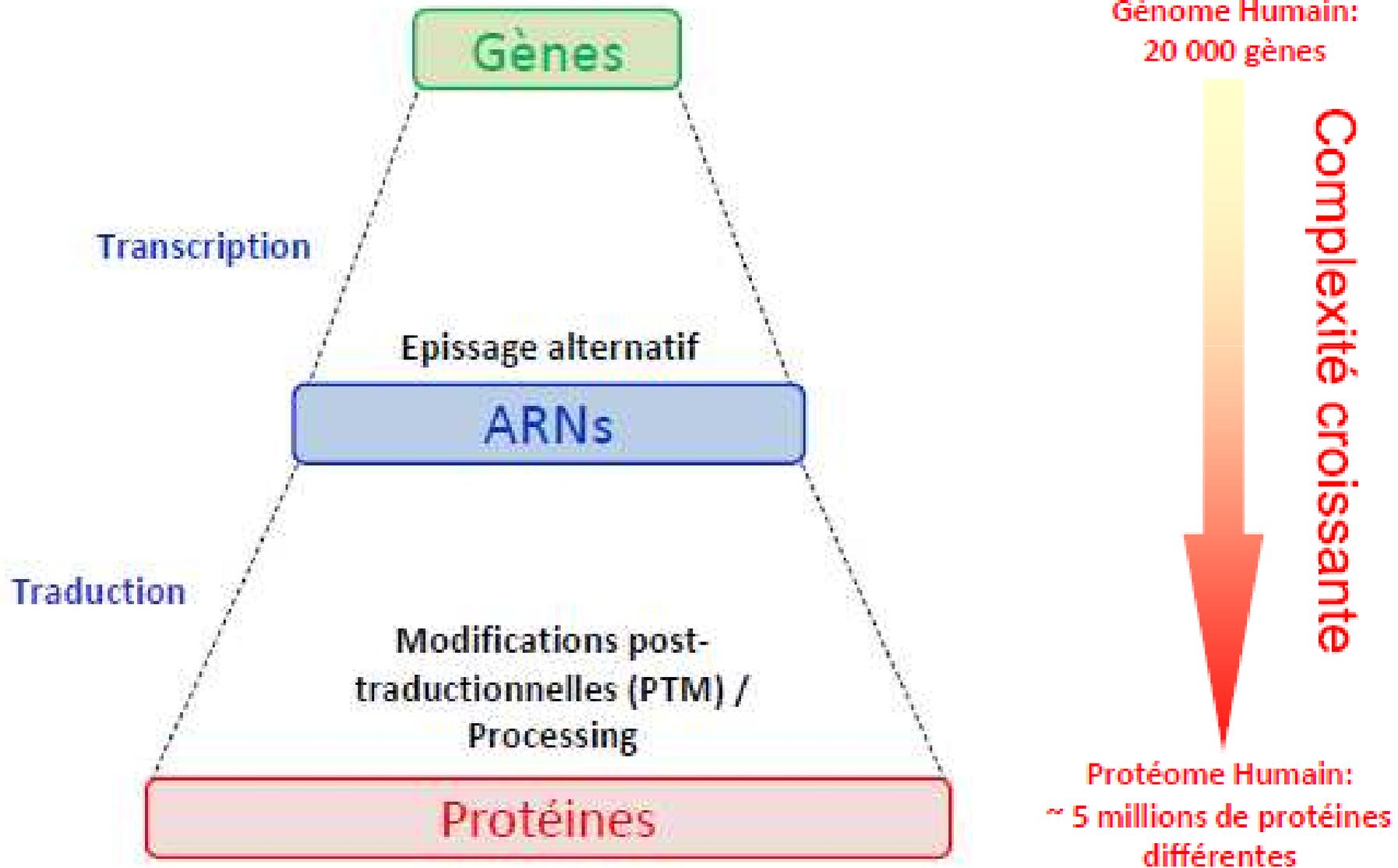


- Séquençage :

	Bases/expé	Nbre de bases / 1\$
< 1990	1200	12
1990-2000	77 000	300
2000-2010	400 000 000	40 000
	10 500 000 000	1 million
	200 000 000 000	6,5 millions
> 2010	1 génome en 15 min pour 1 000 \$	



La problématique des -omiques : **une complexité croissante**



Le codage de l'information génomique

- macromolécule d'ADN \approx
enchaînement d'acides nucléiques
 - adénine : A
 - thymine : T
 - cytosine : C
 - guanine : G
- génome \approx texte écrit dans l'alphabet
de ces quatre lettres

Organisation générale des génomes

Organisés en chromosomes

Chaque chromosome eucaryote contient un ADN linéaire

Répartition des gènes, variable le long des chromosome et région intergénique non codante

Gènes:

séquences codantes continues: cas général chez les procaryote (mais pas exclusif) ou discontinues (exon-intron): cas général chez les eucaryotes (mais très variable d'un organisme à l'autre, rare chez la levure)

Séquences répétées

Dans tous les génomes mais % très varié qui peut être très élevé

Analyses bioinformatique des génomes

Reconnaissance de gènes et autres éléments du génome

Syntaxe des séquences

Recherche de similarité

La bioinformatique permet d'extraire l'information des séquences génomiques.

Le génome comme un langage :

- Support = polymère linéaire
- Alphabet = molécules
- Mots = 3 lettres parmi 4
- Syntaxe = en cours de déchiffrage

Superposition de signaux :

Pas seulement quel est le produit du gène, mais aussi où est-il exprimé, en quelle quantité et quand ???

La nature du contenu informationnel de l'ADN

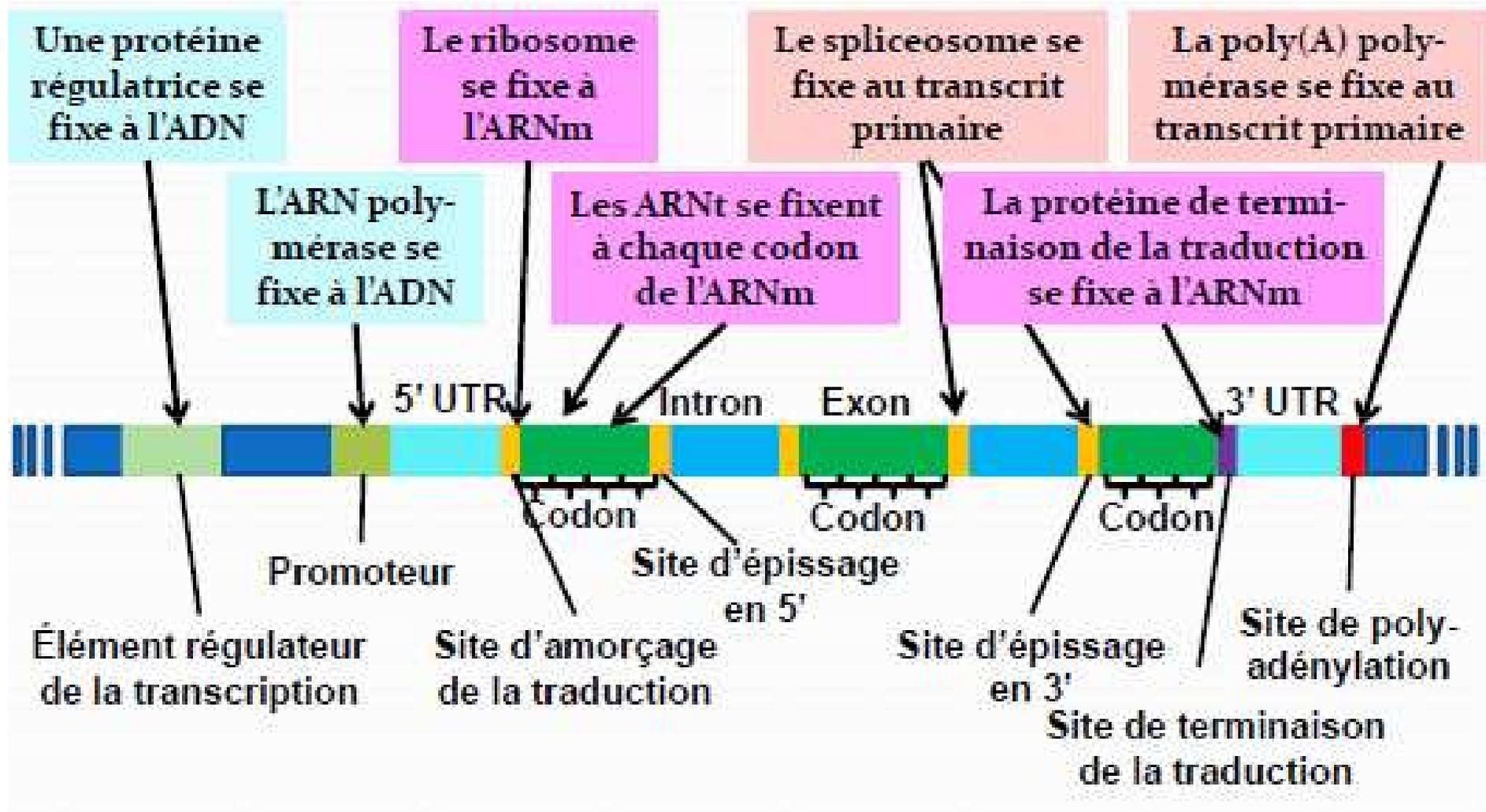
Par convention, on considère que l'information est la somme de tous les produits de gènes, c.à.d. des protéines et des ARN.

* L'ADN contient l'information, mais de quelle façon celle-ci est-elle codée ?

* De nombreuses protéines se fixent au niveau de sites présents sur l'ADN lui-même, tandis que d'autres protéines et des ARN se lient à des sites présent sur l'ARNm.

* La séquence et les positions relatives de ces sites permettent aux gènes d'être transcrits, épissés et traduits correctement au moment adéquat et dans le tissu approprié.

La nature du contenu informationnel de l'ADN



Le contenu informationnel du génome comprend les sites de liaison

Détection de gènes eucaryotes

La recherche de gène dans les génomes eucaryotes est plus complexe.

Les régions codantes sont morcelées sur d'énormes distances par les introns et les régions codantes représentent ainsi moins de 5% du génome eucaryote.

Détection de gènes eucaryotes

Trouver les gènes à partir de la séquence

5' CCGTCGGGCTGAAGGTCCGCGCCGACGTCAACGCCGTGCGGCCCGAGGGCATAGCCGGCGTGGCGTGATGGACGATGCCAAGCCGCTGGC
CGGCACCCAGGGCGTGGGCATCGGGCGGTGGCCATCGGGCAACGTCAAGTACCAGACCCAGCACCCGGCTGCTGCAGCGCATGGGGGAAG
CCGAGAAGGGCGGTCTGCTACAGCTTCGGCGACGCATTCGAGAACCCGACCGCGCTGGCTGGCCCGAGAAGGCCCGCGGGCGGGCTGATGC
TGGCGGTGCGCGCCCTGTCCGCGCGGGCGGTGGTTCGACCTGGCGGGCGCTCGACCGGCATGGCGGTCTGTCGGCTCGACGTCTTCGGCCAC
GCCGACACCGTCAAGCGCGCCGCGCACTGGCATCCGATCGGCACGCCGGCCGGCTGGAGATCGACGGCACCGCGGTGCTGGCGGCGCT
GGAAGCGTGGCGCGGACGGCGACGTGACGGCTGGATCGCGGGCGGGTTTTCGACGGCCGCCCGACCTGCTCGACGCCGGCGCCG
AGCGCTGCGGTGCTGGGCACCGCGGGCGCGGTGCGCCGGCTGGCGGACCCCGCGCGCTTCTTCGCCGCCCTCGACGACCTGGGC
CTGCCGCATCCGGCGGTGAGCTTCGAGCCGCGCGCCGACCCCGCGGGCTGGCTGGAGAAGGACGCCCGGCGGACGCGGGCTGGCACGT
GCAGGACCGGGCGCCGAGCGCCCCCGCGCCCGGCCGTA CTGGCAGCGCTGGCGCCCGGGCCAGGCATGTCGGCGACGCTGGTCC
CCAACGGCCACGACCGCGTCTGCTCGGCTTCAACCTGCAGACCGTGGCGCCGGTGGCCGGCCGGCGCTGGGTCTTCGCCGGCATCGTCC
GCCCGCTGCGCGTGGCGCGGCGGTTCGATEGAACCTCGTGGGTGGGGCTGGTGTCTGGCACGGCGTTTCGACCTGCACGGCTGGCCA
GCCTGGACTTCTGCTCGACGGCGAACACGCCGAGCTGCTGGAACTCAACGCCCGCGCCCGCGGCAGCGCGGAGCTTACCCCGAGGTCC
GACGCGGGCGGCGCTTGGCGGCCCACTGGGTGGGTGACACGGCGCGAGCTGGCGCCGCCCGCGCGCGCGGGTGTGAACGGC
CACGAGATCGTCTTCGCACGCCCGCGCTGGTGTCTCGACGACCTCGCCGACGGCGCATCGCCGCCACGCCGCTGGCGCGGACTGGCCG
CGTGGCGGCCAGCGTTTCGACGTGGCGGACCCCATCTGCAGCCTGGCGGTGGCCGGCCGATGCCCGCGAGGTGCTGGCGGGCGCTGGC
CACACGCCCGGAGGCCCTTGTTCGCTTCTCTGGAGAACCGATGAACGACCGCTGTGGCGCGCCGCTGCCCGGGCGCCACGATCGCGCTAAC
GAGCACGTCCGACCCCTGGTTCGAACGCTGTGTGCCGACGCCCGGGCGCTGGGGTTCGAGGTCTCGCGGACGAACGGCGGTGGGCT
CGTCCAGCCCGGCATCGCCGCGCCGGCCAGCGTCCCGCCGGGTGCTGGTCCGGGAGATCTGCCTCGGCGGCTGGGCCGCGTCCGAGC
TCCGCGCCCGGCCCGACTGGCGGACCTGGGTGCAGTGGCGGAGCTCGCTGCCGGTGTCTGGCTGCTGGGCTCGCAGTACGCCGGCTGG
AGCTGGCGGCCAGCAAGGAAGAGACCGCGCGCAAGAAGTTCTTCGCGCTGGGCTCGGGGCCGGCGCGTGGCTGGCGGCCAAGGAGG
CGCTGTACGGCGAACTCGATTCGGCGGACCCCGCCAGCCCGGGCGTGTGGTGTATGGAGTTCGACCGGCCCGCCCGCGCCGCTCGTCTC
GACAAGATCTTCGGGACTGGCGCTGGCGCCCGAGGGCGCTGACGATCGTGTGACGCGCCGACCCGACGCGCCCGCGCACGACCGATGA
ACGACCGCTTCGCGCGCGGCTGCCCGCGCCACGATCGCGCTCAACGAGCACGTCCGACCCCTGGGTCCGACCGCTGTGTGCCGACGCCG
CGGCGCTGGGCGTCGAGGTCTCGCGCGACGAACCGCGCGTGGCGCTCGTTCGACCGCCGGCATCGCCCGCGCGGGCAGCGTCCCGCCCGG
CTGCTGGTGGCGAGATCTGCCTCGGCGGCTGGGCGCGCTCGAGTCCGCGCCCGGCCCGACTGGCCGACCTGGGTGCAGGTGGCCAG
CTCGCTGCCGCTGTGGCTGCTGGCTCGCAGTACGCCGGTGGAGCTGGCGGCCAGCAAGGAAGAGACCGCGCGCAAGAAGTTCT
TCCGCTGGGCTCGGGCGCGCGCGTGGCGTGGCGGCCAAGGAGCGCTGTACGGCGAACTCGATTGGCGCGACCGCGCCAGCCCGCGC
GTGCTGGTGTATGGAGTCCGACCGCGCCCGCGGCCGCTCGTCTCGACAAGATCTGGCGGACTGCCCGCTGG3'

Détection de gènes eucaryotes

Organisation et structure des gènes « Protéiques » chez les eucaryotes

LagénomeodcbighdccoehchiquezhvbzdcizqhcokqsikeiutrzevuzeidcvbCifdésigneladis
fxqghklmpojqsiaiohcsbcoiohsodjsqjixchcqyxnlqsqshsnchgdqqsoqqpCqpcCcdgjlCj
sjpciplinescienqshxhxqxioXIIItifiquebcjqoqpchhizpps,xqioqsogjydsguipgvaddiXI
XXIOISQIfsdfrittykylibvqhsduzisklxlxhjhchghgchhchsksn,ndoidopezpsmskqcgq
ucvwwwxwdtyhcentréesurjqpcjjcqocccokqsikeiuzjqsiaio,qddzaztrykjdkoljtvlac
rtogccqscqvfg,hk;bscqfjiilopjsdhhjdcizeodcbighdcqsqsazdzaphiedgdjqspqqsiqs
opqpscqpjdiksoaoqjknssndshvsdfsfshhgloqksdgsauaqrwnwsschediokcjcjcds
dfghkcohhchqhbcsbcoiohsodjsqjixchcqyfxqhgqqsoqqpCesgénomepqcCcdgjlCjsj
pdsdvsdvezbnj,uiyterrogjydsguipgvaddiqshxhxqxigdjqqspqqsiqsopqpscqpjdiksoa
oqjknssndshvsdfsfshhgloqkauaqrwnwdfghkhkhjdcizeodcbighdccoehchqhcokqsik
eiuergzaqcqvzjqsiaiohcsbcoiohsodjsswsschediokcjcjcdsdfghkhkhjdcizeodcbighdco
ohchqhcokqsikeiuzjqsiaiohcsbcoiohsodjsqjixchcqyfxqqpCqpcCcdgjlCjsjpvgrgtjyk
ililloleergrrergrrrgerqqqqogjydsguipgvaddiqshxhxqibcoiohsodjsqjixchcqyfxqhg
dqqsoqqpCqpcCsodjsqjixchcqyfxqhgqqsoqksikeiuzjqsiaiohcsbcoiohsodaiobcsbc
oiohsodjsaqrwnwsschediokcjcjcdsdfghkhkhjdcizeodckeiuuzjqsiaipgvaddiqshxhxq
ioXIIIXgfhkhkhjdcizeodcbighdccoehchqsetleséquenceqqpCqdsdfghkhccoehchqhc
odeLADNhcokqsikeiuzjqsiaioh

Détection de gènes eucaryotes

Organisation et structure des gènes « Protéiques » chez les eucaryotes

Lagénome eodcbighdcco hchiquezhvbzdcizqhcokqsikeiutrzevuzeidcvbCIf désigne la dis
fxqghkdm pojqsiaio bcsbcoiohsodjsqjjxchcqyxnl dsqshsnchgdqqsoqqpCqpcCcdgjlCj
sjpciplinescienqshxhxqxioXIItifiquebcjqoqpchhizpps,xqioqsogjydsguipgvaddiXI
XXIOISQIfsdfrittykylibvqhsduzisklxlxhjhchghgchhchsksn,ndoidopezpsmskqcx
ucvwwwxwdtyhcentréesurjqpcjjcqqccccokqsikeiuzjqsiaio,qddzaztrykkloljtvla
rtogccqscqyfg,hk;bscqfjiilopjsdhhjdcizeodcbighdcqsqsazdzraphiedgdjqqsppqqsqs
opqpscqpjdiksoaoqjknshvsdfsfdfshhgloqksdgzsauaqnwnwsschediokcjcjcds
dfghkcohhchqhbcsbcoiohsodjsqjjxchcqyfxqhg dqqsoqqpCesgénomeqpcCcdgjlCjsj
pdsdvsdvezbnj,uiyterrogjydsguipgvaddiqshxhxqxigdjqqspqqqsqsopqpscqpjdiksoa
oqjknshvsdfsfdfshhgloqkauaqnwndfgfhkhhjdcizeodcbighdcco hchqhcokqsik
eiuergzaqcqvzjqsiaio bcsbcoiohsodjswsschediokcjcjcdsdfghkhhjdcizeodcbighdco
ohchqhcokqsikeiuzjqsiaio bcsbcoiohsodjsqjjxchcqyfxqqpCqpcCcdgjlCjsjpvgrgtjyk
ililloleergrrergrrrgerqqqogjydsguipgvaddiqshxhxqxibcoiohsodjsqjjxchcqyfxqhg
dqqsoqqpCqpcCsodjsqjjxchcqyfxqhg dqqsoqqsokqsikeiuzjqsiaio bcsbcoiohsodaio bcsbc
oiohsodjsaqnwnwsschediokcjcjcdsdfghkhhjdcizeodckeiuuzjqsiaipgvaddiqshxhxqx
ioXIXgfhkhhjdcizeodcbighdcco hchqsetleséquenceqqpCqdsdfghkchcco hchqhc
ode l'ADNhcokqsikeiuzjqsiaio b

Détection de gènes eucaryotes

Organisation et structure des gènes « *Protéiques* » chez les eucaryotes

La génomique désigne la discipline scientifique centrée sur la cartographie des génomes et le séquençage de l'ADN



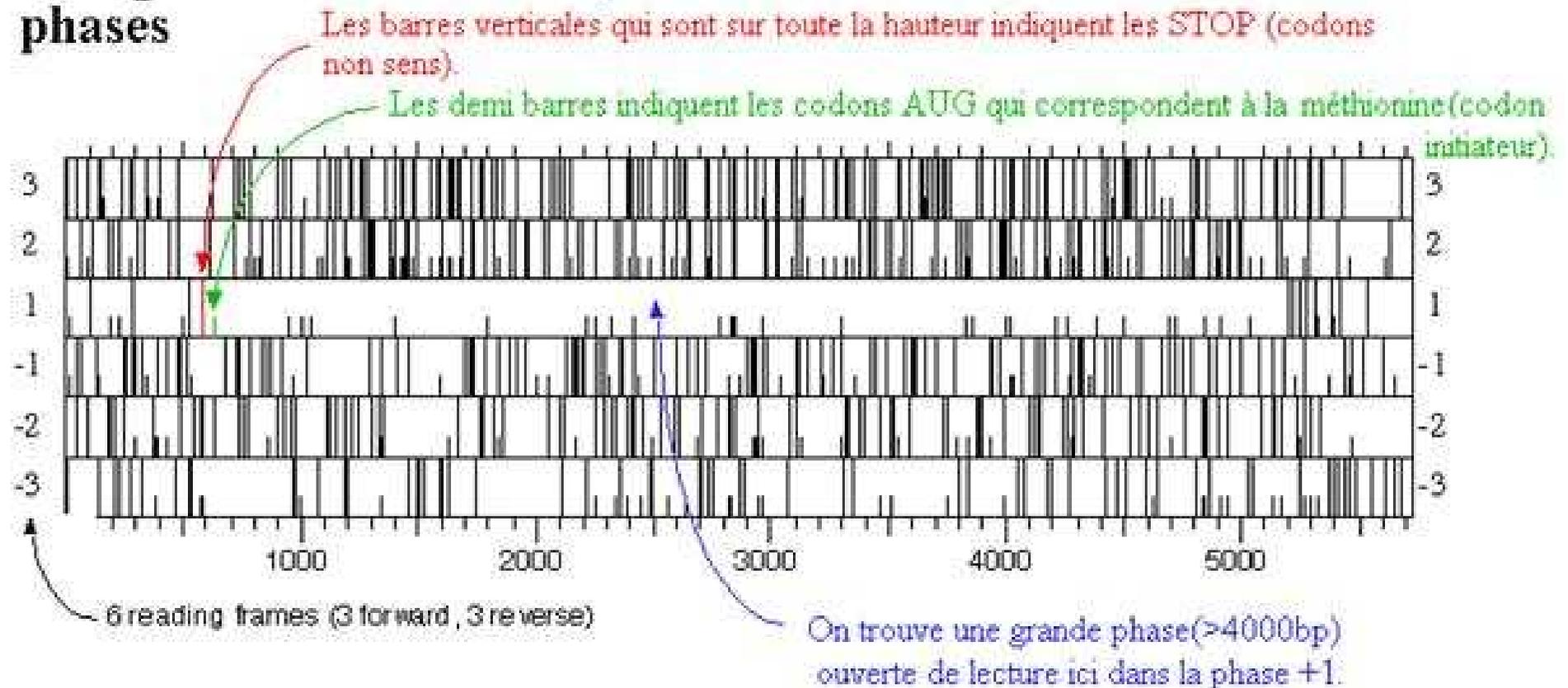
La génomique désigne la discipline scientifique centrée sur la cartographie des génomes et le séquençage de l'ADN

Détection de gènes eucaryotes

Séquence à analyser

```
ttgtatgtataatattcaaegeattttttatgceegtogt agttgetaect accacaegatgatgt attataaatt egga  
actacttcaaegetacgotttcoogcactatgaaaagctataagtatgtgaaaattgggtcattgactgtgattotatttca  
aaaagt aaccaaggagat agacat ogt toggac agaaaatggteogttgtcactcccattccattatgtccttcataa  
gaaattggatattcttgttttcatttccogogaat aaagcaattccogtgoggc agaaagaottatataatatttaactgat  
cttacgctatttatggcaaaacttgtgttacatttttgaagat aaagttacaatcattacggcagcctcaaaacaaaatt  
gggagaaaacat actcaagtgagt actcattttgtgcaagcaaacactgacaattgaagagat ogtcaggATGCCCGAA
```

Le logiciel recherche des phases ouvertes de lecture dans les 6 phases



De très nombreux domaines de recherche en informatique, automatique et mathématiques appliquées sont concernés

- algorithmique sur les séquences, sur les graphes...
- statistique, analyse de données
- apprentissage symbolique et numérique
- visualisation de données
- modélisation et simulation dynamiques
- calcul parallèle
- bases de données et de connaissances

Transcriptome

Transcriptome : ensemble des ARNm ou transcrits présents dans une population de cellules dans des conditions données.

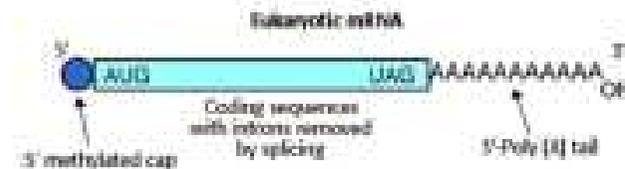
➔ Accès au niveau d'expression de milliers de gènes simultanément (potentiellement l'ensemble des gènes d'un organisme)
= *instantané* de l'état d'une cellule ou d'une population de cellules

Données d'expression des gènes obtenues par :

- qPCR
- Puces à ADN
- Séquençage ultra-haut débit

Transcriptome

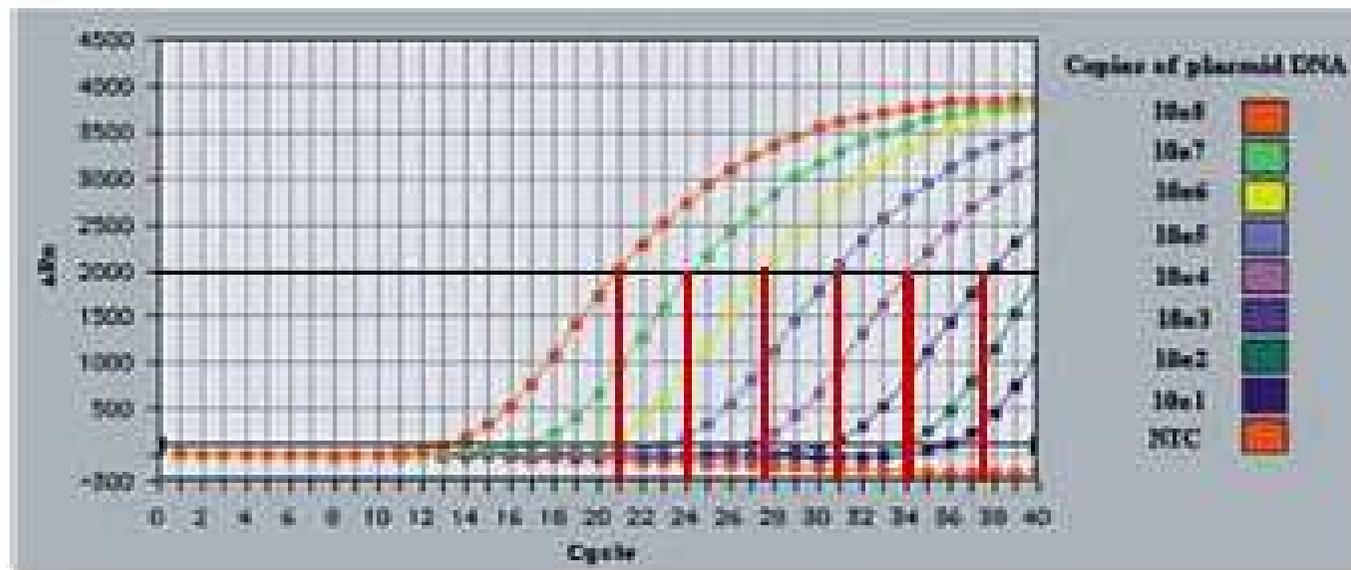
- Seul 5% du génome environ code pour des gènes (Transcrit puis traduit)
- Etudes ciblées à ces 5%



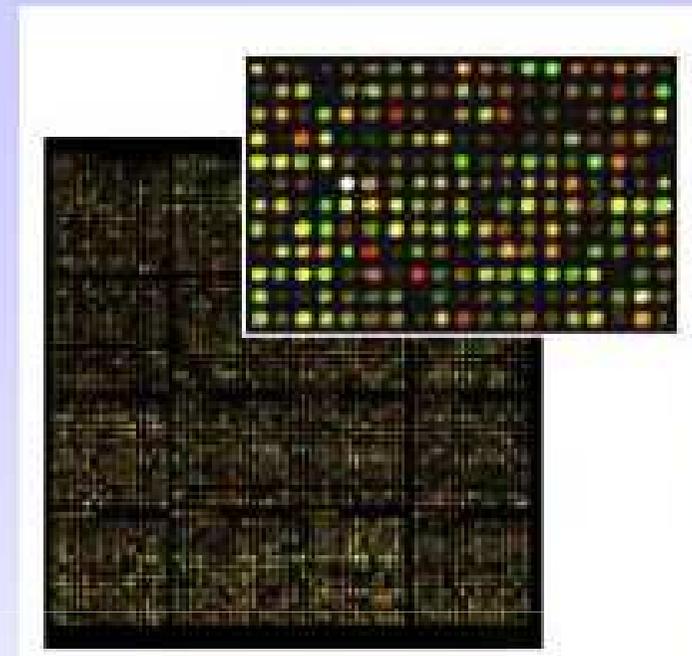
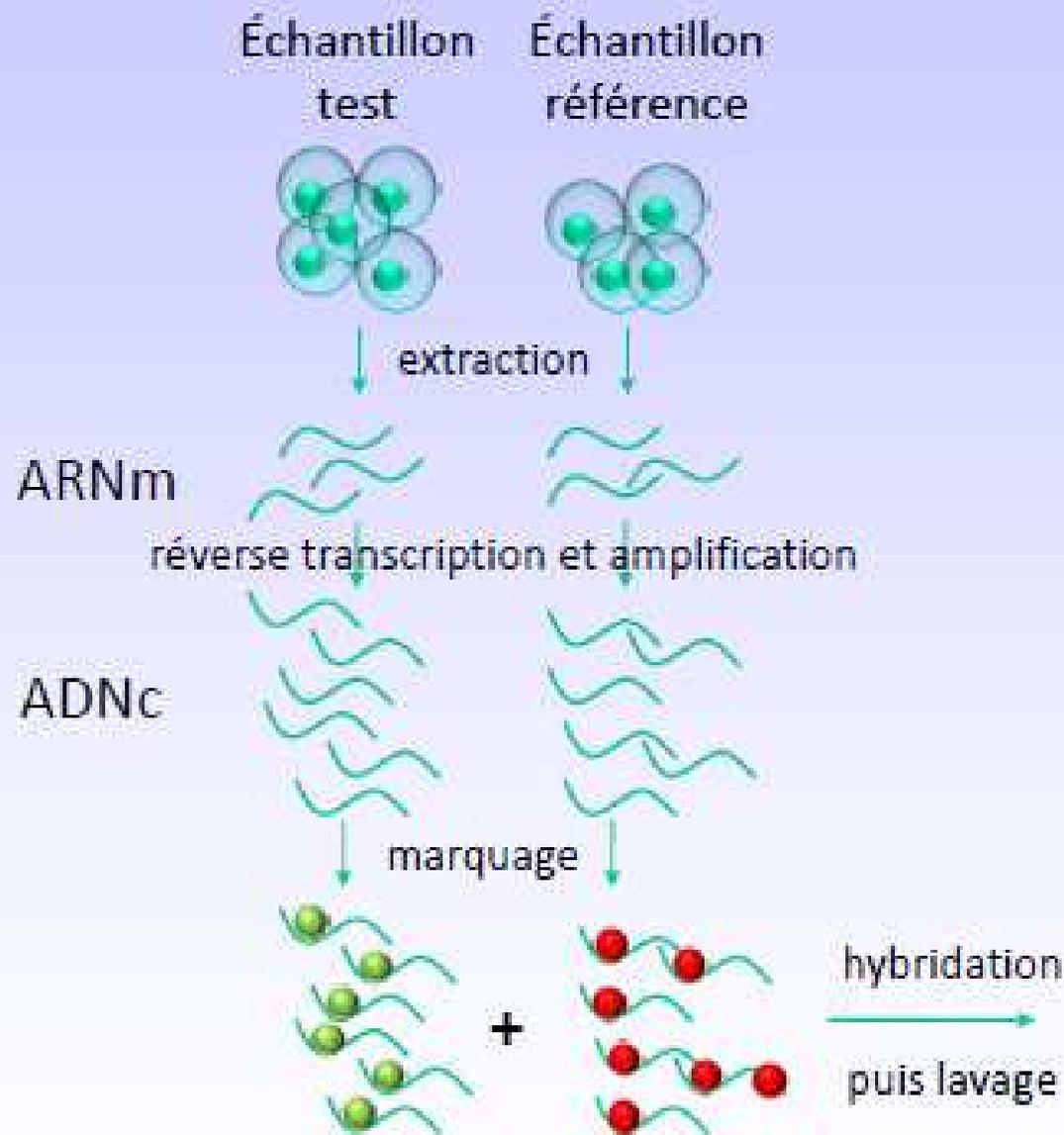
- Banques de cDNA issues de tissus différents :
 - Connaître les gènes (Annotation)
 - Lieu d'expression (Tissus)
 - Abondance (Niveau expression)
- Mise en place d'outils génériques
- Nouvelles technologies d'analyse

Transcriptome

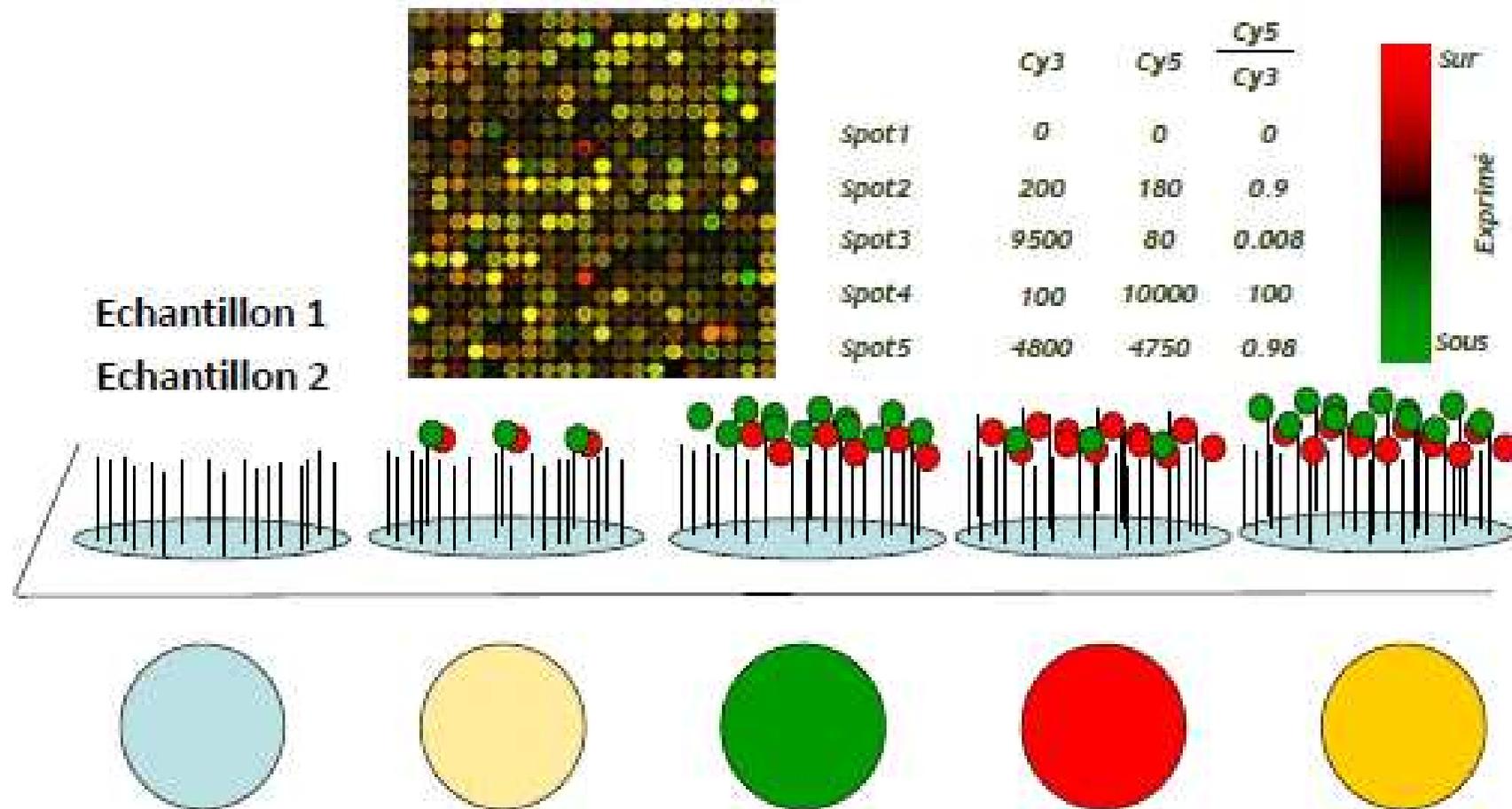
- 1 gènes
 - qPCR (PCR quantitative)



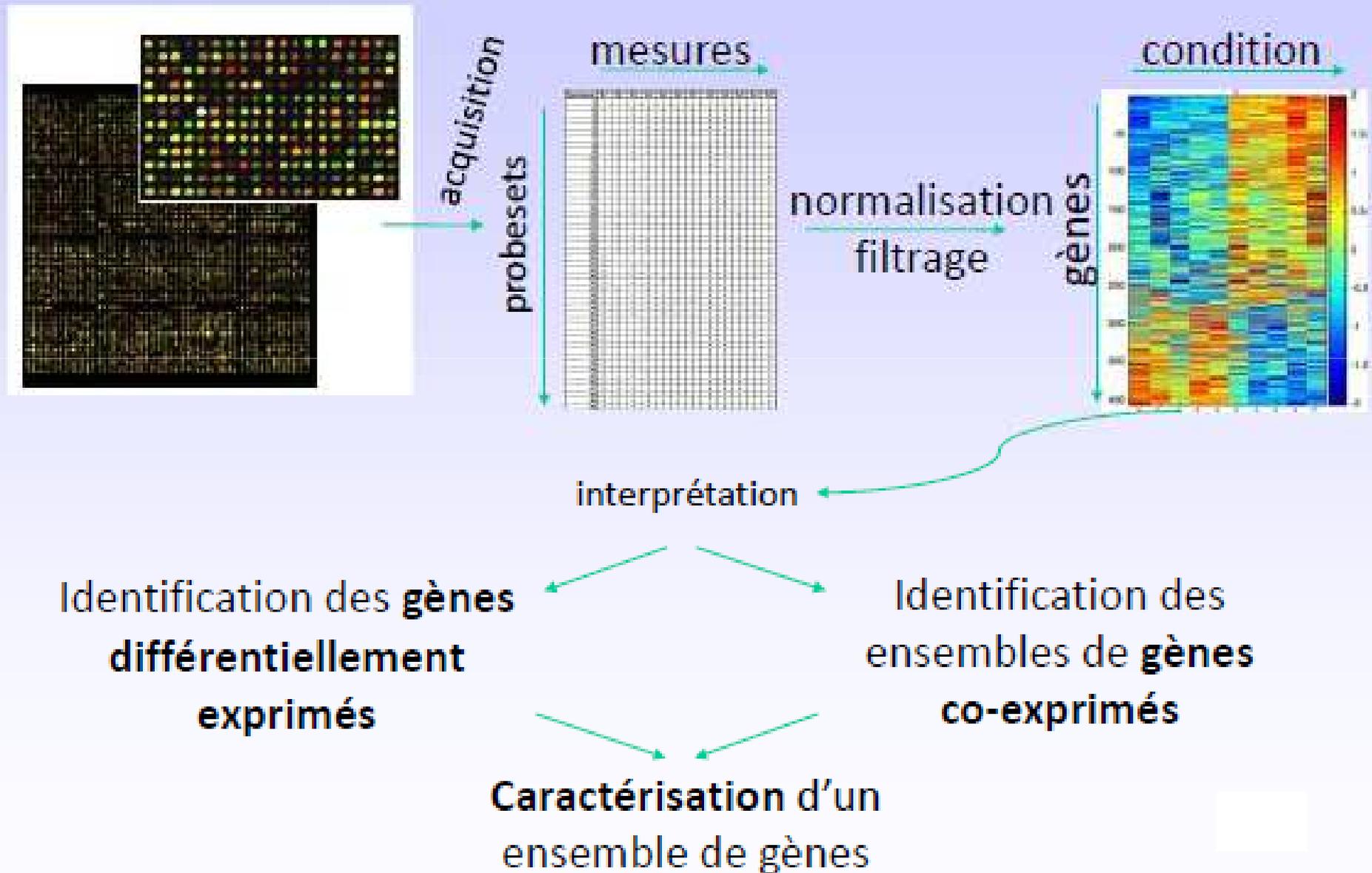
Acquisition des données



Acquisition des données

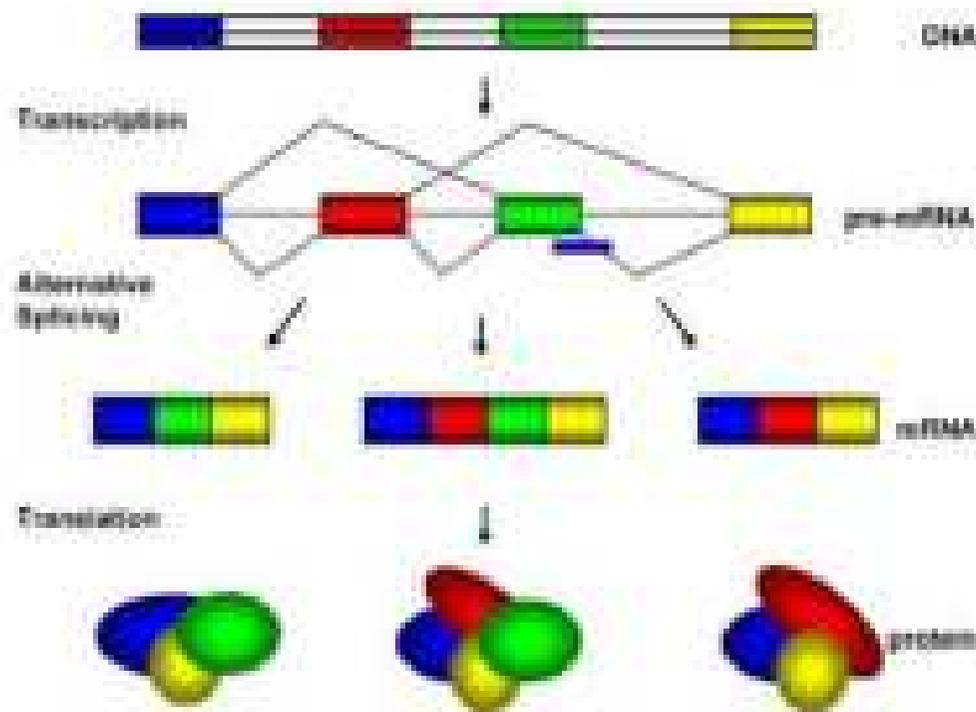


Analyse et interprétation des données



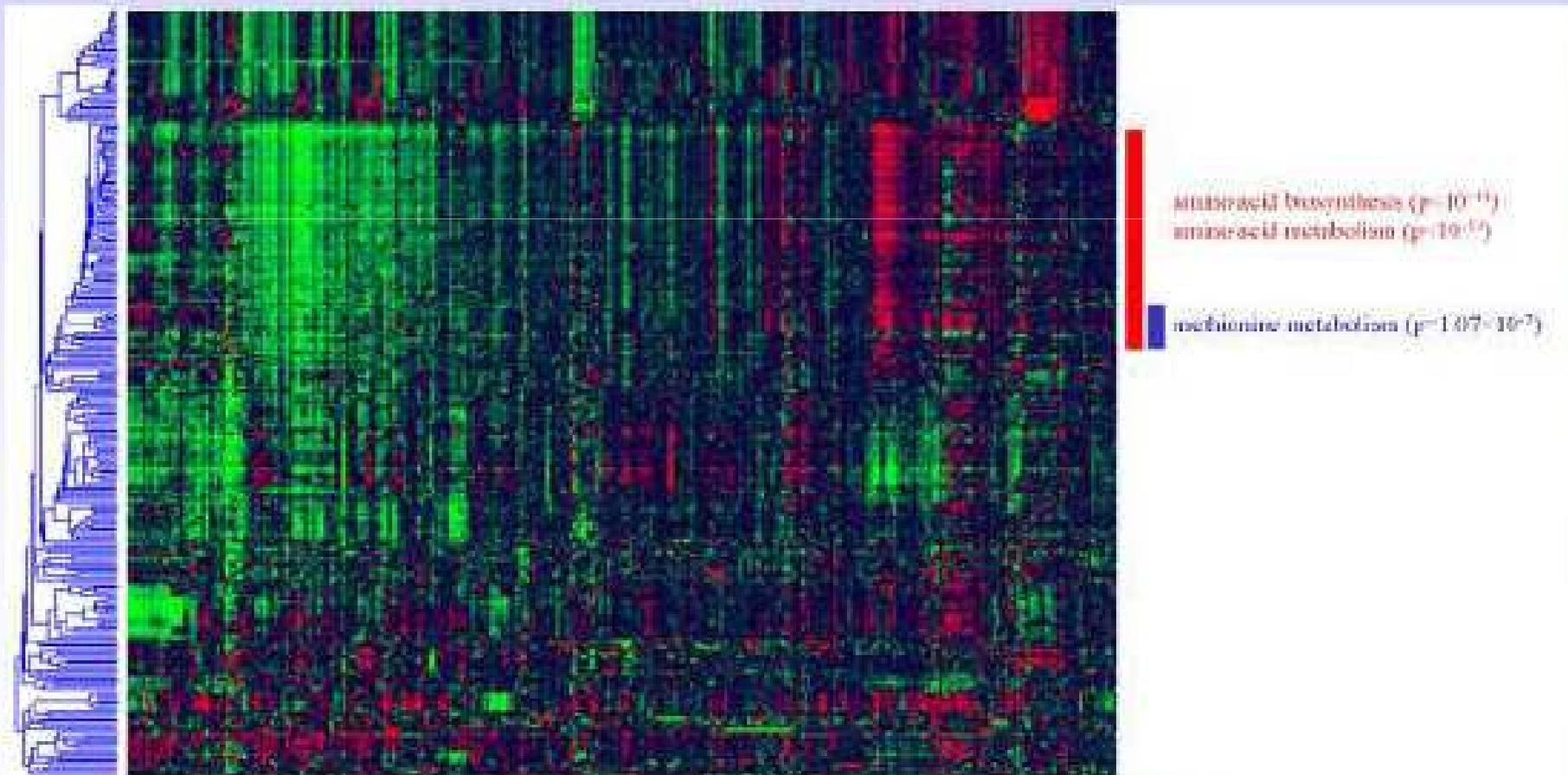
Analyse et interprétation des données

- Ratio de transcription 1/10 000 au sein d'un échantillon
- 30 000 gènes (ADN) = 100 000 ARN
- Un gène peut être transcrit en plusieurs ARN
- Les ARNnc ne sont pas à négliger
- Etre le plus exhaustif possible ... séquençage (RNA-seq)



Gènes co-exprimés

- Motivation : les gènes ayant des profils d'expression similaires sont potentiellement co-régulés et participent à un même processus biologique
- But : regrouper les gènes impliqués dans un même processus biologique



Transcriptome



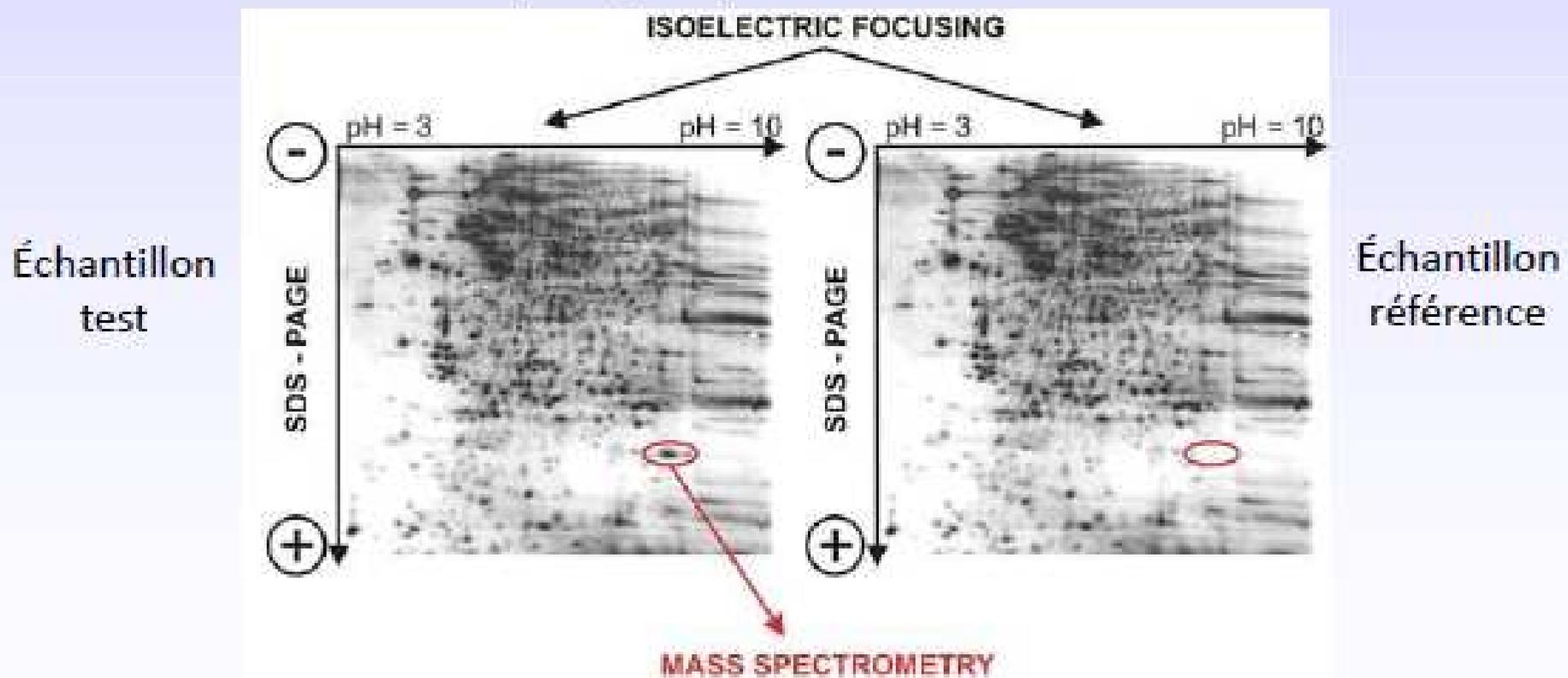
- Etudes comparées de la transcription des gènes
 - Chez des individus différents
 - Caractérisation physiologique d'une voie métabolique
 - « Niveau de transcription d'un gène » est un phénotype (comme le poids d'un animal) : e-QTL
 - Dans des conditions environnementales différentes

Protéomique

Protéome : ensemble des protéines exprimées dans une cellule, une partie d'une cellule (membranes, organites) ou un groupe de cellules (organe, organisme, groupe d'organismes) dans des conditions données et à un moment donné.

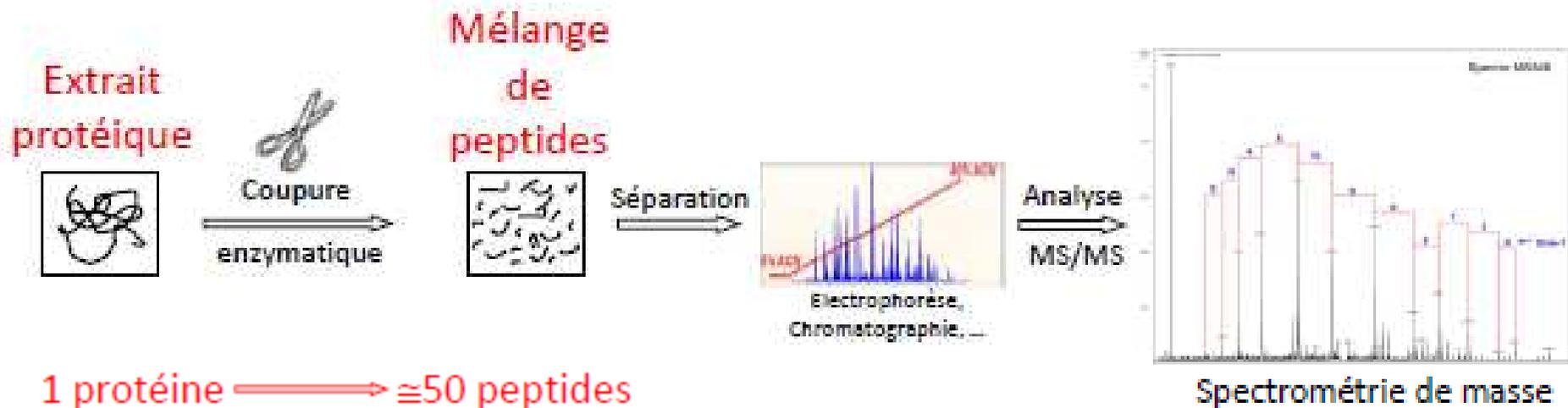
= *instantané* de l'état d'une cellule ou d'une population de cellules

Séparation des protéines par gels d'électrophorèse (1D, 2D) puis identification des spots par spectrométrie de masse



Identification des protéines

L'analyse protéomique repose sur l'interprétation des données de Spectrométrie de Masse.



10 000 protéines par type cellulaire \longrightarrow 500 000 peptides



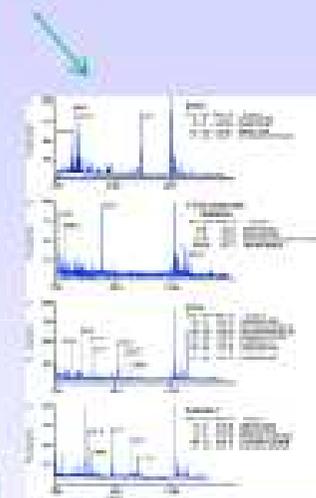
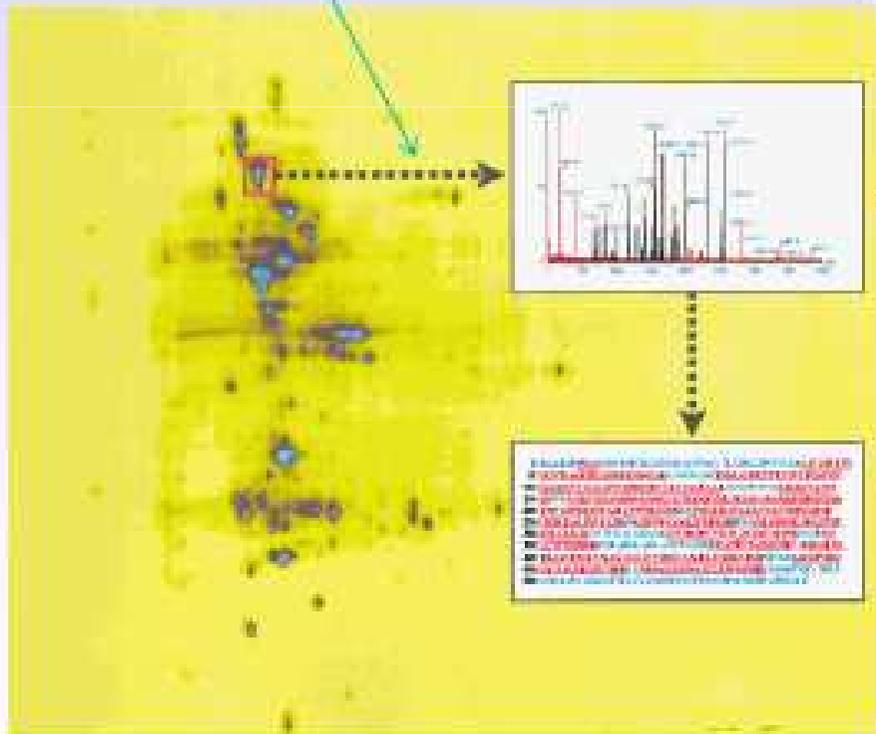
Instruments de type
Quadrupole-TOF, Triple-TOF, Orbitrap, Triple quadrupoles, ...

Identification des protéines

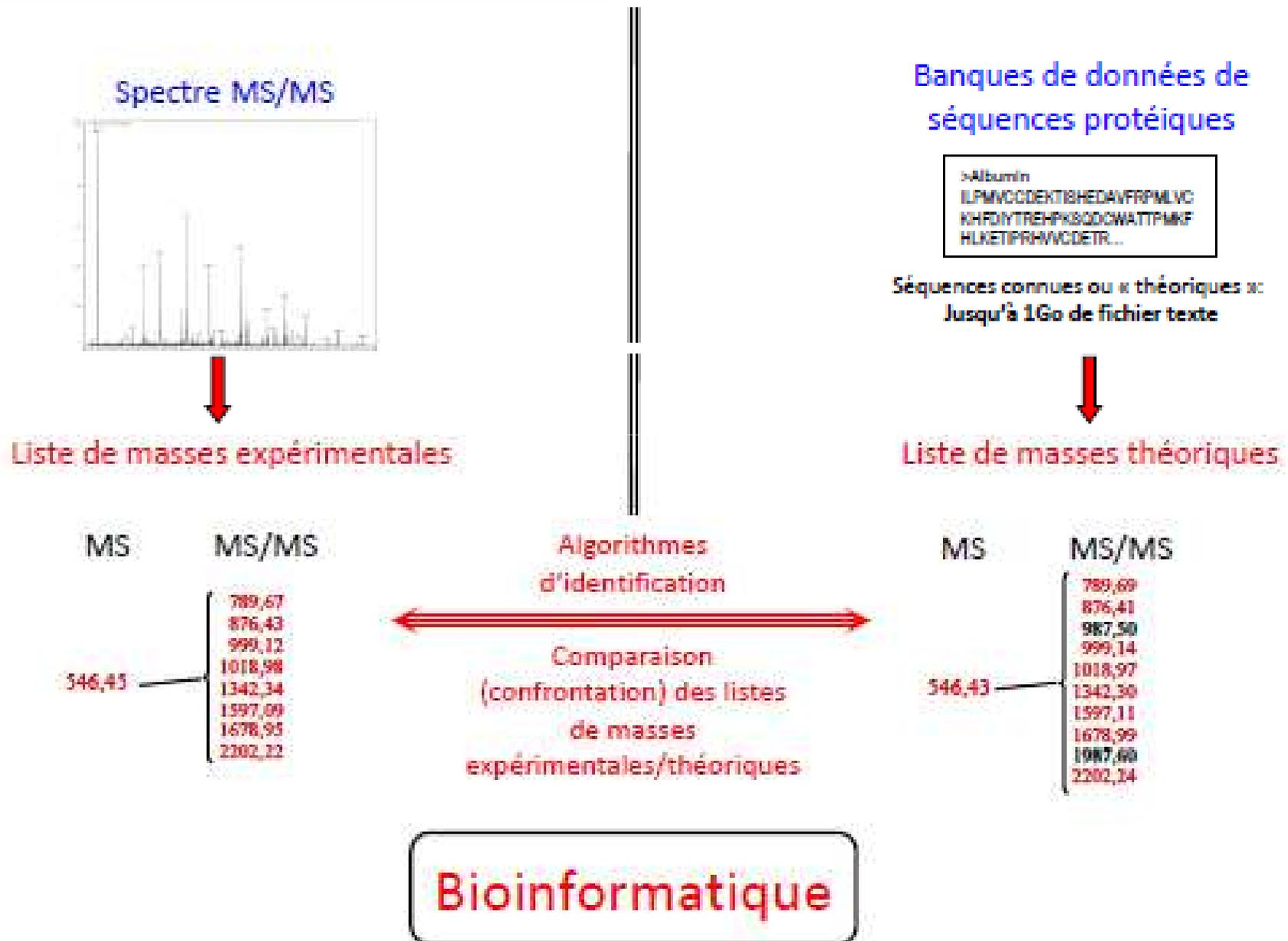
Digestion du spot par une enzyme
(ex: trypsine) et mesure du poids
des peptides obtenus

Digestion *in silico* du protéome

Recherche des
protéines
correspondant au
profil observé



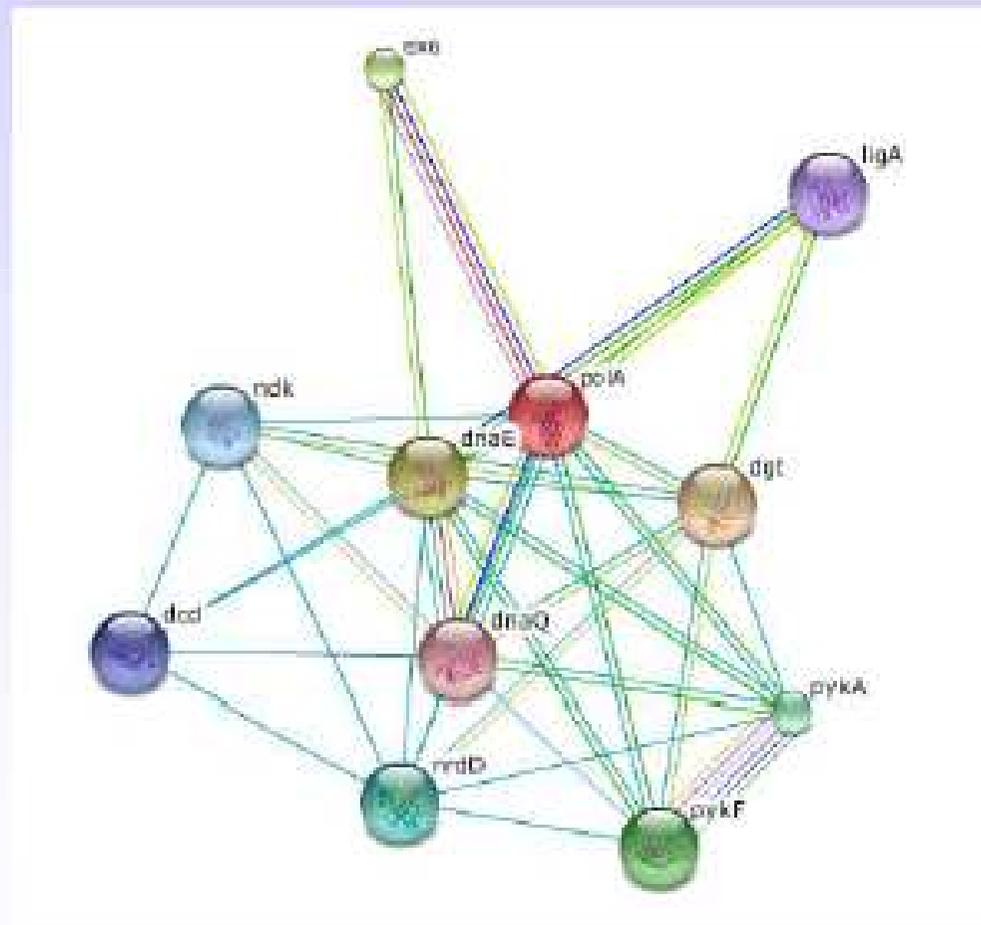
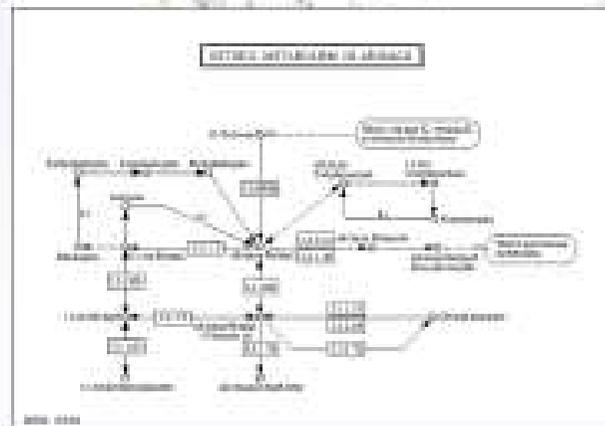
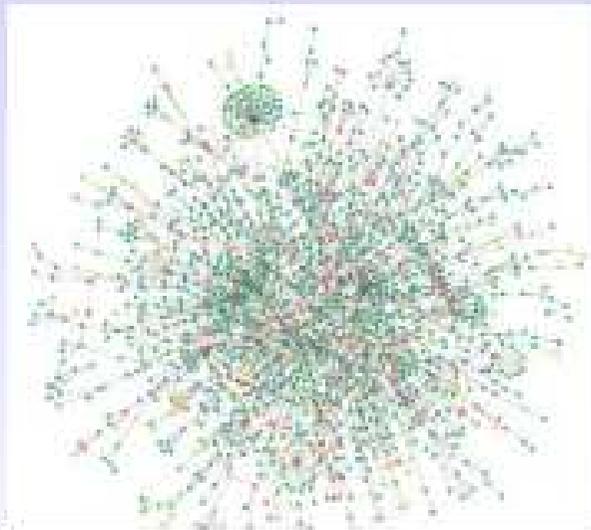
Identification des protéines



Réseaux de gènes, de protéines

Réseaux :

- d'interactions protéine - protéine, génétiques, fonctionnelles, ...
- de régulation des gènes
- métabolisme (enzymes - substrats)
- transduction du signal



Biologie structurale

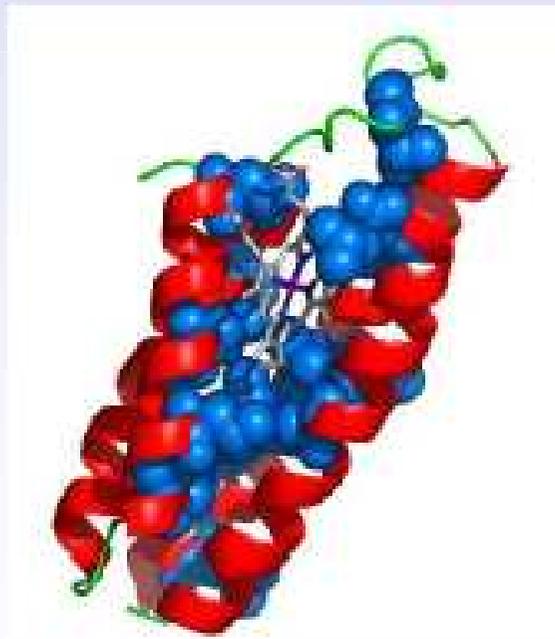
Séquence protéique

->gi|5624211|gb|AAD44166.1| cytochrome b

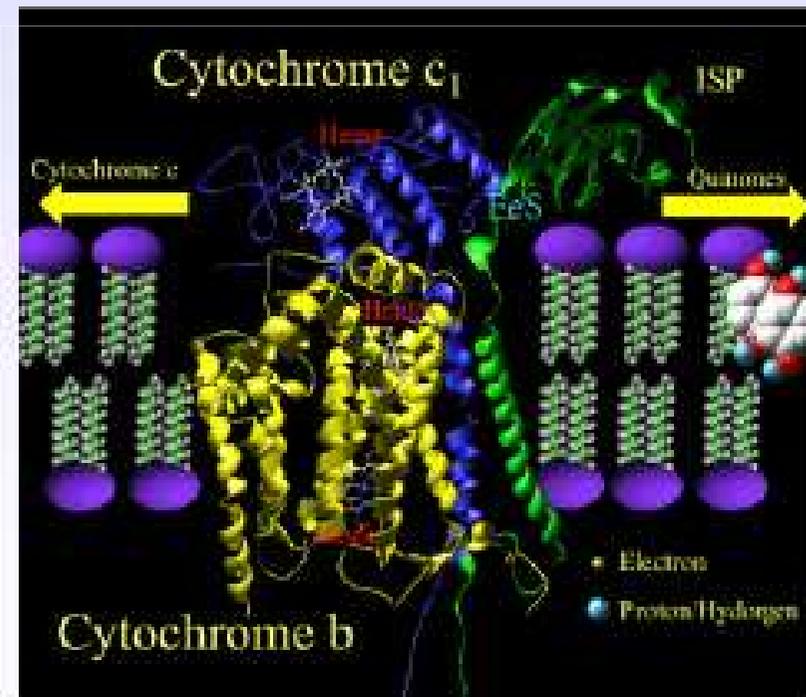
```
LCLYTHIGRNIYYGSYLYSETWNTGIMLLLITMATAFMGYVLPWGQMSFWGATVITNLFSAIPYIGTNLV  
EWWGGFSVDKATLNRFFAFHFILPFTMVALAGVHLTFLHETGSNNPLGLTSDSDKIPFHPYYTIKDFLG  
LLILLLLLLLALLSPDMLGDPDNHMPADPLNTPLHIKPEWYFLFAYAILRSVPNKLGGVLAFLSIVIL  
GLMPFLHTSKHRSMMLRPLSQALFWLTMDLLTLTWIGSQPVEYPYTIIGOMASILYFSIILAFLPIAGX  
IENY
```



Prédiction ou résolution
de la structure tridimensionnelle



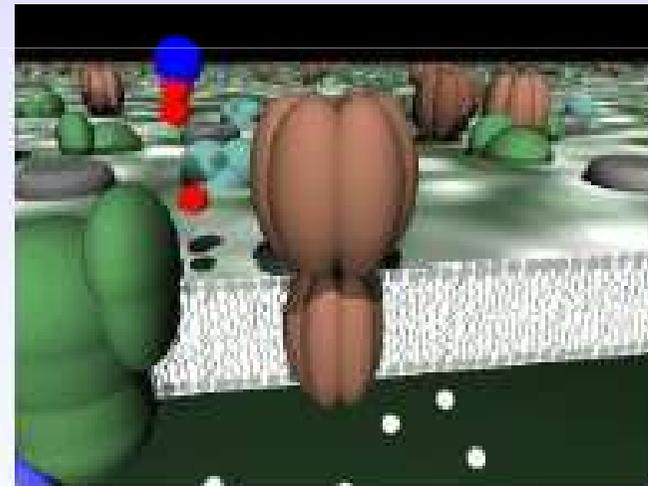
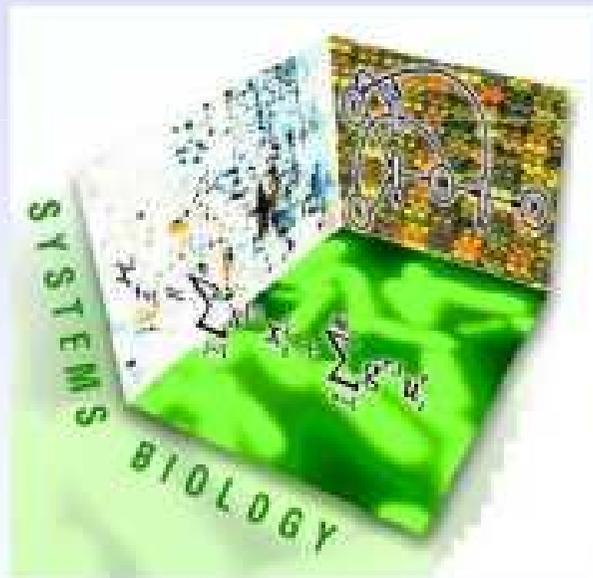
Prédiction des interactions
protéine - protéine ou
protéine - ligand



Biologie des systèmes

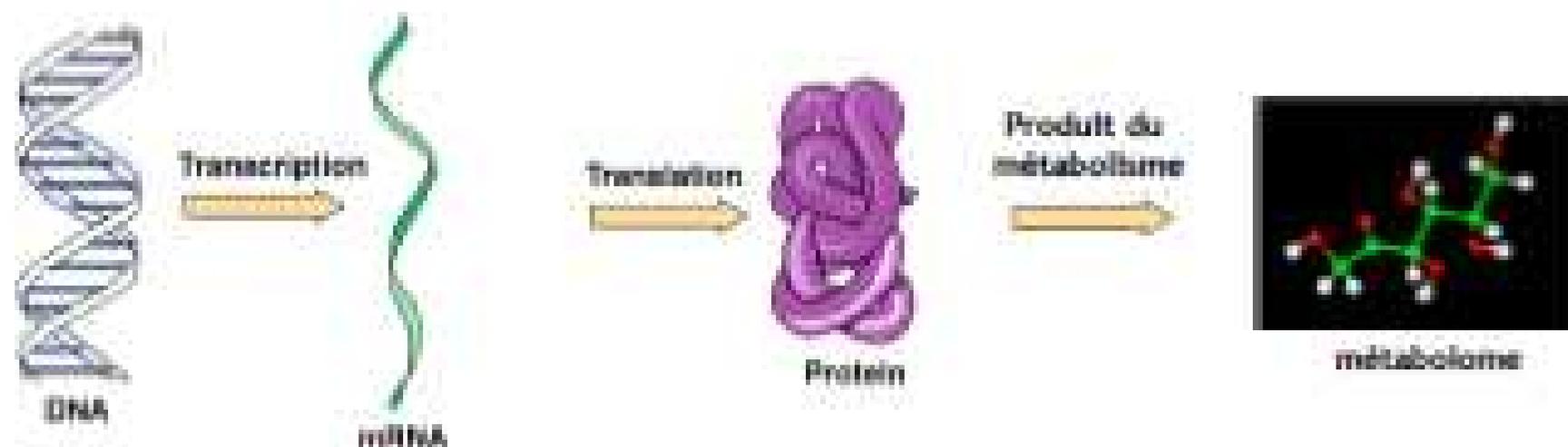
Intégration et synthèse des connaissances

- modélisation d'un système
 - processus biologique (respiration)
 - organite (mitochondrie)
 - cellule
 - population
 - écosystème



À terme : simulation d'une cellule virtuelle et prédiction de son comportement

Métabolomique



- Utilisation comparable au transcriptome :
 - Caractérisation d'un type cellulaire
 - Caractérisation d'un état
 - Analyse comparative de deux états
 - (recherche de Biomarqueurs)

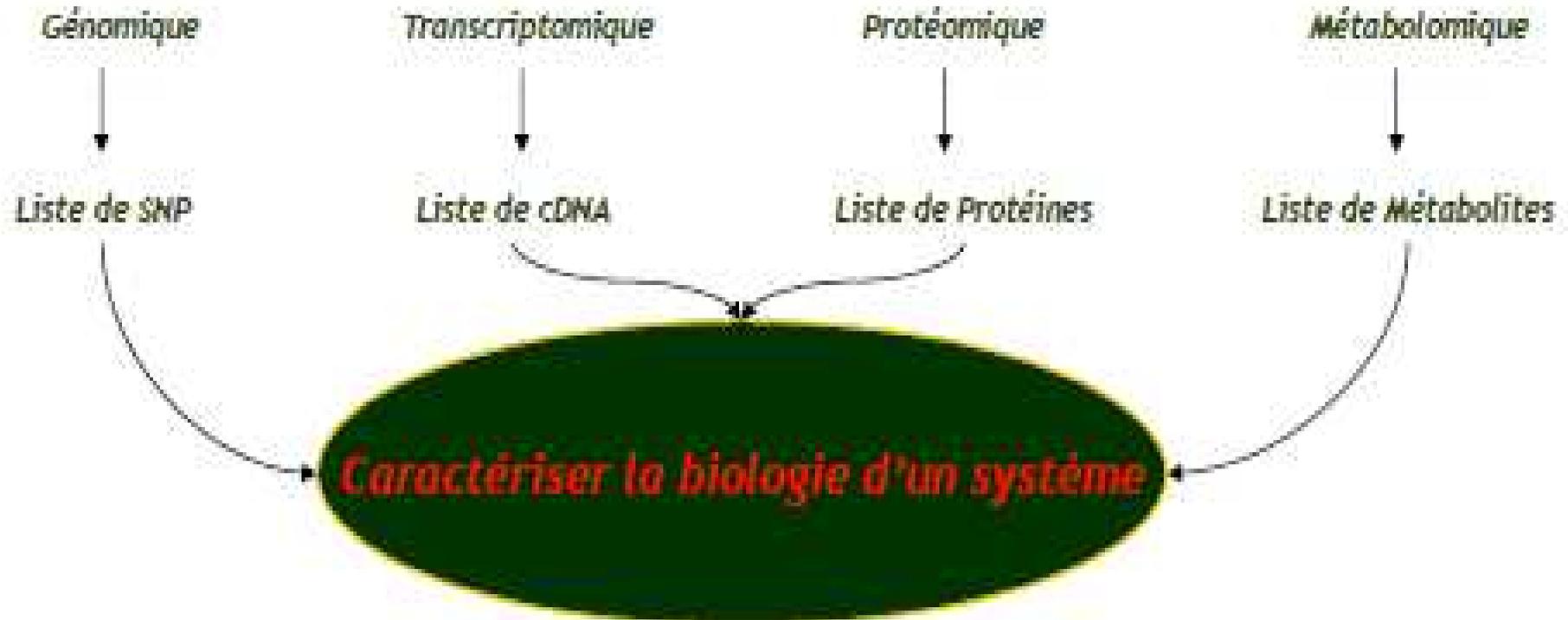
Analyse des données

Bioinformatique

Trois grands domaines où intervient la bioinformatique



Biomathématiques & Statistiques



Université des Frères Mentouri Constantine 1
Faculté des Sciences de la Nature et de la Vie
Département de Biologie Appliquée

Master en : Bioinformatique

Cours de : Génomique Fonctionnelle

Éléments transposables :
Organisation génomique

Introduction

Les séquences d'ADN répétées

- **ADN hautement répété** /10 à 15% du génome
 - Peu codant, hétérochromatique (fortement condensé)
 - Séquences répétées en tandem = les satellites / centromériques, télomériques type TTTAGGG
- **ADN moyennement répété** /25 à 40% du génome
 - Séquences répétées dispersées = **les éléments transposables** avec un certain taux de diversité
 - Les gènes des ARN ribosomiques ARNr, de transfert ARNt (>200 copies)
- **ADN unique ou faiblement répété** Séquences uniques ou à faible nombre de copies = la fraction codant des protéines

Introduction

Les éléments transposables ou éléments mobiles, sont présents dans tous les organismes eucaryotes et procaryotes.

Une propriété commune à tous les éléments transposables est en effet de pouvoir s'insérer en différents endroits du génome

et d'y générer des mutations qui détruisent ou modifient une fonction, lorsque par exemple ils s'insèrent dans ou à proximité d'un gène.

Introduction

Du fait de leur mobilité ou de leur nature répétée, ils entraînent des remaniements chromosomiques à plus grande échelle.

Pour toutes ces raisons, ils sont une source importante de diversité génétique.

Introduction

Éléments transposables : la découverte

- Dès 1940s : maïs / **Barbara McClintock**
- **Altération** de la **pigmentation de grains de maïs**



1902-1992

- ET **inséré** dans un gène de la chaîne de biosynthèse des anthocyanes
⇒ pas de pigmentation
- Expression du gène et pigmentation quand **excision**
- Taille des tâches en fonction du **moment** où a eu lieu l'excision



Pas de produit du gène



Pigmentation



Introduction

Éléments transposables : la découverte

- Modélisation avec **deux éléments** :
 - Ds *Dissociator*
 - Ac *Activator*
- Premières notions de l'existence **d'éléments de contrôle** de l'expression des gènes / prémices des mécanismes de la régulation génique
- Prix nobel en 1983



1902-1992



Introduction

Ces séquences sont capables de se déplacer dans le génome et de s'y insérer à différents endroits.

Leur activité transpositionnelle se traduit donc par de la variabilité génomique, deux génotypes pouvant différer par l'intégration d'un élément en un endroit donné (appelé site d'insertion) dans l'un des deux génotypes mais pas dans l'autre.

Introduction

Il existe alors un polymorphisme d'insertion à ce site entre les deux génotypes.

De nouvelles stratégies ont ainsi vu le jour, afin d'exploiter cette variabilité et d'utiliser les polymorphismes d'insertion d'éléments transposables comme outils de marquage moléculaire.

Eléments transposables : définition

- Caractéristiques :
 - Fragment d'ADN **mobile**
 - Capable de se déplacer de façon **autonome** dans le génome
- Terme « **élément transposable** » dans les **années 70s**
/ génétique moléculaire
- Constituant **majeur** du génome eucaryote
 - Arabidopsis* : 15%
 - maïs : 80%
 - blé : 90%
 - drosophile : 22%
 - homme : 45%

Les éléments transposables

Les éléments transposables sont des fragments d'ADN capables de se déplacer le long des chromosomes. En s'insérant dans un gène, ils modifient sa fonction et **créent une mutation**.

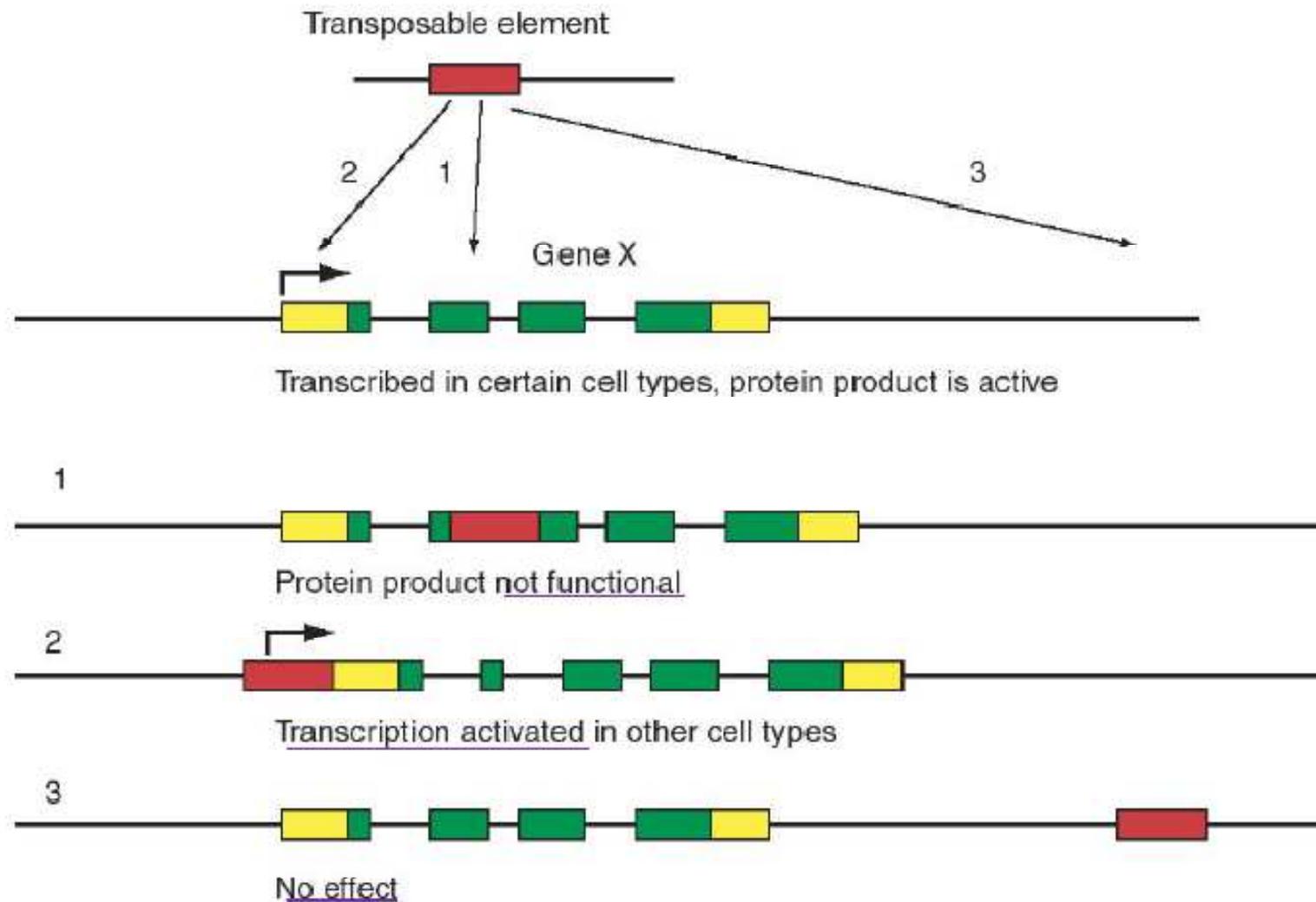
Ils contiennent leurs propres gènes codant pour les protéines nécessaires à leur mobilité.

Ils ne se déplacent pas en permanence, et **leur activité est contrôlée** par eux-mêmes et par l'organisme qui les contient.

Les éléments transposables

- Se déplacent d'un endroit à un autre du génome
- Existents dans tous les organismes étudiés
 - Effets :
 - Insertion à proximité ou dans un gène avec activation ou inactivation du gène cible.
 - Induction de délétions, inversions et translocations de l'ADN
 - Induction de cassures chromosomiques.

Les effets de l'insertion d'un élément transposable sur une séquence codante



Les éléments transposables

Les éléments transposables ont une autre caractéristique commune : bien qu'utilisant la machinerie cellulaire pour leur expression,

Ce sont des entités génétiques bien définies, qui codent généralement pour les fonctions spécifiques nécessaires à leur propre transposition.

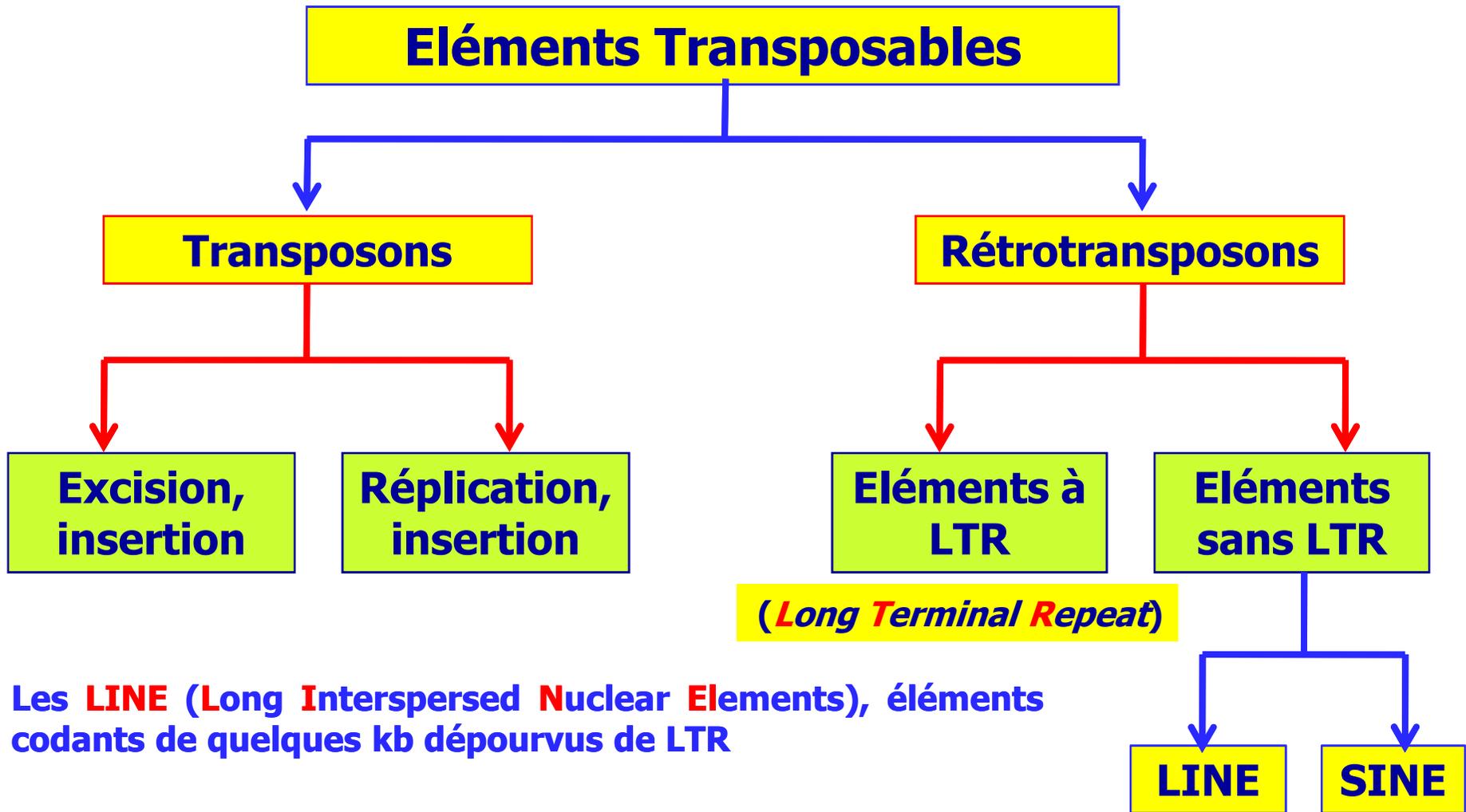
Les éléments transposables

Les éléments transposables sont regroupés en deux grandes classes dont les mécanismes de transposition diffèrent radicalement.

Les éléments de classe I, ou rétroéléments, transposent par l'intermédiaire d'un ARN qui est rétrotranscrit en une copie fille d'ADN,

Tandis que les éléments de classe II, ou transposons à ADN, transposent par l'intermédiaire d'un ADN, généralement par excision suivie de réinsertion dans le cas des eucaryotes.

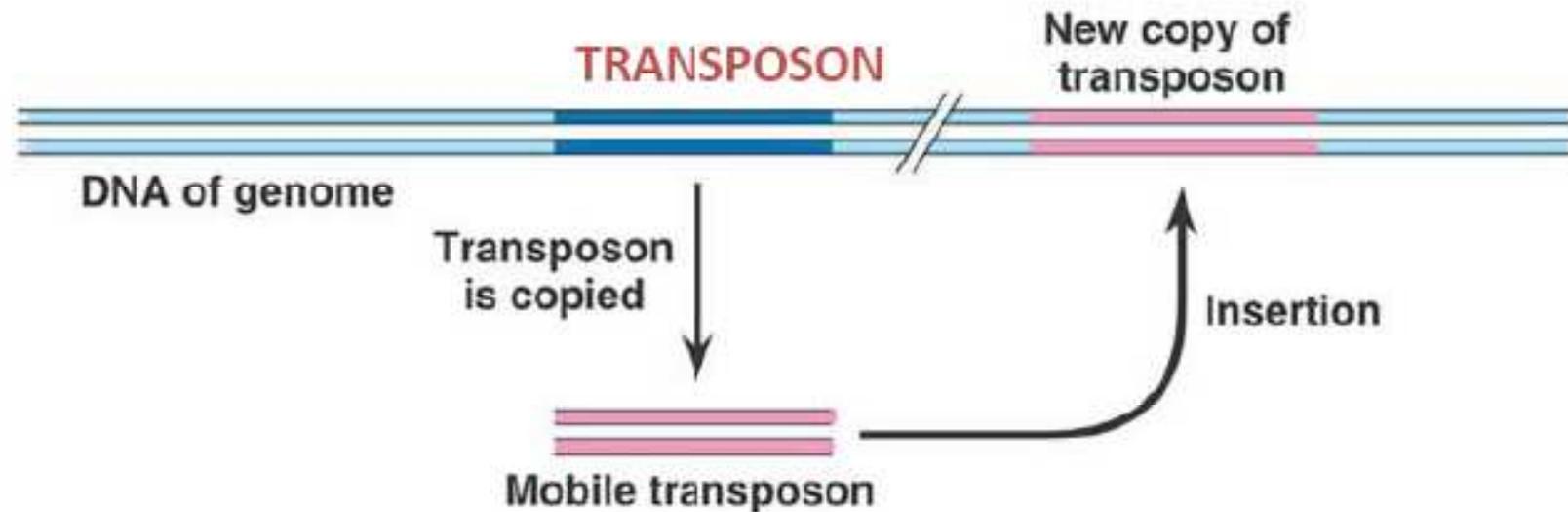
Types de transposons



Les **LINE** (**L**ong **I**nterspersed **N**uclear **E**lements), éléments codants de quelques kb dépourvus de LTR

Les **SINE** (**S**hort **I**nterspersed **N**uclear **E**lements), petits éléments non codants de quelques centaines de paires de bases, qui utilisent la machinerie de transposition des LINE pour assurer leur amplification.

Mécanismes de la transposition

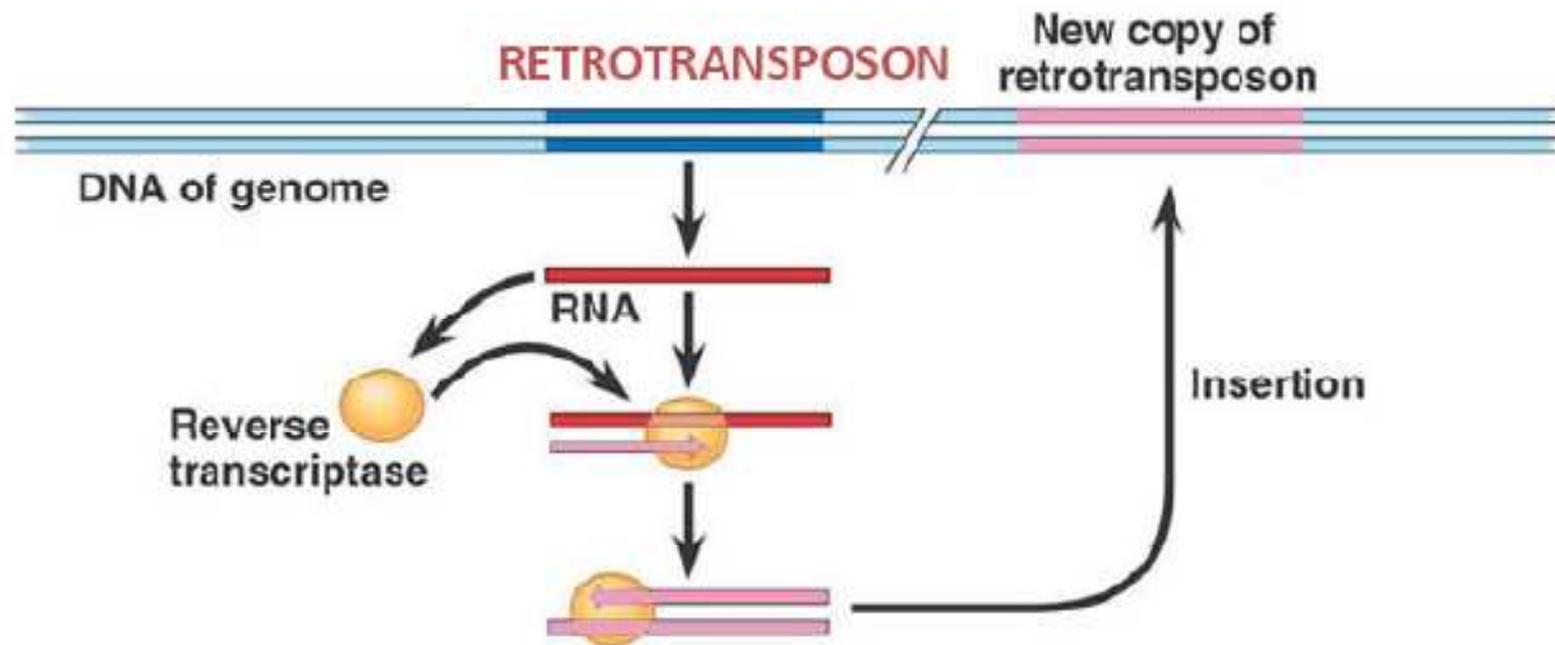


(a) Transposon movement ("copy-and-paste" mechanism)

a - Les transposons ADN : intermédiaire ADN (activité transposase). Exp : Ac (maïs), Slide (tabac) et P (Drosophile). Ils créent des mutations instables.

b - Les rétrotransposons : intermédiaire ARN (activité réverse transcriptase et insertion). Exp : Tnt1 (tabac) et copia (Drosophile).

Ils ressemblent aux rétrovirus mais NE SONT PAS INFECTIEUX



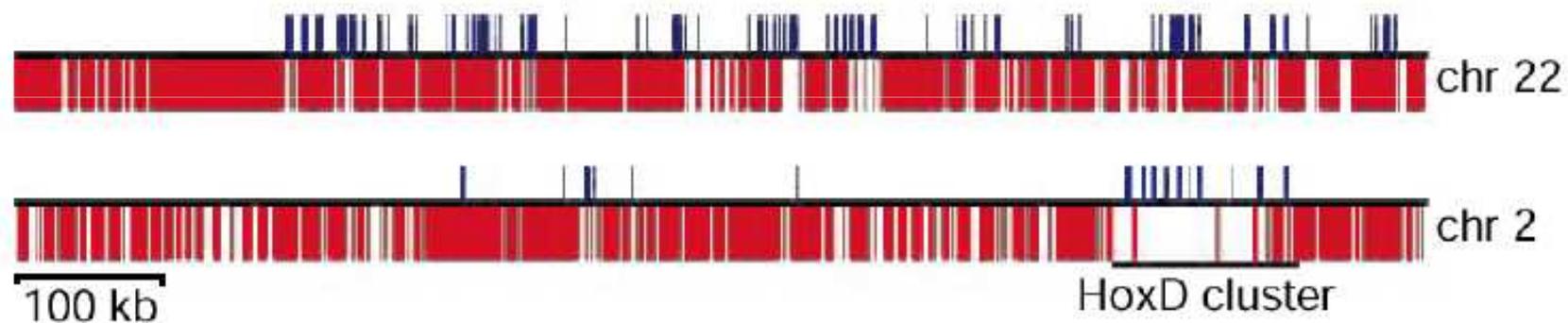
(b) Retrotransposon movement

Mécanismes de la transposition

Ce mécanisme de rétrotransposition implique que les insertions sont en théorie **irréversibles** et que les rétroéléments sont capables de **s'amplifier en très grand nombre de copies**.

Ils peuvent en effet être présents en plusieurs milliers de copies dans les génomes eucaryotes complexes.

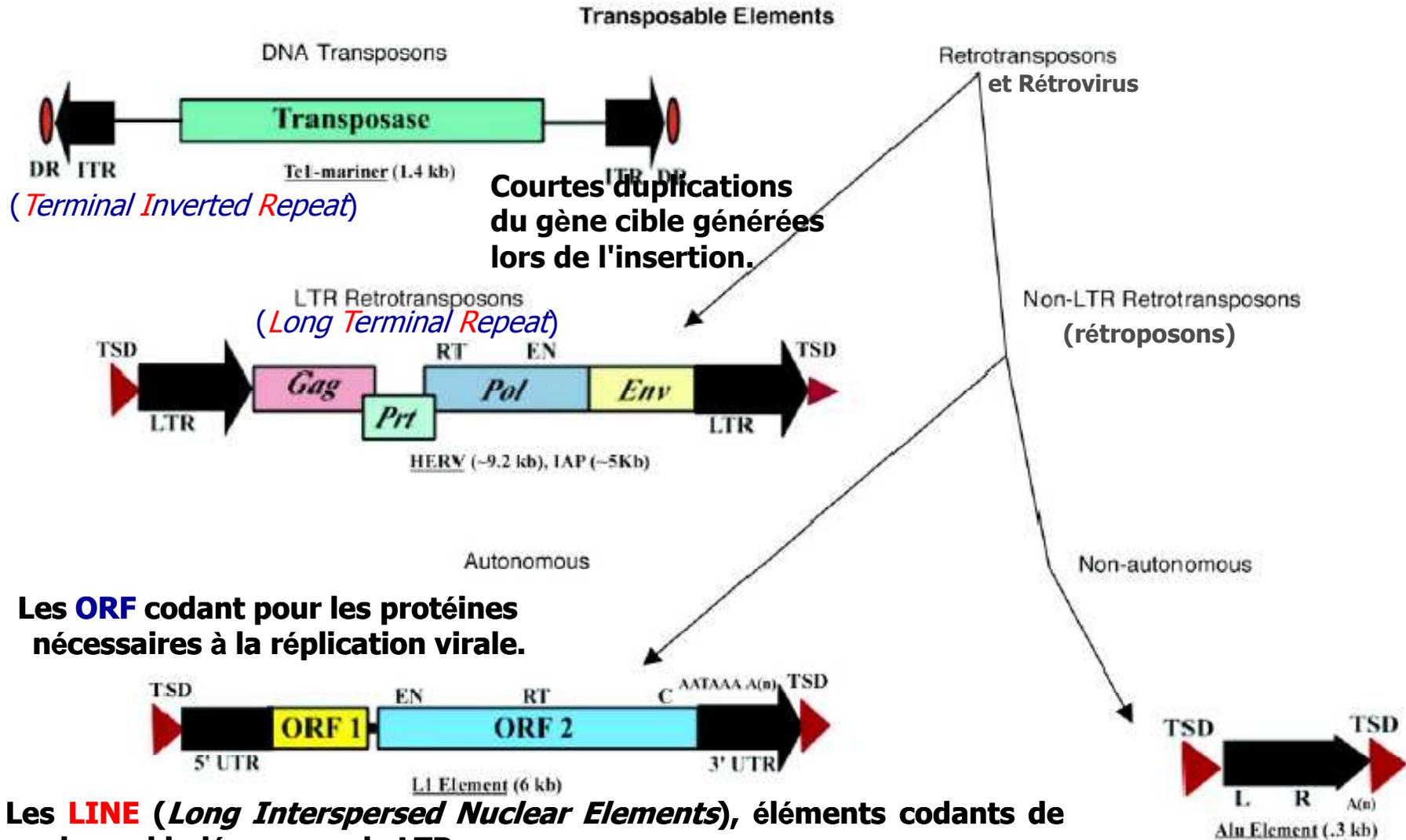
Les données de séquençage montrent qu'ils forment près de 42 % du génome humain et on estime qu'ils représenteraient plus de 60 % du génome chez certaines espèces végétales.



|:Exons
|:Répétitions

Les répétitions sont absentes de certaines régions

La structure des transposons

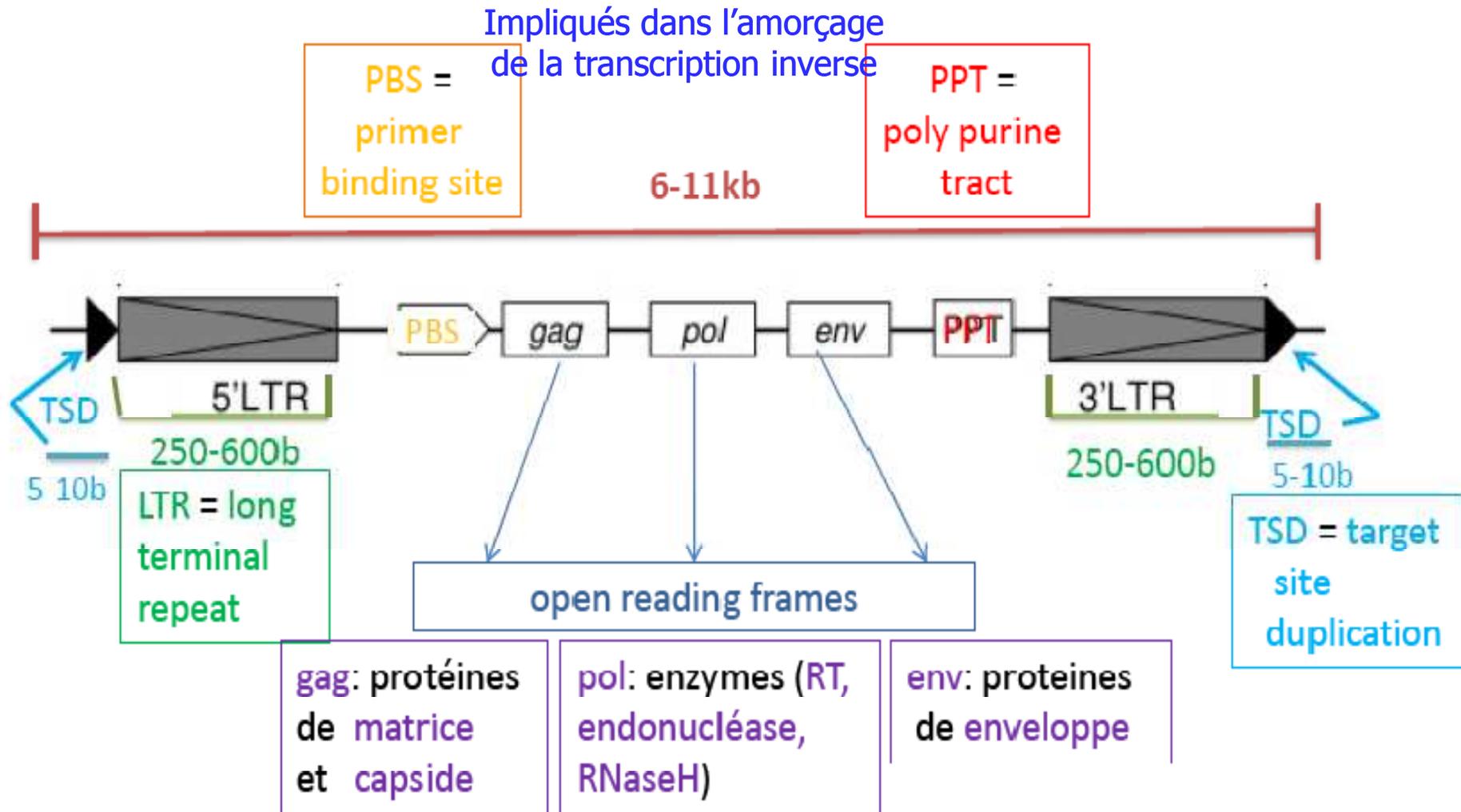


Courtes duplications du gène cible générées lors de l'insertion.

Les **ORF** codant pour les protéines nécessaires à la réplication virale.

Les **LINE** (*Long Interspersed Nuclear Elements*), éléments codants de quelques kb dépourvus de LTR
 Les **SINE** (*Short Interspersed Nuclear Elements*), petits éléments non codants de quelques centaines de paires de bases, qui utilisent la machinerie de transposition des LINE pour assurer leur amplification.

Structure of a typical LTR retrotransposon/retrovirus



LTR:



Promoter (U3 dans LTR 5')

Polyadenylation signal (U5 dans LTR 3')

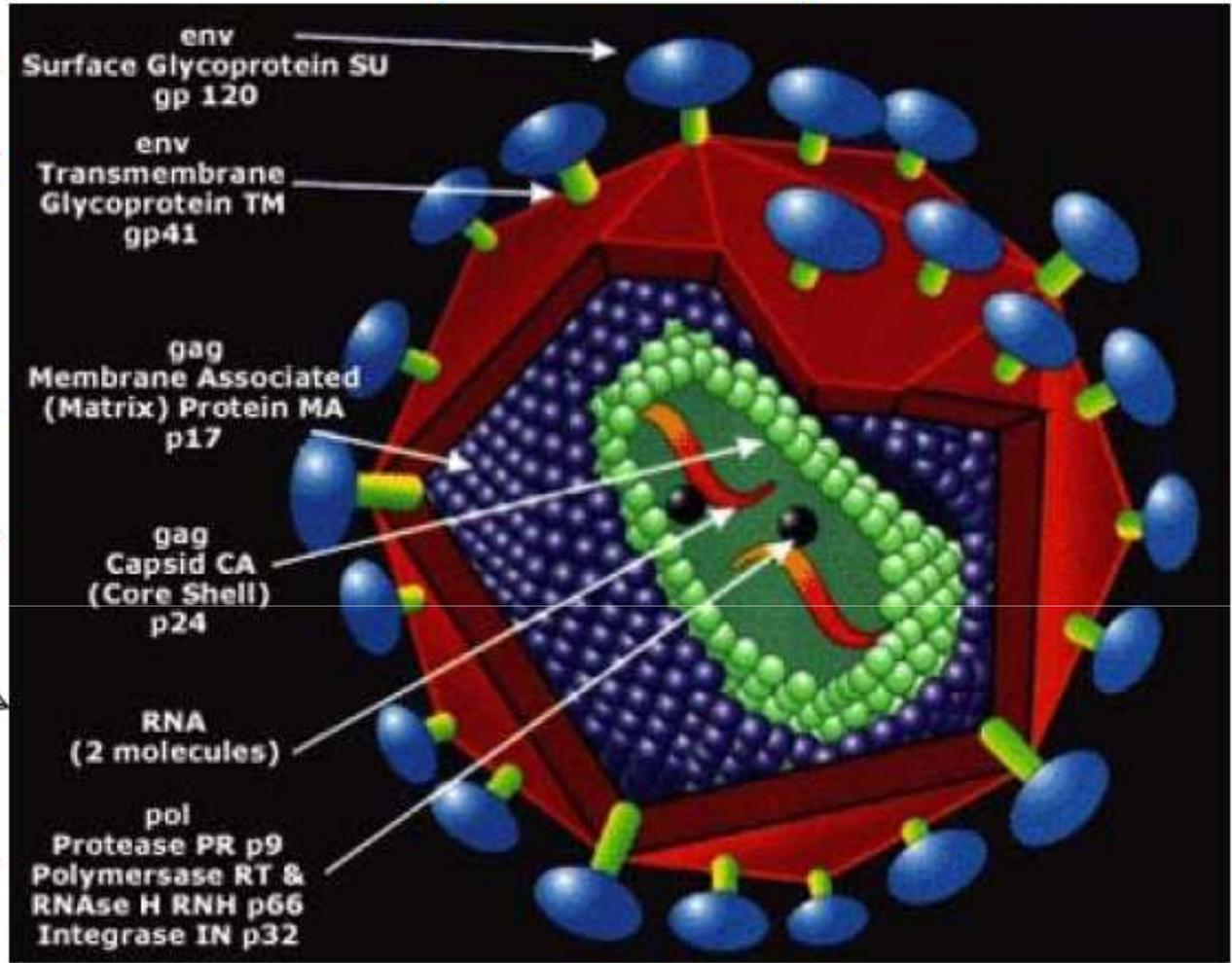
Rétrovirus

env: protéines de
enveloppe

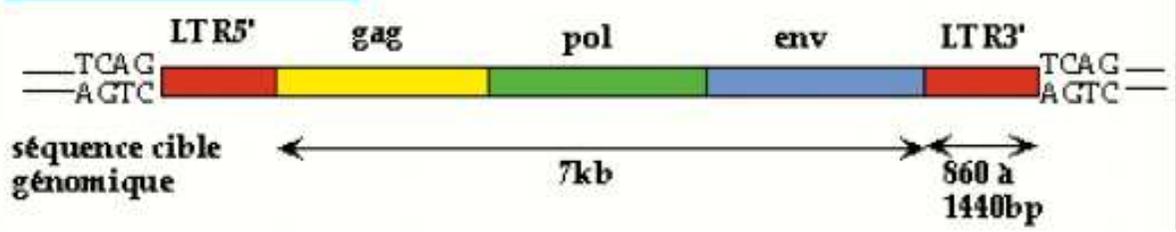
gag: protéines de
matrice
et capside

2 RNA molecules
(single strand)

pol: enzymes (RT,
RNaseH,
intégrase)



ADN rétroviral



Pourquoi les étudier ?

Savoir comment ils fonctionnent, et tenter de comprendre à quoi ils servent ;

- Ils sont souvent activés par des stress (exp ; l'expression de Tnt1 (tabac) est activé par des infections parasitaires et par la blessure.
- Ils ont très probablement un rôle évolutif. En créant des mutations, ils pourraient participer aux phénomènes de formation de nouvelles espèces.
- Leur activation par des stress pourrait permettre à l'espèce hôte de s'adapter à de nouvelles contraintes.

Les organismes hôtes ont appris à vivre avec eux et à en tirer profit

Pourquoi les étudier ?

En fait des outils de biotechnologie, ils peuvent être utilisés pour le clonage de gènes par **étiquetage**.

Une mutation du gène à isoler est créée par insertion d'un élément transposable connu.

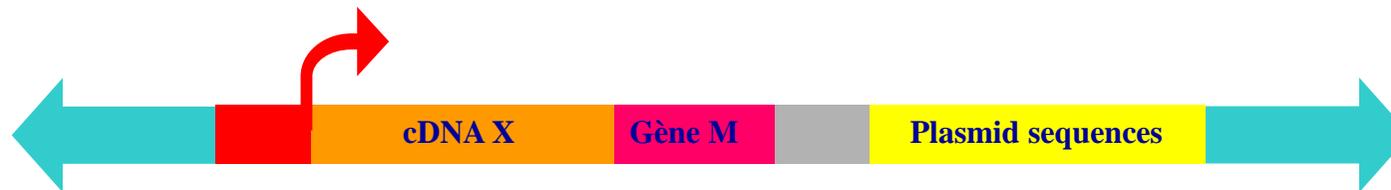
Une sonde moléculaire de l'élément est utilisée pour isoler l'élément étiquette

... et, par la même occasion, les séquences du gène qui lui sont associées.

On peut ainsi isoler un gène d'intérêt agronomique, pour l'introduire dans une espèce végétale qui en est dépourvue.

TRANSGENESE

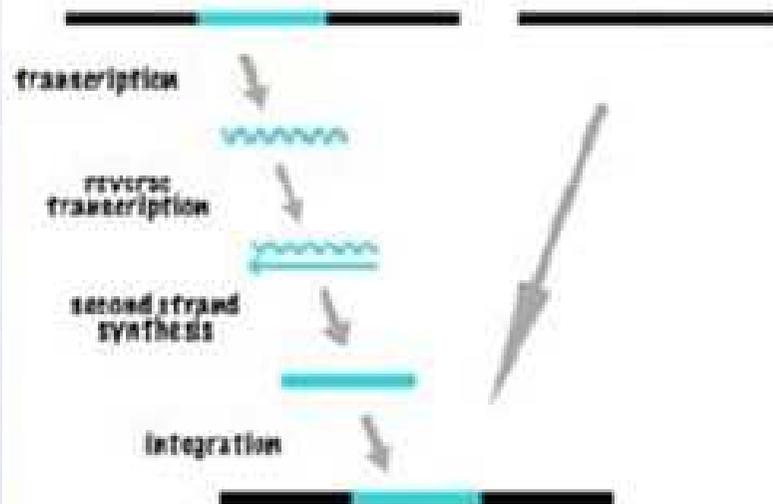
Utilisation d'un élément transposable (élément Ac), pour introduire de façon stable et héritable un transgène dans le génome.



-  Pieds (5'et 3') de l'élément Ac
-  Promoteur minimum transposase
-  Gène X
-  Gène marqueur de la transgénèse: GFP ou GUS
-  Origine de réplication
-  Gène de résistance (i.e. AmpR) } Clonage

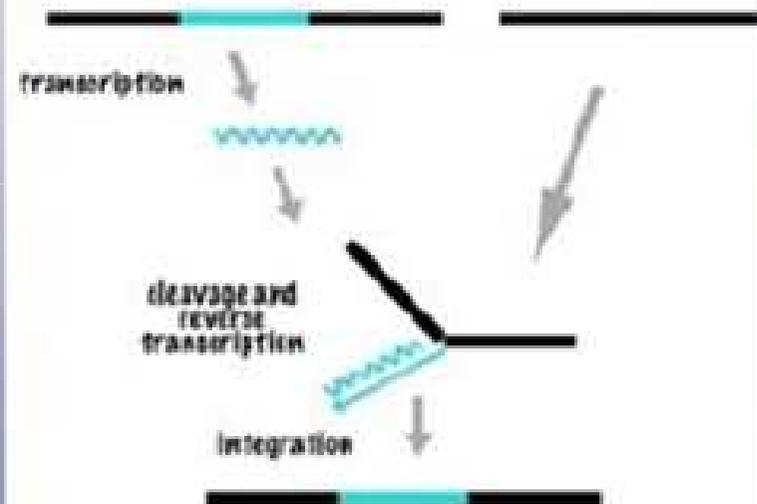
Mechanisms of Retrotransposition

LTR Retrotransposition tRNA primed reverse transcription



Type retrovirus like

non-LTR Retrotransposition Target Primed Reverse Transcription (TPRT)



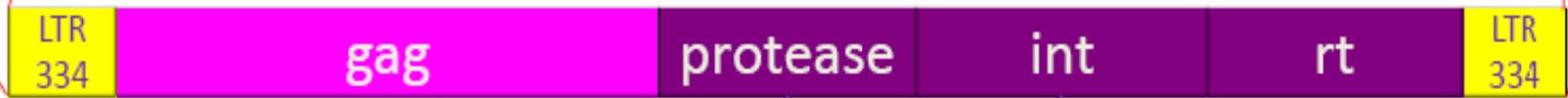
Type LINE

Retrotransposons: copier/coller

Exemple du Ty de levure (6 kb)

Intégration dans un nouveau site du génome

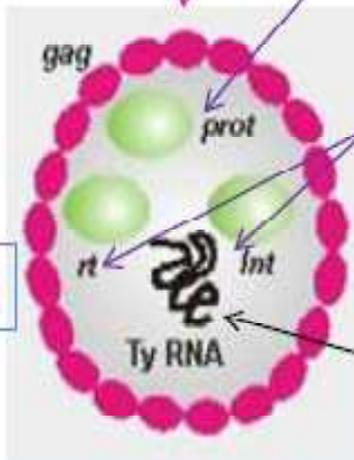
ADN génomique



cDNA

Transcription

Reverse transcription



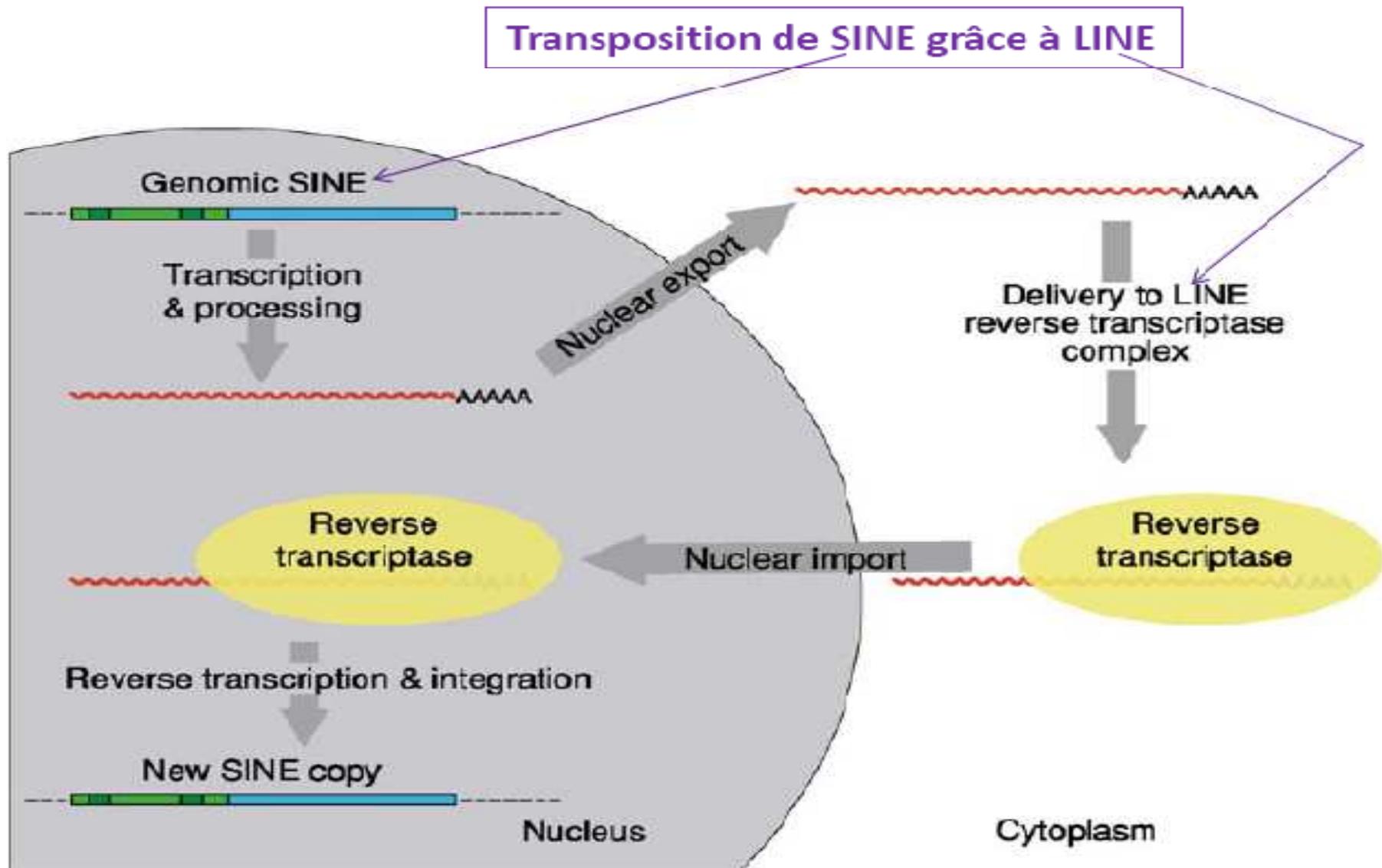
Traduction
Assemblage

mRNA



Pseudo particules cytoplasmiques

Comment les éléments transposables dépourvus d'activité enzymatique se transposent ?



Conséquences génomiques de la présence de ces éléments transposables :

Événements de recombinaison au niveau de séquences répétées (ex: transposons, LINEs, SINEs)

Réarrangements chromosomiques : insertions, délétions, inversions, duplications en tandem, translocations dynamique et évolution des génomes.

ROLE DANS L'EVOLUTION DES GENOMES

