

# Génomique comparative et approches phylogénomiques chez les insectes, à partir de l'étude de collections de séquences courtes transcrites (EST) – exemple du groupe des pucerons



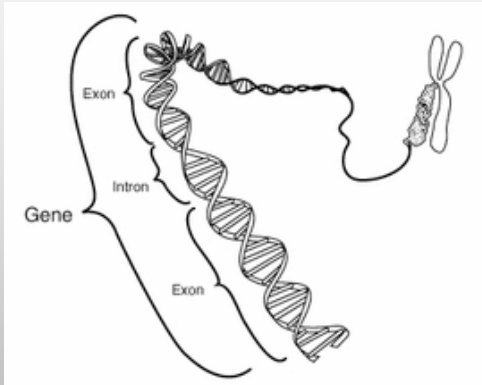
Equipe « Biologie et génétique des populations d'insectes »  
(J.-C. Simon) –UMR BIO3P INRA, Le Rheu –

- modèles agronomiques : pucerons  
(transmission de virus, dégâts directs)

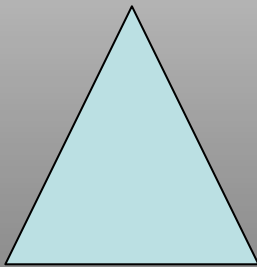
- approches transcriptomiques (Denis Tagu) mode reproducteur, transition clonalité/sexualité

- séquençage en cours du génome complet de *Acyrtosiphon pisum* (puceron du pois)

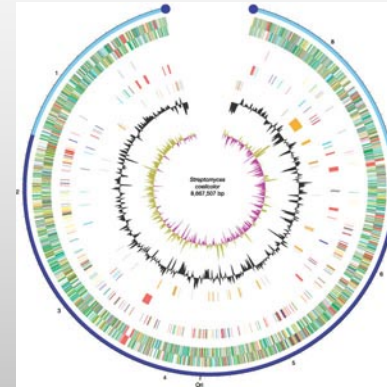
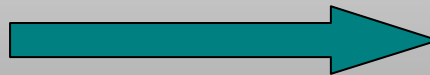
# Biologie moléculaire évolutive: types de données et types d'approches



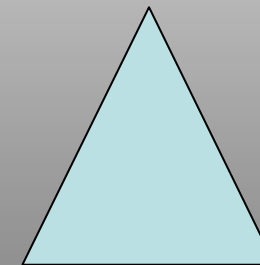
Approche génique



approches comparatives (phylogénie) et populationnelles



Approche génomique



transcriptomique, interactions entre gènes... approches intégratives

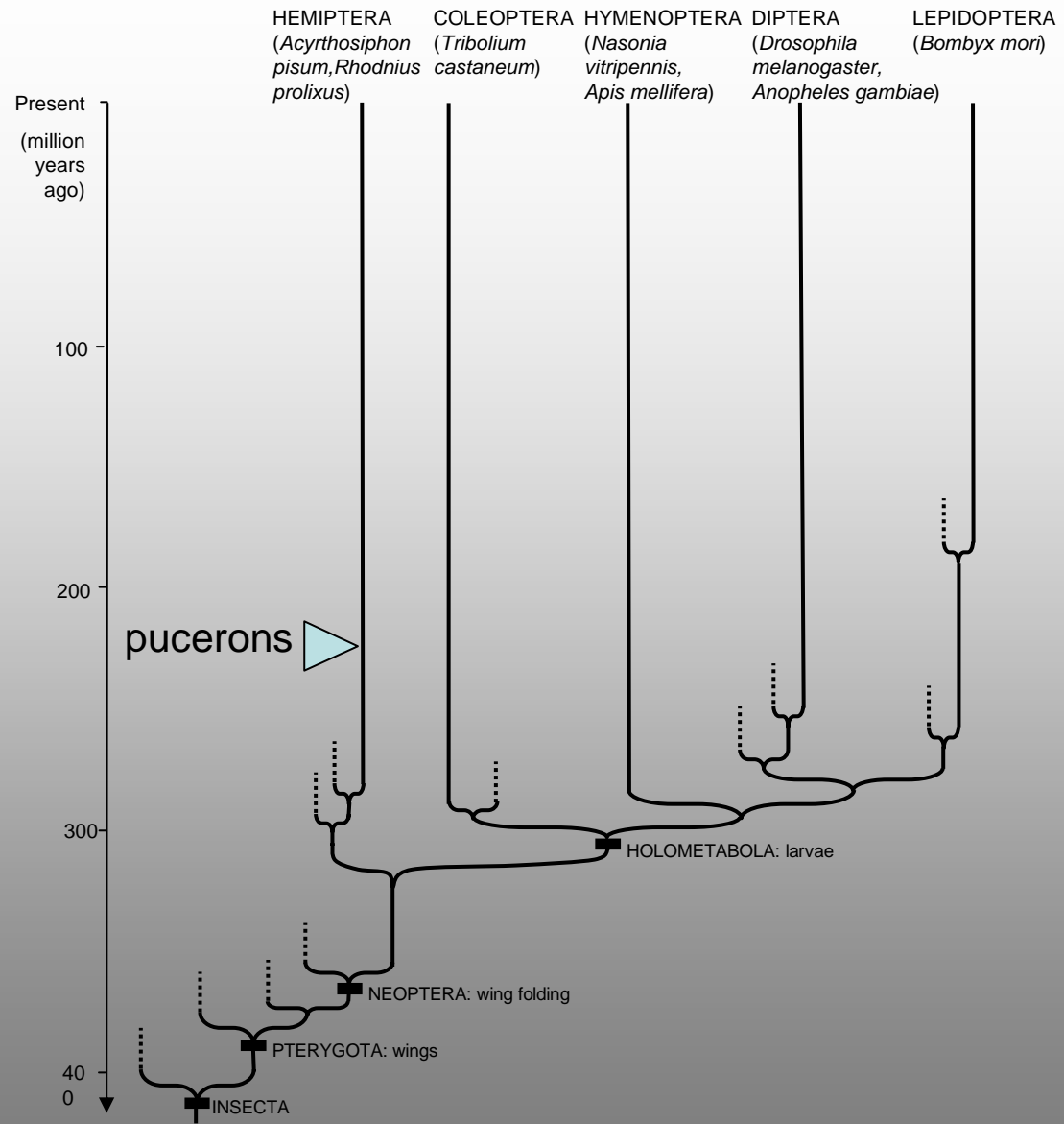
## Une approche intermédiaire pour...

- étudier le génome codant et ses variations d'expression,
- reconstituer des phylogénies robustes,
- décrire les caractéristiques du génome (« style codant »),
- évaluer le niveau de duplication de diverses familles génique ?

• banques d'EST (expressed sequence tags)

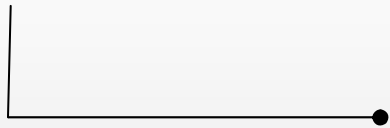
- projets chez les nématodes, reconstitutions de multiples génomes partiels (Blaxter et al.)
- projets chez les insectes (coléoptères, Hughes et al. 2005)
- projets pucerons (*A. pisum* : 60.000 ESTs)

# relations phylogénétiques



## traitement bioinformatique

60.000 ESTs de *A. pisum*



filtrage: qualité, contaminants (rRNA, mtDNA, bactéries)



Assemblage: contigs / singletons



Annotation (BlastX), comparaison avec génome *D. melanogaster*



Extraction des gènes codants ou cds (FrameD)

collabs F. Legeai (URGI,  
INRA Evry)

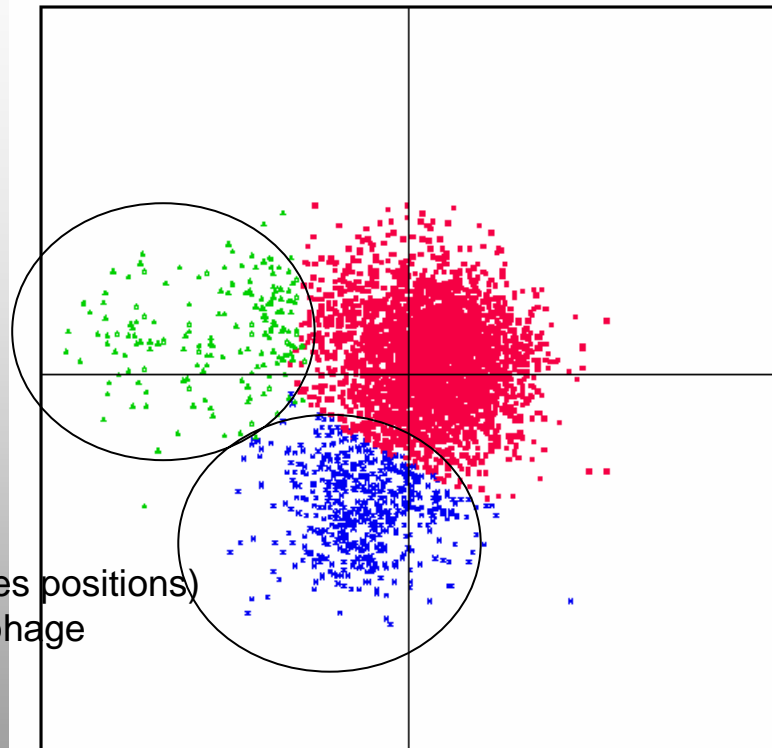
JP Gauthier (INRA Le Rheu)  
-Aphidbase

# Analyse de caractéristiques globales du génome: composition nucléotidique et utilisation des codons

*Bacillus subtilis*

machinerie de la  
transcription et de  
la traduction

gènes riches en AT (3èmes positions)  
– séquences de type prophage



**Figure 4** Factorial correspondence analysis of codon usage in the *B. subtilis* CDSs. Red dots, genes from class 1; green triangles, genes from class 2; blue crosses, genes from class 3. Class 2 contains genes coding for the translation and transcription machineries, and genes of the core intermediary metabolism. Class 3 genes correspond to codons strongly enriched in A or T in the wobble position; they generally belong to prophage-like inserts in the genome.

Kunst et al., 1997.  
Nature

**hétérogénéité  
d'utilisation:**

génome global

biais mutationnel

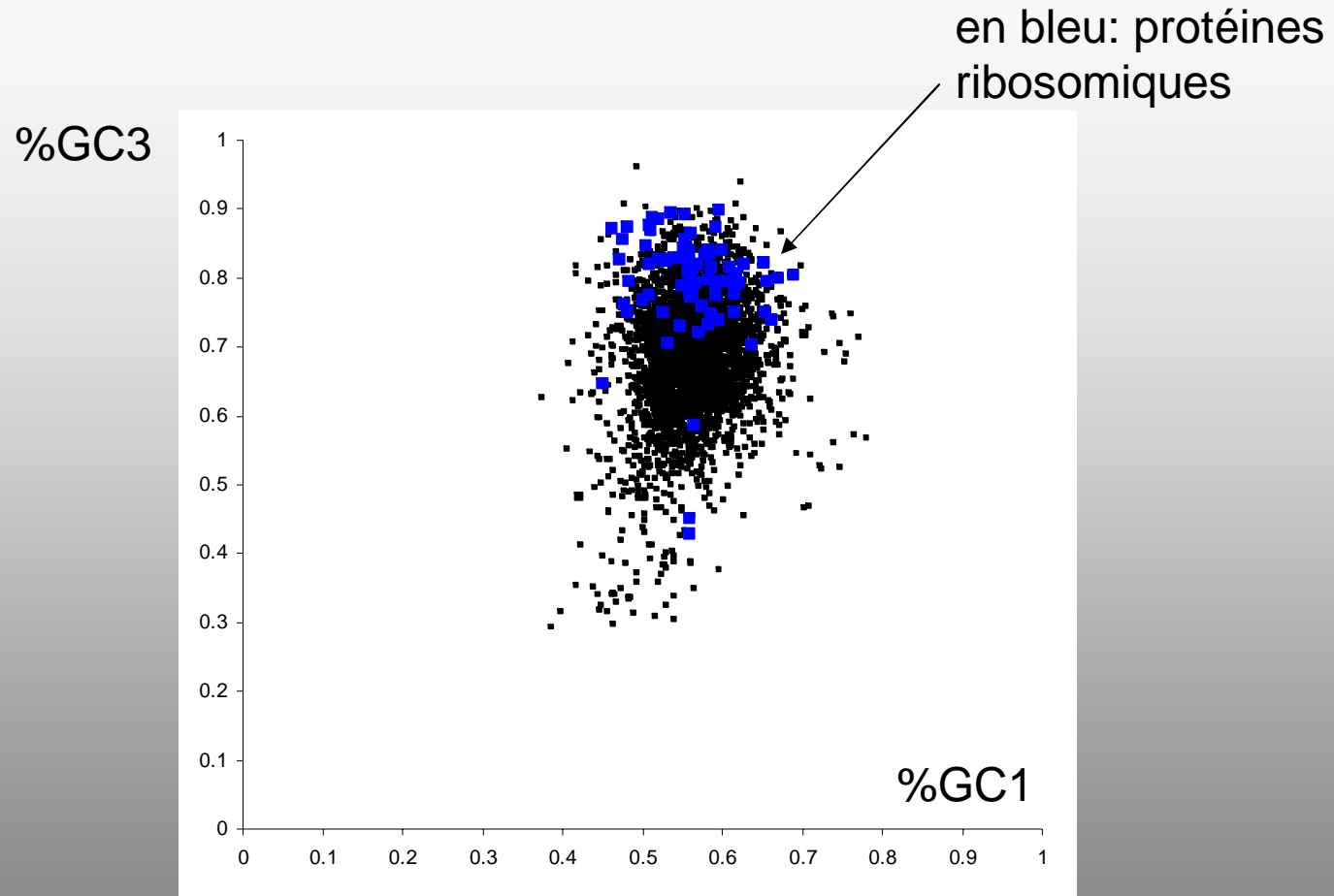
intragénomique

-transfert  
horizontal

-biais locaux  
(isochores,  
orientation)

- biais adaptatifs  
(co-adaptation  
tRNA-codons)

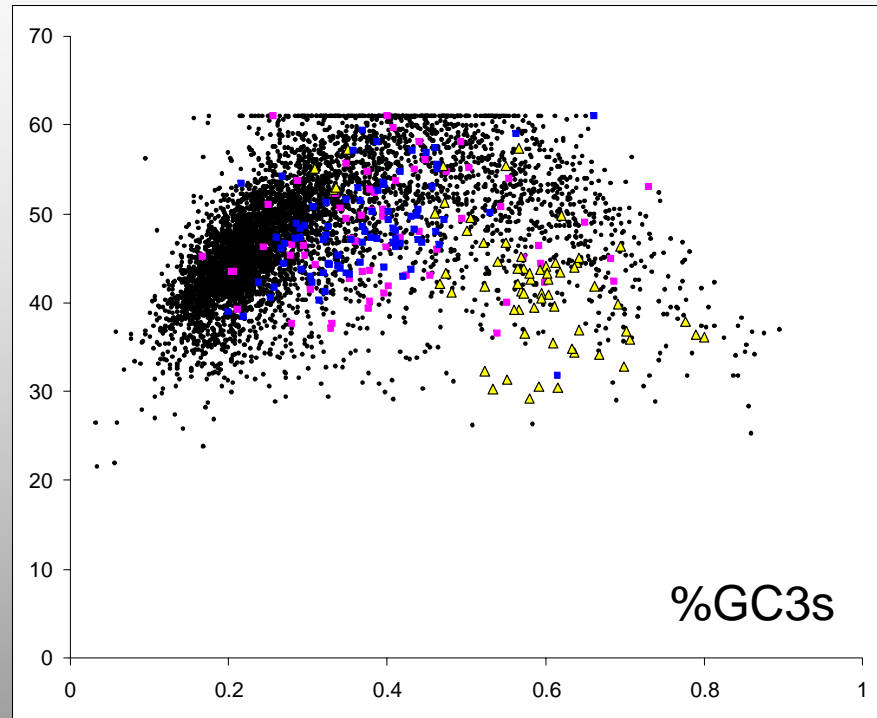
## Analyse de caractéristiques globales du génome: composition nucléotidique et utilisation des codons



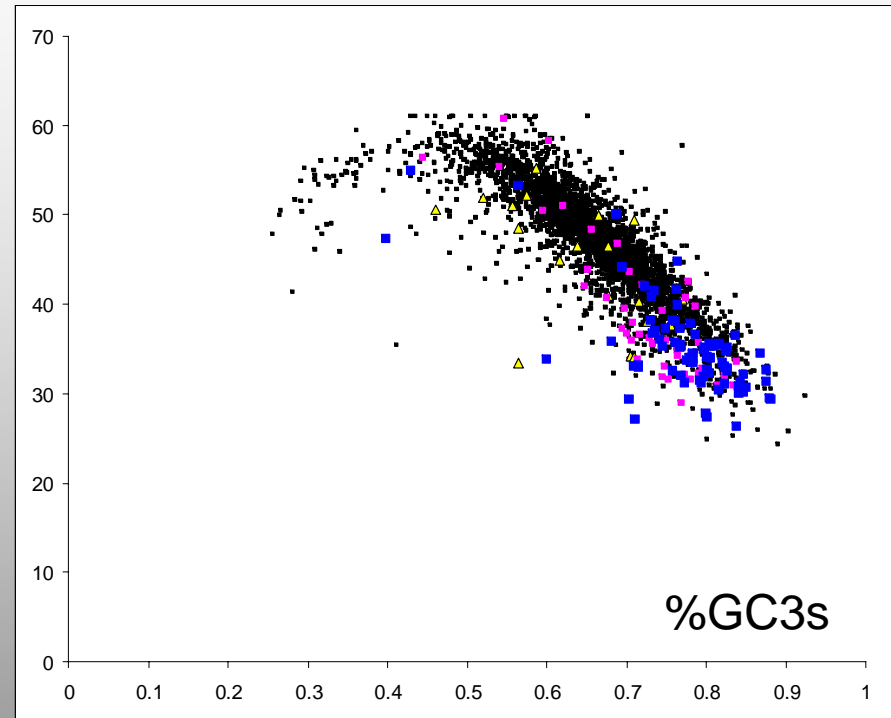
*D. melanogaster* possède un génome codant GC-riche, particulièrement aux troisièmes positions, surtout pour les gènes fortement exprimés

## Biais d'utilisation des codons

↑ Nc, nombre effectif de codons



*A. pisum*

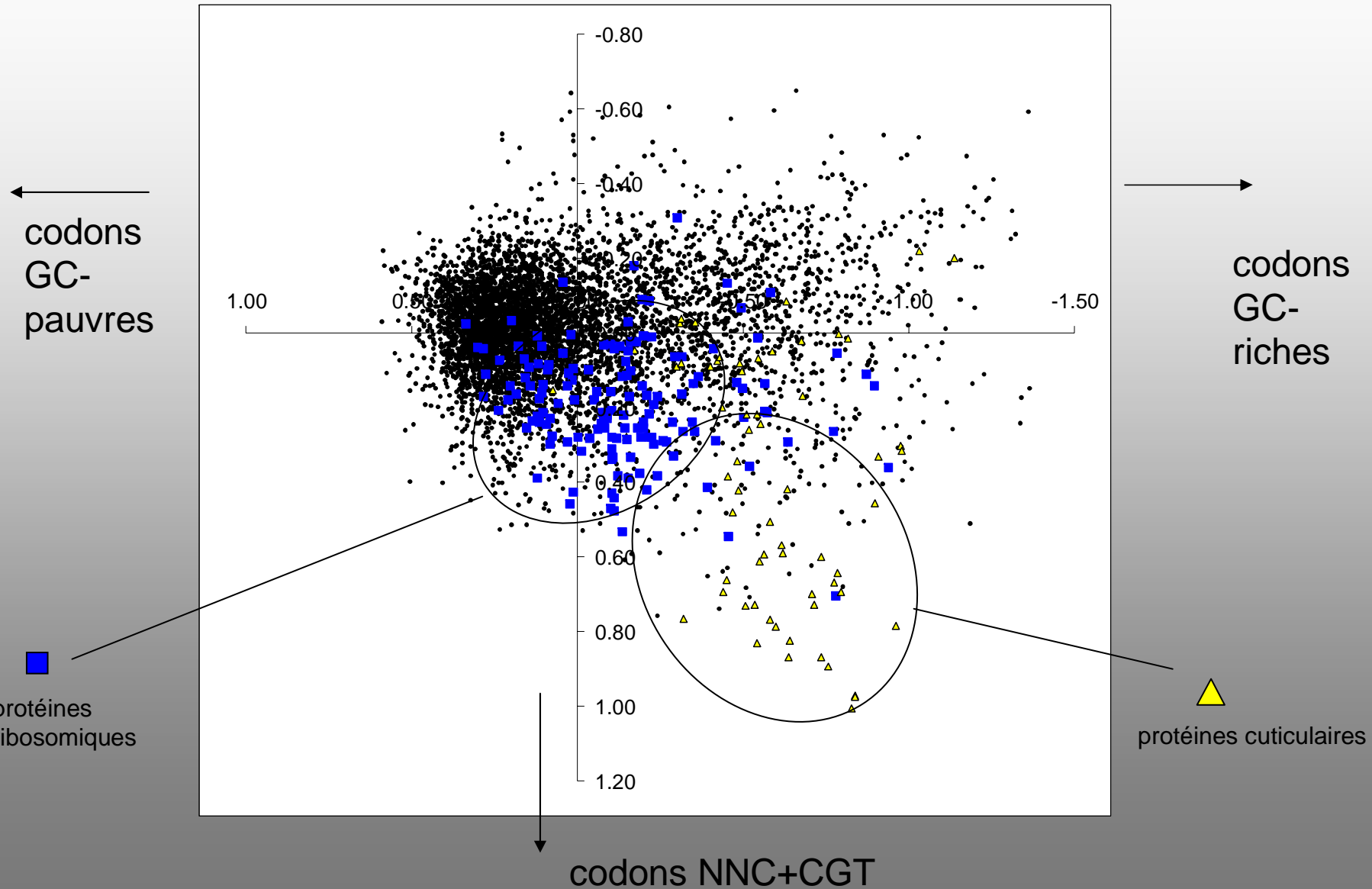


*D. melanogaster*

basculément majeur de composition entre diptères et pucerons



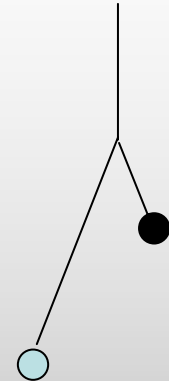
# AFC sur l'utilisation des codons (relative synonymous codon usage =RSCU)



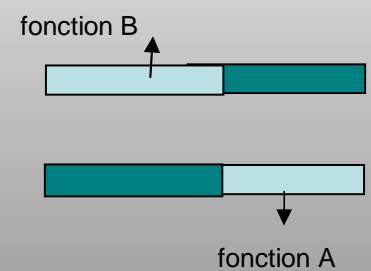
## Conclusions sur l'utilisation des codons

- changement compositionnel majeur puceron (AT-riche) / drosophile (GC-riche)
- spécificité d'usage pour les gènes putativement fortement exprimés (démarcation moindre que chez *D. melanogaster*)
- hétérogénéité très forte : quel facteur pour les protéines cuticulaires ???
- co-adaptation avec les tRNA ? variations locales ?

# Amplification génique et évolution des gènes dupliqués

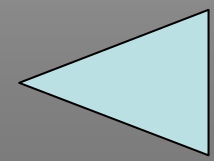


Modèle de Ohno: évolution plus rapide d'une des séquences dupliquées >>> extinction du gène ou néo-fonctionalisation



Modèle de Lynch: évolution accélérée pour les deux copies >>> sub-fonctionalisation

Détection des duplications dans les contigs d'EST



biais possibles: pb de qualité, shimérisme, transcription alternative

## Reconstitution des protéines ribosomiques

*D. melanogaster*

76 protéines  
ribosomiques  
+ 11 paralogues

*A. pisum*

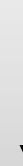
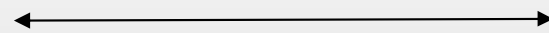
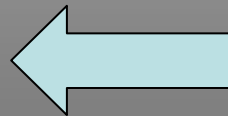
200 contigs

Analyse Blasts réciproques +  
phylogénétiques

extraction des cds

76 protéines ribosomiques (set quasi-complet)  
+ 20 paralogues - 4000 EST

Taux élevé de  
paralogie chez  
*A. pisum* pour  
cette catégorie  
de gènes

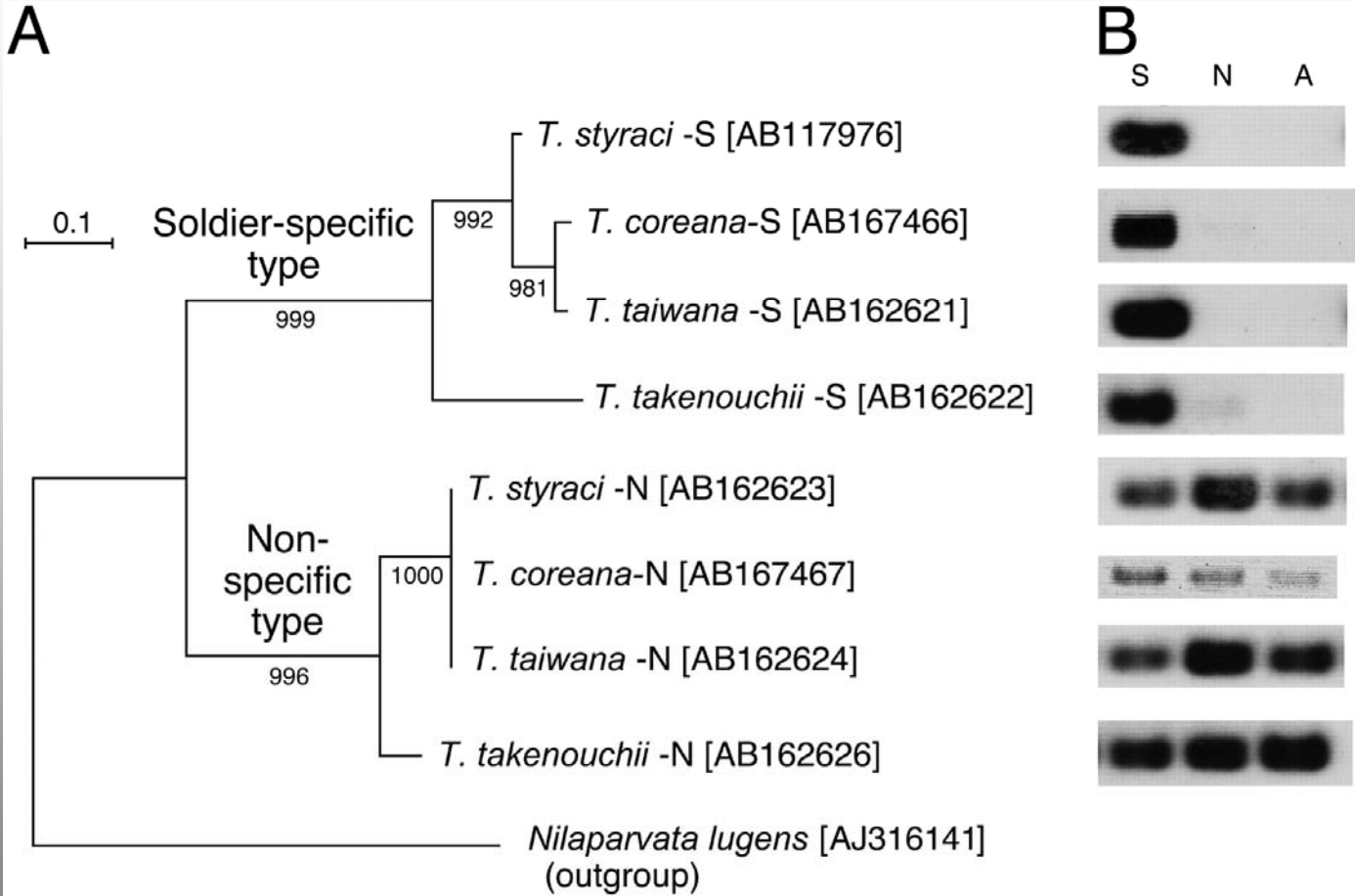


## Une histoire particulière : duplication des cathepsines

protéases dont la fonction est généralement digestive (lysosomes, dégradation des protéines)

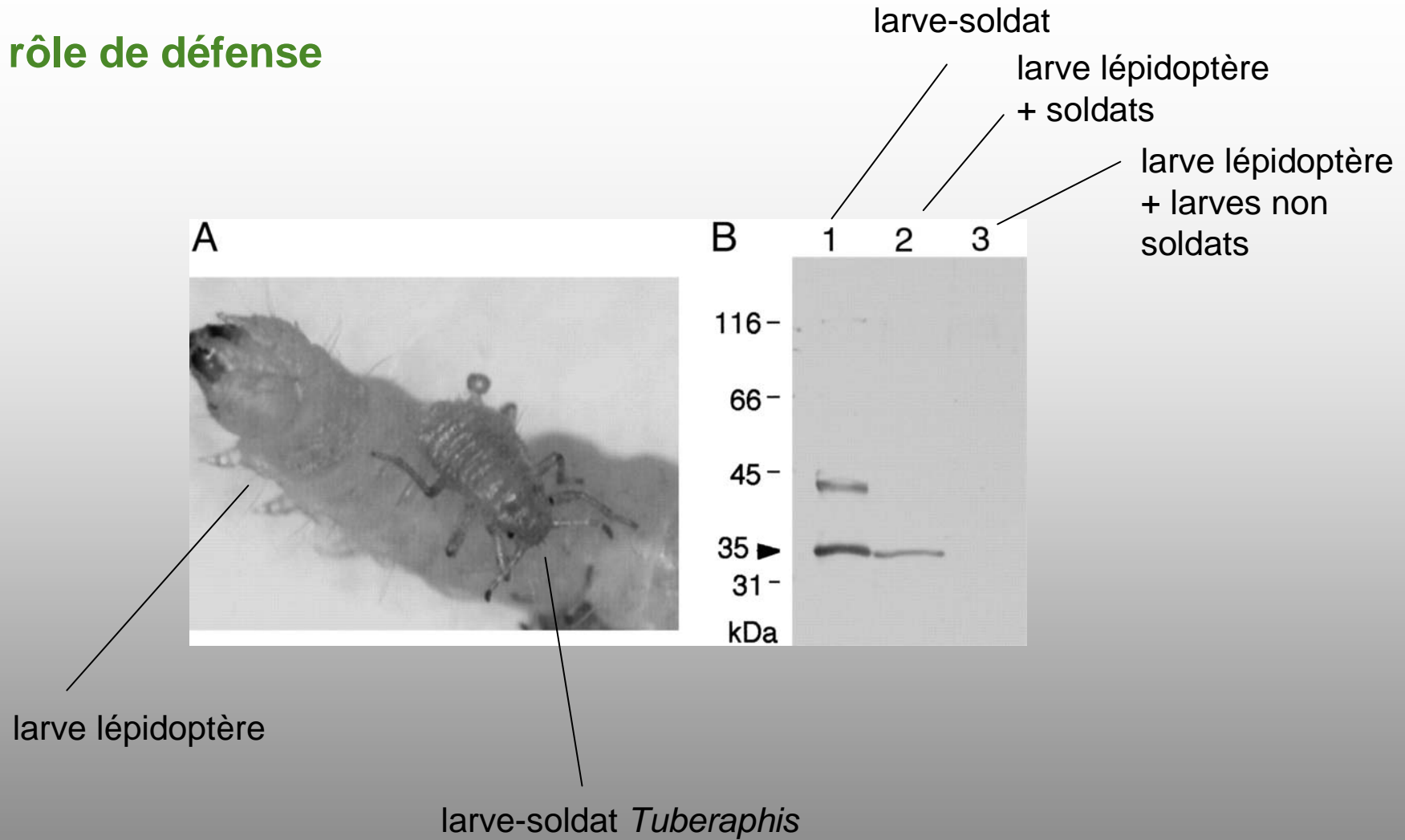
amplification génique des cathepsines L (nématodes, diptères, lépidoptères) ou B (coléoptères) : rôle adaptatif supposé

Phylogenetic relationship of soldier-specific and nonspecific cathepsin B genes from four *Tuberaphis* species



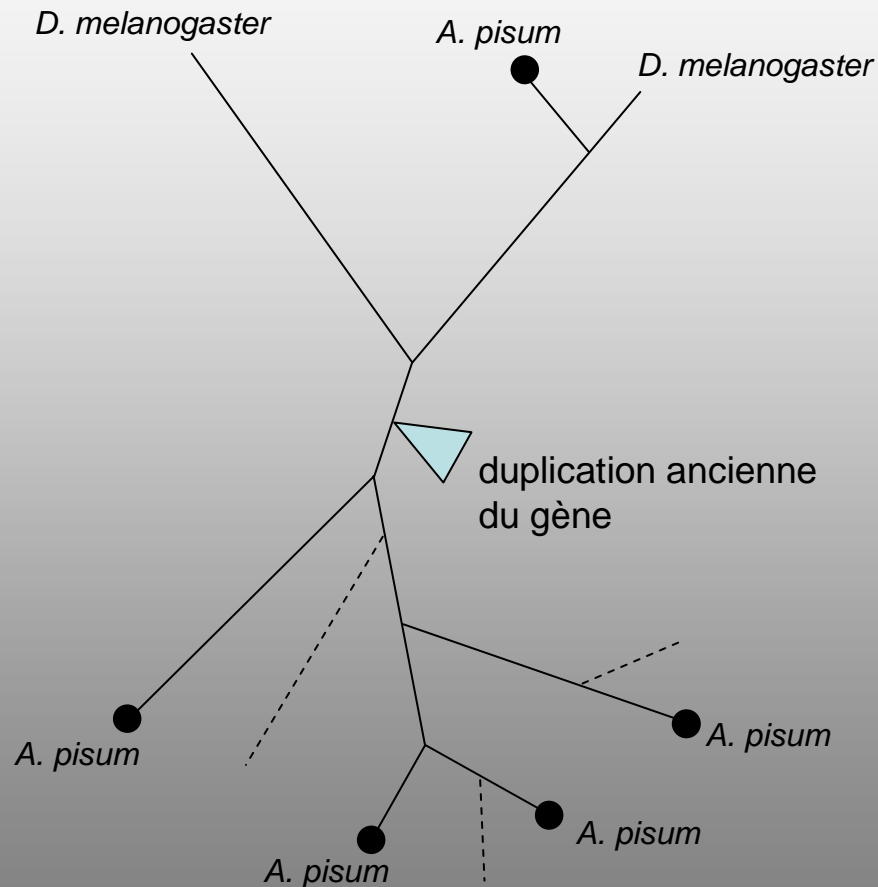
Kutsukake, Mayako et al. (2004) Proc. Natl. Acad. Sci. USA 101, 11338-11343

## rôle de défense



Les soldats injectent une des deux formes de cathepsines aux insectes agresseurs, entraînant leur mort

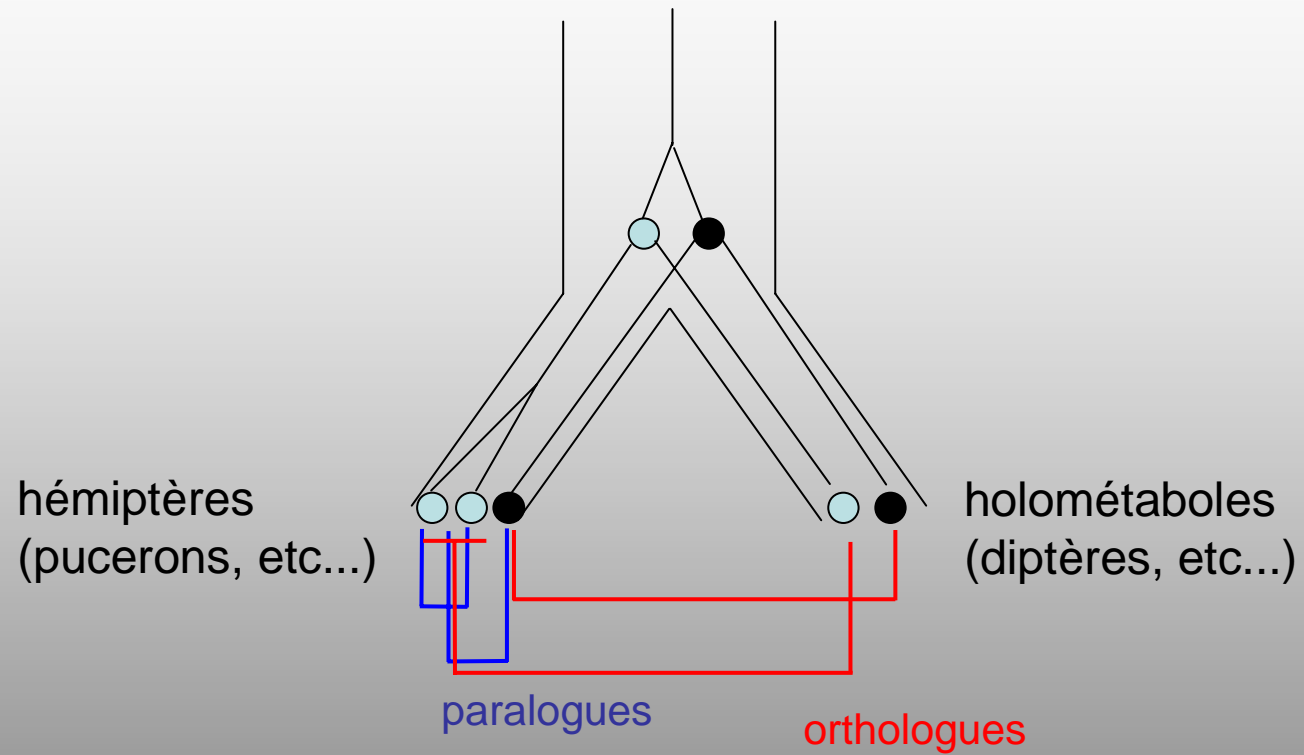
## Schéma interprétatif des relations phylogénétiques entre copies du gène *catB*



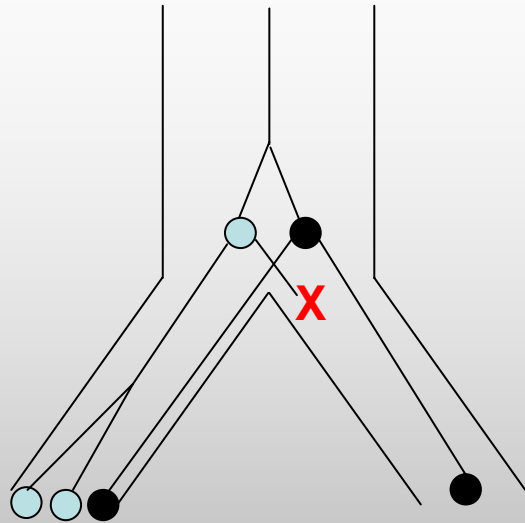
- *D. melanogaster* possède 2 paralogues.
- Au moins 15 paralogues de cathepsine B chez *A. pisum* – duplications répétées



## Scénarios évolutifs pour gènes dupliqués



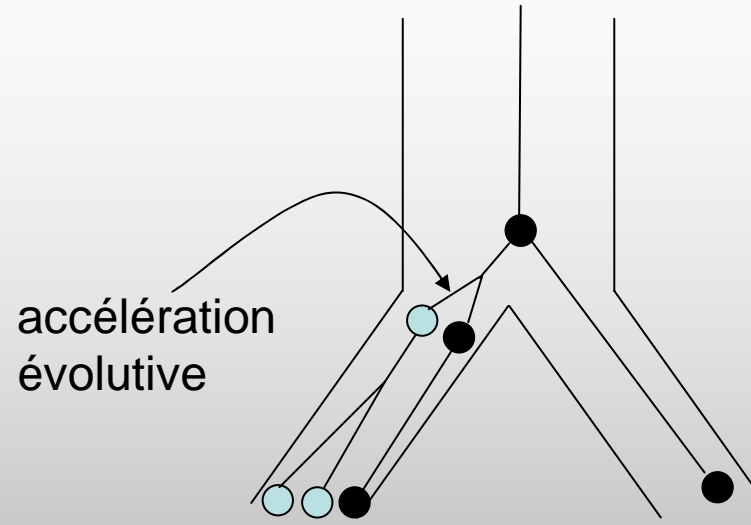
## Scénarios évolutifs pour gènes dupliqués, suite



hémiptères

holométaboles  
(diptères, etc...)

paralogie ancienne + perte  
chez holométaboles



accélération  
évolutive

hémiptères

holométaboles  
(diptères, etc...)

paralogie « récente » +  
différences de taux d'évolution

## Degré et signification des phénomènes de duplication

Amplification globale (WGD=whole genome duplication) ?

Génome *A. pisum* = 530 Mb - Génome *D. melanogaster* = 180 Mb

Amplifications spécifiques au groupe des pucerons ou des hémiptères ?

séquence complète

acquisition de multiples collections d'EST