

# Exploration des génomes



Université de Biskra  
Janvier 2012

Daniel Petit

**Unité de Génétique Moléculaire Animale ; UMR1061  
INRA/Université de Limoges**

- Directeur: A. Maftah, Professeur



# Plan

- Recherche des gènes dans un génome
  - bornes intron-exon
  - algorithme
- Relation entre structure modulaire des protéines et introns
  - exon shuffling
- Modifications de l'expression
  - les sites de reconnaissance de MicroARN
  - expression tissulaire et transcrits
- Evolution des gènes par duplication
  - doublement de génome
  - datation moléculaire
  - synténie et paralogie
  - changements de fonctions

# Recherche d'une séquence nucléotidique dans Genbank : site NCBI

blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE\_TYPE=BlastSearch&

BLAST® Basic Local Alignment Search Tool

Home Recent Results Saved Strategies Help

NCBI/ BLAST/ blastn suite

blastn blastp blastx tblastn tblastx

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s)

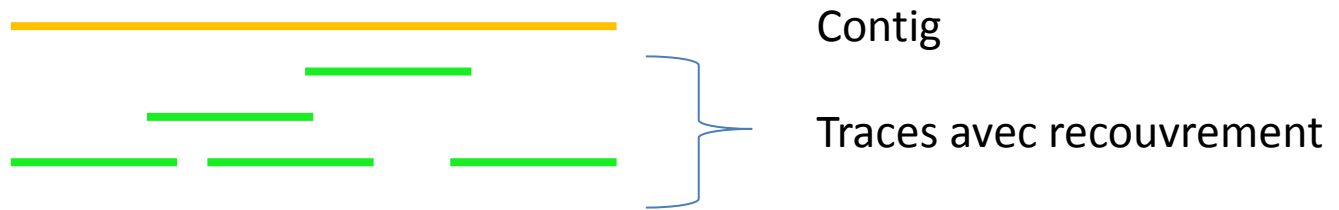
```
gccagtcgctgacagccgcagagctgagagcgtcttctctctctcgcagaagcaggacttctgaat
atatttgaaaactgaacagtttcaaccaagccgaagcatctgtcttcccagagacacaaatcca
acttgagctgaatcacagcagatataagtcacatcggtgaaaagccacataggcagttggatcc
t
```

Options de recherche

	Source nucléotidique	Source peptidique
Recherche nucléotidique	blastn	tblastn
Recherche polypeptidique	tblastx	blastp

## Types de ressources dans GenBank

- Génome assemblé : organismes modèles (Homme, Souris, ... Abeille, Puceron, Tribolium, ... Anémone de mer, Eponge)
- Scaffold
- Contig
- Trace (read)



Gènes avec intron, séquences intergéniques

- EST: expressed sequence tags (ARNm épissés en principe)

Transcrits de gène, sans intron

## Identification des gènes dans le génome assemblé du bovin

gccagtcgctgacagccgcagagctgagagcgtcttctctctcgcagaagcaggacttctgaatatatttgaaaact  
gaacagtttcaaccaagccgaagcatctgtcttcccagagacacaaatccaacttgagctgaatcacagcagatat  
aagtcatcatggtgaaaagccacataggcagttggatcctggttctctttgtggccatgtggagtgacgtgggcctct  
gcaagaagcgaccaaaccctggaggaggatggaacactggggggagccgatacccaggacagggcagtcctgg  
aggcaaccgttatccacctcaggaggggggtggctggggctagccccatggagggtggctggggccagcctcatgg  
aggtggctggggccagcctcatggagggtggctggggctagccccatggtgggtggctggggacagccacatggtgg  
tggaggctgggggtcaaggtggtaccacgggtcaatggaacaaaccagtaagccaaaaaccaatgaagcatg  
tggcaggagctgctgcagctggagcagtggtagggggccttggtggctacatgctgggaagtgccatgagcaggc  
ctcttatacattttggcagtgactatgaggaccgttactatcgtgaaaacatgcaccgttaccccaaccaagtgtact  
acaggccagtggtatcagtagtaaccagaacaactttgtgcatgactgtgtcaatatcacagtcaaggaacacac  
agtcaccaccaccaaggggggagaacttcaccgaaactgacatcaagatgatgaagcgagtggtggagcaaa  
tgtgcattaccagtagcagagagaatcccaggcttattaccaacgaggggcaagtgtgatcctcttcttcccctc  
ctgtgatcctcctcatctcttctcatttttctcatagtaggataggggcaaccttctgttttcattatcttcttaatctt  
accaggttgggggagggagtagtctacctgcagccccgtagtggtggtgtctcatttcgtgcttctcttcttggttacctgt  
atgctaatacccttggcgcttatagcactgggaaatgaagagcagacatgagatgctgtttattcaagtcgtagc  
tcagtagtcaatgccccatcttagcagtgattttgtagcaattttctcatttgtttcaagaacacgtgactacatttccc  
tttggaatagcatttctgccaagtctgga

# The GENSCAN Web Server at MIT

## Identification of complete gene structures in genomic DNA

Organism:  Suboptimal exon cutoff (optional):

Sequence name (optional):

Print options:

Upload your DNA sequence file (upper or lower case, spaces/numbers ignored):

Or paste your DNA sequence here (upper or lower case, spaces/numbers ignored):

```
gccagtcgctgacagccgcagagctgagagcgtcttctctctcgcagaagcaggacttctgaatatatttga
aaactgaacagtttcaaccaagccgaagcatctgtcttcccagagacacaaatccaacttgagctgaatcac
agcagatataagtcacatggtgaaaagccacataggcagttggatcctggttctctttgtggccatgtgga
gtgacgtgggcctctgcaagaagcgacaaaacctggaggaggatggaacactggggggagccgataccag
gacagggcagtcctggaggcaaccgttatccacctcagggaggggggtggctgggggtcagccccatggaggtg
gctggggccagcctcatggaggtggctggggccagcctcatggaggtggctgggggtcagccccatggtggtg
gctggggacagccacatggtggtggaggctgggggtcaaggtggtacccacggtcaatggaacaaaccagta
agccaaaaaccaacatgaagcatgtggcaggagctgctgcagc
>/tmp/01_03_12-04:04:19.fasta|GENSCAN_predicted_peptide_1|264_aa
MVKSHIGSWILVLFVAMWSDVGLCKKRPKPGGGWNTGGSRYPGQGSPGGNRYPPQGGGGW
GQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGGWGQGGTHGQWNKPSKPKTNM
KHVAGAAAAGAVVGGGLGGYMLGSAMSRPLIHFGSDYEDRYRENMHRYPNQVYYRPVDQY
SNQNNFVHDCVNITVKEHTVTTTTTKGENFTETDIKMMKRVVEQMCITQYQRESQAYYQRG
ASVILFSSPPVILLISFLIFLIVG
```

## Interprétation du GenScan

ATG = codon start --> Méthionine

```
gccagtcgctgacagccgcagagctgagagcgtcttctctctcgagaagcaggacttctgaatatttgaaaactgaacagttcaaccaagccg
aagcatctgtctcccagagacacaaatccaacttgagctgaatcacagcagatataagtcatcatggtgaaaagccacataggcagttggatcct
ggttctctttgtggccatgtggagtgacgtgggcctctgcaagaagcgacaaaaacctggaggaggatggaacactggggggagccgataccag
gacagggcagtcctggaggcaaccgttatccacctcagggaggggtggctggggtcagccccatggaggtggctggggccagcctcatggaggt
ggctggggccagcctcatggaggtggctggggtcagccccatggtggtggctgggacagccacatggtggtggaggctggggtcaaggtggtac
ccacggtcaatggaacaaaccagtaagccaaaaaccaatgaagcatgtggcaggagctgctgcagctggagcagtggtagggggccttggt
ggctacatgctgggaagtgccatgagcaggcctcttatacattttggcagtgactatgaggaccgttactatcgtgaaaacatgcaccgttaccca
accaagtgtactacaggccagtggtatagtaaccagaacaactttgtgcatgactgtgtcaatatcagtcaaggaacacacagtcacca
ccaccaccaagggggagaactcaccgaaactgacatcaagatgatgaagcgagtggtggagcaaatgtgcattaccagtaggagagaatc
ccaggcttattaccaacgaggggcaagtgtgatcctctctctccctcctgtgatcctcctcatctcttctcattttctcatagtaggataggggc
aaccttctgttttcattatcttcttaatctttaccaggttgggggagggagtatctacctgcagccccgtagtggtggtgtctcatttcgtgcttctctct
ttgttacctgtatgctaatacccttggcgcttatagcactgggaaatgaagagcagacatgagatgctgtttattcaagtcocgttagctcagtatgct
aatgccccatcttagcagtgattttgtagcaattttctcatttgtttcaagaacacgtgactacatttccttttggaaatagcatttctgccaagtctgga
aggaggccacataatattcattcaaaaaacaaaccgaaatccttagttcatagaccaggggtccacctggttgagagcttgtgtcctgtgtctgc
agagaactataaaggatattctgcattttgaggttacatttgaggtaacacagccagctattgcatcaagaatggatattcatgcaacctttgact
tatgggtagaggacattttcacaaggaatgaacataatacgaaggctctgagactaaaaaattccaacatatgggagaggtgcccttgggtggca
gccttccattttgatgtttaaagcaccttcaagtggtattcctttctttagtaacaaagtatagataattaagttaccttaatttaactaaccttct
```

TAG = codon stop

Dans l'exemple, pas d'intron : séquence monoexonique

# Séquences consensus aux bornes des introns

Règle de Chambon : GT /// AG

Gène codant pour la Thrombospondine 1 de Souris (chromosome 2)

- 1171 AA
- 3516 nucléotides pour ARNm entre ATG et TGA
- 12387 nucléotides pour l'ARN pré-messager entre ATG et TGA
- 20 introns

GTGAGTTTGTG ..

GTAAATCTTCT ..

GTGAGTACCCC ..

GTCAGTGGCTT ..

GTAAGAGACAC ..

GTAAGTTCCTA ..

.....

Partie 5'

..TCTCTCCTTACAG

..CCTCTGCCACAG

..TCTTATTTCTTAG

..TCTCCCCTTTCAG

..TCTTGCCTTGCAG

..TCCCCACAGCCAG

.....

Partie 3'

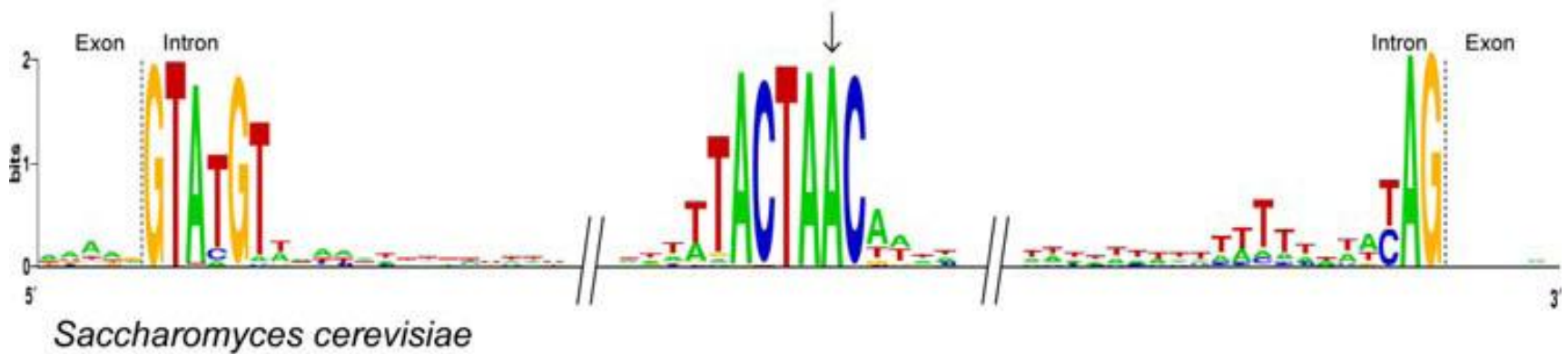


# Généralisation des consensus chez les introns des eucaryotes

Partie 5'

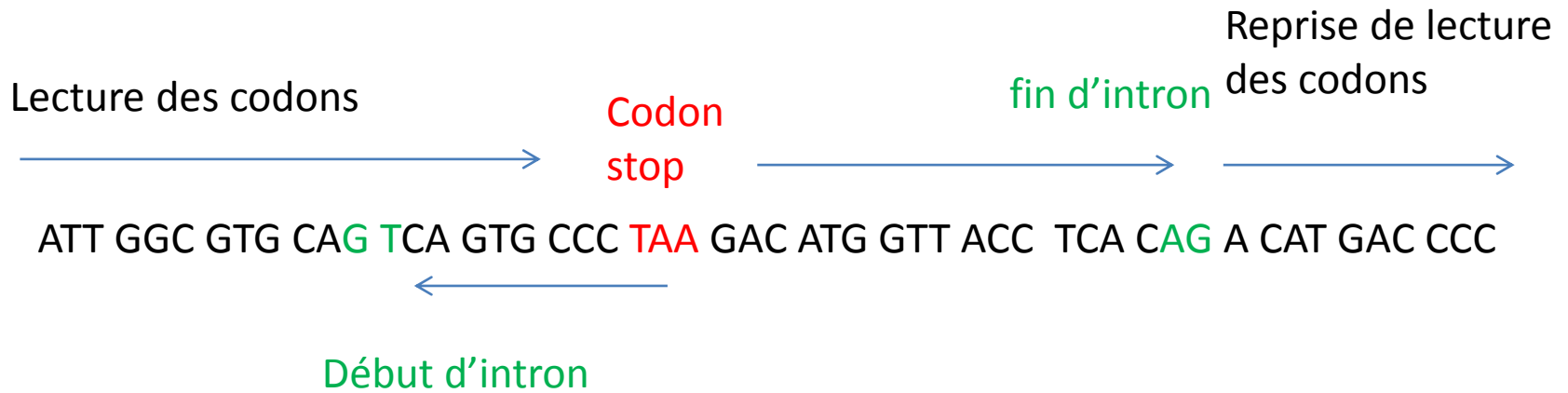
Initiation d'épissage  
(branch point)

Partie 3'



90 % des introns chez *Saccharomyces*  
20 % chez l'Homme

# Reconstitution automatique des gènes épissés



Programmation d'algorithmes avec valeurs seuil de probabilité

Application à l'initiation de transcription, de traduction, du site de polyadénylation

# Relation entre structure modulaire des protéines et introns

Reconnaissance des modules par analogies avec des bases de données dans l'option cdart de NCBI

← → ↻ [www.ncbi.nlm.nih.gov/Structure/lexington/lexington.cgi](http://www.ncbi.nlm.nih.gov/Structure/lexington/lexington.cgi)

Galerie de composa... Sites suggérés CAS – Central Authe... AOL AOL.fr | Email gratui...

**CONSERVED Domain Architecture Retrieval Tool**

HOME SEARCH GUIDE Structure Home 3D Macromolecular Structures Conserved Domains Pubchem

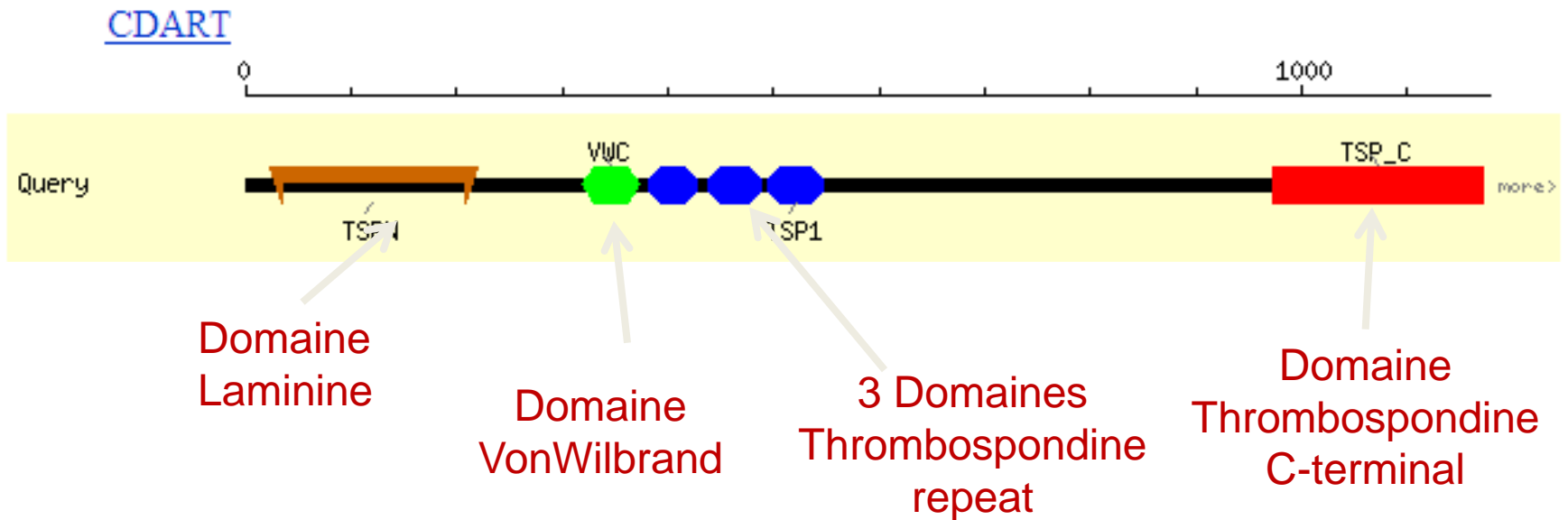
**Launch a new search**

Enter query protein sequence [?](#)

```
mwalrgcrsg srwgqgaall llllgvpprg lalpplyrsh agicpndmnp
nlwvdaqstc
krecetdqc etyekccpnv cgtkscvaar ymdvkkgkkgp vgmpeatcd
hfmclqggse
cdiwdgqpvc kcrdrcekep sftcasdgl yynrcymdae acskgitlav
vtcryhftwp
ntspspett vhpttappet pglatapal lnhpahqsvt vgetvsflcd
vgrprpeit
wekqledren vvmrpnhvirg nvvtvniaql viynaqqda giytotarna
agvlradfpl
```

Submit Reset

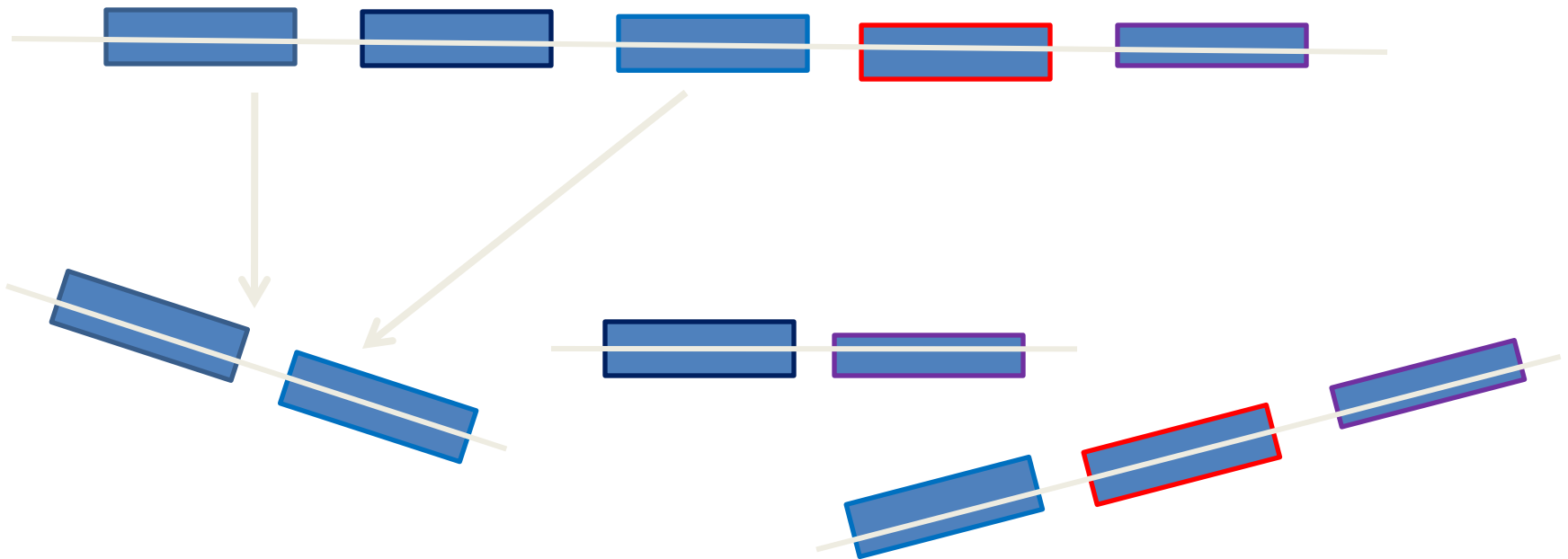
# Exemple de la THROMBOSPONDINE 1 (1171 AA) *Mus musculus*



Comment a pu se faire l'assemblage de ces modules dans différentes familles de protéines ?

# Théorie du mélange d'exons = Exon shuffling = grand bricolage)

Gènes codants pour des modules indépendants



Recombinaison de gènes portant une association particulière de module

Nécessité que chaque module soit codé par un seul exon, entouré par une séquence flanquante

# Prédiction : chaque module de Thrombospondine compris dans un exon

```
1 M E L L R G L G V L F L L H M C G S N R
1 ATGGAGCTCCTGCGGGGACTAGGTGTCTCTGTTCTGTTCATATGTGTGGAAGCAACCGC

21 I P E S G G D N G V F D I F E L I G G A
61 ATTCAGAGTCTGGGGGAGATAACGGTGTGTTGACATCTTTGAACTCATTGGAGGTGCA

41 R R G P G R R L V K G Q D L S S P A F R
121 CGAAGGGGCCCCGGTCGCCGACTGGTGAAGGGCCAAGATCTATCC AGCCCCGCCTTCCGC

61 I E N A N L I P A V P D D K F Q D L L D
181 ATTGAGAATGCCAACCTGATCCCCGCTGTGCCGGATGACAAGTTCCAAGACCTACTGGAC

81 A V W A D K G F I F L A S L R Q M K K T
241 GCTGTGTGGCCGACAAAGGCTT CATCTTCTGGCTTCCCTTGAGGCAGATGAAGAAGACC

101 R G T L L A V E R K D N T G Q I F S V V
301 CGGGGCACACTCCTGGCTGTGGAACGGAAAGACAACACTGGCCAGATCTTCAGTGTGGTC

121 S N G K A G T L D L S L S L P G K Q Q V
361 TCCAACGGCAAAGCTGGCACCCCTGGACCTGAGCCTGAGCCTGCCAGGGAAGCAACAAGTG

141 V S V E E A L L A T G Q W K S I T L F V
421 GTGTCAGTGGAG GAAGCTCTCTGGCCACTGGCCAGTGGAAAGAGCATCACGCTGTTTGT

161 Q E D R A Q L Y I D C D K M E S A E L D
481 CAAGAGGACCGGGTCAACTCTACATAGACTGTGATAAGATGGAGAGCGCGGAGCTGGAT

181 V P I Q S I F T R D L A S V A R L R V A
541 GTACCCATCCAGAGCATCTTCACCAGGGATCTGGCCAGCGTTGCCAGGCTCCGAGTTGCA

201 K G D V N D N F Q G V L Q N V R F V F G
601 AAGGAGATGTCAATGACAATTTTCAGG GGGTGTGCAGAATGTGAGGTTTGTCTTTGGA

221 T T P E D I L R N K G C S S S A T N V L
661 ACCACCCGAGAAGACATTCTCAGGAACAAAGGCTGTCCAGCT CAGCTACCAACGTCCTT

241 L T L D N N V V N G S S P A I R T N Y I
721 CTTACCTTGCACAACAACGTGGTGAACGGTTCAGCCCTGCTATCCGCACCAACTACATC
```

Module 1 = laminine

# Modules thrombospondine repeat

261 G H K T K D L Q A I C G L S C D E L S S  
781 GGCCACAAAACAAGGACCTCCAAGCTATCTGTGGCCTCTCCTGTGATGAAC TATCCAGC

281 M V L E L K G L R T I V T T L Q D S I R  
841 ATGGTCCTGGAACTGAAGGCCTGCGCACCATCGTGACCACTCTGCAGGACAGCATCCGA

301 K V T E E N R E L V S E L K R P P L C F  
901 AAAGTGACGGAAGAGAACAGAGAGCTGGTC AGTGAGCTGAAGCGGCCTCCCCTCTGCTTT

321 H N G V Q Y K N N E E W T V D S C T E C  
961 CACAATGGAGTCCAGTACAAGAACAACGAGGAGTGGACTGTAGACAGTGCACAGAGTGT

341 H C Q N S V T I C K K V S C P I M P C S  
1021 CACTGCCA GAACTCGGTTACCATCTGCAAAAAGGTGCTCCTGTCCCATCATGCCCTGCTCC

361 N A T V P D G E C C P R C W P S D S A D  
1081 AACGCCACAGTTCCTGATGGTGAATGCTGCCACGGTGGTGGC CCAGCGACTCTGCTGAC

381 D G W S P W S E W T S C S A T C G N G I  
1141 GATGGCTGGTCTCCCTGGTCTGAGTGGACCTCCTGCTCTGCCACATGTGGCAATGGAATT

401 Q Q R G R S C D S L N N R C E G S S V Q  
1201 CAGCAACGTGGTTCCTGTGACAGCCTCAACAACAGATGCGAGGGCTCTTCGGTACAG

421 T R T C H I Q E C D K R F K Q D G G W S  
1261 ACAAGGACCTGCCACATTCAGGAGTGTGACAAAAGAT TTAAACAGGATGGTGGCTGGAGT

441 H W S P W S S C S V T C G D G V I T R I  
1321 CACTGGTCTCCATGGTCTGCTCTGTGACCTGTGGTGGACGGTGTGATCACAAGGATC

461 R L C N S P S P Q M N G K P C E G E A R  
1381 CGGCTCTGCAACTCCCCAGCCCCAGATGAACGGGAAGCCCTGTGAAGGTGAAGCCCGG

481 E T K A C K K D A C P I N G G W G P W S  
1441 GAGACCAAAGCCTGCAAGAAGACGCCTGCCAA TTAATGGAGGCTGGGGTCCCCTGGTCA

501 P W D I C S V T C G G G V Q R R S R L C  
1501 CCATGGGACATCTGCTCTGTACCTGTGGAGGAGGAGTGCAGAGACCGCAGCCGACTCTGT

Module 2,3,4 = TSR

901 N D G I P D D R D N C R L V P N P D Q K  
 2701 AATGACGGCATCCCTGATGACAGAGACAACTGCAGGCTGGTGCCCAATCCTGACCAGAAG  
  
 921 D S D G D G R G D A C K D D F D H D N V  
 2761 GACTCTGATGGTGATGGCCGAGGTGACGCCTGCAAAGACGACTTTGACCATGACAATGTG  
  
 941 P D I D D I C P E N F D I S E T D F R Q  
 2821 CCAGATATTGATGACATCTGTCTGAGAATTTGACATCAGTGAACCGATTTCCGACAA  
  
 961 F Q M I P L D P K G S Q N D P N W V V  
 3001 TTCCAGATGATTCCCTAGATCCCAAAGGAACCTCCCAAATGACCCTAACTGGGTTGTC  
  
 981 R H Q G K E L V Q T V N C D P G L A V  
 3061 CGCCATCAGGGCAAAGAACTTGTCCAGACTGTAACTGTGACCCTGGACTTGCTGTAGGT  
  
 1001 Y D E F N A V D F S G T F F I N T E R D  
 3001 TATGATGAGTTAATGCTGTGGACTTCAGCGGTACCTTCTTCATCAACACCGAGAGAGAT  
  
 1021 D D Y A G F V F G Y Q S S S R F Y V V  
 3061 GGTGACTACGCTGGCTTTGTTTTCGGCTACCAGTCCAGCAGCCGCTTCTACGTTGTGATG  
  
 1041 W K Q V T Q S Y W D T N P T R A Q G Y  
 3121 TGGAAACAAGTCACCCAGTCTACTGGGACACCAACCCACAAGGGCTCAGGGATACTCA  
  
 1061 G L S V K V V N S T T G P G E H L R N  
 3181 GGCCGTCTGTAAAGGTTGTGAACTCCACCACCGGCCCTGGCGAGCACCTGCGGAATGCA  
  
 1081 L W H T G N T P G Q V R T L W H D P R H  
 3241 CTGTGGCACACAGGAACACCCCTGGCCAGGTGCGCACCCCTGTGGCATGACCCTCGCCAC  
  
 1101 I G W K D F T A Y R W R L S H R P K T  
 3361 AGCGGCTGGAAAGATTTCACTGCATACAGATGGCGTCTCAGCCACAGGCCAAAGACCGGT  
  
 1121 Y I R V V M Y E G K K I M A D S G P I Y  
 3361 TATAFCAGAGTGGTGATGTATGAAGGAAAGAAAATCATGGCTGACTCGGGACCCATCTAT  
  
 1141 D K T Y A G G R L G L F V F S Q E M V F  
 3421 GACAAAACCTACGCCGGCGGTAGACTAGGCCTGTTCGTCTTCTCTCAGGAAATGGTGTTC  
  
 1161 F S D M K Y E C R D S -  
 3481 TTCTCAGACATGAAATACGAGTGTGAGATTCCTAA

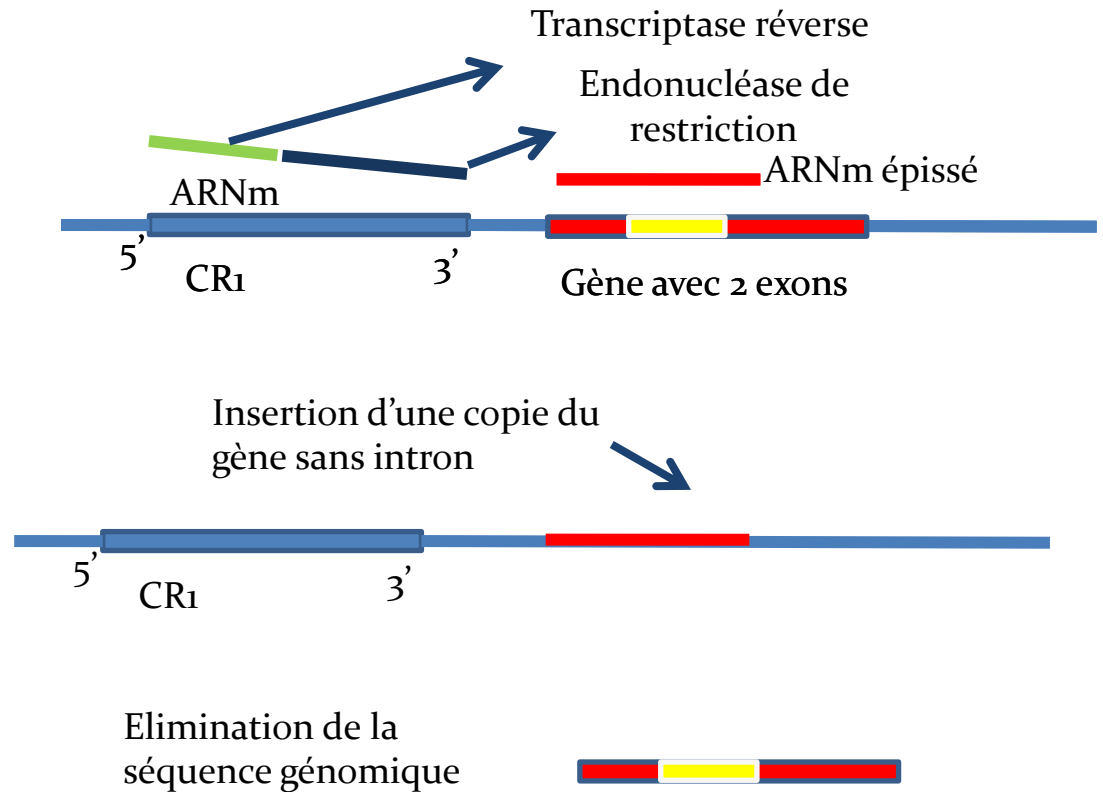
# Module thrombospondine C-ter



## Hypothèse : perte d'introns aux 2 bornes de l'exon

Il peut arriver qu'un (ou plusieurs modules) d'une protéine ne soit pas codé par un exon

Rétrotransposition et recombinaison partielle ou complète

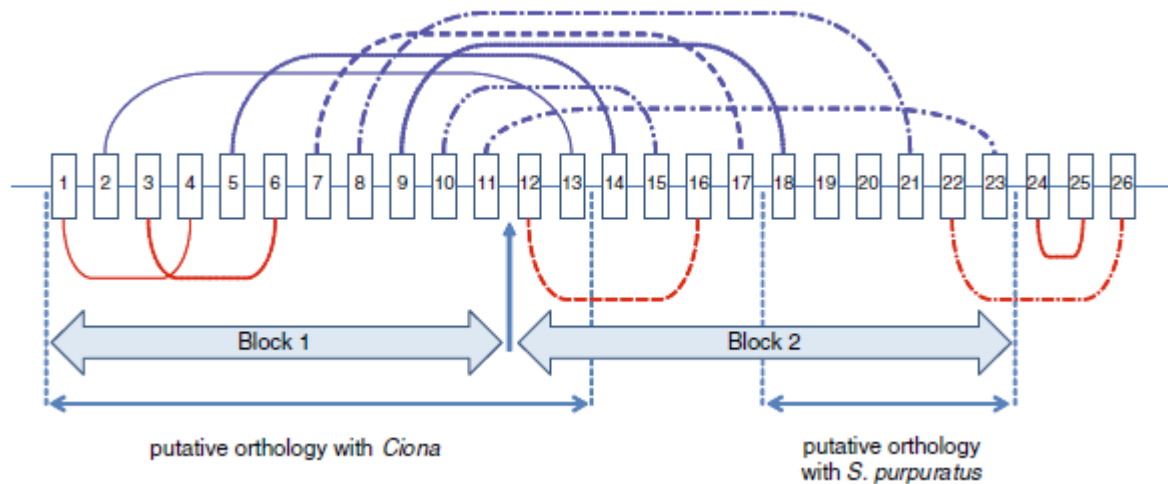
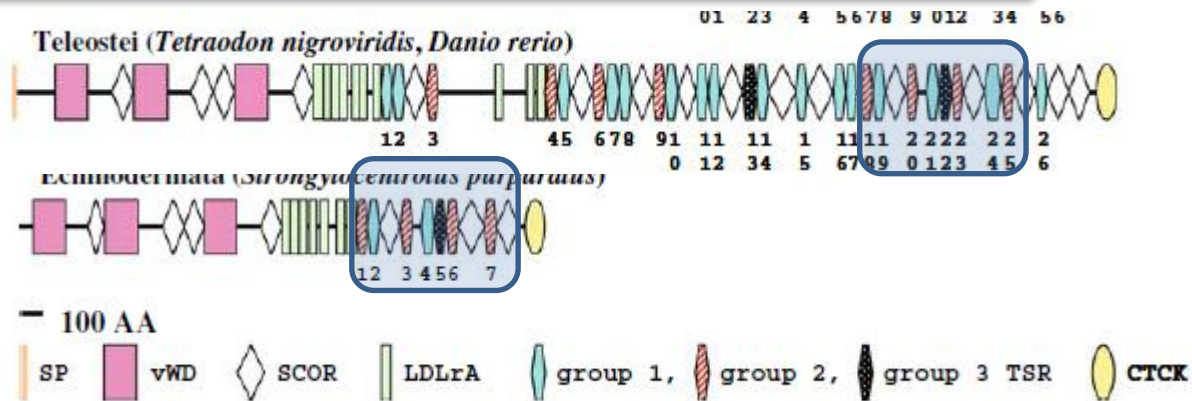


Subcommissural organ  
 (SCO)-spondin

5000 amino acids

The Lengthening of a Giant Protein: When, How, and Why?

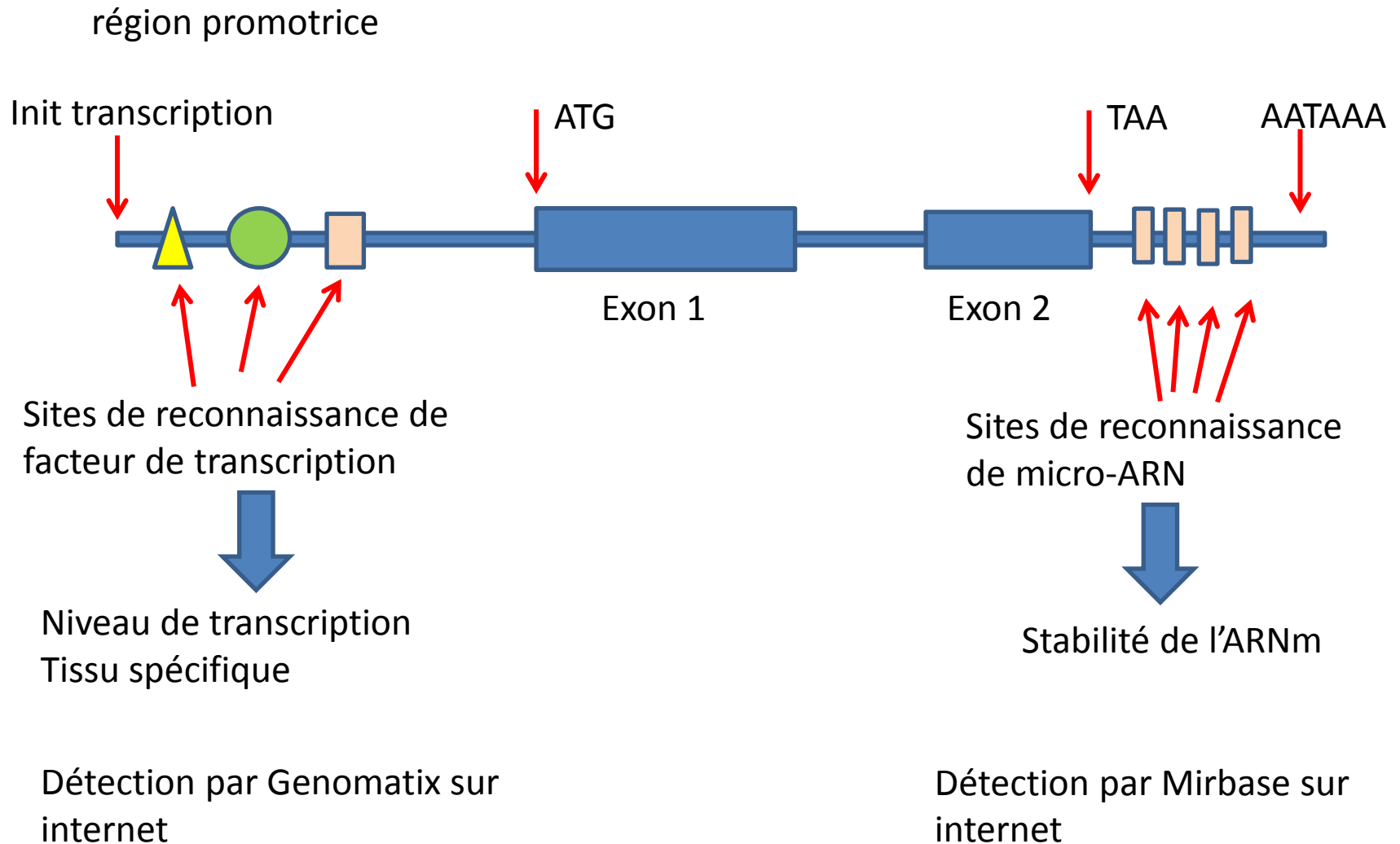
Olivier Meiniel · Robert Meiniel · Fabrice Lalloué · Robert Didier ·  
 Marie-Odile Jauberteau · Annie Meiniel · Daniel Petit



# Modifications de l'expression

- Modification de la quantité d'ARNm
- Modification des tissus où s'expriment les gènes

# Régulation des ARN produits



# Analysis of 3'UTR of *Prnp* Gene in Mammals: Possible Role of Target Sequences of miRNA for TSE Sensitivity in Bovidae and Cervidae

*Daniel Petit, Jean-Michel Petit and François Gallet,  
LIGMA, UMR 1061 INRA/University of Limoges  
France*

Veterinary medicine,  
2012, InTech edition,  
in press

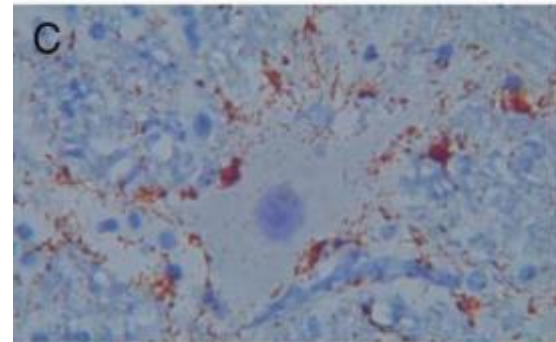
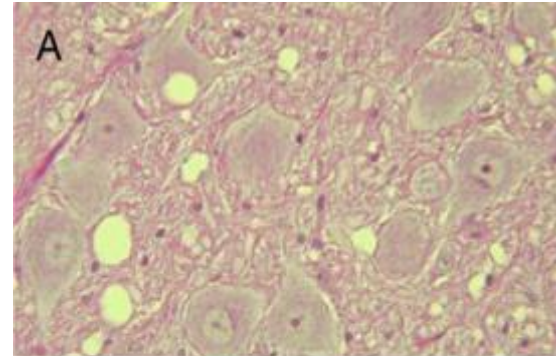


Maladie de la vache folle  
**encéphalopathie spongiforme bovine**

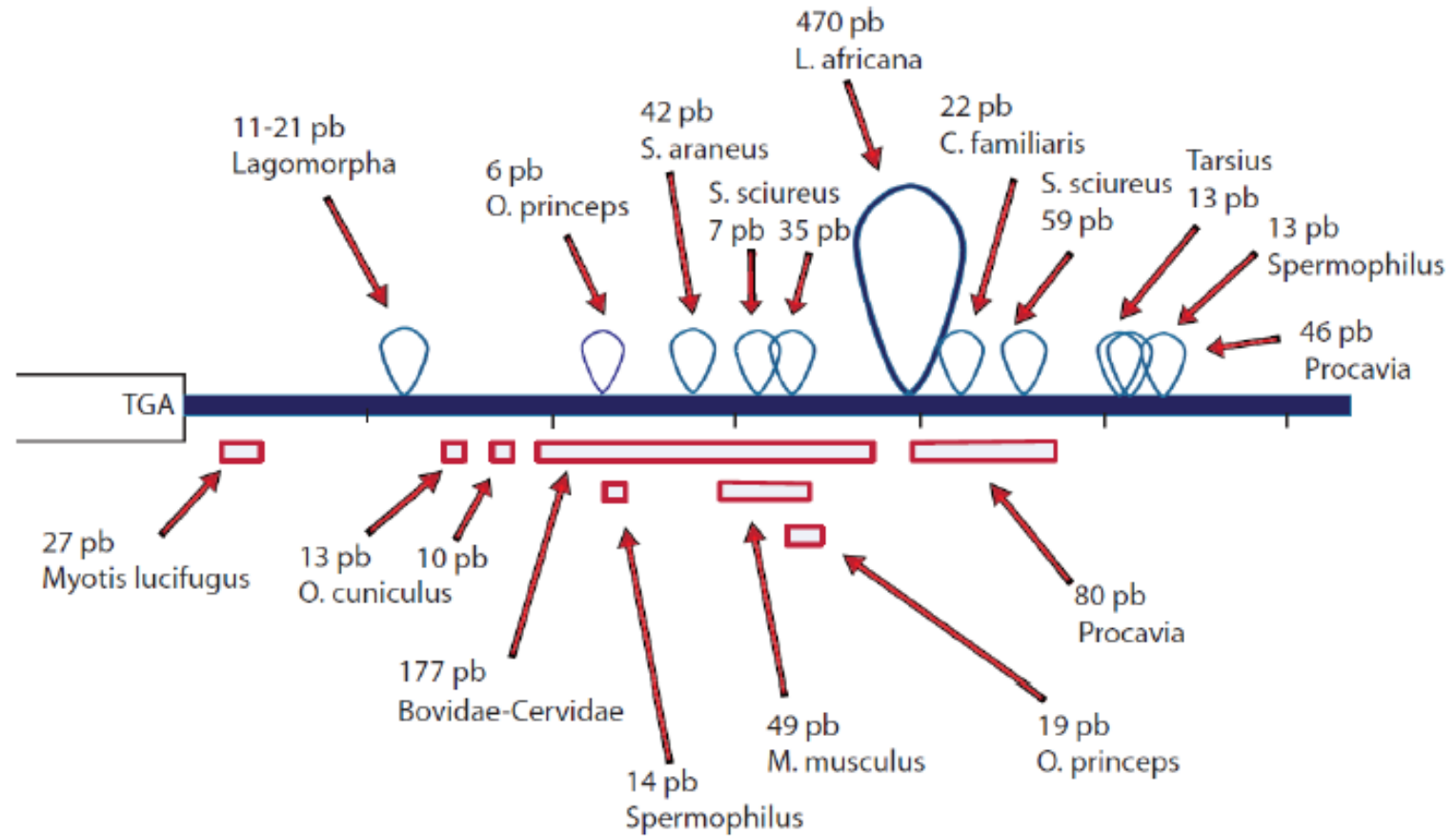


Cerf muet nord  
américain

Accumulation dans le cerveau d'une protéine prion anormalement conformée. Bovidés et Cervidés sont sensibles à la maladie, mais pas le porc, la souris ou le lapin. Pour rendre une souris sensible, il faut la modifier génétiquement pour qu'elle surexprime la protéine prion de mouton.

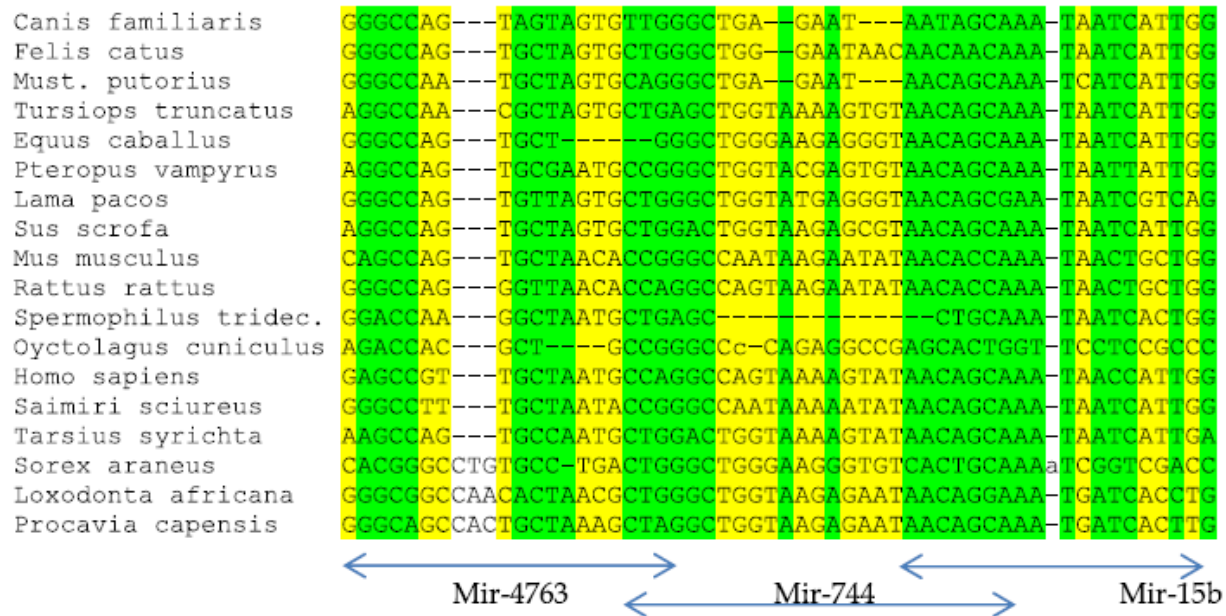


Hypothèse de base : il y a des mutations dans la partie 3' UTR du gène codant pour la protéine prion qui provoquent une surexpression du gène.



1

Alignement des séquences chez 22 mammifères euthériens



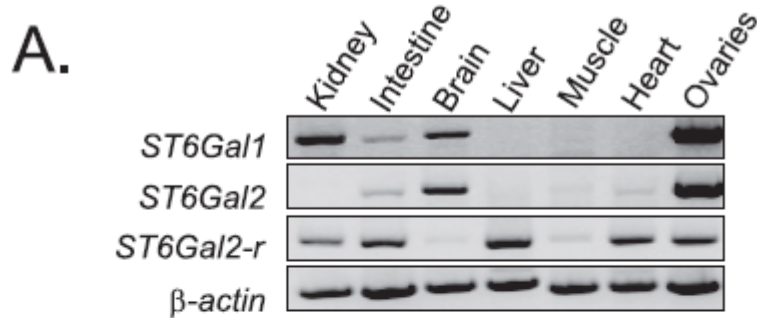
Alignement des séquences de mammifères non Bovidae-Cervidae dans la région correspondant à la grande délétion

Mir name	<i>Canis familiaris</i>	<i>Equus caballus</i>	<i>Homo sapiens</i>	<i>Felis catus</i>	<i>Lama pacos</i>	<i>Mustela putorius</i>	<i>Myodes glareolus</i>	<i>Sus scrofa</i>
4763		X		X	X			X
1587		X		X				
744				X		X	X	
15B/369		X	X	X				X
3613	X	X		X				
603					X			X
4723	X			X				

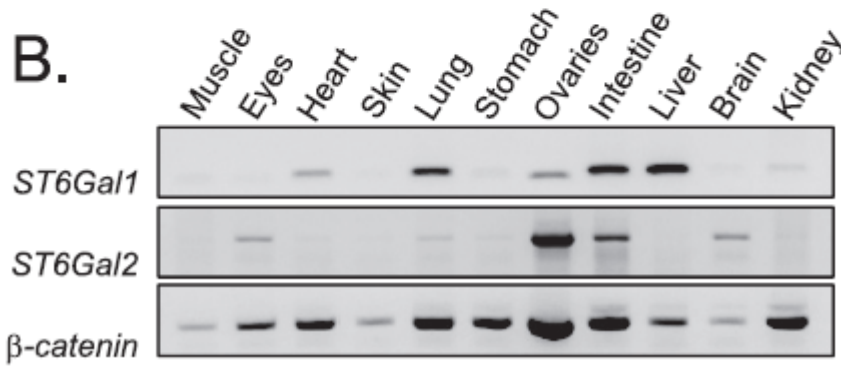
Table 3. Micro RNAs targets common to several species in the large deletion region

# Evolution de l'expression des ST6gal chez les Vertébrés

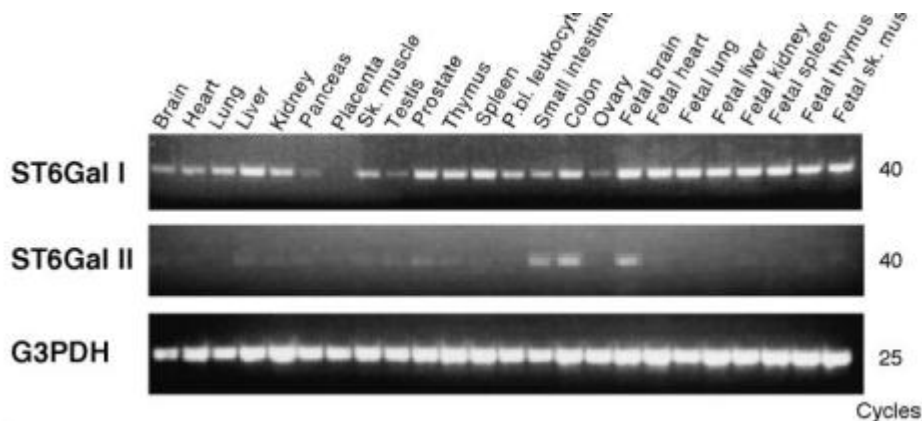
## RT-PCR



A, fish *st6gal* genes.



B, amphibian *st6gal* genes



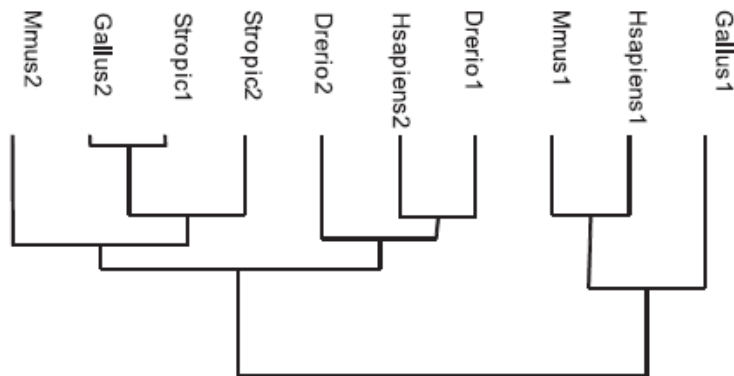
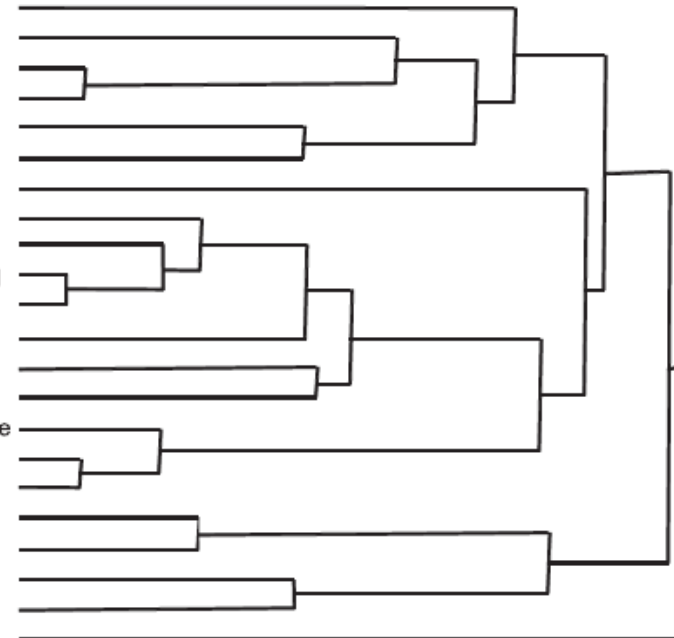
**C, Humain**

Takashima et al., JBC 2002

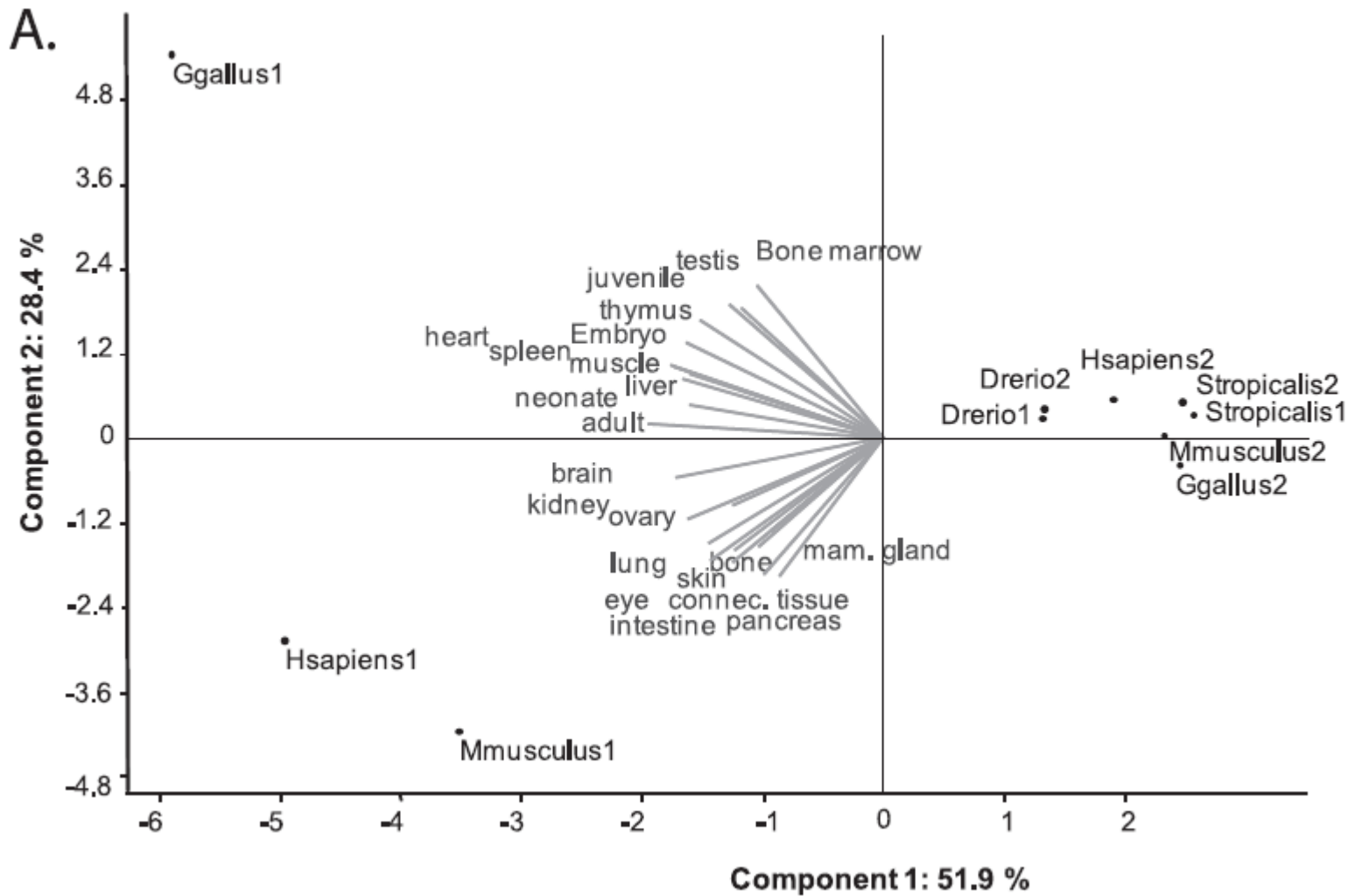


# Exploitation des bases de données d'EST dans NCBI

0	0	0	3,201	?	3,695	?	2,886	2,95	4,475	testis
0	0	0	0	0	3,409	3,671	4,303	4,77	3,806	ovary
0	0	0	0	?	2,589	?	2,859	4,601	5,658	thymus
0	?	0	0	3,681	2,343	3,012	3,226	4,815	?	kidney
3,902	0	0	0	3,53	3,43	3,139	4,541	4,356	4,338	brain
2,279	0	0	1,664	2,728	2,373	3,44	4	4,198	4,423	adult
0	0	0	0	0	0	0	0	3,533	4,149	heart
0	?	0	0	0	0	0	5,082	4,78	?	skin
0	0	0	0	?	0	?	3,763	3,638	4,616	spleen
0	?	?	?	?	0	?	4,658	3,685	?	mammary gland
0	?	0	0	?	0	?	4,799	3,937	?	lung
0	0	0	0	0	0	0	2,95	5,185	5,51	liver
0	0	0	0	2,856	0	0	3,621	2,327	4,117	muscle
0	0	?	?	?	0	?	2,729	3,062	6,024	bone marrow
0	0	?	?	?	2,039	?	3,934	4,311	0	connective tissue
0	0	0	0	0	2,25	0	3,19	3,865	0	intestine
0	?	0	0	0	2,703	0	3,425	3,756	?	bone
3,213	0	?	?	0	0	0	3,65	4,004	5,991	juvenile
3,864	0	?	?	0	0	0	3,864	4,577	4,709	neonate
2,427	2,816	2,207	0	?	3,51	?	?	3,857	4,156	embryo
3,318	?	?	?	0	2,347	0	4,771	3,97	?	eye
3,372	0	0	0	?	0	?	3,869	3,512	0	pancreas

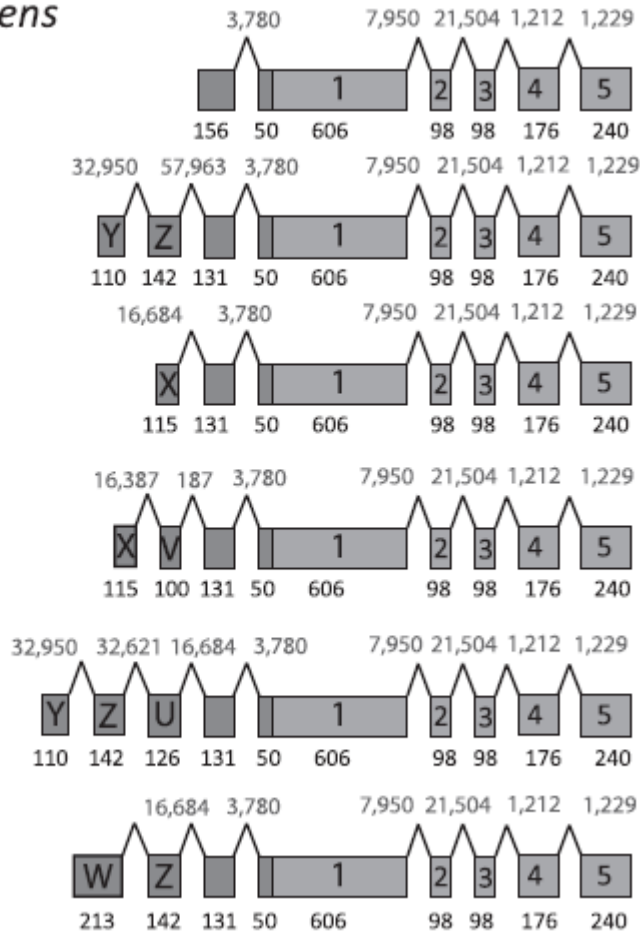


# Analyse en Composantes principales

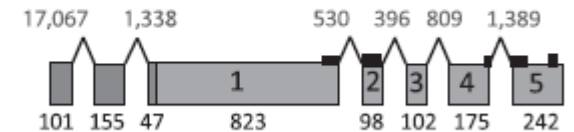


# Analyse des transcrits de ST6gal 1

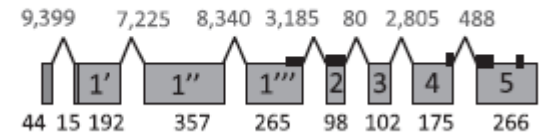
## *Homo sapiens*



## *Silurana tropicalis*



## *Danio rerio*



## Interprétation

Délétion dans la région 3'UTR chez l'ancêtre de Bovidae-Cervidae

Perte de sites de fixation de microARN

Plus grande stabilité de l'ARNm de *PrP*

Plus forte expression de PrP

Plus grande sensibilité aux  
encéphalopathies spongiformes

# Evolution des gènes par duplication

## BMC Evolutionary Biology

BioMed Central

Research article

Open Access

### Evolutionary history of the alpha2,8-sialyltransferase (ST8Sia) gene family: Tandem duplications in early deuterostomes explain most of the diversity found in the vertebrate ST8Sia genes

Anne Harduin-Lepers<sup>\*1</sup>, Daniel Petit<sup>2</sup>, Rosella Mollicone<sup>3</sup>,  
Philippe Delannoy<sup>1</sup>, Jean-Michel Petit<sup>2</sup> and Rafael Oriol<sup>3</sup>

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 285, NO. 49, pp. 38399–38414, December 3, 2010  
© 2010 by The American Society for Biochemistry and Molecular Biology, Inc. Printed in the U.S.A.

### Molecular Phylogeny and Functional Genomics of $\beta$ -Galactoside $\alpha$ 2,6-Sialyltransferases That Explain Ubiquitous Expression of *st6gal1* Gene in Amniotes<sup>\*[5]</sup>

Received for publication, July 15, 2010, and in revised form, September 13, 2010 Published, JBC Papers in Press, September 20, 2010, DOI 10.1074/jbc.M110.163931

Daniel Petit<sup>†</sup>, Anne-Marie Mir<sup>§</sup>, Jean-Michel Petit<sup>†</sup>, Christine Thisse<sup>§</sup>, Philippe Delannoy<sup>§</sup>, Rafael Oriol<sup>||</sup>,  
Bernard Thisse<sup>§</sup>, and Anne Harduin-Lepers<sup>§1</sup>

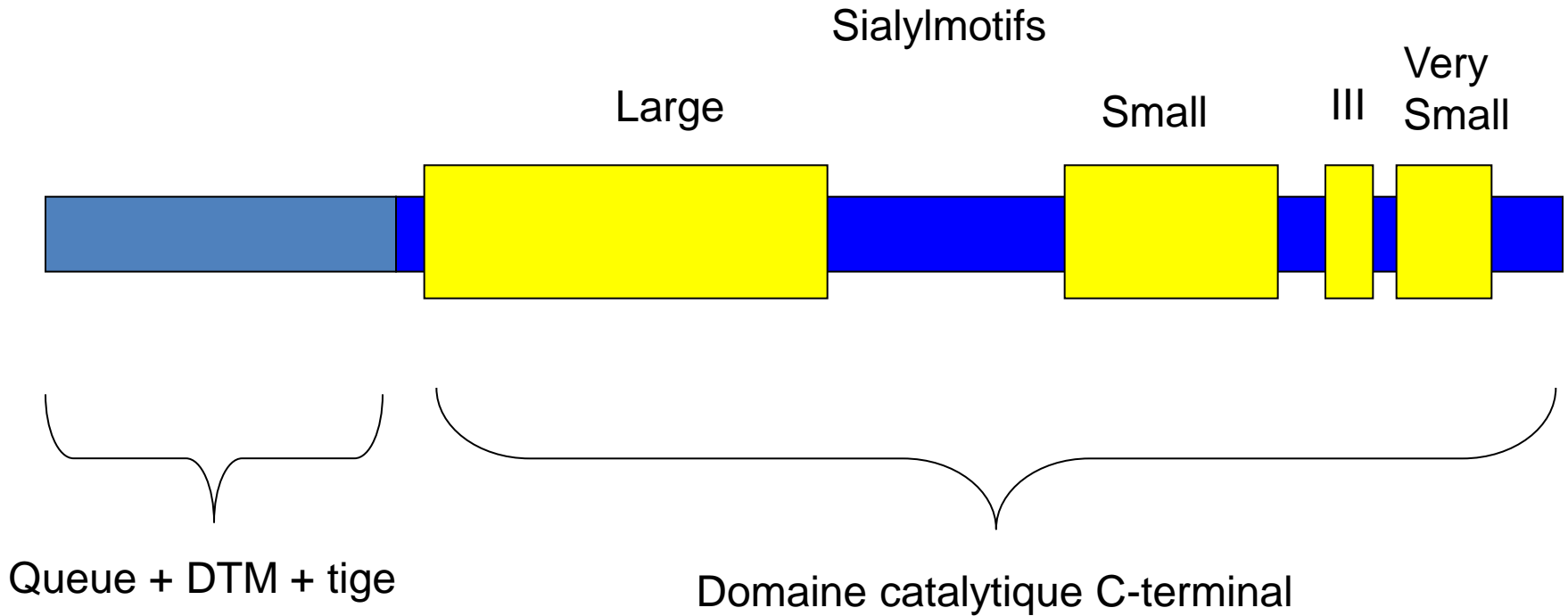
### En Bloc Duplications, Mutation Rates, and Densities of Amino Acid Changes Clarify the Evolution of Vertebrate $\alpha$ -1,3/4-Fucosyltransferases

Daniel Petit, Abderrahman Maftah, Raymond Julien, Jean-Michel Petit

JOURNAL OF MOLECULAR EVOLUTION

Datation et mécanismes des duplications  
Evolution des fonctions

# Sialyltransférases : Glycoprotéine transmembranaire de type II



4 superfamilles

- ST3Gal
- ST6Gal
- ST6GalNAc
- **ST8Sia**

# Alignement au format PHYLIP Interleaved

```
7 200
Hsapiens RRCVVVNSG NLRGSGYGQD VDGHNFMIR- MNQAPTGVGE QDVGSRTTHH
Ggallus RRCVVVNSG NLRGSGYGHE IDGHDFIMR- MNQAPTGVGE GDVGSRTTHH
Drerio LRCVVVNSG NLRGAGYGFV IDGHDFIMR- MNLAPTGVGE EDAGSRTTHH
Btaurus RRCVVVNSG NLRGSGYGPD VDGHNFMIR- MNQAPTGVGE QDVGSRTTHH
Tguttata RRCVVVNSG NLRGSGYGFV IDGHDFIMR- MNQAPTGVGE GDVGSRTTHH
Trubripes LRCVVVNSG NLRGAGYGAT IDGHNYIMR- INLAPTGVGE EDAGSRTTHH
Xlaevis RRCVVVNSG NLRGSSYGKD IDGHEFIMR- INQAPTGVGE MDVGSRTTHH

FMYPESAKNL PANVSFVLVP FKVLDLLWIA SALSTGQI-R FTYAPVKSPFL
FMYPESAKNL PANVSFVLVP FKTLDLLWIA SALSTGQI-R FTYAPVKPFL
FMYPESAKNL AANVSFVLVP FKTLDLLWIT SALSTGQI-R FTYAPVKQFL
FMYPESAKNL PANVSFVLVP FKALDLLWIA SALSTGQI-R FTYAPVKSPFL
FMYPESAKNL PANVSFVLVP FKTLDLLWIA SALSTGQI-R FTYAPVKPFL
FMYPESAKNL AANVSFVLVP FKTLDLVWIT SALSTGQI-R LTYAPVKQFL
FMYPESAKNL PANVSFVLVP FKALDLLWIT SALSTGQI-R FTYAPVKAPL

RVDKKVKQIY NPAFFKYIHD RWTERRHGRYP STGMLVLFFA LHCDEVNVVY
RVDKKVKQIY NPAFFKYIHD RWTERRHGRYP STGMLVLFFA LHCDEVNVVF
RVDKKVKQIF NPAFFKYIHD RWTERRHGRYP STGMLVLFFA LHCDEVNVVF
RVDKKVKQIY NPAFFKYIHD RWTERRHGRYP STGMLVLFFA LHCDEVNVVY
RVDKKVKQIY NPAFFKYIHD RWTERRHGRYP STGMLVLFFA LHCDEVNVVF
RVDKKVKQIY NPAFFKYVHD RWTERRHGRYP STGMLVLFFA LHCDEVNVVF
RVDKKVKQIY NPAFFKYIHD RWTERRHGRYP STGMLVLFFA LHCDEVNVVF

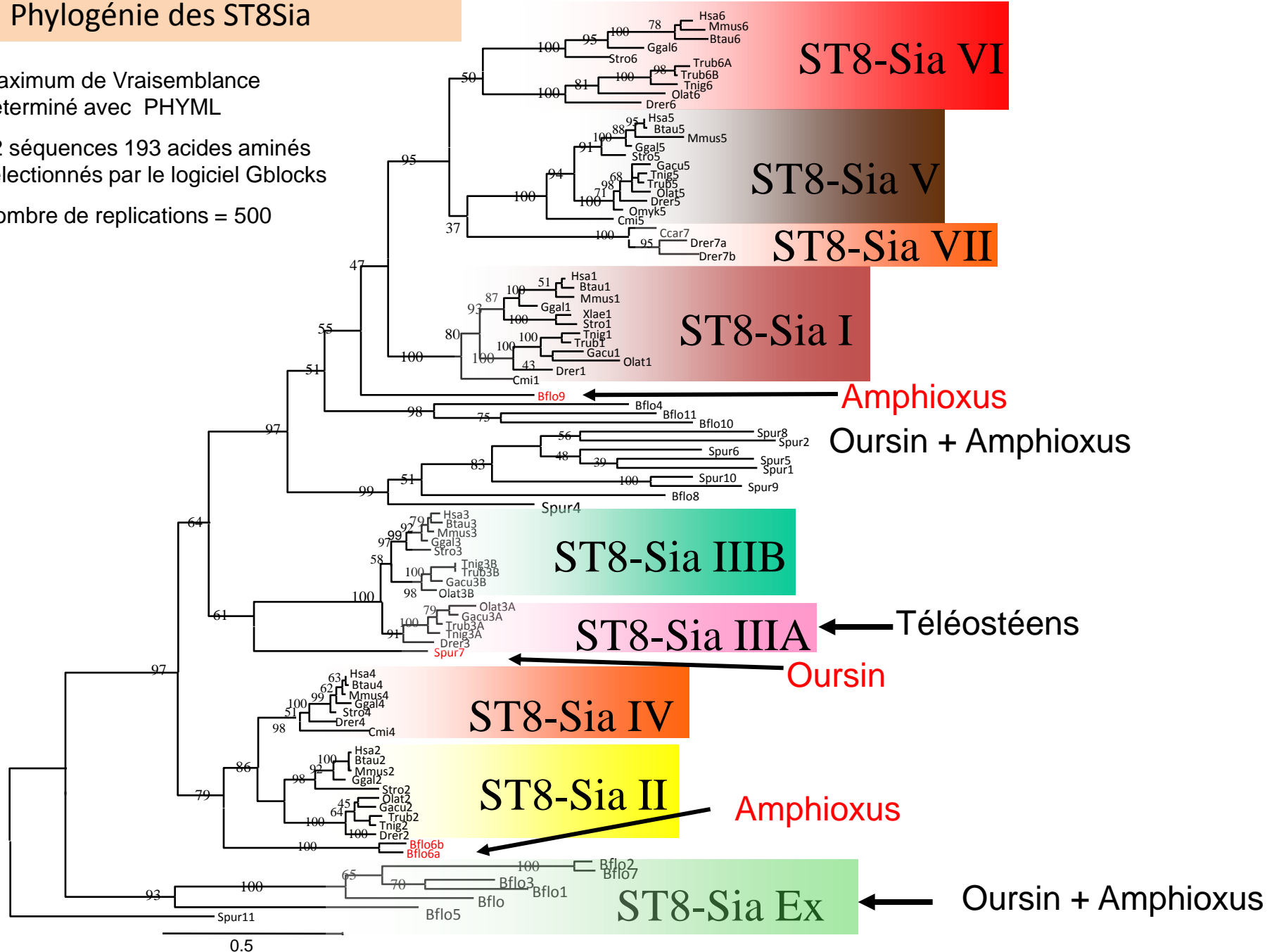
GFGADSRGNW HHYWEENRYA GEFKRTGVHD ADPEAHIIDM LAKAKSIEVY
GFGADSRGNW HHYWEENRYA GEFKRTGVHD ADPEAHIIDM LAKTSRIEGV
GFGADSRGNW HHYWEQNRYS GEFKRTGVHD ADYEAQIIDK LAKAGKISVF
GFGA
GFGA
GFGA
GFGA
```

# Phylogénie des ST8Sia

Maximum de Vraisemblance  
déterminé avec PHYML

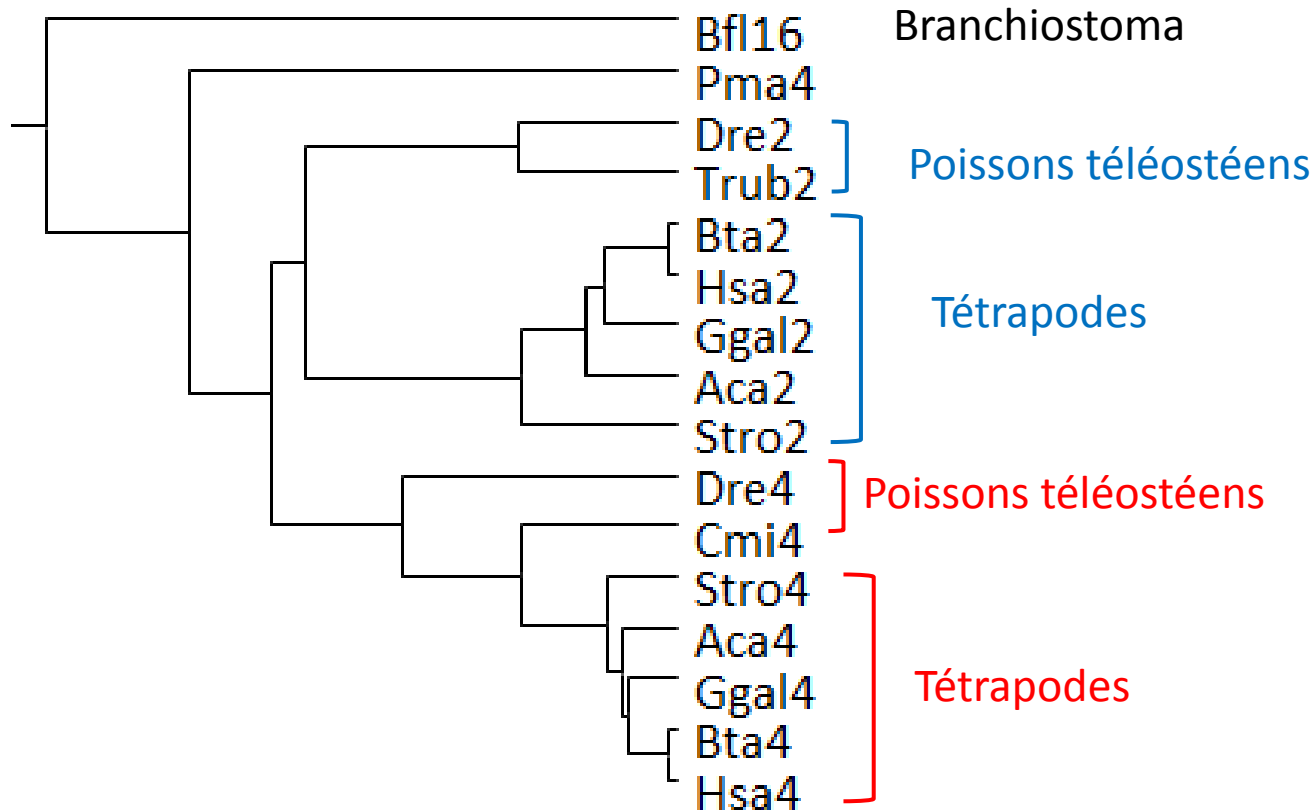
92 séquences 193 acides aminés  
sélectionnés par le logiciel Gblocks

Nombre de replications = 500



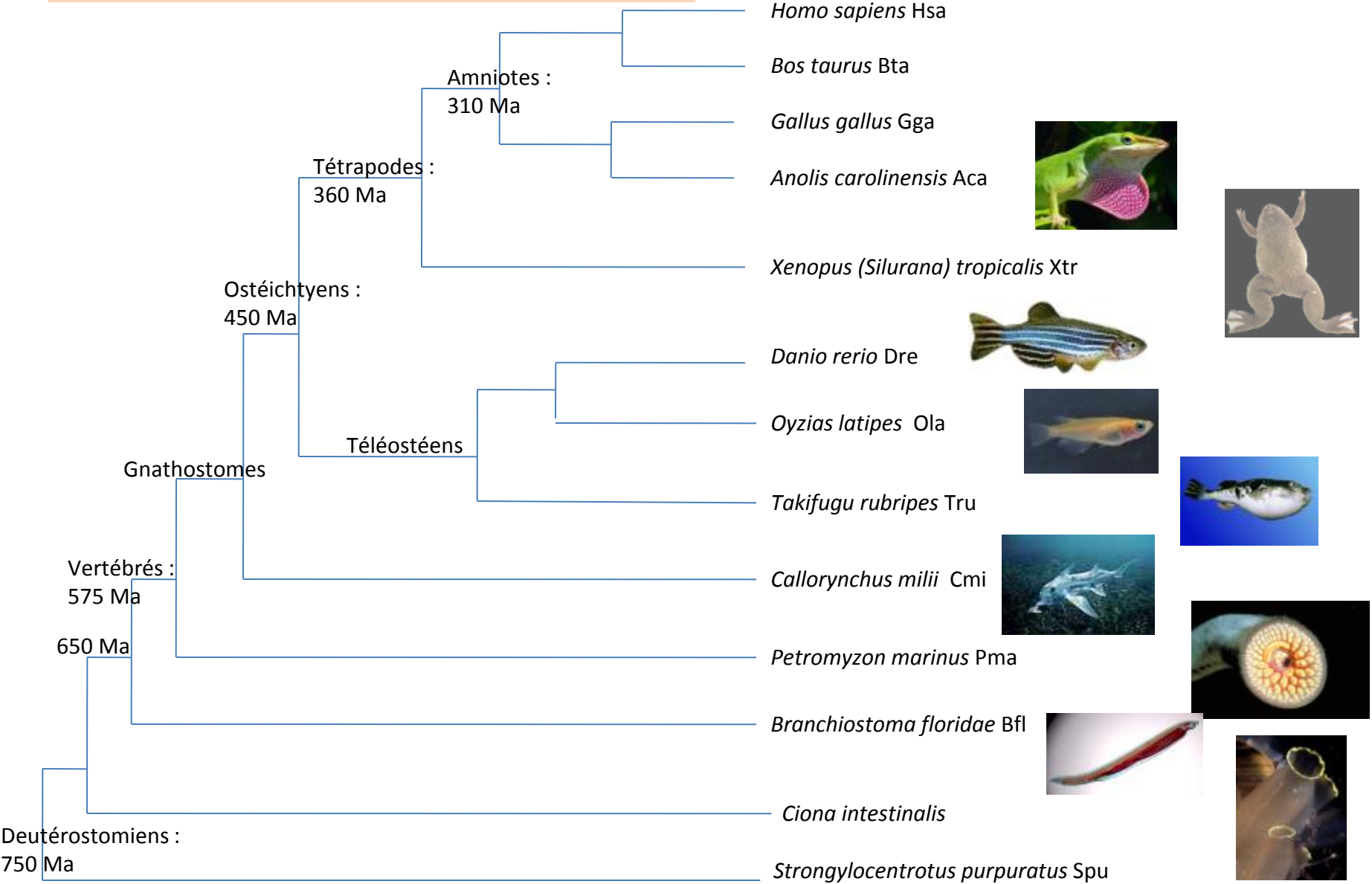


## Zoom sur les STII et STIV

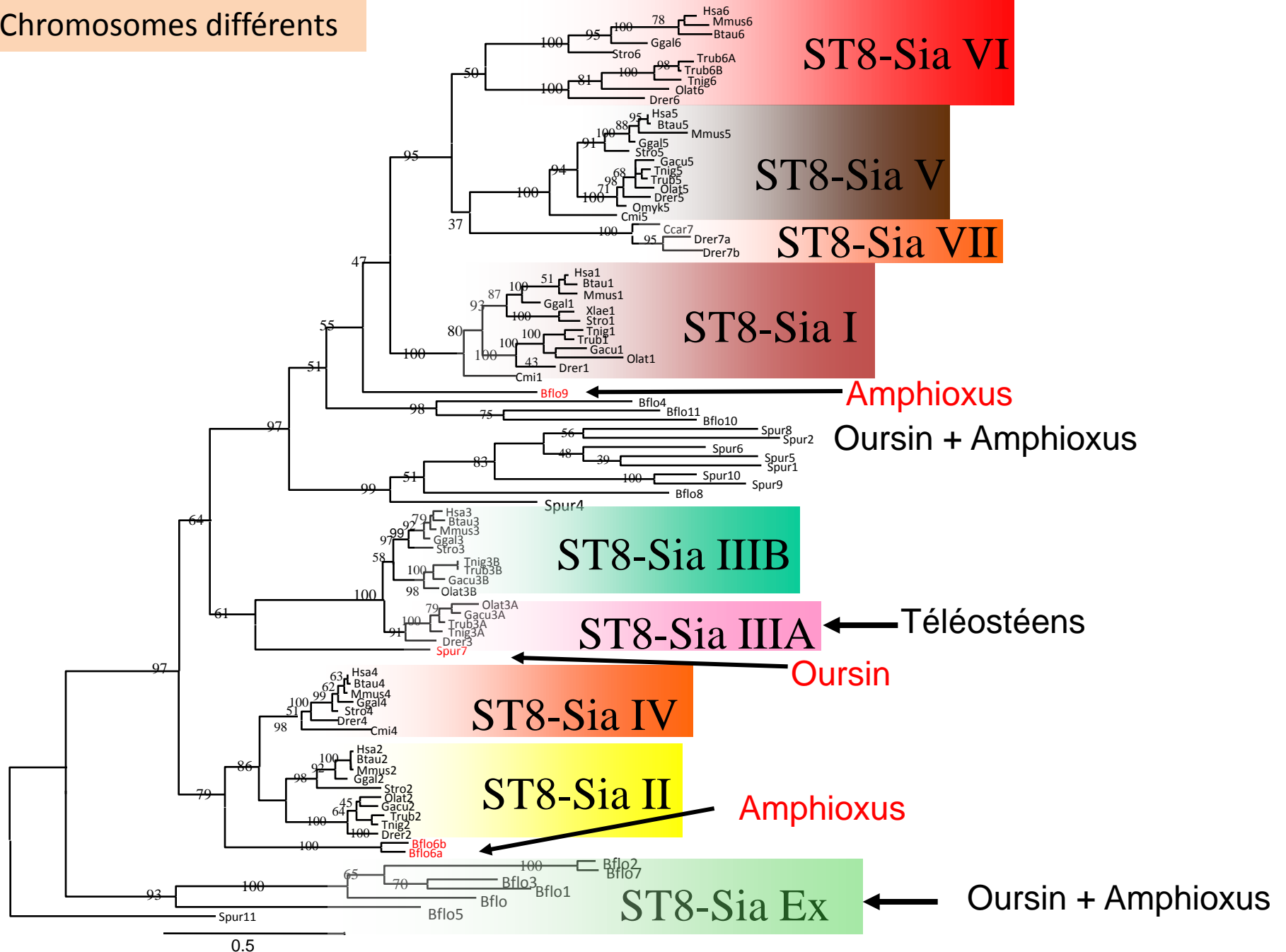


La duplication qui a engendré les STII et STIV s'est produite entre Branchiostoma et les premiers vertébrés à mâchoire (Gnathostomes)

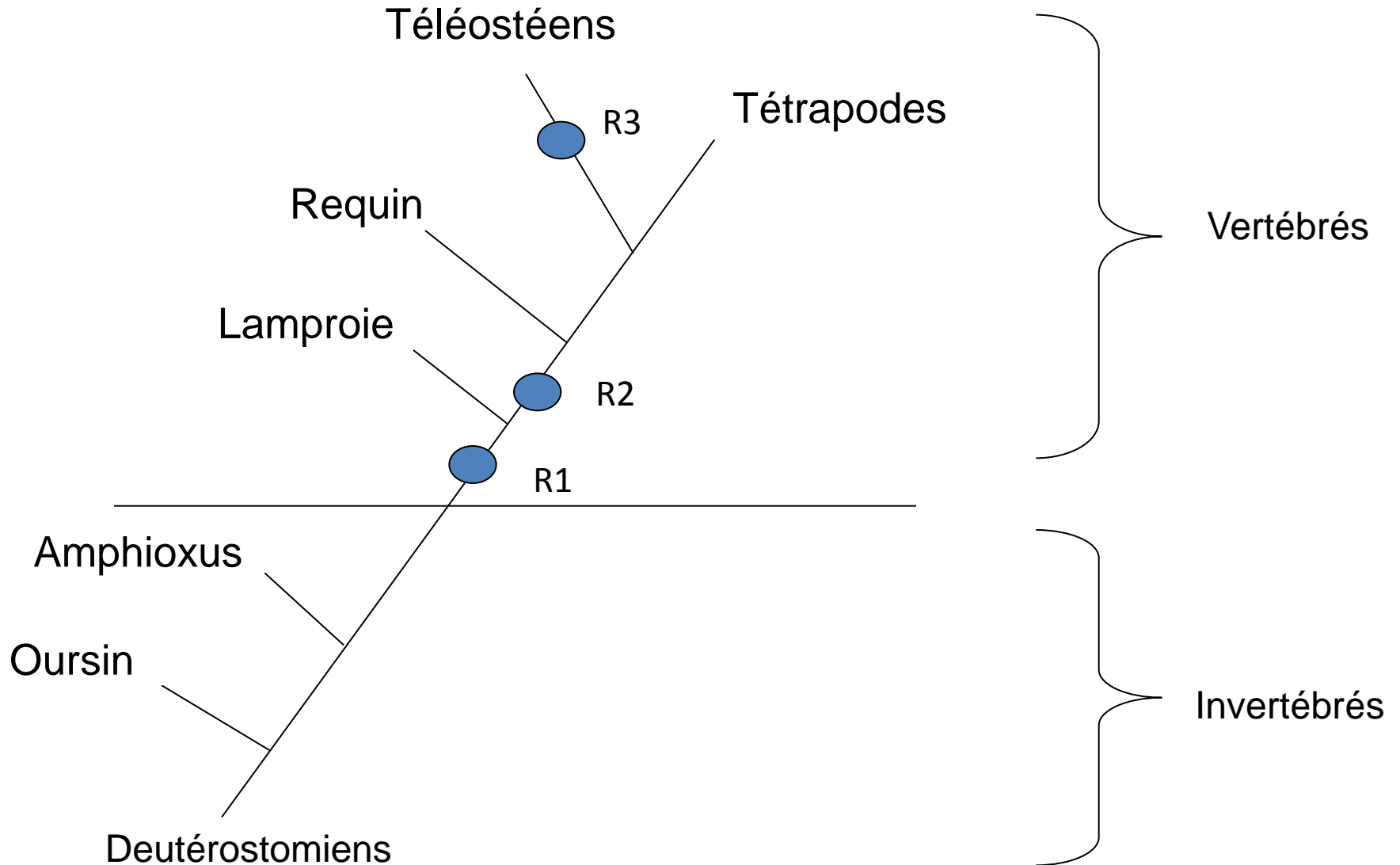
# Phylogénie des Vertébrés



Chromosomes différents

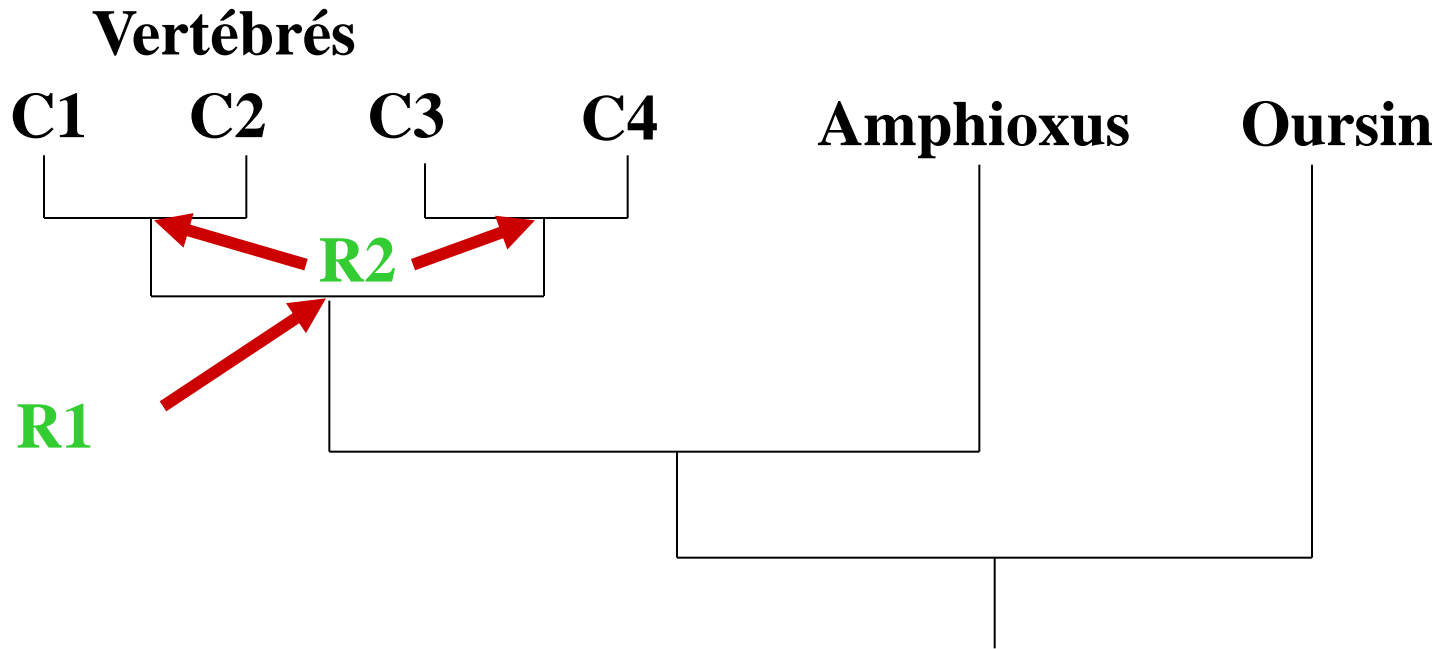


# Rappel sur l'hypothèse 2R : 2 rounds de doublements de génome chez les premiers Vertébrés



# Comment reconnaître si des gènes dupliqués chez Vertébrés sont issus de R1+R2 ?

## Critère 1 : topologie symétrique

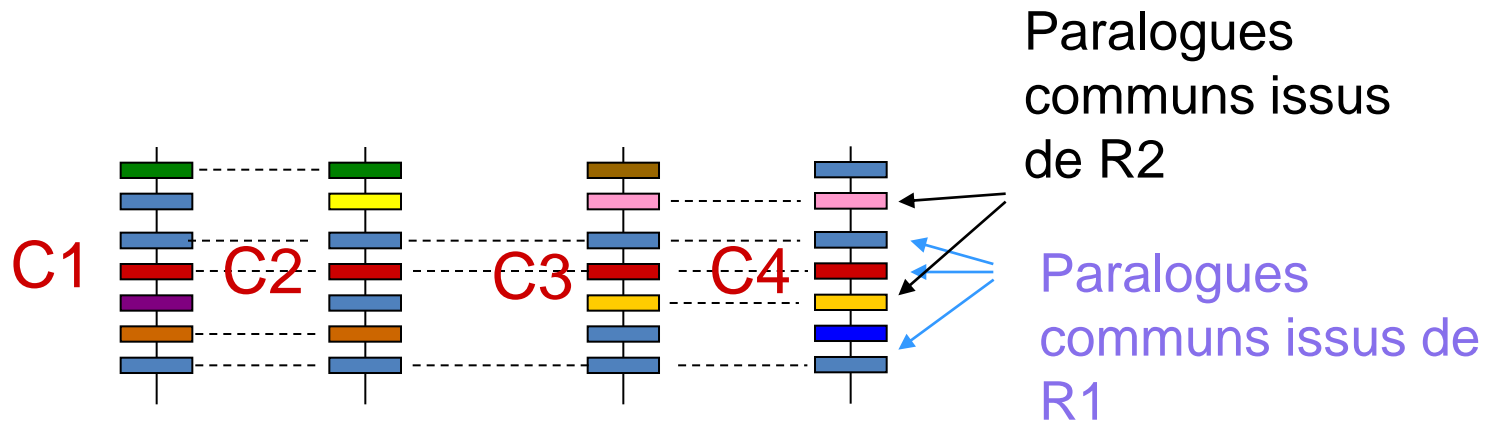


Gène ancestral en simple copie

## Critère 2 : reconnaissance de blocs de gènes (paralogons)

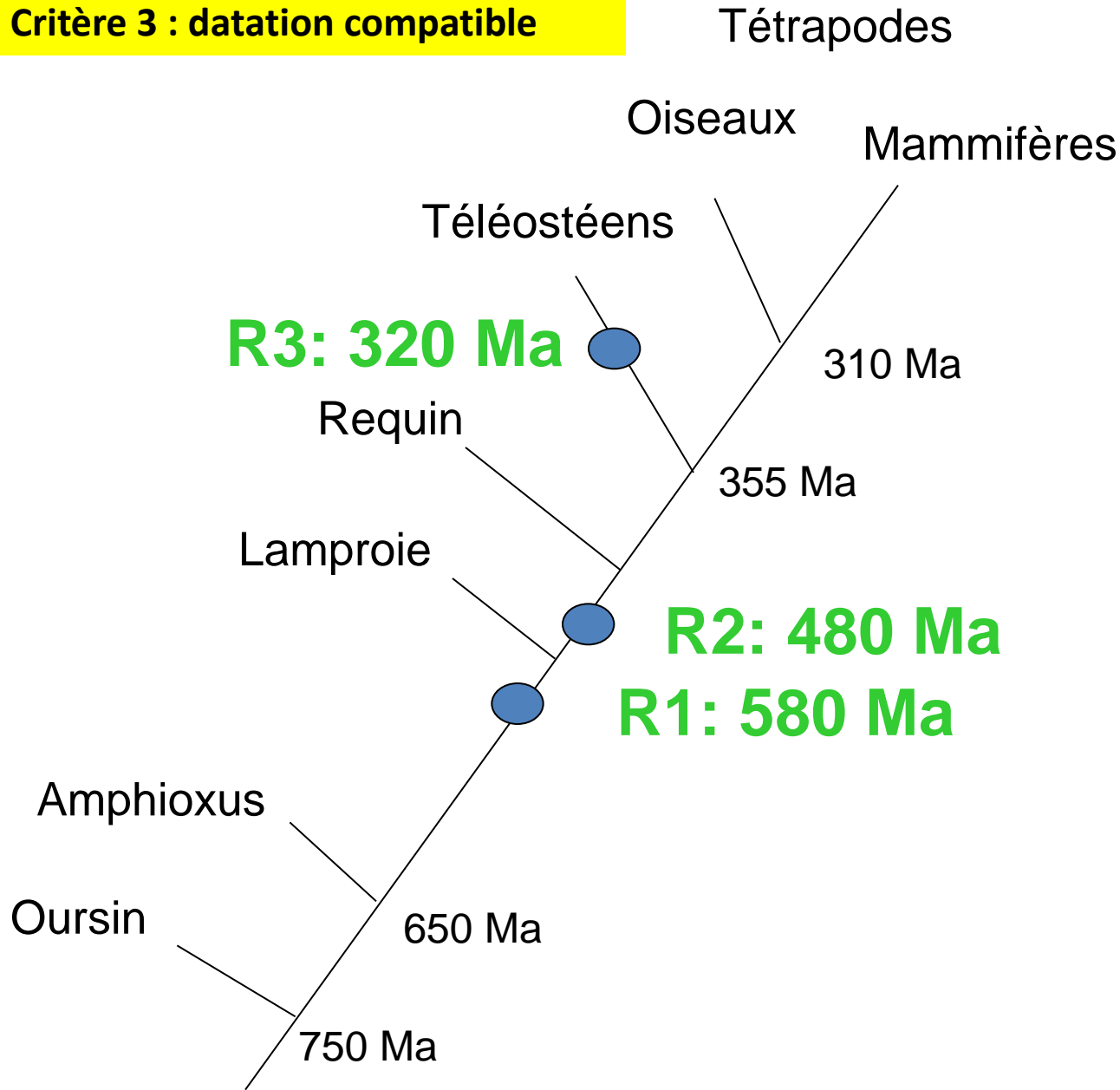
Blocs de gènes orthologues conservés : synténie

Blocs de gènes paralogues conservés : paralogon



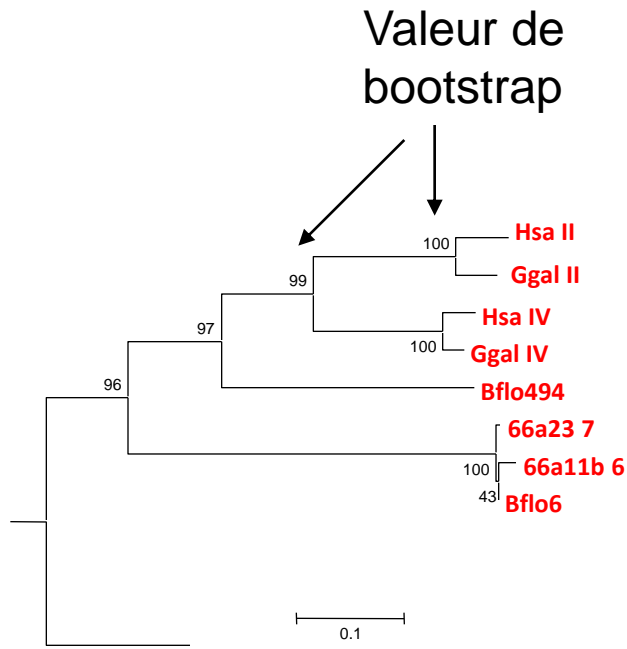
Quatre paralogons issus de R1 + R2, situés sur des chromosomes différents

**Critère 3 : datation compatible**

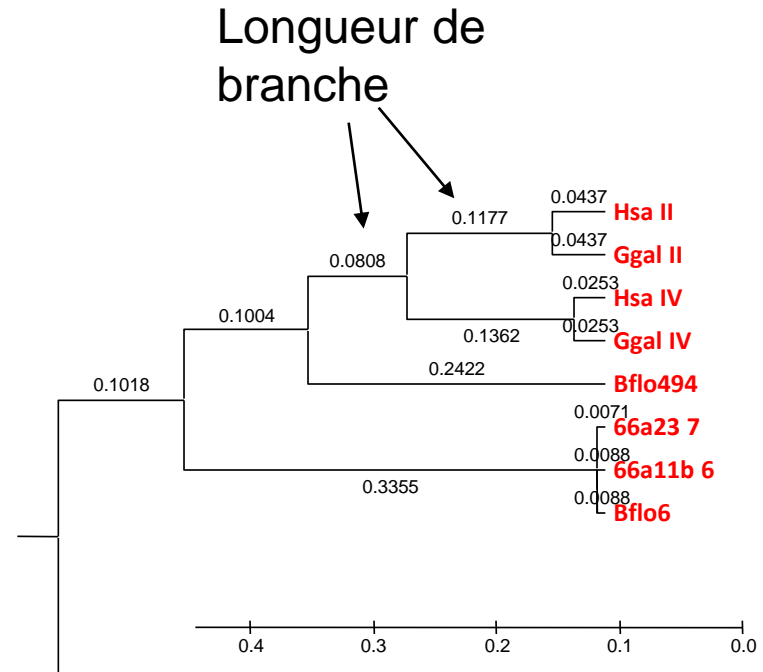


# Calculs de datations : méthode des arbres linéarisés, appliquée à chaque famille

## Exemple : famille II-IV avec MEGA3.1



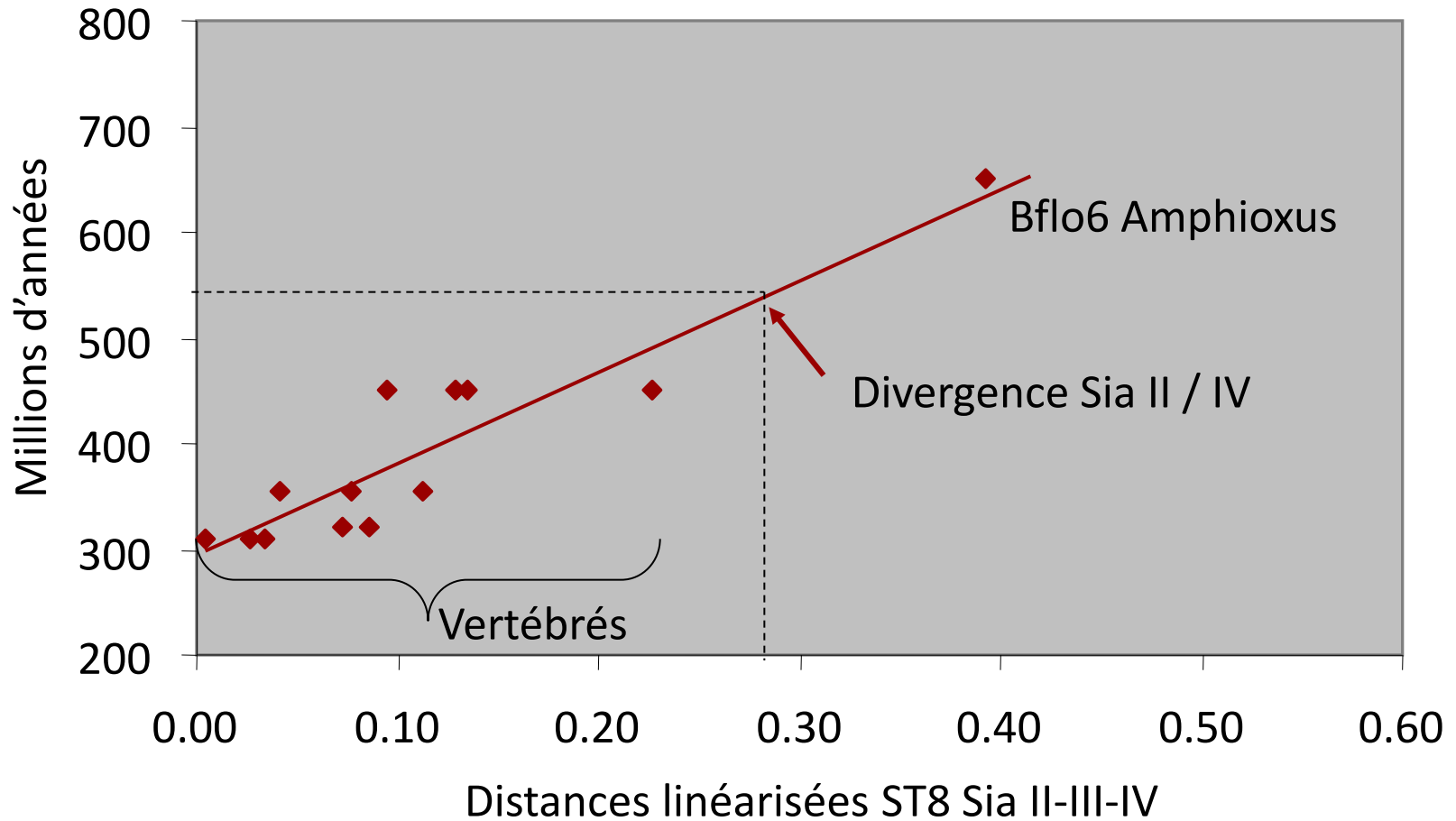
Longueurs de branches proportionnelles aux distances



Longueurs de branches recalculées pour que les taxons terminaux soient alignés



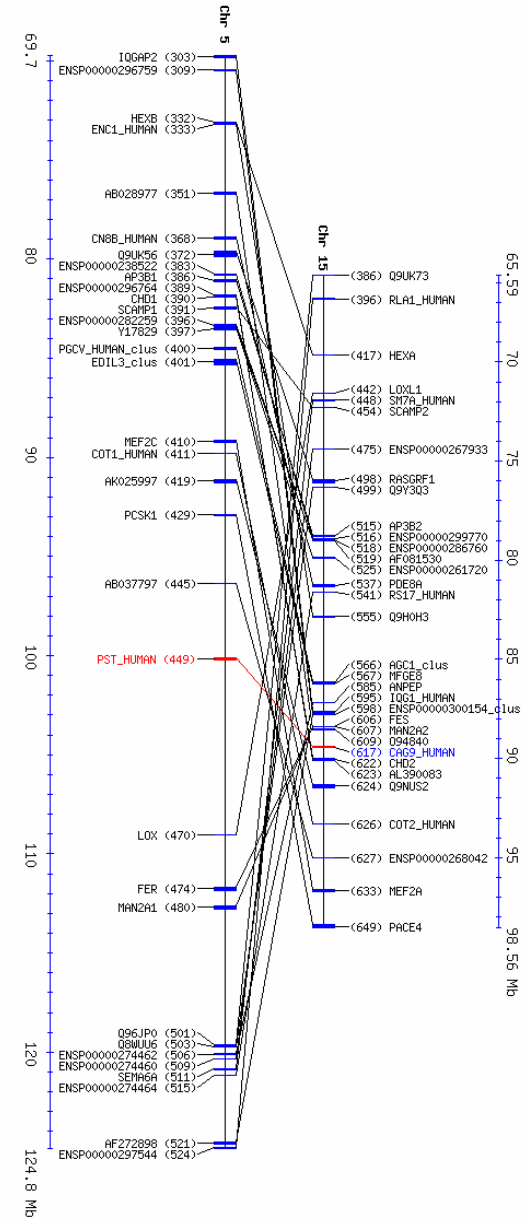
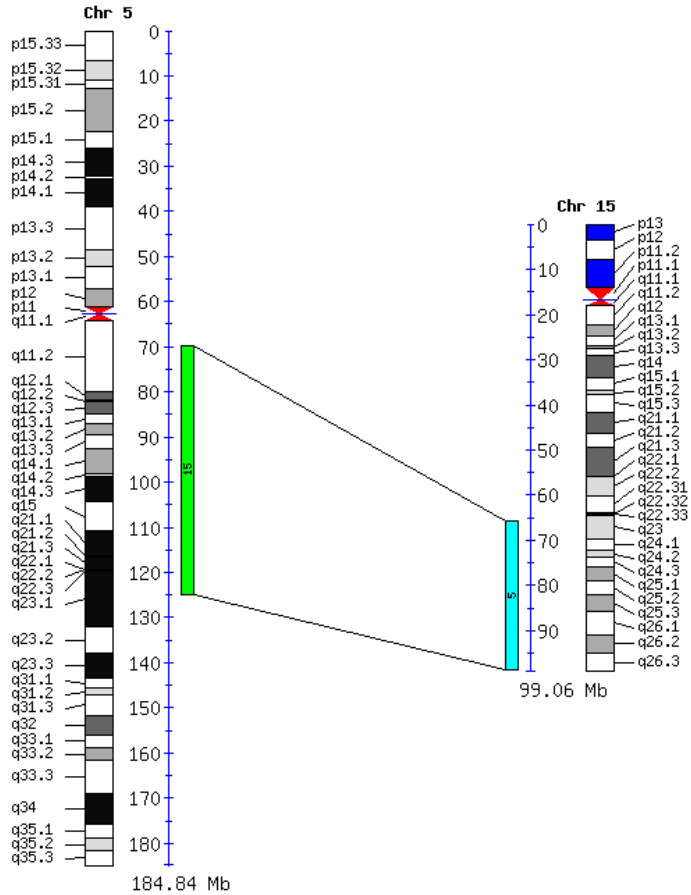
## Hypothèse : horloge moléculaire



La duplication *ST8 Sia II / ST8 Sia IV* remonterait à 544 Ma, d'où une attribution à R1

# Etude des localisations chromosomiques et paralogs

## ST8Sia II and ST8Sia IV du bloc de paralogon

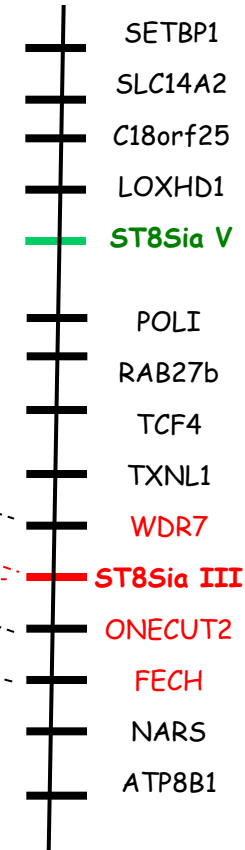


# Recherche de blocs synténiques chez Amphioxus

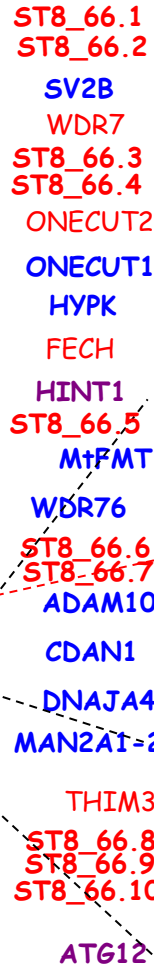
*Hsa* 18q21

*Bflo* Scaff 66

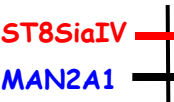
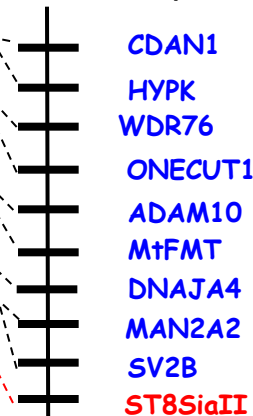
Groupe VIII absent  
chez Vertébrés



*Hsa* 5q

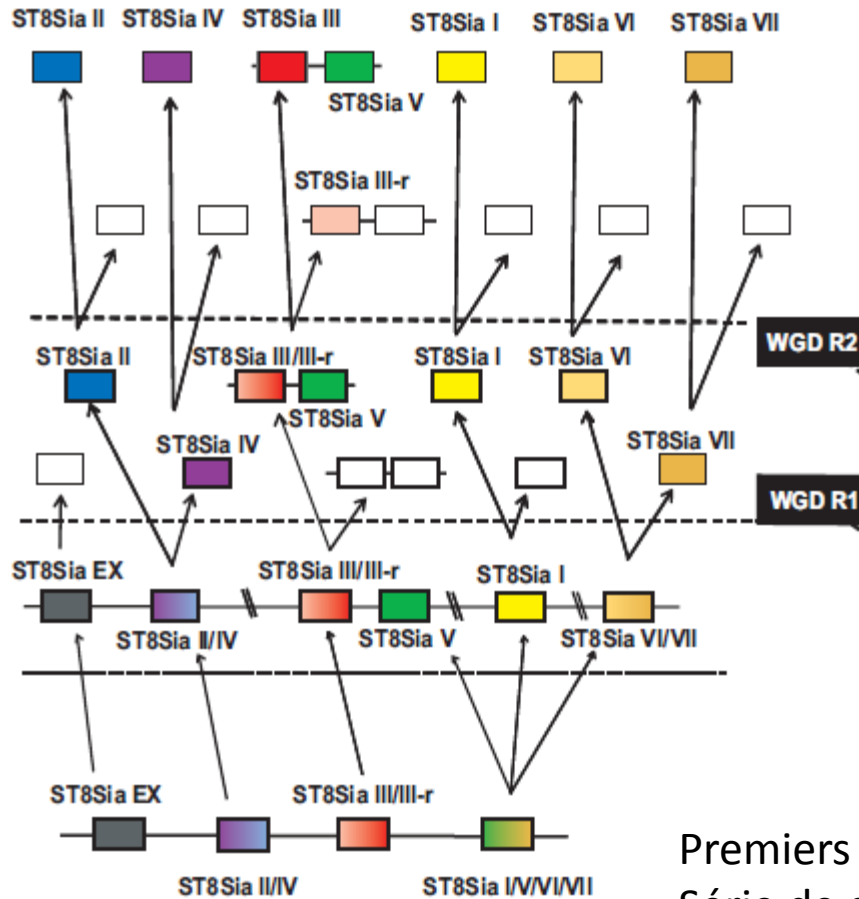


*Hsa* 15q



Un seul bloc chez l'Amphioxus  
contient les synténies de 3  
chromosomes humains

# Interprétation des blocs de synténie



Vertébrés  
gnathostomes

Premiers deutérostomiens  
Série de duplications en tandem

# Changement des fonctions

## Cas des Polyialyltransférases ST8Sia II et Sia IV

### ST8Sia-II

```

-----PSTD-----
AM422133bta2  ..KH--HVNVRTAYPSLRLHHA VRGYWLTNKVHIKR
U33551Hsa  2  ..KH--HVNVRTAYPSLRLHHA VRGYWLTNKVHIKR
AY055462dre2 ..LH--TVNVRTAFPSLRLHHA VRGYWLTNHVQIKR
    
```

### ST8Sia-IV

```

AJ699421bta4 ..KN--KLKVRTAYPSLR LIHAVRGYWLTNKVPIKR
L41680Hsa  4  ..KN--KLKVRTAYPSLR LIHAVRGYWLTNKVPIKR
AJ715545dre4 ..KH--KLKVRTAYPSLR LIHAVRGFWLTNKINIKR
Bfl6 II IV   ..KH--NINVRTAYPS PDMEKWIRRYWKS TEYRVAR
    
```

### ST8Sia-III

```

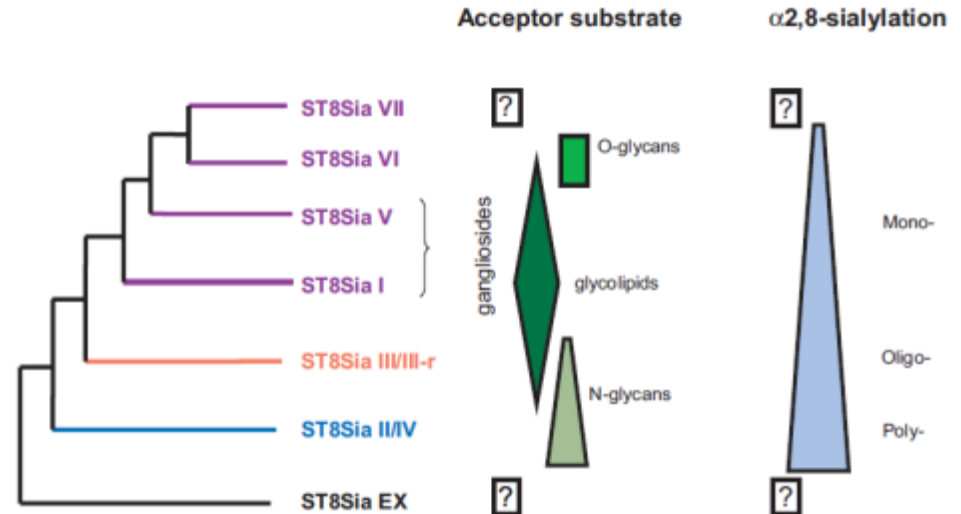
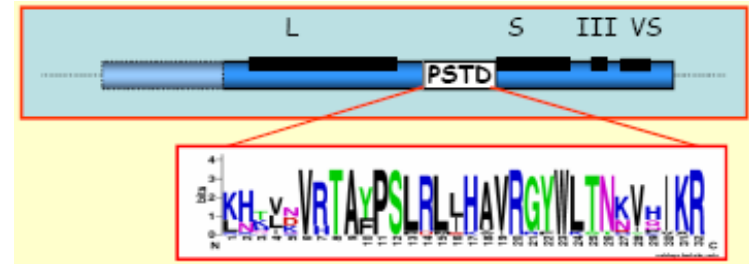
AF004668Hsa3 ..EHRGQLKVQLAWPGNIMQHVN R-YWKNKHLSPKR
AJ704563bta3 ..EHRGQLKLQLAWPGNIMQHVN R-YWKNKHLSPKR
    
```

### Other ST8Sia

```

SI8A_HUMAN 1  ..DV--GANQTVL FANPNFLRS IGKFWKSRGIHAKR
SI8E_HUMAN 5  ..DF--ESPQAVYY FHPQYLVNVSRYWLS LGVRAKR
SI8F_HUMAN 6  ..ES--KARQKVL FFHPKYLKDLALFWRTKGV TAYR
B.floridae 9  ..EL--KSDNEIIF ANPVFMGKINGYWKKKGV KAWR
B.floridae 0  ..VSFVLLTSASS FTSAEKINPIASKIARKRYMKKH
    
```

polysialylation domain (PSTD) 32 aa, Troy et al.



## Conclusions

L'exploration des génomes prend de plus en plus d'intérêt pour étudier :

- Les gènes et leur entourage dans des espèces voisines (séquences répétées, modifications des séquences régulatrices en 5' et en 3')
- Les duplications et pertes de gènes
- L'exon shuffling et la compréhension des protéines modulaires
- L'évolution des fonctions au travers des motifs conservés
- L'évolution de l'expression par l'étude des EST

## Recent publications

1. **PETIT, D.**, MAFTAH, A., JULIEN, R., & PETIT, J.M. 2006. - En bloc duplications, mutation rates and densities of amino acid changes clarify the evolution of vertebrate  $\alpha$ 1,3/4 fucosyltransferases. *Journ. Molecular Evolution* 63(3): 353–364
2. LAUTRETTE, C., LOUM-RIBOT, E., **PETIT, D.**, VERMOT-DESROCHES, C., WIJDENES J. & JAUBERTEAU M.O., 2006. - Increase of fas-induced apoptosis by inhibition of extracellular phosphorylation of fas receptor in jurkat cell line. *Apoptosis* 11: 1195-1204
3. GIRAUD, S., BESSETTE, B. BODA, C., LALLOUE, F., **PETIT, D.**, MATHONNET, M. & JAUBERTEAU, M.O. 2007. In vitro apoptotic induction of human glioblastoma cells by Fas Ligand plus etoposide and in vivo antitumour activity of combined drugs in xenografted nude rats. *Int J Oncol.* 30(1): 273-81.
4. MEINIEL, O., MEINIEL, R., LALLOUE, F., DIDIER, R., JAUBERTEAU, M.-O., MEINIEL, A. & **PETIT, D.** 2008. The lengthening of a giant protein: When, How and Why? *Journ. Molecular Evolution*, 66(1): 1-10.
5. HARDUIN-LEPERS, A., **PETIT, D.**, MOLLICONE, R., DELANNOY, P., PETIT, J.-M., & ORIOL, R. 2008. Evolutionary history of the alpha2,8-sialyltransferase (ST8Sia) gene family: tandem duplications in early deuterostomes explain most of the diversity found in the vertebrate ST8Sia genes. *BMC Evolutionary Biology* 8: 258.
6. ERMONVAL, M., **PETIT, D.**, LE DUC, A., KELLERMANN, O., & GALLET, P.-F. 2009. Glycosylation-related genes are variably expressed depending on the differentiation state of a bioaminergic neuronal cell line: implication for the cellular prion protein. *Glycoconj. Journal* 26(4): 477-493.
7. MARTIN, R. GALLET, P.-F., ROCHA, D. & **PETIT, D.** 2009. Polymorphism of the Prion Protein in Mammals: A Phylogenetic Approach. *Recent Patents on DNA & Gene Sequences* 3(1): 1-9.
8. TROUTAUD, D., B. PETIT, C. BELLANGER, B. MARIN, M.-P. GOURIN-CHAURY, **D. PETIT**, A. OLIVRIE, J. FEUILLARD, M.-O. JAUBERTEAU, D. BORDESSOULE. 2009. Bad and AIF in diffuse large B cell lymphoma. Prognostic significance of Bad and AIF apoptotic pathways in diffuse large B-cell lymphoma. *Clinical Lymphoma, Myeloma & Leukemia*, 10(2) : 118-124.
9. FAUCHAIS A.L, OUATTARA B., GONDRAN G., LALLOUE F., **PETIT D.**, LY K., LAMBERT M., LAUNAY D., LOUSTAUD-RATTI V., BEZANAHARY H., LIOZON E., HACHULLA E., JAUBERTEAU M.O., VIDAL E., HATRON P.Y., 2010. Articular manifestations in primary Sjögren syndrome. Clinical Significance and prognosis of 188 patients. *Rheumatology (Oxford)*, 49(6): 1164-72
10. **PETIT, D.**, MIR, A.-M., PETIT, J.-M., THISSE, C., DELANNOY, P. ORIOL, R., THISSE, B. & HARDUIN-LEPERS, A., 2010. Molecular phylogeny and functional genomics of  $\beta$ -galactoside  $\alpha$ 2,6-sialyltransferases to explain ubiquitous expression of st6gal1 gene in Amniotes. *J. Biol. Chem.*, 285(49): 38399-38414.
11. PERRAUD, A., NOUAILLE, M., AKIL, H., **PETIT, D.**, LABROUSSE, F., JAUBERTEAU, M.O., & MATHONNET, M. 2011. Retinoid acid receptors in human colorectal cancer: An unexpected link with patient outcome. *Experimental and Therapeutic Medicine* 2: 491-497.
12. PERRAUD, A., AKIL, H., NOUAILLE, M., **PETIT, D.**, LABROUSSE, F., JAUBERTEAU, M.O., & MATHONNET, M. 2011. Expression of p53 and DR5 in normal and malignant tissues of colorectal cancer: Correlation with advanced stages. *Oncology reports* 26: 1091-1097.
13. CHERAY M., **PETIT, D.**, FORESTIER, L., KARAYAN-TAPON, L., MAFTAH, A., JAUBERTEAU, M.O., BATTU, S., GALLET, F.P., LALLOUÉ, F. 2011. Glycosylation-related gene expression is linked to differentiation status in glioblastomas undifferentiated cells. *Cancer Letters*. 312(1):24-32.

## Book chapter

14. **PETIT D.**, PETIT J.-M. & GALLET P.F. 2011. Analysis of 3'UTR of Prnp gene in Mammals: possible role of target sequences of miRNA for TSE sensitivity in Bovidae and Cervidae. In *Veterinary Medicine*", ISBN 979-953-307-413-8