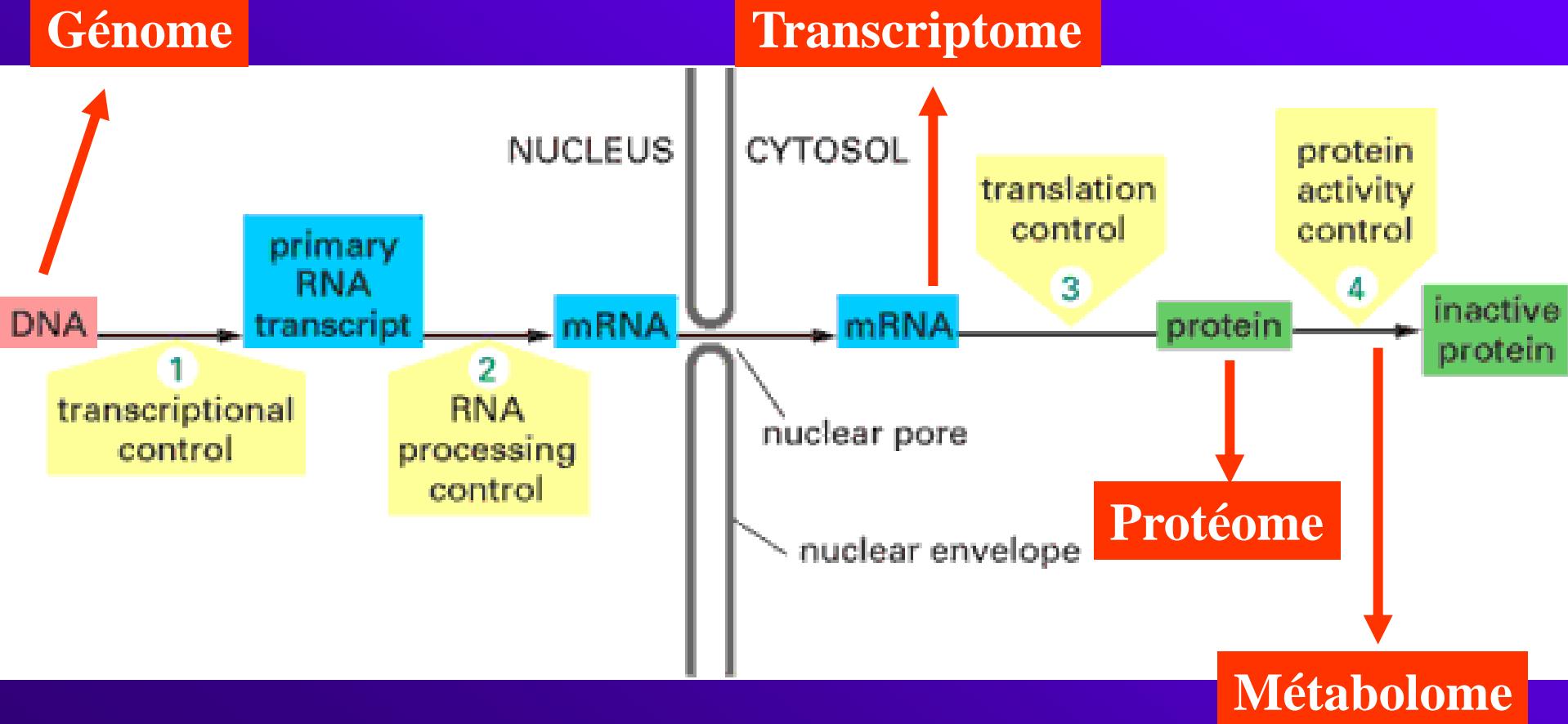


Le Transcriptome

- Introduction
- Méthodes d'analyse du transcriptome
- Puces à ADN : principe et méthodologie
- CGH-array

Définitions



Evolutions technologiques et nouvelles approches

- Les avancées technologiques ont entraînées une baisse significative des coûts de séquençage.
- La robotisation permet de traiter plusieurs dizaines ou centaines de milliers de clones avec peu d'interventions humaines et en réduisant les erreurs de manipulation



Développement d'outils de génomique

Organismes procaryotes entièrement séquencés et assemblés (05/06)

- 1703 virus (dont 91 rétrovirus)
- 382 bactéries + 643 en cours
- 933 plasmides naturels
- 459 organelles (81 chloroplastes, 1001 mitochondries)

Organismes eucaryotes entièrement séquencés et assemblés (05/06)

- 9 champignons dont *Saccharomyces cerevisiae* (12.07 MB, 10/1996), *Schizo. Pombe*
- 4 animaux : drosophile (180 MB, 03/2000), *C.elegans* (12/1998), homme (3038 MB, 12/1999), souris (2500, 07/2005)
- 6 parasites dont *plasmodium falciparum* (23 MB, 11/1998), *trypanosoma cruci* (34 MB, 07/2005)
- 2 plantes : *Arabidopsis thaliana* (119 MB, 12/2000), riz (430 MB, 12/2002)

A ceux-la viennent se rajouter 112 organismes en assemblage et 181 en cours dont rat, chien, chat, vache, cochon, abeille, blé, tomate, orge, maïs, peuplier, poisson zèbre, fugu.....

Nombre de gènes identifiés dans différents organismes

• HIV-1	9
• Bactériophage Lambda	80
• <i>Escherichia coli</i>	4300
• <i>Saccharomyces cerevisiae</i>	6300
• <i>C. Elegans</i>	19000
• Drosophile	13600
• <i>Arabidopsis thaliana</i>	25500
• <i>Hommo sapiens</i>	~35000

Etape suivante: génomique fonctionnelle

Functional genomics: The development and application of global (genome-wide) experimental approaches to assess gene function by making use of the information provided by structural genomics.

It is characterized by high throughput experimental methodologies combined with statistical and computational analysis of the results. The fundamental strategy is to expand the scope of biological investigation from studying single genes or proteins to studying all genes or proteins at once in a systematic fashion. [*Phil Hieter and Mark Boguski "Functional Genomics: It's All How You Read It" Science 278: 601- 602, October 24, 1997*]

L'objectif des approches de génomique fonctionnelle est de répondre à la question suivante:

quel est le rôle des gènes identifiés par les programmes de séquençage?

Comment assigner une fonction à un gène?

Protéine déjà étudiée par des approches classiques.

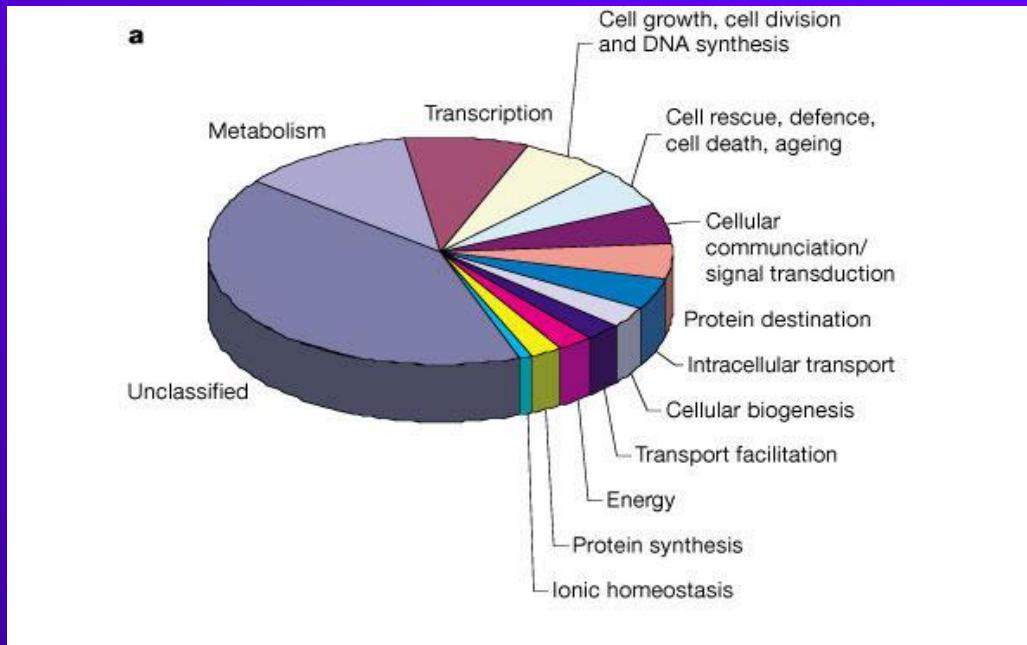
Gène dont la séquence, ou des motifs dans la séquence, présente des homologies avec des gènes de fonctions connues (dans le même organisme ou ailleurs!). C'est l'annotation par homologie de séquence (bioinformatique)

Génétique inverse (on inactive le gène et on observe le résultat). Pb avec les gènes essentiels et les familles de gènes (compensation)

Etude de l'expression des gènes (expression spatio-temporelle). Des gènes présentant un profil d'expression similaire peuvent être impliqués dans des processus commun. Transcriptome/protéome (annotation fonctionnelle)

Localisation intracellulaire des protéines correspondantes (immunohistochimie, couplage GFP...)

Exemple chez *Arabidopsis thaliana*



*The Arabidopsis Genome Initiative, 2000
Nature, Vol. 408, p. 796-815*

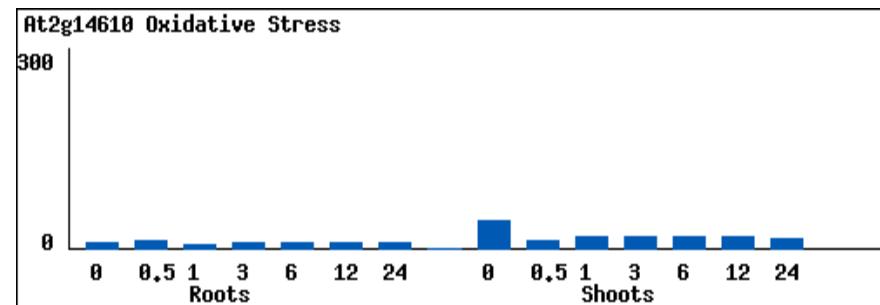
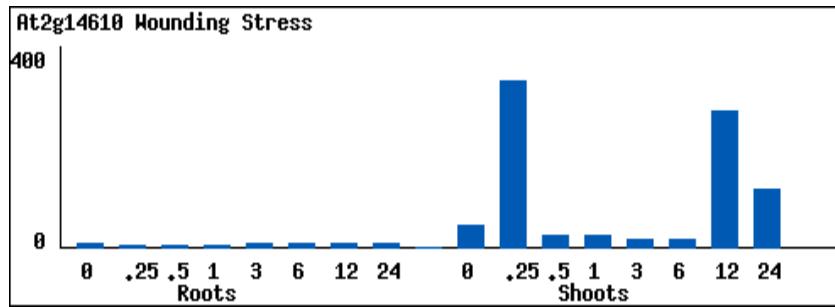
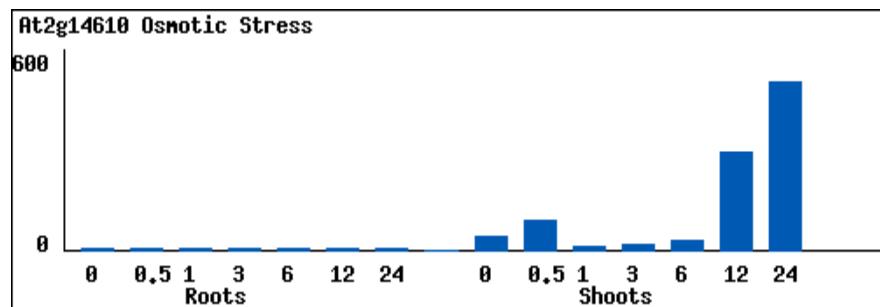
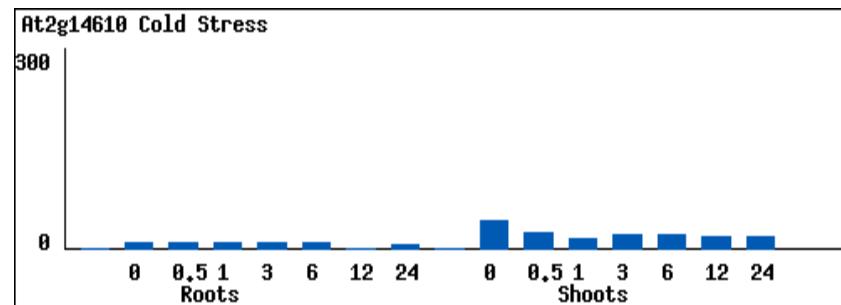
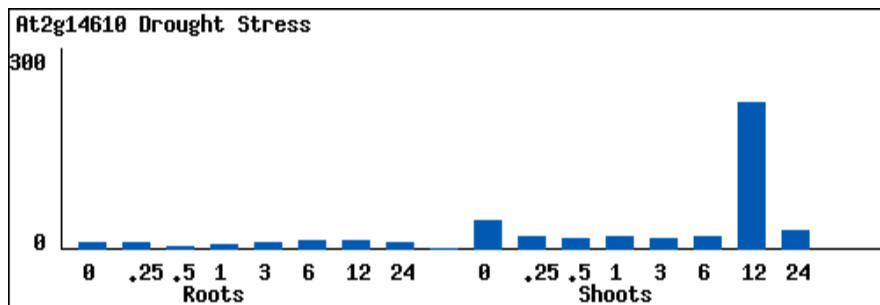
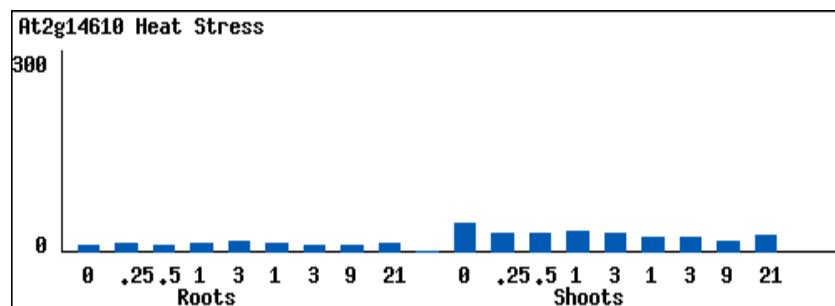
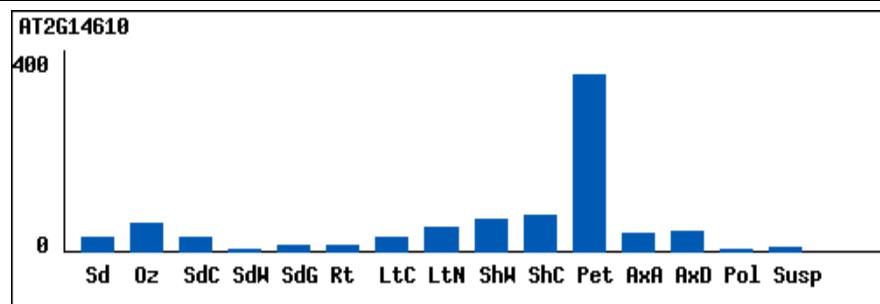
- Près de 40% des gènes d'*Arabidopsis* n'ont pas de fonction biologique qui puisse leur être attribuée a priori sur la base de leur séquence (Multinational coordinated *Arabidopsis thaliana* functional genomics project, Juin 2002)
- Moins de 10% des gènes étiquetés par T-DNA (KO du gène) donnent lieu à une variation visible du phénotype de la plante

D'où ...

Nécessité d'étudier à grande échelle l'expression des gènes

- pour savoir où et quand ils s'expriment (1ère annotation fonctionnelle)
- pour étudier l'effet de conditions variées, de mutations connues ou inconnues sur des réseaux de régulation, des voies métaboliques complètes

Développement d'outils d'étude du transcriptome et du protéome *c-à-d.* de l'expression des gènes par des approches globales.



Le Transcriptome

- Introduction
- Méthodes d'analyse du transcriptome
- Puces à ADN : principe et méthodologie
- PCR en temps réel

II- METHODES D'ETUDE DU TRANSCRIPTOME

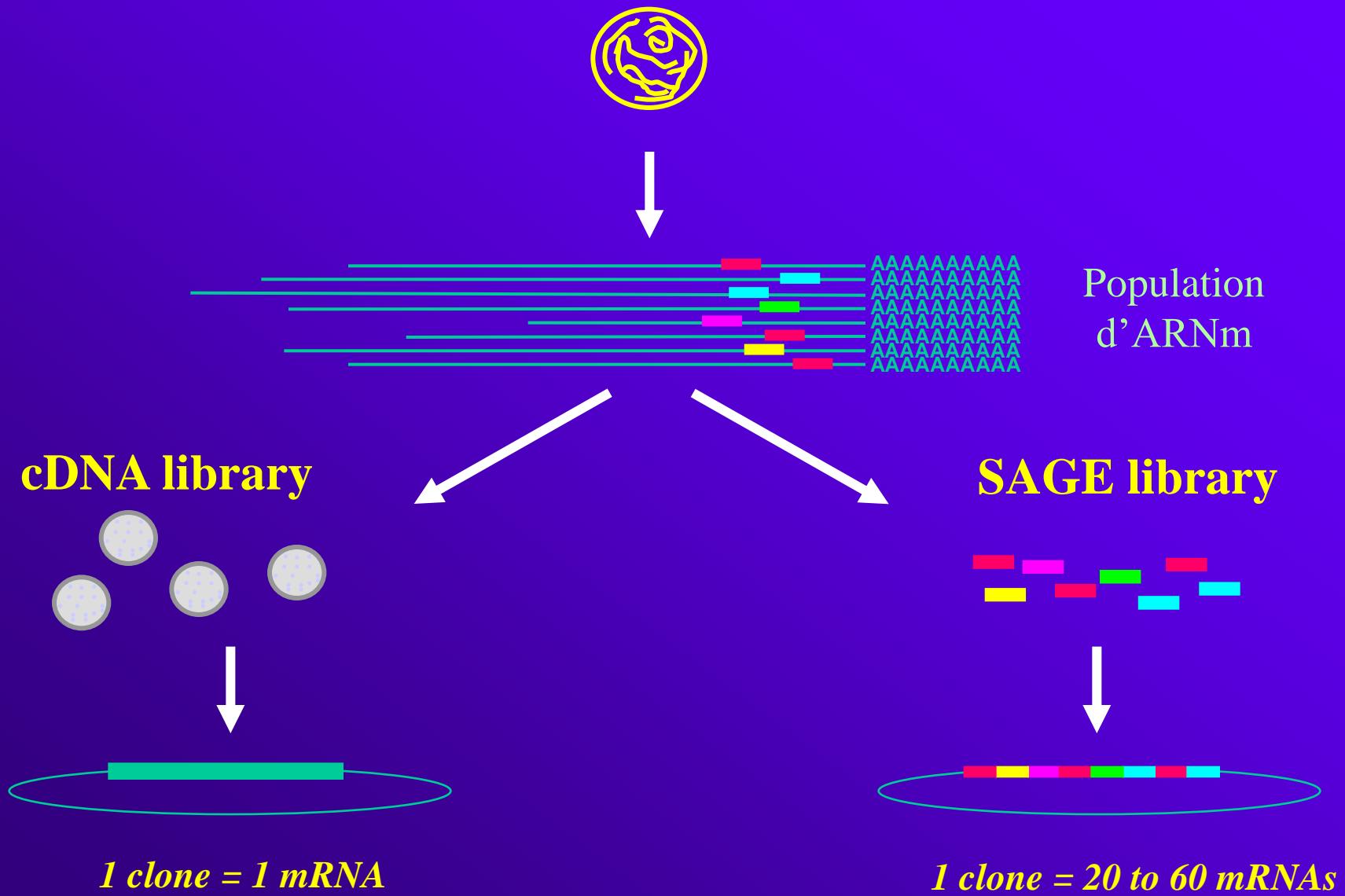
- L'objectif de ces techniques est de mesurer simultanément dans un échantillon biologique (tissus normal, tumeurs) les espèces d'ARNm exprimés et leurs quantités respectives.
- Il s'agit de méthodes globales (de 1000 à 40 000 gènes étudiés simultanément) qui permettent d'avoir des résultats qualitatifs (quels sont les gènes exprimés dans telle situation biologique) et quantitatifs (à quels niveaux sont-ils exprimés)
 - SAGE et séquençage d'EST
 - Séparation de fragments : DDRT-PCR
 - Macroarrays et microarrays

La technique SAGE

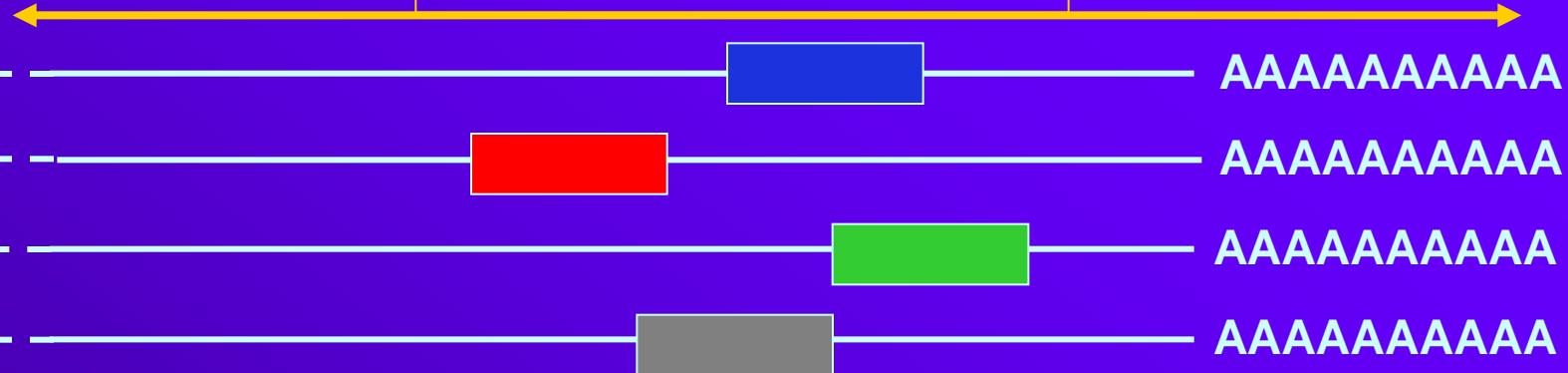
- Serial Analysis of Gene Expression
- Technique basée sur deux principes
 - La représentation de chaque espèce d'ARNm par des étiquettes (tags) correspondant à de petites séquences (10-15 pb)
 - La concanétation de ses tags avant clonage et séquençage



Principe de la technique



Extrémités 3' des ARNm



Extraction tags, ligation, clonage



Vecteur

Séquençage

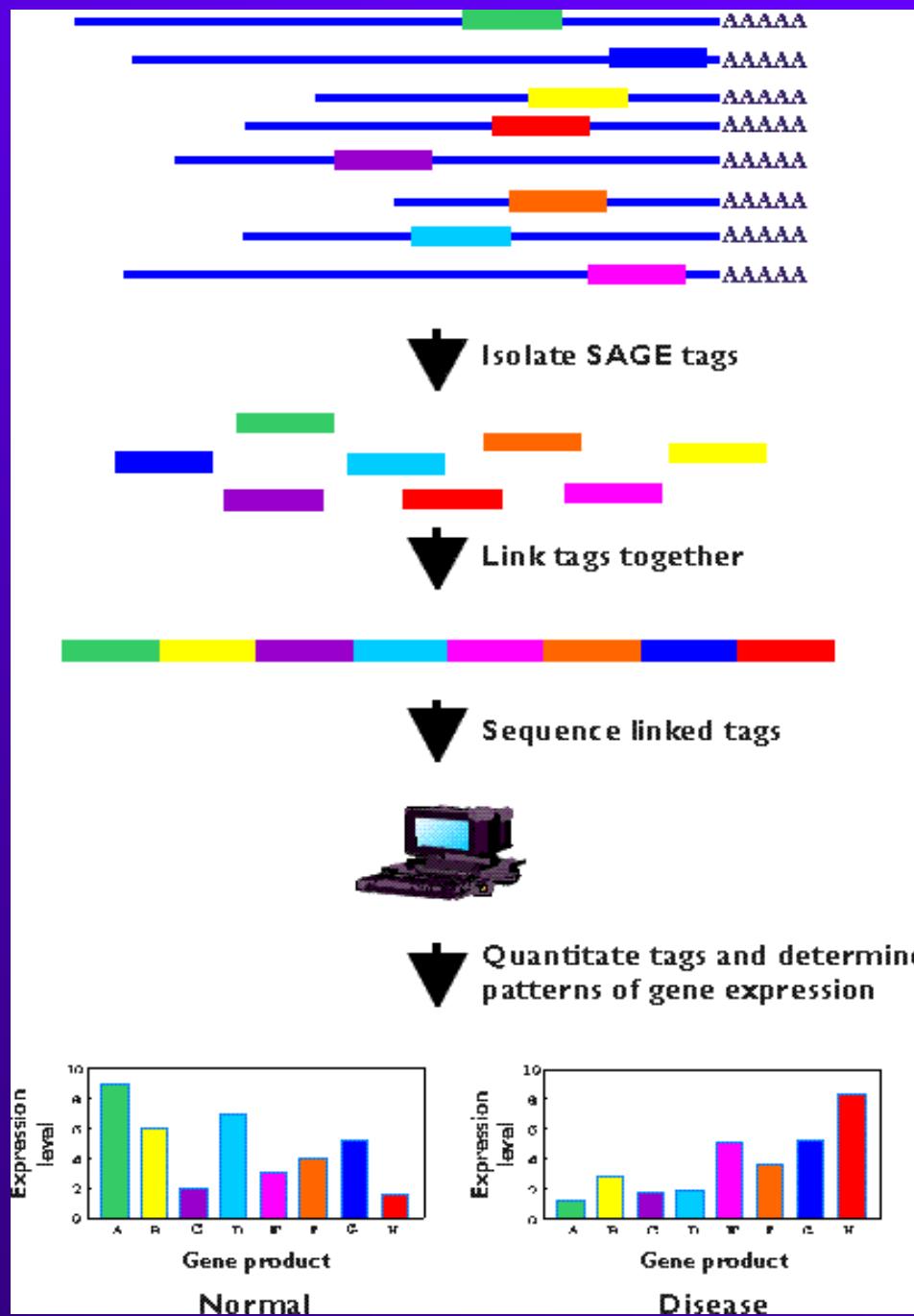


3

2

1

1



- Les tags font entre 10 et 15 pb. Probabilité est de 4^{10} (1 chance sur 1 million) et 4^{15} (1 chance sur 1 milliard) si les bases sont distribuées au hasard dans les séquences
- 1 séquençage permet de lire environ 500 à 600 bases ce qui correspond à 40-60 tags (donc 40-60 molécules d'ARNm)

Population représentative des ARNm
présents dans l'échantillons de départ

AAAAAAA

Synthèse ADNc
Amorce oligodT biotinylée

AAAAAAA biotine
T TT TTT

Coupure par Enzyme d'Ancre (*Nla*III)
Reconnaissance 4 bases
(1 coupure toutes les 256 pb en moyenne
 $= 4^4$)

GTAC

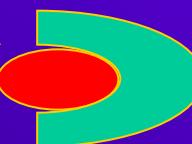
GTAC

GTAC

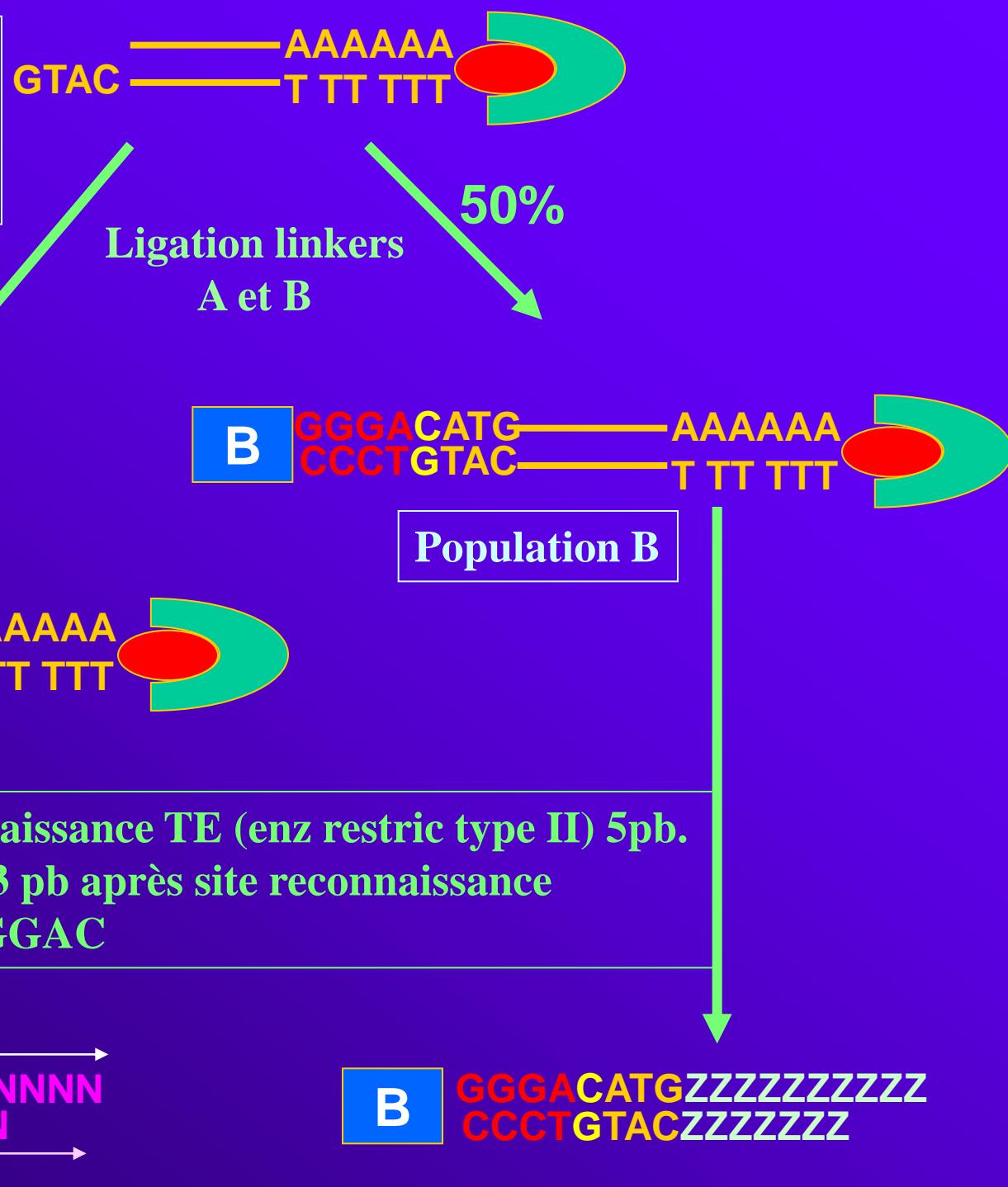
Fragments 3'
de tous les ARNm présents
Dans l'échantillons

GTAC ————— AAAAAA
 T TT TTT

Purification sur billes
de streptavidine



**Population de fragments
représentative des parties 3'
de tous les ARNm présents
dans l'échantillons de départ**



A

GGGACATGNNNNNNNNNN
CCCTGTACNNNNNNNN

B

GGGACATGZZZZZZZZZZZ
CCCTGTACZZZZZZZZZZZ

Création bords francs
T4 DNA pol

13

A

GGGACATGNNNNNNNNNN
CCCTGTACNNNNNNNNNN

B

GGGACATGZZZZZZZZZZZ
CCCTGTACZZZZZZZZZZZ

10

Ligation

A GGGACATG NNNNNNNNNN ZZZZZZZZZZ CATG TCCC
CCCTGTAC NNNNNNNNNN ZZZZZZZZZZ GTAC AGGG **B**

A GGGACATG VVVVVVVVVV YYYYYYYY CATG TCCC
CCCTGTAC VVVVVVVVVV YYYYYYYY GTAC AGGG **A**

B GGGACATG KKKKKKKKKK SSSSSSSSSS CATG TCCC
CCCTGTAC KKKKKKKKKK SSSSSSSSSS GTAC AGGG **B**

PCR amorce A et B

A GGGACATG NNNNNNNNNN ZZZZZZZZZZ CATG TCCC
CCCTGTAC NNNNNNNNNN ZZZZZZZZZZ GTAC AGGG **B**

Population ditags type AB majoritaire (AA et BB donnent des « hairpins » lors PCR)

A GGGACATG NNNNNNNNNN ZZZZZZZZZZ CATG TCCC
CCCTGTAC NNNNNNNNNN ZZZZZZZZZZ GTACAGGG **B**

Population de ditags purifiés

Coupe AE
Purification ditags

NNNNNNNNNN ZZZZZZZZZZ CATG
GTACNNNNNNNNNN ZZZZZZZZZZ

Ligation
Clonage

CATG NNNNNNNNNN ZZZZZZZZZZ CATG WWWWWWWWWWWYYYYYYYYYY CATG...
GTAC NNNNNNNNNN ZZZZZZZZZZ GTAC WWWWWWWWWWWYYYYYYYYYY GTAC...

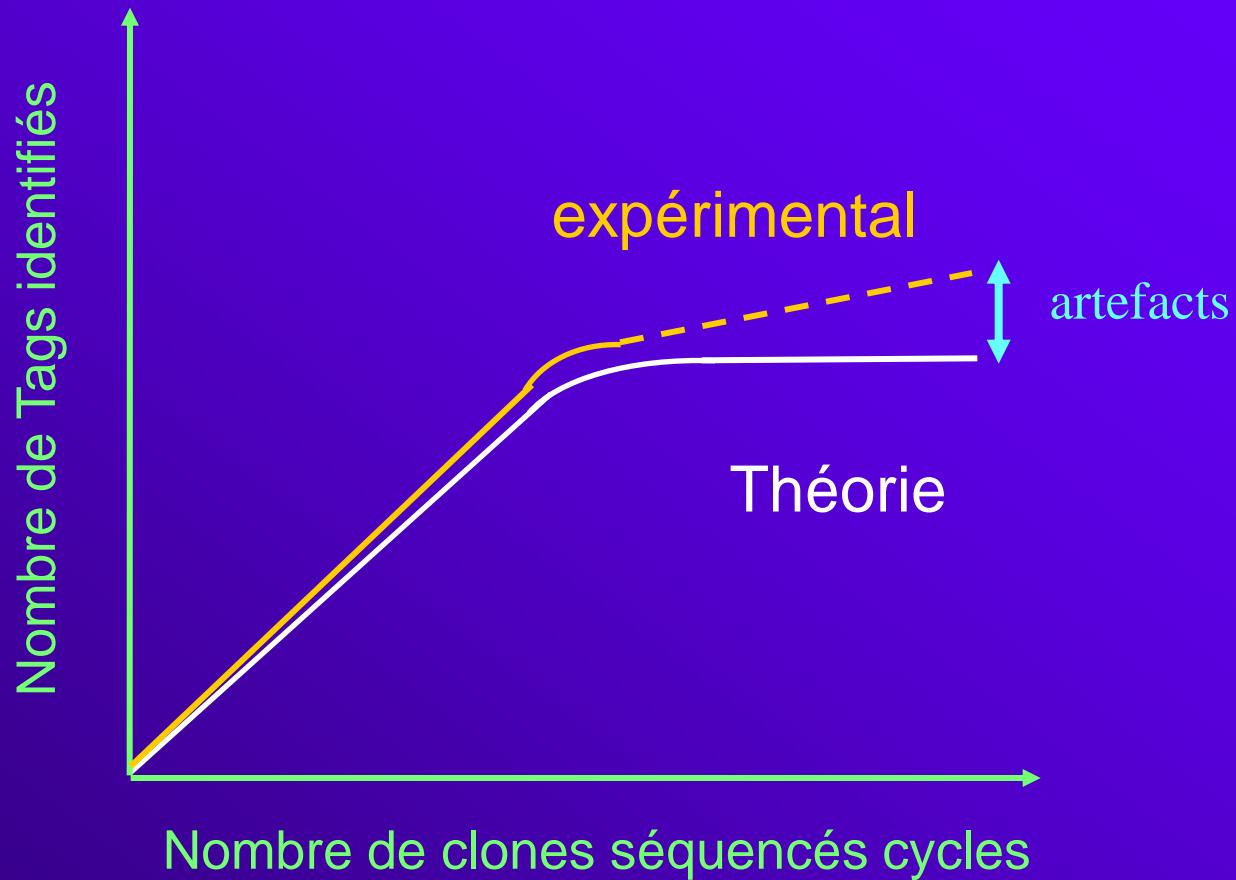
← →
Tag A Tag B

Séquençage

Analyse bioinformatique

Avantages et inconvénients la technique SAGE

- Permet une quantification absolue des ARNm (contrairement aux puces à ADN où la quantification est relative)
- Ne nécessite pas d'investissements lourds
- Technique compliquée et difficile à maîtriser
- Assez peu reproductible (pour un même échantillon, les résultats peuvent être significativement différents)
- Ce n'est pas une technique exhaustive (on peut « louper » des gènes)
- Un même gène peut donner des tags différents (si polyadénylation différentielle)
- Théorie: ditags = 10+10. En réalité, cela peut être 9+11, 8+12 car TE ne coupe pas à 100% à position +10. Pb pour identifier ensuite les transcrits!!!!



Séquençage d'ESTs

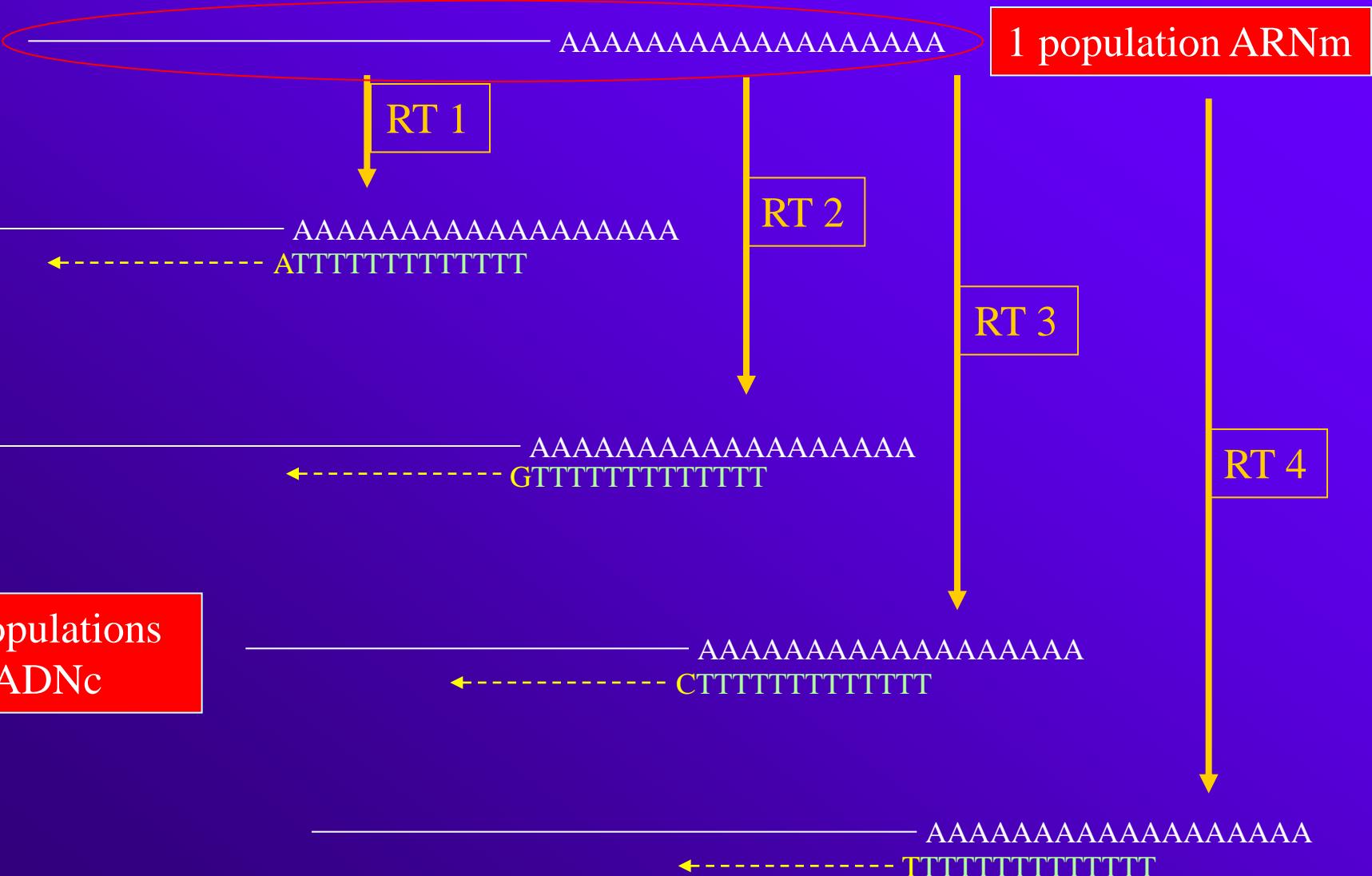
- Expressed sequence tag. Séquençage systématique de tous les clones d'ADNc d'une banque
- Biais liés à la sous-représentation de certains gènes (absence des transcrits les moins abondants)
- Coûteux mais très informatif
- Base de travail pour l'utilisation d'autres méthodes d'étude du transcriptome (ex: macro- ou microarrays) chez des espèces dont le génome n'a pas été séquencé (plantes d'intérêt agronomique par exemple)

La technique de Differential Display Reverse Transcription PCR DDRT-PCR

Principe de la technique

- Synthèse d'ADNc à partir d'une population d'ARNm en utilisant des amorces ancrées oligo dT dont l'extrémité 3' est A, T, G ou C
- Amplification PCR des ADNc en utilisant ces mêmes amorces oligo dT (amorces inverses complémentaires) et des amorces directes aléatoires (10-mers)
- Séparation des produits de PCR sur gel de séquence (acrylamide) et comparaison des profils obtenus avec un même couple d'amorce pour différents échantillons biologiques
- Clonage et identification des bandes d'intérêt

Etape 1: synthèse des 4 populations d'ADNC



Etape 2: PCR a partir des 4 populations d'ADNC

Exemple avec la population obtenue avec ATTTTTTTTTTTTTT

5'

Messager type X

5' TAAAAAAA.....AAA
----- ATT TT TT TT TT TT TT 5'

ATTTTTTTTTTTTTT
----->

Amplifiats correspondant à l'ARNm X

.....

5'

Messager type Y

5' TAAAAAAA.....AAA
----- ATT TT TT TT TT TT TT 5'

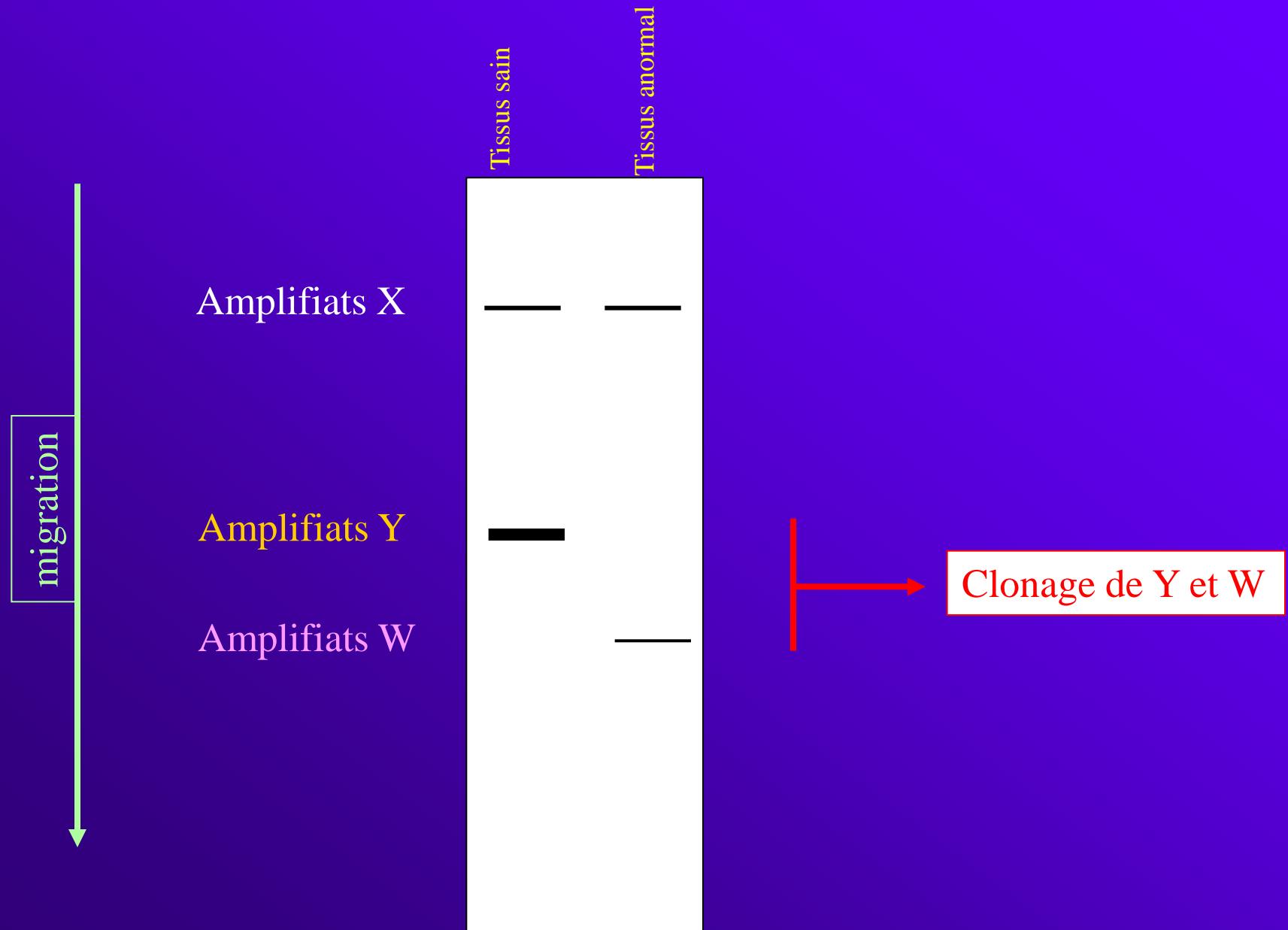
ATTTTTTTTTTTTTT
----->

Amplifiats correspondant à l'ARNm Y

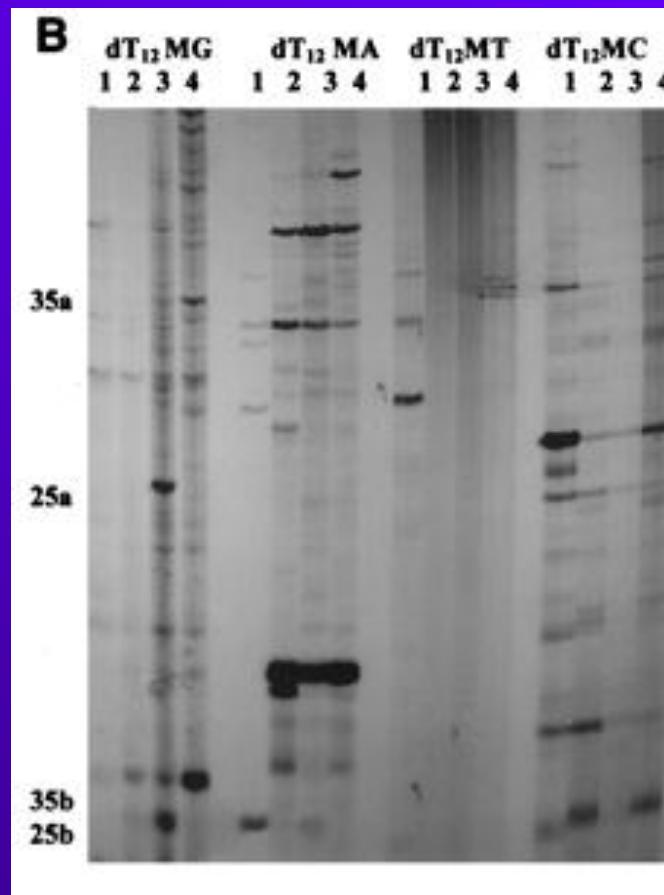
.....

La taille et la quantité des produits de PCR obtenus sont spécifiques de chaque mRNA

Etape 3: Séparation des produits de PCR



Dans le monde réel



Sturtevant, 2000
Clinical Microbiology Reviews
Vol.13, N°3, p.408-427.

Puis clonage et séquençage des bandes d'intérêt: permet l'identification des transcrits différentiellement exprimés

Avantages et inconvénients DDRT-PCR

Mise en œuvre relativement simple

Beaucoup de faux positifs

- Hybridation dans des conditions de faible stringence
- Co-migration de produits de PCR
- Amplification avec les amorces aléatoires (et pas les amorces ancrées)