



**HAL**  
open science

## A propos de la génomique fonctionnelle et du xénope

Nicolas Pollet

► **To cite this version:**

Nicolas Pollet. A propos de la génomique fonctionnelle et du xénope. Biochimie [q-bio.BM]. Université Paris Sud - Paris XI, 2005. tel-00159672

**HAL Id: tel-00159672**

**<https://theses.hal.science/tel-00159672>**

Submitted on 3 Jul 2007

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**Mémoire d'**  
**HABILITATION A DIRIGER DES RECHERCHES**  
**Université PARIS-XI**

*Nicolas POLLET*

**2 0 0 4**

---

***A Sophie, Anne et Lucie***

<b>Cartographie physique, génétique et génique de la région p12 du chromosome 20 humain contenant le locus du syndrome d'Alagille.</b>	<b>5</b>
Introduction : le syndrome d'Alagille	5
Caractérisation de la région 20p12	5
Cartographie physique	7
Cartographie génétique	9
Cartographie génique	10
Conclusions	12
<b>Analyse des profils d'expression génique au cours du développement embryonnaire du xénope.</b>	<b>14</b>
Introduction	14
L'hybridation in situ sur embryons entiers à grande échelle	16
Les groupes de synexpression	20
Axeldb: A Xenopus laevis database	21
Analyse in silico des profils d'expression	22
Etude de l'expression cyclique des gènes <i>esr9</i> et <i>esr10</i> dans le mésoderme présomitique du xénope.	26
Constitution d'un atlas de profils d'expression géniques pour l'étude de l'embryogenèse sur le modèle xénope.	27
<b>Génomique fonctionnelle du modèle <i>Xenopus tropicalis</i></b>	<b>32</b>
Introduction	32
Etude du transcriptome du système nerveux chez <i>Xenopus tropicalis</i>	36
<b>Xénopuces : Etude globale des transcriptomes au cours du développement chez le modèle xénope à l'aide de puces oligonucléotides (Thèse de Raphaël Thuret)</b>	<b>43</b>
Centre de Ressources Biologiques pour la cartographie génique et physique du génome de <i>Xenopus tropicalis</i>	48
Les séquences répétées du génome de <i>Xenopus tropicalis</i>	49
Caractérisation des microARN exprimés au cours de la métamorphose chez le xénope.	52
<b>Références</b>	<b>54</b>
<b>CURRICULUM VITAE</b>	<b>56</b>
<b>PRODUCTION SCIENTIFIQUE</b>	<b>57</b>
<b>Annexes</b>	<b>63</b>

## TRAVAUX ANTÉRIEURS

Mon expérience de recherche débuta en 1991 avec mon DEA puis ma thèse en 1996 sous la direction de Michèle Meunier-Rotival à l'unité 347 de l'INSERM " Génétique et mécanismes des maladies du foie de l'enfant " au Kremlin-Bicêtre. Mon travail a consisté à identifier par clonage positionnel le gène dont les dysfonctions sont responsables du syndrome d'Alagille.

J'ai ensuite effectué mon séjour post-doctoral de 1997 à 2000 à Heidelberg, au centre allemand de recherche sur le cancer (le DKFZ) au sein du département "Molecular Embryology" dirigé par Christof Niehrs. J'ai travaillé plus particulièrement sur l'analyse systématique des profils d'expression génique, l'étude de l'anatomie moléculaire de l'embryon de vertébré et l'identification de gènes impliqués dans le contrôle du développement embryonnaire.

J'ai été recruté au CNRS en 2000 et ai rejoint le laboratoire de Transgénèse et Génétique des Amphibiens avec André Mazabraud de l'UMR 8080 Développement et Evolution dirigée par Maurice Wegnez à Orsay. Depuis, mon attention se porte sur des études de génomique fonctionnelle appliquées au modèle *Xenopus tropicalis* visant à déchiffrer le programme transcriptionnel de la métamorphose.

Ce mémoire comporte trois volets correspondant à chacune de ces périodes et où je décris les recherches que j'ai pu mener à bien. Dans le troisième volet, je présente les perspectives et mes projets de recherche.

## Cartographie physique, génétique et génique de la région p12 du chromosome 20 humain contenant le locus du syndrome d'Alagille.

### ***Introduction : le syndrome d'Alagille***

Le syndrome d'Alagille (AGS, MIM 118450) est un syndrome malformatif complexe, non évolutif, caractérisé dans sa forme complète par l'association de cinq éléments majeurs : une cholestase intra-hépatique chronique liée à une paucité des canaux biliaires interlobulaires; une sténose périphérique des branches de l'artère pulmonaire; un défaut de fusion au niveau des arcs antérieurs d'une ou plusieurs vertèbres dorsales, réalisant l'aspect radiologique de „vertèbres en ailes de papillon“; un embryotoxon postérieur (malformation de la chambre antérieure de l'oeil); une dysmorphie faciale. Le diagnostic clinique de l'AGS peut être établi si au moins 3 de ces 5 éléments sont présents, on parle alors de syndrome d'Alagille partiel. Il n'y a pas d'associations ou de combinaisons préférentielles entre les symptômes.

L'incidence de l'AGS a été estimée entre 1/70 000 et 1/100 000 naissances. Le syndrome d'Alagille représente la deuxième cause de cholestase chronique chez l'enfant. Les observations de familles au sein desquelles le syndrome d'Alagille se transmet sur plusieurs générations ont permis d'établir une origine génétique. L'observation d'anomalies du bras court du chromosome 20 associées à ce syndrome a permis d'assigner le locus AGS au niveau des bandes chromosomiques 20p11.23-p12.2. Une analyse de ségrégation a mis en évidence un mode de transmission autosomique dominant avec pénétrance quasi-complète, et a permis d'estimer la proportion de cas sporadiques à 15 %. L'une des plus petites délétions du chromosome 20 associée à l'AGS a été observée au laboratoire chez un patient suivi à l'hôpital de Bicêtre (Zhang et coll., 1990).

### ***Caractérisation de la région 20p12***

Une méthode de sélection du chromosome 20 dans les cellules de hamster, fondée sur la sélection d'un gène de transport de la leucine à température semi-permissive, a permis d'obtenir au laboratoire un hybride somatique homme-hamster, retenant le chromosome 20 délété d'un patient. Cet hybride somatique a été utilisé pour analyser des gènes dits „candidats“ pour le syndrome d'Alagille. Le choix d'un

gène candidat a été effectué selon 3 critères: sa localisation dans la région 20p11.23-p12.2, sa localisation sur le chromosome 2 de la souris (dans les régions F, G et H conservées en synténie), et son implication dans des processus embryonnaires de type interaction mésenchyme-épithélium. Les gènes *BMP2A* (Bone morphogenetic protein 2A), *CST3* (Cystatine C3), *PAX1* (Paired box gene 1) localisés en 20p11-p12 et *HNF3β* (Hepatic nuclear factor 3β) localisé dans la région H du chromosome 2 de la souris ont été cartographiés en dehors de la délétion (Deleuze et coll., 1994).

**Deleted chromosome 20 from a patient with Alagille syndrome isolated in a cell hybrid through leucine transport selection : Exclusion of three candidates genes.**

Deleuze, J.-F., Dhone, S., Hazan, J., Borghi, E., Raynaud, N., **Pollet, N.**, Meunier-Rotival, M., Deschatrette, J., Alagille, D. and Hadchouel, M.

**Mamm. Genome 5 :663-669 (1994).**

Alagille syndrome (AGS) is a well-defined genetic entity assigned to the short arm of Chromosome (Chr) 20 by a series of observations of AGS patients associated with microdeletions in this region. By fusing lymphoblastoid cells of an AGS patient that exhibited a microdeletion in the short arm of Chr 20 encompassing bands p11.23 to p12.3 with rodent thermosensitive mutant cells (CHOtsH1-1) deficient in-leucyl-tRNA synthetase, we isolated a somatic cell hybrid segregating the deleted human Chr 20. This hybrid clone, designated NR2, was characterized by several methods, including PCR, with eight pairs of oligonucleotides mapped to Chr 20: D20S5, D20S41, D20S42, D20S56, D20S57, D20S58, adenosine deaminase (ADA), and Prion protein (PRIP); Restriction Fragment Length Polymorphism (RFLP) analyses with four genomic anonymous probes (D20S5, cD3H12, D20S17, D20S18); and fluorescent in situ hybridization (FISH) with total human DNA and D20Z1, a sequence specific to the human Chr 20 centromere, as probes. The NR2 hybrid allowed us to exclude three candidate genes for AGS: hepatic nuclear factor 3 beta (HNF3 beta), paired box 1 (PAX1), and cystatin C (CST3) as shown by their localization outside of the deletion. The NR2 hybrid is a powerful tool for the mapping of new probes of this region, as well as for obtaining new informative probes specific for the deletion by subtractive cloning of the region. Such markers will be useful for linkage analysis and screening of cDNA libraries.

L'objectif de mon travail de thèse a été de rechercher le ou les gène(s) responsable(s) de ce syndrome par clonage positionnel. L'identification et la caractérisation du ou des gènes dont la dysfonction est à l'origine de l'AGS devant permettre de mieux comprendre les mécanismes du développement embryonnaire impliqués. Par ailleurs, ces résultats ont contribué à cartographier davantage le chromosome 20 qui était l'un des moins bien étudiés du génome humain au début de

ces travaux. Le projet génome humain était en plein essor avec le développement des cartes génétiques de haute résolution au Généthon et la cartographie physique du génome à l'aide des YACs au Centre d'Etude du Polymorphisme Humain (CEPH) mais également le séquençage systématique d'ADNc (ESTs). La bioinformatique était également une discipline effervescente.

### ***Cartographie physique***

Au début de cette étude, l'observation d'une microdélétion chez 9 patients atteints du syndrome d'Alagille avait permis de localiser cytogénétiquement le défaut génétique responsable sur le bras court du chromosome 20 en 20p11-23 p12-2. Seuls quatre marqueurs de type RFLP (D20S5, D20S6, cD3H12 et D20S18) étaient connus pour être localisés dans la région délétée d'un patient (Zhang et coll., 1990).

L'acquisition de nouvelles données pour identifier la région contenant le locus du syndrome d'Alagille nécessitait l'établissement d'une carte physique des microdélétions associées. La construction d'une carte physique en 20p12 représentait la première étape vers une caractérisation précise, en termes moléculaires, de cette région délétée commune. Elle devait fournir de nouveaux marqueurs et les ordonner avec ceux qui étaient répertoriés dans la région. Cette étape de cartographie nous était indispensable pour rechercher le ou les gènes responsables de l'AGS, et permettait de contribuer à caractériser davantage le chromosome 20 qui était alors l'un des moins bien cartographié du génome humain.

Une analyse de liaison 4 points<sup>1</sup> utilisant les marqueurs D20S5, D20S6 et IP20M57 (microsatellite) avait donné un lod score maximum de 2,65 lorsque le locus morbide était confondu avec D20S5, ceci indiquant que cette sonde était la plus proche génétiquement du locus de l'AGS (Dhorne, 1991). Cette sonde avait également été localisée par hybridation in situ en 20p12.2. D'autre part D20S5 avait les caractéristiques d'un fragment d'ADN bien unique dans le génome et avait donc été séquencé pour définir un STS.

Etant donnée la taille de la région emportée par la délétion (plusieurs mégabases), le système des YAC a été choisi pour construire la carte physique.

---

<sup>1</sup> L'approche probabiliste utilisée dans une étude de cartographie génétique est la procédure des lod scores (lod est l'abréviation de « log of the odds »). La notion de lod score, originellement introduite pour la liaison entre deux marqueurs génétiques, peut être généralisée à la liaison entre un marqueur et un ensemble de locus liés entre eux.



Nous décrivons dans cet article une carte de 3,7 Mb composée de clones de YAC ordonnés et d'une carte de restriction utilisant cinq enzymes de restriction à site de coupure rare.

La construction de cette carte s'est faite en combinant plusieurs méthodes: comparaisons d'empreintes, synthèse et cartographie des extrémités des YAC, contenu en STS et cartographie de restriction. Le chimérisme a été détecté par des résultats anormaux dans la caractérisation de l'ADN des YAC par plusieurs méthodes et par hybridation *in situ* fluorescente sur chromosome métaphasique. Dans l'ensemble, 18 marqueurs, dont 7 nouvellement définis d'après la séquence d'extrémités de chromosomes artificiels, ont été ordonnés sur les YAC. Onze révèlent des polymorphismes parmi lesquels 6 ont un taux d'hétérozygotie supérieur à 0,70. Ces marqueurs potentiellement très informatifs dans une analyse de liaison s'avéreront utiles dans la cartographie génétique du locus morbide.

La construction de cette carte a permis de mesurer la zone délétée commune associée au Syndrome d'Alagille. Celle-ci s'étend de D20S5 à D20S61 sur 2250 kb, et contient cinq îlots CpG. Plusieurs marqueurs de polymorphisme sont localisés dans la région et devraient être utiles pour cartographier les points de cassure de plusieurs délétions dans l'espoir de réduire la taille de la région minimum délétée. L'intégration de nos données de cartographie physique avec celle de cartographie génétique (carte de Gyapay et coll., 1994) permettait d'observer une correspondance de 70 kb par cM entre les marqueurs D20S189 et D20S186, alors que pour l'ensemble de la carte cette correspondance était environ de 300 kb par cM. Ces deux marqueurs distants de 5 cM dans la carte de Gyapay et coll. ne sont distants que de 1 cM dans la carte de Dib et coll. (Dib et coll., 1996) , et ce qui apparaissait initialement comme un point chaud de recombinaison n'était en fait qu'une appréciation erronée de la distance séparant ces deux marqueurs sur la carte génétique de Gyapay et coll. en 1994.

**Construction of a 3.7-Mb physical map within human chromosome 20p12 ordering 18 markers in the Alagille syndrome locus.**

**Pollet, N.**, Dhorne-Pollet, S., Deleuze, J.-F., Boccaccio, C., Driancourt C, Raynaud, N., Le Paslier, D., Hadchouel, M. and Meunier-Rotival, M.

**Genomics 27 :467-474 (1995).**

Alagille syndrome (AGS, MIM 118450) is associated with human chromosome band 20p12. To study this region, we constructed a 3.7-Mb physical map using 36 YACs isolated from the CEPH YAC library with three sequence-tagged sites (STS): D20S503, D20S41, and D20S188. New STSs were obtained from 6 isolated YAC end-fragments. Eighteen markers were ordered on the contig as follows: 20ptel-D20S177-D20S175-D20S509- D20S5/D20S503-D20S506-D20S162-D20S504-D20S505-D20S507-D20S188-(D20S6-D20S27- D20S189)-D20S186-D20S41-D20S61-D20S492- D20S508-20pcen. A restriction map with the enzymes *Ascl*, *MluI*, *NotI*, *SacII*, and *SfiI* was generated, revealing seven putative CpG islands. We established a YAC contig that spans the AGS region and thus will be valuable for cloning candidate genes and searching for DNA polymorphisms segregating with this syndrome.

### ***Cartographie génétique***

Dans la stratégie de clonage positionnel appliquée au syndrome d'Alagille, la cartographie génétique du locus morbide devait être un des éléments clé. Devant la rareté des délétions chez les patients, une question d'importance était de savoir si tous les patients atteints de syndrome d'Alagille sont porteurs d'une anomalie génétique localisée en 20p12.

L'étude d'une famille où le syndrome d'Alagille ségrègeait depuis trois générations avait permis de mettre en évidence une liaison entre un marqueur localisé en 20p12 et le locus morbide ( $Z = +2,96$  à  $\theta = 0$  pour D20S27, et  $Z = + 3,56$  à  $\theta = 0$  pour l'haplotype D20S27-D20S61). Les haplotypes définis par sept marqueurs avaient permis de localiser le locus morbide entre les marqueurs D20S59 et D20S65, dans un intervalle de 40 cM englobant dans son intégralité la région commune à toutes les délétions décrites (Hol et coll., 1995). Cette analyse ne permettait pas de réduire l'intervalle à étudier pour rechercher le locus impliqué dans le syndrome d'Alagille. Une cartographie génétique du locus morbide a été menée au laboratoire sur un nombre croissant de familles et en utilisant le maximum de marqueurs localisés en 20p12 et au-delà. Les résultats obtenus sur 15 familles n'ont pas permis d'obtenir un

Iod-score supérieur à 3 avec des marqueurs localisés dans la zone délétée commune (ZDC). En revanche, les données de cartographie génétique permettaient d'envisager l'existence d'un autre locus localisé sur le chromosome 20 en position distale par rapport à l'haplotype D20S5-D20S162 ( $Z = +3,21$  à  $q = 0,05$ ), donc distal à la région délétée commune. Aucune perte d'hétérozygotie n'avait été observée chez les patients génotypés.

Ce résultat posait la question de savoir quelles sont les relations entre le syndrome d'Alagille et la ZDC, et entre la ZDC et le locus mis en évidence par l'analyse de liaison. Une explication était que le syndrome d'Alagille aurait pu être une maladie génétiquement hétérogène, et que plusieurs gènes soient impliqués. Il y aurait eu un gène dans la région délétée commune, et un autre plus distal. Néanmoins, aucune hétérogénéité génétique n'avait pu être mise en évidence dans l'échantillon analysé. Une autre explication avancée était que le syndrome d'Alagille pouvait résulter d'une mutation dans un seul gène distal à cette région délétée, et un effet de position pourrait être le mécanisme sous-jacent chez les patients présentant une délétion en 20p12 n'impliquant pas physiquement le locus AGS. Une autre explication envisagée était qu'il existe des phénocopies du syndrome d'Alagille.

Aucune de ces explications n'était correctes, nous devons apprendre par la suite qu'il s'agissait d'une erreur de cartographie génétique due à l'accumulation de plusieurs difficultés: absence de familles sur plus de deux générations compliquant l'analyse des haplotypes et surtout problèmes de génétique clinique sur le diagnostic de l'AGS. Ces travaux n'ont donc jamais été publiés.

### ***Cartographie génique***

Dans un troisième temps, une carte génique de la région 20p12 a été construite afin de proposer des gènes candidats pour le syndrome d'Alagille (Pollet et coll., 1997). La carte physique a été étendue pour couvrir toute la bande p12 du chromosome 20, puis nous avons localisé des marqueurs géniques (eSTS) du consortium IMAGE préalablement assignés au chromosome 20 ou à la région p12 (Auffray et coll., 1995). La carte intégrée qui a été construite localise 74 STS répartis

en 50 groupes ordonnés les uns par rapport aux autres sur une région de 10 Mb couverte par 65 YAC. Une carte de restriction avec les enzymes *NotI* et *Ascl* sur de l'ADN génomique et de l'ADN de YAC a également été construite pour valider la carte de clones. Nous avons utilisé 28 marqueurs eSTS qui nous ont permis d'identifier 20 gènes. Cinq de ces gènes ont été cartographiés dans les régions associées à l'AGS. Le gène *JAGGED1* humain (*HJ1* ou *JAG1*), un ligand de Notch, est le seul gène localisé dans la ZDC. Ce gène est considéré comme un excellent candidat pour l'AGS étant donné ses fonctions connues au cours du développement, et son expression dans certains tissus touchés par le syndrome d'Alagille (Pollet et coll., 1997). Ceci sera d'ailleurs confirmé par la suite (Oda et coll., 1997; Li et coll., 1997).

**Construction of an integrated physical and gene map of human chromosome 20p12 providing candidate genes for Alagille syndrome.**

**Pollet, N.**, Boccaccio, C., Dhorne-Pollet, S., Driancourt, C., Raynaud, N., Auffray, C., Hadchouel, M. et Meunier-Rotival, M.

**Genomics** 42 :489-498 (1997).

Physical mapping and localization of eSTS markers were used to generate an integrated physical and gene map covering a approximately 10-Mb region of human chromosome 20p12 containing the Alagille syndrome (AGS) locus. Seventy-four STSs, 28 of which were derived from cDNA sequences, mapped with an average resolution of 135 kb. The 28 eSTS markers define 20 genes. Six known genes, namely CHGB, BMP2, PLCB1, PLCB4, SNAP, and HJ1, were precisely mapped. Among the genes identified, one maps in the smallest region of overlap of the deletions associated with AGS and could therefore be regarded as a candidate gene for Alagille syndrome.

Pour connaître quels types d'anomalies dans la structure de *JAG1* sont rencontrés, une recherche de mutations a été entreprise. Il en ressort que la plupart des mutations autres que non sens sont de nature à donner naissance à une protéine tronquée. Toutes les mutations recensées se situent dans le domaine extracellulaire. L'analyse de la transmission des mutations a révélé que 70% des cas de patients AGS étudiés sont sporadiques (Crosnier et coll., 1999).

**Mutations in JAGGED1 gene are predominantly sporadic in Alagille syndrome. Gastroenterology. 116 :1141-1148 (1999).**

Crosnier, C., Driancourt, C., Raynaud, N., Dhorne-Pollet, N., **Pollet, N.**, Bernard, O., Hadchouel, M. et Meunier-Rotival, M.

**BACKGROUNDS & AIMS:** Mutations in the JAGGED1 gene are responsible for the Alagille syndrome, an autosomal dominant disorder characterized by neonatal jaundice, intrahepatic cholestasis, and developmental disorders affecting the liver, heart, vertebrae, eyes, and face. We screened a large group of patients for mutations in JAGGED1 and studied transmission of the mutations. **METHODS:** The coding sequence of the JAGGED1 gene was searched by single-strand conformation polymorphism and sequence analysis for mutations in 109 unrelated patients with the Alagille syndrome and their family if available. **RESULTS:** Sixty-nine patients (63%) had intragenic mutations, including 14 nonsense mutations, 31 frameshifts, 11 splice site mutations, and 13 missense mutations. We identified 59 different types of mutation of which 54 were previously undescribed; 8 were observed more than once. Mutations were de novo in 40 of 57 probands. **CONCLUSIONS:** Most of the observed mutations other than the missense mutations in JAGGED1 are expected to give rise to truncated and unanchored proteins. All mutations mapped to the extracellular domain of the protein, and there appeared to be regional hot spots, although no clustering was observed. Thus, the sequencing of 7 exons of JAGGED1 would detect 51% of the mutations. Transmission analysis showed a high frequency of sporadic cases (70%).

## **Conclusions**

Les recherches actuelles sur le syndrome d'Alagille visent à comprendre la physiopathologie de cette maladie. Ceci se traduit par l'étude de l'expression de *JAGGED1* au cours de l'embryogenèse chez l'homme, et l'analyse des mécanismes d'action des différentes mutations recensées.

Au cours de ma thèse, les avancées des projets génomes avaient mis en évidence un important nombre de gènes de fonction inconnue qui révélaient « l'étendue de notre ignorance ». Les domaines de l'après-séquençage, de l'après-gène et de l'analyse fonctionnelle étaient à explorer. Un thème récurrent était de savoir dans quelle mesure l'analyse à l'aide d'outils de traitement de l'information serait suffisante pour permettre le passage de la séquence à la fonction moléculaire et au rôle biologique (Miklos et Rubin, 1996). La génomique fonctionnelle se développait (par exemple Eurofan pour la levure; Dujon, 1998) avec comme justification que l'expérimentation sur les organismes modèles pourrait permettre une analyse fonctionnelle détaillée et à grande échelle de l'ensemble des gènes identifiés. En revanche, les approches expérimentales qui devaient permettre

l'analyse en parallèle de la fonction de nombreux gènes restaient pour la plupart à définir. L'étude des patrons d'expression des transcrits était celle qui se prêtait le mieux à une telle analyse. La fabrication plus ou moins dirigée de mutants et leur caractérisation, pour les organismes chez lesquels les études génétiques sont les plus aisées (drosophile, levure, souris, nématode, poisson-zèbre) était toujours du plus grand intérêt, l'étude des interactions géniques nécessitant de disposer de modèles génétiques très perfectionnés. Enfin, cette analyse fonctionnelle devait passer également par l'étude des protéines (« protéome ») afin de relier les données d'ordre génétique à celles d'ordre biochimique.

Dans cette perspective, on peut voir la biologie du développement comme un moyen puissant permettant d'identifier de nombreux gènes, d'appréhender leur rôle et de comprendre leur fonction au cours du développement. Intégrée dans l'ère actuelle de la génomique, la biologie du développement peut trouver des applications dans le choix des cibles pour la mise au point de médicaments traditionnels, le diagnostic moléculaire, la production de protéines thérapeutiques voire la thérapie génique (Ko, 2001).

J'ai décidé d'effectuer mon stage post-doctoral dans le domaine de la biologie du développement „à grande échelle“. Christof Niehrs avait constitué un jeune groupe de recherche au DKFZ à Heidelberg et m'avait proposé de continuer un criblage par hybridation in situ systématique sur des embryons de xénope. Ce projet m'ayant enthousiasmé dans le cadre de la génomique fonctionnelle, je suis parti en Allemagne où j'ai travaillé quatre ans à développer l'approche WISH („Wholemout In Situ Screen“).

# 2

## Analyse des profils d'expression génique au cours du développement embryonnaire du xénope.

### *Introduction*

L'amphibien anoure *Xenopus laevis* est l'un des organismes modèles favoris des embryologistes (Beck et Slack, 2001). Les raisons en sont sa facilité d'élevage, ses embryons de relativement grande taille produits en grandes quantités et au développement externe assez rapide. En termes moléculaires, de nombreux gènes codant des protéines jouant des rôles importants au cours du développement embryonnaire précoce ont été clonés et caractérisés initialement chez le xénope (*gooseoid*, *siamois*, *noggin*, *cerberus*, *dickkopf* ...) et des contributions importantes ont été faites dans les domaines de l'étude de la réplication, du cycle cellulaire, de la chromatine et de la transcription.

De nombreux mécanismes du développement mettent en jeu une expression génétique différentielle: certains gènes s'expriment dans des cellules données, à des moments spécifiques et selon des intensités particulières. Je me suis intéressé aux gènes activés différentiellement au cours du développement embryonnaire précoce, qui sont donc susceptibles de jouer des rôles régulateurs majeurs.

Je me proposais d'étudier systématiquement l'expression des gènes au cours de l'embryogenèse chez *Xenopus laevis*. La grande majorité du programme génétique s'exprime pendant le développement embryonnaire. L'embryon puis la larve se prêtent particulièrement bien à l'expérimentation, permettant d'étudier la fonction des gènes *in vivo* et dans leur contexte. Il est en effet vraisemblable que l'ensemble des gènes et des voies de signalisation possibles sont en oeuvre au cours du développement. Le modèle murin reste coûteux et lourd à établir pour étudier les mécanismes moléculaires du développement précoce. En ce sens, *Xenopus laevis* est un bon modèle vertébré.

De nombreuses méthodes ont été développées pour caractériser à grande échelle les profils d'expression génique de différents types cellulaires ou tissus. Le séquençage partiel d'ADNc (production d'EST) est l'approche la plus ancienne, et a largement contribué à identifier de nombreux gènes, en particulier dans l'espèce humaine. Cependant, cette méthode n'est pas optimale pour quantifier les niveaux d'expression des transcrits en ce qu'elle requiert une importante capacité de séquençage et qu'elle n'est pas suffisamment sensible pour pouvoir détecter les

ARNm rarement exprimés.

La technique SAGE ("*Serial Analysis of Gene Expression*"), réduisant chaque gène exprimé à une information minimum, un *tag* d'une quinzaine de nucléotides, reste difficile à mettre en oeuvre et nécessite de disposer préalablement d'une importante collection de séquences d'ADNc pour passer du *tag* au gène.

Le *differential display* et ses variantes sont bien adaptés à la recherche d'un ou deux gènes dont l'expression est fortement modulée entre deux situations, mais leur productivité est trop faible et leur fiabilité trop aléatoire pour des études à grande échelle.

Actuellement, les méthodes fondées sur l'hybridation de sondes complexes provenant de la rétrotranscription d'une population d'ARN avec de grands ensembles de segments d'ADN (oligonucléotides, produits de PCR issus de clones d'ADNc) fixés à haute densité sur un support solide (membranes de Nylon dans le cas des *macroarrays*, lames de verre pour les *microarrays* encore appelés puces à ADN) sont les plus utilisées, les plus sensibles et les plus productives pour caractériser en masse l'expression de milliers de gènes différents.

L'utilisation d'ADNc préalablement caractérisés par séquençage partiel ou empreintes d'oligonucléotides est nécessaire pour disposer d'ensembles non redondants de milliers de cibles représentant chacune un gène différent. La puissance de cette approche repose sur son parallélisme en ce que chaque hybridation donne des renseignements pour chacun des gènes représentés. Rien ne s'oppose à l'obtention de tels *micro* ou *macroarrays* sur lesquels la presque totalité des gènes exprimés dans l'embryon serait représentée. L'utilisation de supports solides en verre est une technique intéressante du fait qu'elle permet une comparaison directe des profils d'hybridation obtenus avec des sondes complexes marquées avec différents fluorophores. Cependant, la quantité de cible fixée sur les supports de type Nylon est beaucoup plus importante que sur lame de verre, la sensibilité intrinsèque de la radioactivité est meilleure et les filtres sont facilement réutilisables. Aujourd'hui, de nombreux systèmes disponibles dans le commerce permettent de produire et d'utiliser ses propres *micro* ou *macroarrays*.

L'information apportée par ce type d'étude est essentiellement quantitative, et très indirecte. Elle gagne à être complétée par des approches telles que l'hybridation *in situ* sur embryons entiers qui donne des informations quant à la distribution spatiale des ARNm, *in vivo*. Contrairement à la souris ou au poulet, on peut obtenir aisément des milliers d'embryons de xénope à partir d'une seule femelle. La petite taille des embryons (0,5 à 1,5 mm de diamètre) permet la manipulation de nombreux échantillons en parallèle (chambre à puits multiples, plaques de microtitration, robot), rendant possible une étude à grande échelle. Le même travail effectué avec des



embryons de poulet ou de souris serait techniquement extrêmement laborieux et très coûteux.

Contrairement à la plupart des travaux actuels d'analyse systématique de profils d'expression qui faisaient appel aux puces à ADN chez la levure ou l'homme, j'ai utilisé une approche d'hybridation *in situ* à grande échelle couplée à un séquençage d'ADNc chez un organisme modèle vertébré non-mammifère, l'amphibien anoure *Xenopus laevis*.

### ***L'hybridation in situ sur embryons entiers à grande échelle***

A mon arrivée au Laboratoire d'embryologie moléculaire, Volker Gawantka terminait une étude pilote par hybridation *in situ* systématique pour identifier des gènes exprimés au cours du développement précoce. Avant son départ (6 mois après mon arrivée), j'ai repris et achevé cette étude. Tout d'abord j'ai réalisé un travail bioinformatique d'analyse et de représentation des données puis j'ai étudié la régionalisation du bourgeon caudal à l'aide des marqueurs de différenciation identifiés.

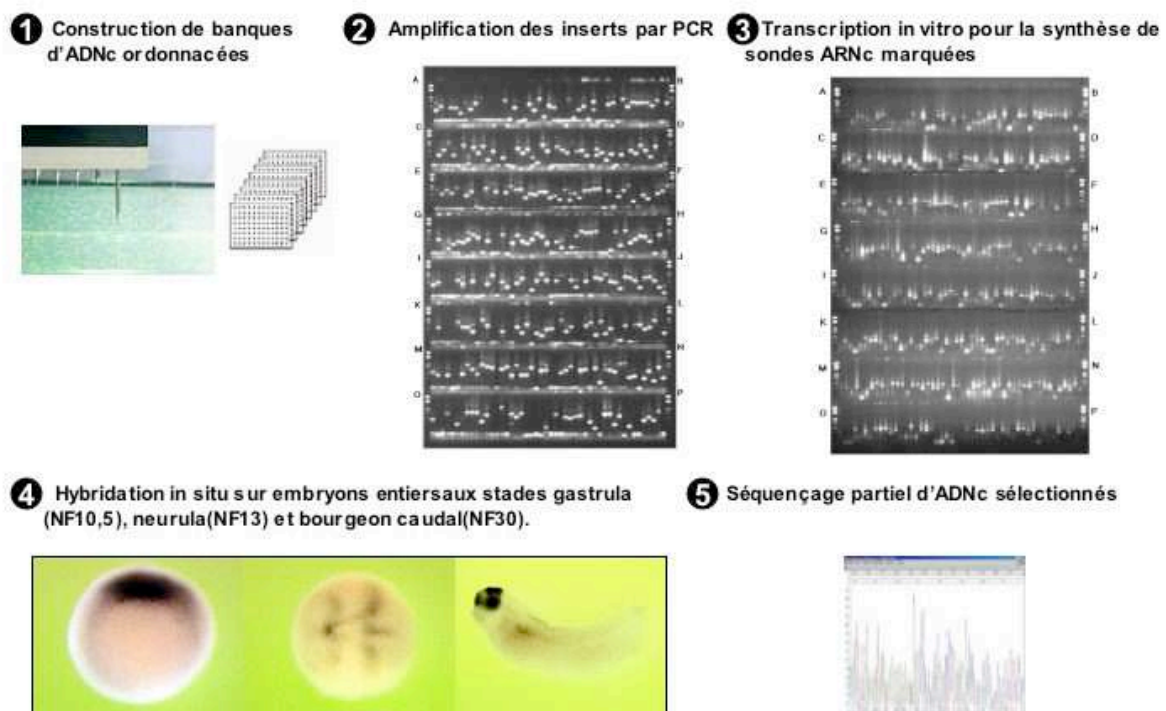


Figure 1 : Principe du criblage par hybridation *in situ* systématique.

Dans cette première étude, 1765 ADNc pris au hasard dans une banque d'ADNc de xénope au stade neurula ont été caractérisés par hybridation *in situ* sur embryons entiers aux stades gastrula, neurula et bourgeon caudal. Ces trois stades embryonnaires ont été choisis afin de dévoiler l'expression de gènes lors de la régionalisation du mésoderme (stade gastrula NF10,5) de la neurogenèse (stade neurula NF13) et de l'organogenèse (stade bourgeon caudal NF 30). La banque d'ADNc utilisée avait été construite par Volker Gawantka à partir d'ARN extraits d'embryons au stade NF13, moment où de nombreux gènes impliqués dans la régionalisation du mésoderme sont encore exprimés (chordin, noggin, gooseoid ...) et où ceux impliqués dans la neurogenèse commencent à l'être.

Les résultats obtenus peuvent être résumés ainsi : 26 % des ADNc correspondent à des gènes d'expression régionalisée; 51 % des ADNc correspondent à des gènes d'expression ubiquitaire; et l'hybridation de 23 % des ADNc utilisés ne donne lieu à aucun signal. Le séquençage partiel des 449 ADNc correspondant aux gènes d'expression différentielle a mis en évidence qu'ils correspondent à 273 gènes distincts (moins de 1 % du génome exprimé pendant l'embryogenèse), dont 207 étaient nouvellement identifiés chez le xénope et 24 % sans homologues connus. Cette recherche systématique est donc un moyen puissant permettant de caractériser de nouveaux facteurs contrôlant le développement normal.

Parmi ces gènes, 65 peuvent servir de marqueurs de différenciation au stade bourgeon caudal. Ces marqueurs permettent d'identifier les différents territoires résultant des premières étapes de régionalisation et ainsi dévoiler l'anatomie moléculaire de l'embryon (la description du développement en terme d'expression génique). Nous avons décrit, à l'aide de gènes nouvellement identifiés, la régionalisation de l'embryon au stade gastrula et celle de l'épiderme et de la région formant la queue au stade bourgeon caudal.

Ce type de criblage par hybridation *in situ* systématique sur embryons était l'un des premiers du genre. Il a permis d'identifier de nombreux gènes régulateurs jouant des rôles essentiels au cours du développement tels les gènes à homéoboites de la famille *Vent* (*Xvent1* et *Xvent2*), le facteur de transcription à doigt de zinc *blimp* ainsi que *bambi*, un régulateur négatif de la voie de signalisation BMP4. Une bonne surprise a été de constater l'abondance (1 clone sur 4) de clones d'ADNc correspondants à des gènes d'expression différentielle et non caractérisés. Il faut

rappeler qu'aucun projet de séquençage partiel d'ADNc n'était en cours chez le xénope et que le jeu de 502 séquences soumises correspondait à un important jeu de données chez le xénope. L'avantage indéniable de ce type d'approche est sa facilité de mise en œuvre, et le fait de disposer immédiatement des informations d'expression et de séquence qui sont un préalable à toute étude de perte et gain de fonction visant à disséquer les rôles biologiques joués par un gène. Le fait également d'identifier des marqueurs de cellules différenciées permet d'aller plus loin dans l'étude de la différenciation.

**Gene expression screening in *Xenopus* identifies molecular pathways, predicts gene function and provides a global view of embryonic patterning.**

**Pollet\*, N.**, Gawantka\*, V. , Delius, H., Vingron, M., Pfister, R., Nitsch, R., Blumenstock, C. and Niehrs C. (1998)

**Mech. Dev. 77 :95-141.**

In a large scale gene expression screen 1765 randomly picked cDNAs were analysed by whole-mount in situ hybridization in *Xenopus* embryos. 273 unique, differentially expressed genes were identified, 204 of which are novel in *Xenopus*. Partial DNA sequences and expression patterns were documented and assembled into a database, "AXeIDB". About 30 % of cDNAs analysed represent differentially expressed genes and about 5 % show highly regionalized expression. Novel marker genes and potential developmental regulators were found. Differential expression of mitochondrial genes was observed. Marker genes were used to study regionalization of the entire gastrula as well as the tail forming region and the epidermis of the tailbud embryo. Four "synexpression" groups representing genes with shared, complex expression pattern that predict molecular pathways involved in patterning and differentiation were identified. According to their probable functional significance these groups are designated as Delta1, Bmp4, ER-import and Chromatin group. Within synexpression groups, a likely function of genes without sequence similarity can be predicted. The results indicate that synexpression groups have strong prognostic value. A cluster analysis was made by comparing gene expression patterns to derive a novel parameter, "tissue relatedness". In conclusion, this study describes a semi-functional approach to investigate genes expressed during early development and provides global insight into embryonic patterning.

\*co premier auteurs

Les travaux présentés dans cet article ont fait l'objet de plusieurs présentations dans des congrès, dont deux (Genomics and Proteomics, Heidelberg, Germany, 4-7 octobre 1998 et Transcriptome 2000 Institut Pasteur, France, 6-9 Novembre 2000) ont publiés les actes après relecture.

**Large scale expression screening identifies molecular pathways and predicts gene function.**

**Pollet, N.**, Gawantka, V., Delius, H. and Niehrs, C. (2000)

"**Genomics and Proteomics**". Ed. Sandor Suhai, Kluwer Academic/ Plenum Publishers, New-York.

-----  
**Analyse In Situ de l'expression des gènes dans les embryons de Xénope.**

**Pollet, N.**, Delius, H. and Niehrs, C. (2003)

**Comptes Rendus Biologies.** 326 :1011-1017.

Nous avons établi plusieurs collaborations fructueuses au cours de cette étude, dans lesquelles nous partageons nos informations sur des patrons d'expression particuliers pour des gènes connus ou non. Avec le laboratoire de Chris Kintner, le gène *esr4* (enhancer of split related 4) codant un facteur de transcription a été étudié plus en détail pour ses rôles au cours de la somitogenèse.

**Periodic repression of Notch pathway genes governs the segmentation of Xenopus embryos.**

Jen, W.C., Gawantka, V., **Pollet, N.**, Niehrs, C. and Kintner, C. (1999).

**Genes Dev.** 13 :1486-1499.

During the development of the vertebrate embryo, genes encoding components of the Notch signaling pathway are required for subdividing the paraxial mesoderm into repeating segmental structures, called somites. These genes are thought to act in the presomitic mesoderm when cells form prospective somites, called somitomeres, but their exact function remains unknown. To address this issue, we have identified two novel genes, called ESR-4 and ESR-5, which are transcriptionally activated in the somitomeres of *Xenopus* embryos by the Su(H)-dependent Notch signaling pathway. We show that the expression of these genes divides each somitomere into an anterior and posterior half, and that this pattern of expression is generated by a mechanism that actively represses the expression of the Notch pathway genes when paraxial cells enter a critical region and form a somitomere. Repression of Notch signaling during somitomere formation requires a negative feedback loop and inhibiting the activity of genes in this loop has a profound effect on somitomere size. Finally we present evidence that once somitomeres form, ESR-5 mediates a positive feedback loop, which maintains the expression of Notch pathway genes. We propose a model in which Notch signaling plays a key role in both establishing and maintaining segmental identity during somitomere formation in *Xenopus* embryos.

Cet ensemble de gènes dont l'expression est caractérisée a été utilisée par la suite par le groupe de Nancy Papalopoulou pour réaliser un criblage par étude de perte de fonction à l'aide d'oligonuclotides anti-sens de type morpholinos (Kenwrick et coll., 2004).

## **Les groupes de synexpression**

L'étude de l'ensemble des gènes transcrits, le transcriptome, peut fournir des informations inédites sur la fonction des gènes pour lesquels la recherche de similitude de séquence dans les bases de données n'a pas donné d'indices probants. Nous avons comparé les profils d'expression obtenus, ce qui a permis de constituer des ensembles de gènes en fonction des similitudes de leur profil d'expression, dénommés « groupes de synexpression ».

Quatre groupes de "synexpression" ont été définis dans notre étude : Delta, Bmp4, réticulum endoplasmique et chromatine. Un examen de la littérature montre qu'il en existe de nombreux autres (glycolyse, biosynthèse du cholestérol, neurotransmission par le GABA ...) qui ont été révélés chez la levure, le rat et l'homme par l'examen systématique des profils d'expression génique à l'aide de puces à ADN ou de RT-PCR quantitative. Ainsi, on peut considérer que les groupes de synexpression correspondent à une forme de régulation génétique eucaryote comparable à l'opéron procaryote. Il est probable que cette organisation de l'information génétique facilite les changements au cours de l'évolution qui sont à l'origine de la diversité du monde animal (Niehrs et Pollet, 1999).

Au sein d'un même groupe de synexpression, il existe au moins un gène dont la fonction et le rôle biologique sont connus. Pour le groupe Bmp4 il s'agit du gène *Bmp4* codant un facteur de croissance de type Transforming Growth Factor  $\beta$ . Pour le groupe Delta il s'agit du gène *Xdelta1*, un ligand de Notch. Pour le groupe réticulum endoplasmique, il s'agit de gènes codant des protéines impliquées dans la translocation des protéines néo-synthétisées vers le réticulum endoplasmique (sous-unités du translocon, par exemple). Enfin, pour le groupe chromatine, il s'agit de gènes codant des protéines telles que les "High Mobility Group proteins" impliquées dans l'organisation de la chromatine.

Ces regroupements se révèlent très instructifs à plus d'un titre. On retrouve ensemble des gènes impliqués dans un même processus cellulaire, métabolique ou dans une même voie de signalisation. Il est ainsi possible d'attribuer une fonction potentielle à un gène inconnu lorsque celui-ci est associé au sein d'un groupe avec des gènes dont la fonction a déjà été identifiée. L'étude détaillée du gène en question permet dans un second temps de vérifier l'hypothèse de sa fonction (ce qui a été fait pour les gènes *Xvent2* et *Bambi*, membres du groupe de synexpression BMP4; (Onichtchouk et coll., 1996; Onichtchouk et coll., 1999) ainsi que pour NRARP, membre du groupe de synexpression delta1 (Lamar et coll., 2001 ; Lahaye et coll., 2002).

### **Synexpression groups in eukaryotes.**

Niehrs, C. and **Pollet, N.** (1999)

**Nature** 402 :483-487.

In 1960, Jacob and Monod described the bacterial operon, a cluster of functionally interacting genes whose expression is tightly coordinated. Global expression analysis has shown that the highly coordinate expression of genes functioning in common processes is also a widespread phenomenon in eukaryotes. These sets of co-regulated genes, or 'synexpression groups', show a striking parallel to the operon, and may be a key determinant facilitating evolutionary change leading to animal diversity.

### **Nrarp is a novel intracellular component of the Notch signaling pathway.**

Lamar, E., Deblandre, G., Wettstein, D., Gawantka, V., **Pollet, N.**, Niehrs, C. and Kintner, C. (2001).

**Genes Dev.** 15 :1885-99.

The Lin12/Notch receptors regulate cell fate during embryogenesis by activating the expression of downstream target genes. These receptors signal via their intracellular domain (ICD), which is released from the plasma membrane by proteolytic processing and associates in the nucleus with the CSL family of DNA-binding proteins to form a transcriptional activator. How the CSL/ICD complex activates transcription and how this complex is regulated during development remains poorly understood. Here we describe Nrarp as a new intracellular component of the Notch signaling pathway in *Xenopus* embryos. Nrarp is a member of the Delta-Notch synexpression group and encodes a small protein containing two ankyrin repeats. Nrarp expression is activated in *Xenopus* embryos by the CSL-dependent Notch pathway. Conversely, overexpression of Nrarp in embryos blocks Notch signaling and inhibits the activation of Notch target genes by ICD. We show that Nrarp forms a ternary complex with the ICD of XNotch1 and the CSL protein XSu(H) and that in embryos Nrarp promotes the loss of ICD. By down-regulating ICD levels, Nrarp could function as a negative feedback regulator of Notch signaling that attenuates ICD-mediated transcription.

### ***Axeldb: A Xenopus laevis database***

En parallèle à ce travail expérimental de caractérisation des profils d'expression génique, nous avons voulu travailler à l'organisation de toutes les informations éparses et diverses que nous avons produites. Cela a demandé un travail de constitution d'une base de données spécialisée et conviviale. Celle-ci a été développée en utilisant Acedb (A *Caenorhabditis elegans* database) comme logiciel générique (Durbin et Thierry-Mieg, 1991). Axeldb (A *Xenopus laevis* database; Pollet et coll., 2000a) recense l'ensemble des données d'expression, de séquence ainsi que les photographies des résultats d'hybridation *in situ*, et a été rendu d'accès

public via Internet (<http://www.dkfz-heidelberg.de/abt0135/axelddb.htm>). Cette forme de communication de nos résultats est la plus appropriée et facilite les approches comparatives. Il est aisé d'observer l'expression pendant le développement de l'homologue chez le xénope d'un gène humain, puis le cas échéant de réaliser ou compléter une étude fonctionnelle en utilisant les approches expérimentales disponibles chez le xénope (par expression ectopique d'ARNm par exemple).

**Axelddb: a *Xenopus laevis* database focusing on gene expression.**

**Pollet, N.**, Schmidt, H.A., Gawantka, V., Vingron, M. and Niehrs, C. (2000)

**Nucleic Acids Res. 28 :139-140.**

Axelddb is a database storing and integrating gene expression patterns and DNA sequences identified in a large-scale in situ hybridization study in *Xenopus laevis* embryos. The data are organised in a format appropriate for comprehensive analysis, and enable to compare images of expression pattern for any given set of genes. Information on literature, cDNA clones and their availability, nucleotidic sequences, expression pattern and accompanying pictures is available. Current developments are aimed toward the interconnection with other databases and the integration of data from literature. Axelddb is implemented using an ACEDB database system, and available through the web at <http://www.dkfz-heidelberg.de/abt0135/axelddb.htm>

### ***Analyse in silico des profils d'expression***

La structuration des données est le préalable indispensable à une analyse *in silico* des profils d'expression génique. Au travers d'une collaboration très fructueuse avec l'équipe de Martin Vingron (département de bioinformatique théorique du DKFZ) il a été possible de développer les outils informatiques permettant d'analyser systématiquement des données d'expression : comparaison des patrons d'expression, et comparaison des tissus d'après leur signature d'expression. Les résultats sont prometteurs et les outils applicables à d'autres données d'expression (Pollet et coll., 2000b).



***In silico* analysis of gene expression patterns during early development of *Xenopus laevis*.**

**Pollet, N.**, Schmidt, H.A., Gawantka, V., Niehrs, C and Vingron, M. (2000)

**Pac. Symp. Biocomput.** 443-454

The information as to where and when a mRNA is present in a given cell is essential to bridge the gap between the DNA sequence of a gene and its physiological function. Therefore, a major component of functional genomics is to characterize the levels and the spatio-temporal domains of gene expression. Currently, there is just a few specialised public databases available storing the data on gene expression while they are needed as a resource for the field. Moreover, there is a need to develop and assess computational tools to compare and analyse expression profiles in a suitable way for biological interpretation. Here we describe our recent work on developing a database on gene expression for the frog *Xenopus laevis*, and on setting up and using new tools for the analysis and comparison of gene expression patterns. We used histogram clustering to compare expression profiles at both gene and tissue levels using a set of data coming from the characterization of the expression of genes during early development of *Xenopus*. This enabled us to draw a tree of tissue relatedness and to identify coexpressed genes by *in silico* analysis.

L'utilisation du xénope comme système modèle apparaît être un moyen rapide et économique pour acquérir une grande quantité d'information utilisable pour mieux comprendre le développement normal ou pathologique chez les vertébrés supérieurs, dont l'homme. Il nous est ainsi possible de construire un index des gènes humains homologues des gènes exprimés pendant l'embryogenèse du xénope dont nous aurons caractérisé l'expression.

Nous avons voulu continuer notre étude à plus grande échelle en tirant avantage de l'utilisation de banques d'ADNc ordonnées en plaques de microtitration, et d'un robot qui automatise les différentes étapes de l'hybridation *in situ*. Un protocole permettant de cribler environ 200 clones d'ADNc par semaine a été mis au point. Dans ce protocole toutes les étapes ont été optimisées pour fonctionner en système de plaques de microtitration à 96 puits.

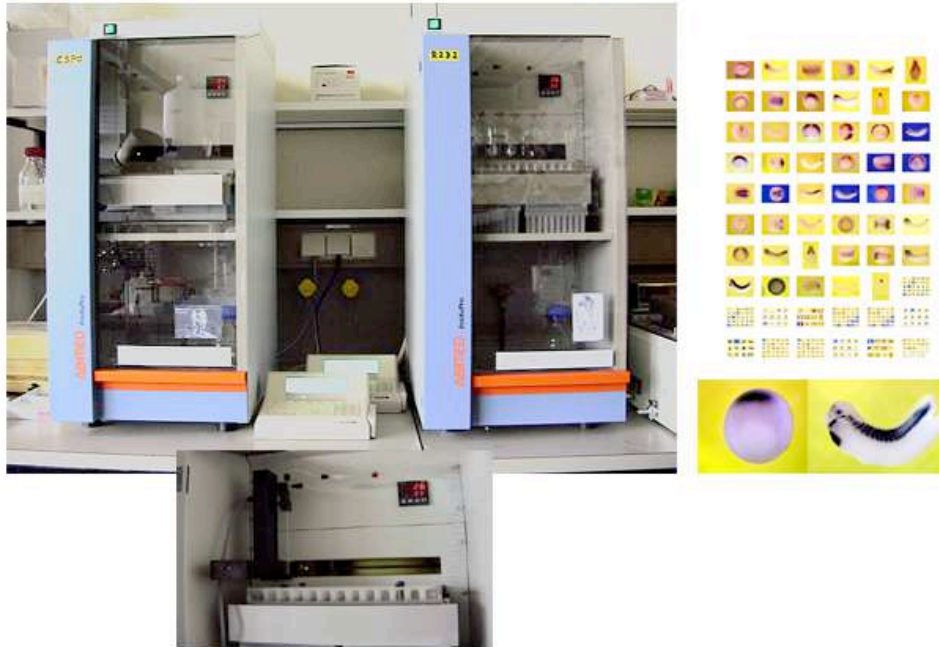
**Expression profiling by systematic high-throughput *in situ* hybridization to whole-mount embryos.**

**Pollet, N.** and Niehrs, C. (2001).

**Methods Mol. Biol.** 175 :309-21.



## Automated *in situ* hybridization Capacity 96 samples/run



**Figure 2: Automatisation de l'hybridation *in situ* sur embryons entiers.**

Afin de minimiser la redondance dans le jeu de clones criblés, nous avons échantillonné plusieurs banques d'ADNc: neurula NFst13 comme précédemment, neurula LiCINFst13, bourgeon caudal st30. Nous avons rapidement identifié des gènes au patron d'expression très régionalisé, dont celui codant la rétinaldéhyde déhydrogénase de type II, une enzyme impliqué dans la synthèse d'acide rétinoïque. Notre collègue Tomas Pieler s'intéressant à cette voie de signalisation, nous avons établi une collaboration afin de caractériser les fonctions de XRALDH2 au cours de la régionalisation du système nerveux:

**Increased XRALDH2 activity has a posteriorizing effect on the central nervous system of Xenopus embryos.**

Chen, Y., **Pollet, N.**, Niehrs, C. and Pieler, T. (2001).

**Mech. Dev.** 101 :91-103.

Retinoic acid (RA) metabolizing enzymes play important roles in RA signaling during vertebrate embryogenesis. We have previously reported on a RA degrading enzyme, XCYP26, which appears to be critical for the anteroposterior patterning of the central nervous system (EMBO J. 17 (1998) 7361). Here, we report on the sequence, expression and function of its counterpart, XRALDH2, a RA generating enzyme in Xenopus. During gastrulation and neurulation, XRALDH2 and XCYP26 show non-overlapping, complementary expression domains. Upon misexpression, XRALDH2 is found to reduce the forebrain territory and to posteriorize the molecular identity of midbrain and individual hindbrain rhombomeres in Xenopus embryos. Furthermore, ectopic XRALDH2, in combination with its substrate, all-trans-retinal (ATR), can mimic the RA phenotype to result in microcephalic embryos. Taken together, our data support the notion that XRALDH2 plays an important role in RA homeostasis by the creation of a critical RA concentration gradient along the anteroposterior axis of early embryos, which is essential for proper patterning of the central nervous system in Xenopus.

Par ailleurs, nous avons été contacté par Erich Nigg qui souhaitait caractériser l'expression de plusieurs gènes codant des kinases, dont plusieurs de type polo-like. De façon surprenante, le gène *plx2* est exprimé de façon très restreinte dans le système vasculaire.

**Cloning and characterization of Plx2 and Plx3, two additional Polo-like kinases from *Xenopus laevis*.**

Duncan, P. I., **Pollet, N.**, Niehrs, C. and Nigg, E. A. (2001)

**Exp Cell Res.** 270:78-87.

Members of the family of Polo-like kinases are implicated in the regulation of cell cycle progression in all eukaryotes. In *Xenopus laevis*, only one member of this family, Plx1, has previously been described. Here we report the cloning and characterization of *X. laevis* Plx2 and Plx3, the likely homologs of mammalian Plk2 (Snk) and Plk3 (Fnk/Prk), respectively. RNA expression studies indicate that all three *Xenopus* Plks are present in both oocytes and unfertilized eggs. Further analysis by in situ hybridization revealed that Plx1 RNA is ubiquitously expressed in early embryos, but shows more restricted expression at later stages. In contrast, Plx2 and Plx3 expression is highly restricted in both early and late-stage embryos. Using Plx-specific antisera, Plx1 and Plx3 polypeptides could readily be detected on immunoblots of oocyte and egg extracts. Both Plx1 and Plx3 protein levels remained virtually constant during oocyte maturation. However, whereas Plx1 is more active in M phase than in I phase (P. Descombes and E. A. Nigg (1998) EMBO J. 17, 1328-1335), Plx3 protein and activity levels remained constant upon release of meiotic metaphase II-arrested egg extracts into interphase. Finally, microinjection of in vitro-transcribed RNAs for Plx1, Plx2, and Plx3 increased the rate of progesterone-induced oocyte maturation, and concomitantly, all three kinases became activated. Conversely, overexpression of the corresponding catalytically inactive kinases delayed maturation. This suggests that, at least in oocytes, all three kinases may be regulated by similar mechanisms, and they may also share common substrates. However, the strikingly restricted pattern of expression of Plx2 and Plx3 observed in embryos strongly suggests that individual Plk family members perform at least partly distinct functions at later stages of development.

***Etude de l'expression cyclique des gènes esr9 et esr10 dans le mésoderme présomitique du xénope.***

Parmi les gènes du groupe de synexpression delta1, plusieurs codent des facteurs de transcription de la famille enhancer of split. Je me suis intéressé à caractériser deux de ces gènes de la famille hairy/Enhancer of Split dont l'expression est extrêmement dynamique, tant dans le neurectoderme que dans le mésoderme présomitique. Ces travaux faisaient suite à la découverte par Isabelle Palmeiri du laboratoire d'Olivier Pourquié du patron d'expression périodique du gène hairy1 chez le poulet en relation avec la somitogenèse. Un manuscrit a été publié début 2003

dans la revue « Differentiation » rapportant les résultats obtenus :

**Cyclic expression of *esr9* gene in *Xenopus* presomitic mesoderm.**

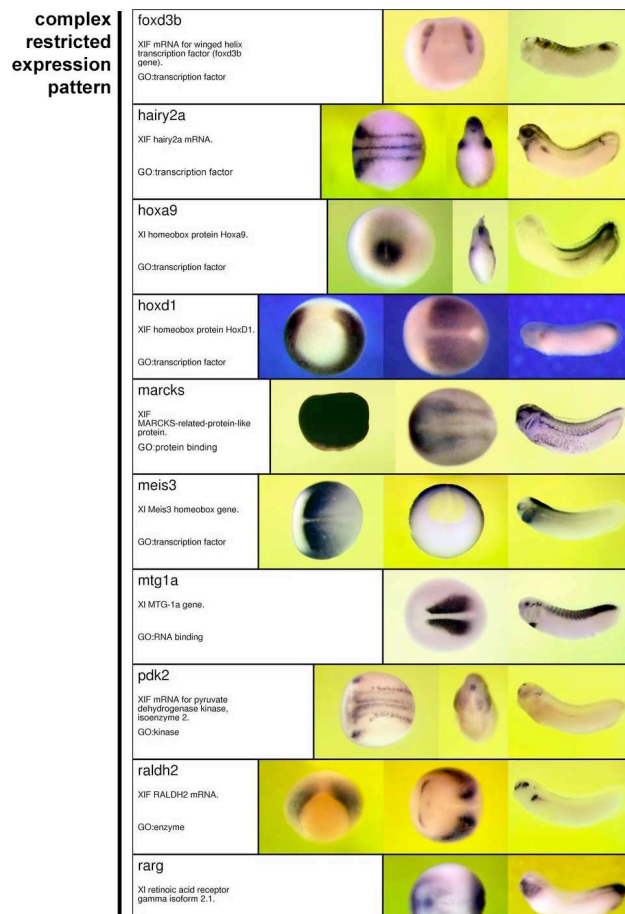
Yan Li, Ursula Fenger, Christof Niehrs and **Nicolas Pollet**. (2003)

**Differentiation**. 71 :83-89.

During somitogenesis, the cycling expression of members of the Notch signalling cascade is involved in a segmentation clock that regulates the periodic budding of somites in chicken, mouse and zebrafish. In frog, genes with cycling expression in the presomitic mesoderm have not been reported. Here, we describe the expression of *Xenopus esr9* and *esr10*, two new members of the Hairy/Enhancer of split related family of bHLH proteins. We show that they are expressed in a highly dynamic fashion, with their mRNA levels oscillating periodically in the presomitic mesoderm during somitogenesis. This dynamic expression is independent of de novo protein synthesis. Thus, expression of *esr9* and *esr10* is an indicator of the segmentation clock in the amphibian embryo. This confirms the evolutionary conservation of a molecular pathway involved in vertebrate segmentation clock.

**Constitution d'un atlas de profils d'expression géniques pour l'étude de l'embryogenèse sur le modèle xénope.**

Au total, 17664 clones d'ADNc ont été criblés par hybridation in situ sur embryons entiers. Ce travail m'a permis d'identifier 723 gènes, de caractériser leur profil d'expression et d'obtenir leur séquence partielle ou complète. Ces données m'ont permis de constituer un atlas de profils d'expression géniques pour l'étude de l'embryogenèse sur le modèle xénope. Au sein de cet atlas, chaque gène a été annoté en utilisant le vocabulaire contrôlé défini par le consortium Gene Ontology ([www.geneontology.org](http://www.geneontology.org)). Chaque profil d'expression a été documenté en



**Figure 3 : Exemple de profils d'expression régionalisés obtenus par WISH sur des embryons de xénope.**

utilisant un vocabulaire contrôlé en association avec les images d'hybridation in situ. Un article décrivant l'intégralité de cette ressource unique est soumis à publication pour un numéro spécial de la revue « Mechanisms of Development ». Les données y sont présentées sous la forme d'un atlas (62 pages de figures d'hybridation in situ du même type que celle présentée Fig. 1, l'article est présenté en annexe). Ce type d'article très descriptif est utile à la communauté scientifique mais difficile à faire publier dans des revues à large audience. C'est pourquoi j'ai travaillé à l'annotation des données et à leur présentation sur un support papier avec l'ambition que cet article serve de référence.

**An atlas of differential gene expression during early *Xenopus* embryogenesis.**

**Nicolas Pollet**, Nadja Muncke, Barbara Verbeek, Yan Li, Ursula Fenger, Hajo Delius and Christof Niehrs.

**Mech. Dev. *Xenopus* special issue**, Accepted (voir annexe)

We have carried out a large-scale, semi-automated whole-mount in situ hybridization screen of 8369 cDNA clones in *Xenopus laevis* embryos. We confirm that differential gene expression is prevalent during embryogenesis since 24 % of the clones are expressed non-ubiquitously and 8 % are organ or cell type specific marker genes. Sequence analysis and clustering yielded 723 unique genes displaying a differential expression pattern. Of these, 18 % were already described in *Xenopus*, 47% have homologs and 35% are lacking significant sequence similarity in databases. Many of them encode known developmental regulators. We classified 363 of the 723 genes for which a Gene Ontology annotation for molecular function could be attributed and found "DNA binding" and "enzyme" the most represented terms. The most common protein domains encoded in these embryonic, differentially expressed genes are the homeobox and RNA Recognition Motif (RRM). 59 homologs of human disease genes, and 254 organ or cell specific marker genes were identified. Markers were found for nasal placode and archenteron roof, organs for which a specific marker was previously unavailable. Markers were also found for novel subdomains of various other organs. The tissues for which most markers were found are muscle and epidermis. Expression of cell cycle regulators fell in two classes, containing proliferation-promoting and anti-proliferative genes, respectively. We identified 66 new members of the BMP4, chromatin, endoplasmic reticulum, and karyopherin synexpression groups, thus providing a first glimpse of their probable cellular roles. Cluster analysis of tissues to measure tissue relatedness yielded some unorthodox affinities besides expectable lineage relationships. In conclusion, this study represents an atlas of gene expression patterns, which reveals embryonic regionalization, provides novel marker genes, and makes predictions about the functional role of unknown genes.

La base de données axelldb constitue l'accès internet à cet atlas, et sera hébergée sur un serveur au laboratoire à Orsay. Par ailleurs, ces données feront partie

intégrante de la base de données de référence sur le xénope, xenbase ([www.xenbase.org](http://www.xenbase.org)).

Parmi les nombreux gènes nouvellement identifiés dans ce crible, *flrt3* a été et étudié plus en détail par Ralph Böttcher, un étudiant en thèse au DKFZ. La raison en a été son expression fortement similaire de celle du facteur de croissance FGF8. *flrt3* code une protéine membranaire riche en leucines (Eibronectin Leucin Rich Transmembrane protein) et fait partie d'une famille de trois gènes identifiés indépendamment dans un criblage pour des protéines de la matrice extracellulaire exprimées dans le muscle. Cette protéine s'est révélée être un modulateur de la signalisation par le FGF, montrant une fois de plus l'intérêt de l'approche de comparaison des patrons d'expression génique au cours du développement pour identifier de nouveaux régulateurs.

**The transmembrane protein XFLRT3 forms a complex with FGF receptors and promotes FGF signalling.**

Ralph Böttcher, Nicolas Pollet, Hajo Delius and Christof Niehrs. (2004)

**Nat Cell Biol** 6:38-44

Fibroblast growth factors (FGFs) are implicated in numerous processes during cell growth, differentiation, migration as well as cell death in normal and cancer cells and they signal through high affinity tyrosine kinase receptors. We have identified XFLRT3, member of a family of leucine-rich repeat transmembrane proteins, as a novel modulator of FGF signalling. *XFLRT3* is coexpressed with FGFs and its expression is induced following activation and is downregulated following inhibition of FGF signalling. *XFLRT3* phenocopies FGF signalling in gain- and loss-of function experiments. *XFLRT3* signalling is blocked by dominant negative ras but not by dominant negative IP3 kinase. *XFLRT3* interacts with FGF receptors in co-immunoprecipitation experiments in vitro as well as in bioluminescence resonance energy transfer assays in live cells. The results indicate that *XFLRT3* is a novel extracellular modulator of FGF-MAP kinase signalling in vertebrates.

En continuité avec mon travail de caractérisation systématique des patrons d'expression, j'avais depuis longtemps réalisé des travaux préliminaires sur la confection de puces à ADN de type macro ou microarrays à partir des cDNAs identifiés dans le crible WISH. J'ai co-encadré une thésarde italienne, Danila Baldessari, dans son projet concernant la construction et l'utilisation de puces à ADN pour l'étude systématique des transcrits chez *X. laevis*.

Dans un premier temps, la question de l'analyse des données de « profiling<sup>2</sup> » s'est posée. En comparant des données provenant de comparaison directe d'échantillons à des comparaisons utilisant une référence commune, différents critères ont pu être validés par le groupe de bioinformatique. Nous avons pu tester la correspondance entre les données puces à ADN et les résultats d'hybridation in situ. Un critère de robustesse a été établi à partir d'un ensemble de profils d'expression géniques dans des embryons de xénope à différents stades de développement, qui est basé sur la comparaison d'hybridations dans différentes conditions. Ce travail s'est concrétisé par un article soumis à publication dans le cadre d'une collaboration avec le groupe de bioinformatique théorique :

**Accuracy estimation and filtering of gene expression signals in multi-conditional experiments.**

Koenig, R., Baldessari, D., **Pollet, N.**, Niehrs, C. and Eils, R. (2004)

**Nucleic Acids Res 32:e29**

In a typical gene expression profiling experiment with multiple conditions, a common reference sample is used for co-hybridisation with the samples to yield expression ratios. Differential expression for any other sample pair can then be calculated by assembling the ratios from their hybridisations with the reference. In this study we test the validity of this approach. Differential expression of a sample pair (i, j) was obtained in two ways: directly, by hybridisations of sample i versus j, and indirectly, by multiplying the expression ratios for hybridisations of sample i versus pool and pool versus sample j. We performed gene expression profiling using amphibian embryos (*Xenopus laevis*). Every sample combination of four different stages and their pool was profiled. Direct and indirect values were compared and used as the quality criterion for the data. Based on this criterion, 82% of all ratios were found to be sufficiently accurate. To increase the reliability of the signals, several widely used filtering techniques were tested. Filtering by differences of repeated hybridisations was found to be the optimal filter. Finally, we compared microarray based gene expression profiles with the corresponding expression patterns obtained by whole mount in situ hybridisations, resulting in a 90% correspondence.

Au cours des quatre années passées au DKFZ j'ai eu l'opportunité de former plusieurs étudiants et d'encadrer deux techniciennes. Ludmilla Oumanskaia puis Ursula Fenger ont travaillé sous ma direction pendant 3 ans. Le travail répétitif formant l'ossature du criblage par hybridation in situ a constitué le sujet de stage de huit étudiants de différents niveaux : Ullrich Dossous, pharmacien souhaitant se réorienter vers la recherche, a travaillé 3 mois sous ma direction. Michael Herrmann

---

<sup>2</sup> Pour des expériences utilisant des puces à ADN, on parle de « profiling » lorsqu'on compare des profils d'expression obtenus dans de nombreuses conditions différentes.

et Marion Morscher, étudiants de deuxième cycle ont passé 6 mois à produire les embryons prêts pour l'HIS en quantités suffisantes. Lars Hildebrandt, un étudiant en médecine souhaitant également se diriger vers la recherche, a passé deux mois d'été pour faire ses premiers pas en laboratoires. Susen Zirker puis Thomas Hüller, étudiants d'une école d'ingénieur de Biotechnologie ont réalisé leur stage de 6 mois avec les deux robots d'ISH (D2R2 et Z6PO) et ont contribué significativement à la production des résultats. Nadja Muncke et Barbara Verbeek, étudiantes de Diplomarbeit (équivalent du M2 actuel et de l'ancien DEA) ont chacune contribué à des développements nouveaux pour le criblage et ont pu défendre brillamment leurs diplômes. Ainsi, j'ai acquis une certaine expérience en ce qui concerne l'encadrement d'étudiants de différents niveaux et la direction du travail de techniciens de laboratoire.



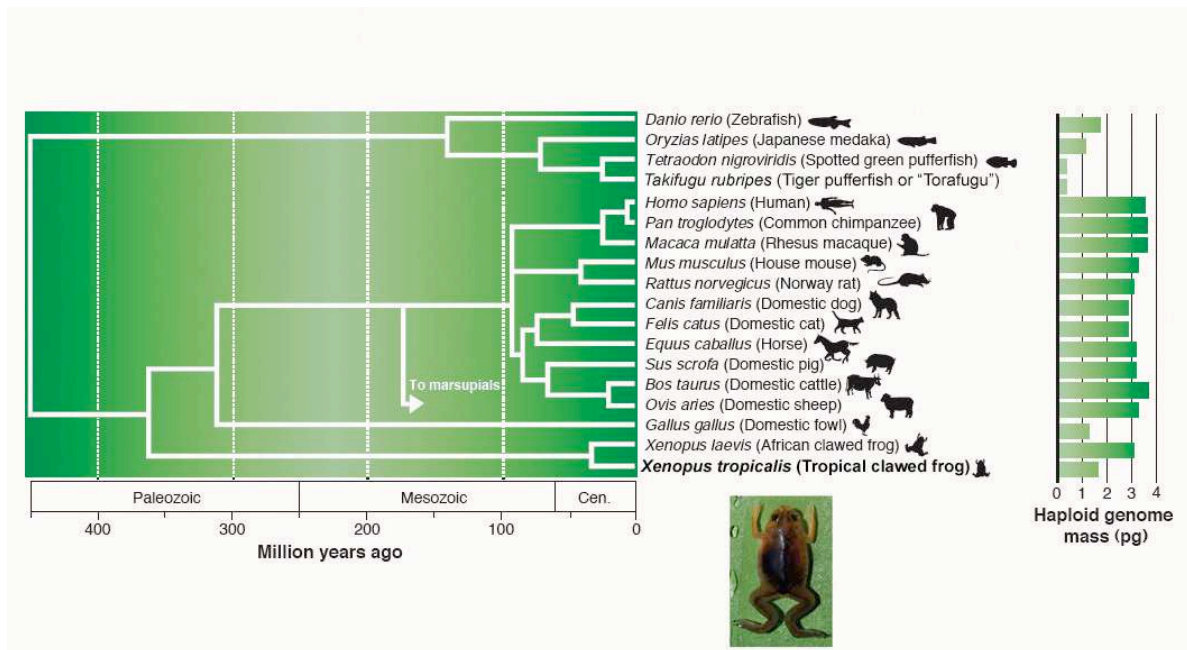
***Introduction***

Dans une perspective d'analyse comparée des génomes de vertébrés, le xénope occupe une place de référence comme modèle amphibien anoure. Aujourd'hui, on peut dire qu'il est regrettable que ce soit l'espèce *X. laevis* qui ait été mise sur le devant de la scène de la biologie moléculaire, alors que son génome est pseudo-tétraploïde. En effet, lorsqu'une pléthore de génomes ont été mis en compétition pour être déchiffrés, c'est celui de *X. tropicalis* qui a été choisi en raison de sa taille deux fois plus petite que celle de son proche parent *X. laevis* (Tymowska, 1973 ; Thiébaud et Fischberg, 1977, Figures 4 et 5). Ainsi, les données de biologie moléculaire accumulées chez *X. laevis* ne sont exploitables dans un contexte génomique qu'à la hauteur de la conservation entre ces deux espèces. Contrairement à l'espèce *X. laevis* qui est allotétraploïde<sup>3</sup>, *X. tropicalis* est diploïde et 30 à 50 millions d'années d'évolution séparent ces deux espèces (Evans et coll. 2004).

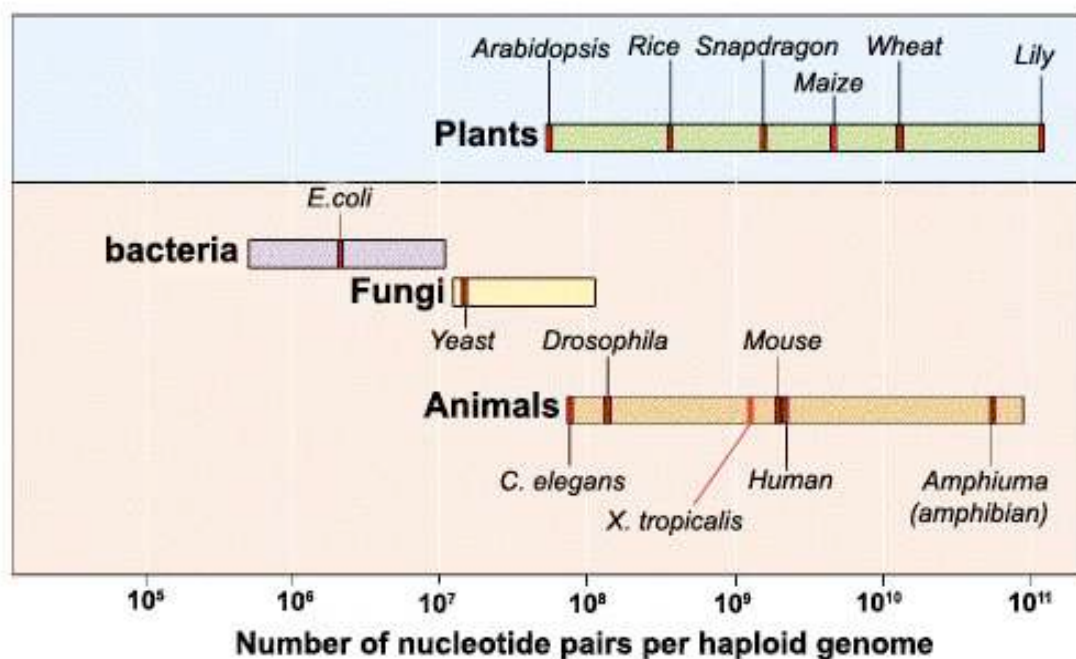
Dès la fin des années 1970-début 1980, Fischberg démontrait la diploïdie de *X. tropicalis* dans des études de biologie moléculaire (Bisbee et coll., 1977, Burki et Fischberg, 1985, Graf et Fishberg, 1986). Jusqu'à la fin des années 1990, peu de travaux se sont ensuivis pour développer l'utilisation de ce modèle en génétique moléculaire. Alors que j'étais à Heidelberg, et après avoir constaté la difficulté de travailler à grande échelle sur les transcrits de *X. laevis*, j'ai décidé de continuer mes travaux sur le modèle *X. tropicalis*. Les outils de base pour faire de la génomique fonctionnelle chez *Xt* étant absents, je me suis engagé dans un travail de fond pour les mettre en place.

---

<sup>3</sup> Une espèce tétraploïde peut se former par simple doublement d'un stock chromosomique de base (autotétraploïde) soit par hybridation de deux espèces assez différentes (allotétraploïde).



**Figure 4. Evolution des génomes chez les crâniates.** L'arbre phylogénétique montre les relations de parenté évolutive des crâniates avec les temps de divergence et la taille des génomes en pg d'ADN. Classiquement, 1 pg d'ADN est considéré équivalent à 1 Gpb, bien que le facteur de conversion entre l'humain et le fugu soit de 0,91. Adapté de Hedges and Kumar (2002), Science 2971:1283-1285.



**Figure 5. Tailles des génomes de quelques modèles en biologie.**

Les espèces de xénope *X. laevis* et *X. tropicalis* sont des modèles de vertébrés non-mammifères particulièrement pertinents pour étudier, entre autres, les mécanismes moléculaires du développement précoce (Table 1). Par comparaison avec le modèle poisson zèbre, la bibliographie sur le xénope est plus abondante

dans plusieurs domaines (embryologie, cancérologie, neurobiologie ...) ce qui suggère que davantage de connaissances ont été accumulées. En revanche, la comparaison avec le modèle poulet souligne les thèmes qui restent à développer en vue d'améliorer l'apport du modèle xénope pour l'étude de la biologie humaine. C'est également dans cette direction que je souhaite porter mes efforts de recherche dans les années à venir.

Les méthodes de la génomique ont été appliquées au xénope, et de nombreuses séquences géniques sont aujourd'hui disponibles pour étudier la biologie de *X. laevis* et de *X. tropicalis* avec les outils de la génomique fonctionnelle (Table 2).

Le séquençage « en vrac » du génome de *X. tropicalis* est entrepris par le Joint Genome Institute aux Etats-Unis (<http://genome.jgi-psf.org/xenopus1/xenopus1.info.html>, Richardson et coll., 2003). Aujourd'hui, 9,3 Gb de séquences brutes sont accessibles ainsi que 99 338 séquences « scaffold » d'une couverture de 1 053 Mpb (soit 59% du génome). Une couverture de 527 Mb du génome est assurée par 15 017 scaffold de plus de 21 kbp. Le modèle xénope dispose donc de nombreux atouts permettant d'envisager des études systématiques et comparatives à l'échelle du génome.

**Table 1. Documents PubMed , de 1995 à nos jours.**

Mots-clés	<i>Xenopus</i>	Poisson zèbre	Poulet
Total documents	17.636	2.885	20.185
<b>And development</b>	<b>5.121</b>	<b>1.577</b>	<b>13.097</b>
And virus	916	74	10.846
And cancer	706	56	986
And neuron	1.276	433	1.183
And immunology	456	76	2.938
And embryo	2.434	1.166	3.319
And pathology	187	84	1.276
<i>And disease</i>	343	98	2.036

**Table 2. Nombre de documents dans différentes bases de données du NCBI par organisme modèle.**

	Poisson zèbre	Poisson Medaka	<i>X. laevis</i>	<i>X. tropicalis</i>	Poulet
Nucleotide	704 202	150 472	449 780	483 574	762 891
Protein	13 652	743	9 126	1 544	10 023
3D Domains	2	0	253	0	1 875
UniSTS	26 922	0	0	0	666
PubMed Central	196	37	2 269	28	180
Gene ( <i>sensu</i> Entrez)	17 167	13	4 127	770	2 047

*X. tropicalis* a été présenté comme un organisme modèle permettant de combiner les approches génétiques aux méthodes sophistiquées de l'embryologie (Amaya et coll., 1998). Les techniques de mutagenèse chimique (Nakayama et Grainger, 2003) et par insertion ("gene trap", Bronchain et coll., 1999) et de transgénèse (permettant par exemple l'expression conditionnelle avec le système UAS-GAL4 ) appliquées au xénope ont été développées et sont utilisées de plus en plus largement (Amaya et coll., 1996 ; Amaya et Kroll, 1999 ; Marsh-Armstrong et coll., 1999 ; Chae et coll., 2002 ; Hartley et coll., 2002). L'approche d'inactivation génique par TILLING (Targeting Induced Lesion In Genomes est également en cours de développement chez *X. tropicalis* (Stemple, 2004)

Le xénope, tout comme le poisson zèbre, le nématode, la drosophile et la levure, fait partie d'une sélection de modèles pour lesquels les approches de génomique fonctionnelle sont désirables. Chez le xénope, cela se traduit par des projets de séquençage d'ADNc qui totalisent environ 400 000 étiquettes séquencées transcrites (EST) chez *X. laevis* et autant chez *X. tropicalis* (Klein et coll., 2002, Peng et coll., 2003 ; Gilchrist et coll., 2004). Par ailleurs, le centre de ressources du projet génome allemand distribue une collection totalisant 200 000 ADNc ordonnés provenant de 6 banques d'ADNc différentes de *X. laevis* (stades oeuf à bourgeon caudal). Ainsi, il existe d'ores et déjà des ressources génomiques accessibles chez le xénope qui faciliteront la recherche en biologie du développement, et elles sont amenées à s'accroître dans le futur.

## *Etude du transcriptome du système nerveux chez Xenopus tropicalis*

Au sein du laboratoire « Transgénèse et Génétique des Amphibiens », je me suis engagé en premier lieu dans une étude systématique de l'expression des gènes dans le système nerveux au cours de l'embryogenèse et pendant la métamorphose chez *X. tropicalis*.

Mon ambition est de mener à bien une étude systématique et à grande échelle de caractérisation de l'expression d'environ 10 000 gènes dans le système nerveux au cours de l'embryogenèse et pendant la métamorphose chez *X. tropicalis* pour identifier des gènes ayant un rôle potentiel dans la croissance et la différenciation chez les vertébrés. Ce projet est fédérateur puisqu'il intéresse directement quatre partenaires : les équipes « Embryologie expérimentale » (M. Wegnez, M. Perron) et « Transgénèse et Génétique des Amphibiens » (A. Mazabraud et N. Pollet) de l'UMR 8080, le laboratoire de Physiologie Générale et Comparée du Muséum National d'Histoire Naturelle (B. Demeneix et L. Coen) et le Centre National de Séquençage.

L'ensemble des résultats obtenus sera compilé dans une base de données d'accès public. Les ADNc identifiés seront rendus accessibles et utilisables pour identifier les homologues d'autres espèces, dont l'homme en premier lieu.

Cette étude me permettra de compléter l'atlas de l'expression des gènes de xénope au cours du développement. Cette ressource est déjà et sera une aide précieuse dans l'étude des réseaux de régulation et pour comprendre les mécanismes par lesquels l'expression différentielle des gènes se met en place au cours de la neurogenèse, de la différenciation neuronale et de la métamorphose.

## **Séquençage d'ADNc de *Xenopus tropicalis***

L'accès direct à une partie importante de l'information contenue dans le génome, celle qui code les protéines, est rendue possible par l'étude des banques d'ADNc. L'étude systématique de collections d'ADNc peut être utilisée pour obtenir une vue globale de l'expression du génome dans un type cellulaire donné en estimant l'abondance de différentes espèces d'ARNm (par l'intermédiaire de signatures, Okubo et coll., 1992). Elles peuvent également servir à établir des catalogues de transcrits (Adams et coll., 1995, Auffray et coll., 1995; Houlgatte et coll., 1995, Mara et coll., 1999). L'approche du séquençage d'ADNc se révèle finalement plus complexe que le séquençage génomique alors que les « ESTs » ont été initialement proposés comme un expédient permettant de caractériser la fraction codante du

génomique. Aujourd'hui encore, l'annotation des séquences génomiques ne peut pas se passer de l'information que recèlent les séquences d'ADNc. Malheureusement, il n'est pas possible expérimentalement d'obtenir des ADNc représentant chaque transcrite possiblement présent dans au moins une cellule au cours de la vie d'un organisme. La notion de gène prend une dimension toute probabiliste, dépendante de la quantité de transcrits correspondant à chaque gène et de la quantité de données de séquence d'ADNc disponibles pour l'annotation d'un génome.

Les gènes participant au développement du xénope sont identifiés par séquençage partiel d'ADNc (obtention d'EST). Il s'agira de participer à un effort international. Le but sera d'obtenir un jeu de clones d'ADNc séquencés partiellement représentant 10 000 gènes.

### *Construction de banque d'ADNc*

Le préliminaire indispensable au projet de séquençage partiel d'ADNc est la construction de banques d'ADNc. Nous avons choisi de construire deux banques :

- l'une à partir de cerveau et de moelle épinière de têtards pré et post-métamorphiques a été construite par Laurent Coen dans le laboratoire de B.

Demeneix au Muséum National d'Histoire Naturelle (banque Xtbs).

- L'autre à partir de têtes et de rétines de têtards au stade du bourgeon caudal (st. 25 à 35 selon Nieuwkoop et Faber) que j'ai construite (banque Xthr).

Nous avons choisi de construire ces banques à l'aide de la technologie SMART de Clontech. Celle-ci permet en théorie d'enrichir la banque en clones d'ADNc complets (définis ici comme ceux recouvrant de la coiffe 5' à la queue polyA en 3' des transcrits). Nous avons opté pour un vecteur plasmidique pSPORT6 que nous avons modifié par introduction de sites *SfiI* asymétriques afin qu'il soit compatible avec la technologie SMART.

La banque que j'ai construite, dénommée Xthr, contient  $7.10^6$  clones dont la taille moyenne est de 1060 bp (sur un échantillon de 53 clones) et la proportion de clones sans inserts est inférieure à 1 %.

Il faut remarquer qu'il existe encore peu de banques d'ADNc spécifiques de *X. tropicalis* et que celles que nous avons construites sont uniques à notre connaissance.

### *Séquençage initial*

Ces deux banques ont été données au Génoscope en février 2001. Dans un premier temps, 116 736 clones ont été repiqués en plaques de microtitration (58368 de chaque banque), puis une série de filtres à haute densité de colonies ont été produits permettant le criblage par hybridation. Dans un deuxième temps, le Génoscope a obtenu au moins une étiquette de séquence pour 1989 clones de la banque Xthr et 1694 de la banque Xtbs. Cette approche a été retenue pour s'assurer de la qualité des banques avant séquençage systématique, et employer une approche d'hybridation soustractive pour normaliser la population de clones échantillonnés par séquençage.

En parallèle, j'ai pu installer différents outils de bioinformatique nécessaires à l'exploitation des séquences. L'analyse des premières séquences a indiqué que les deux banques ne présentent pas de défauts rédhibitoires pour le séquençage systématique. Tout au long du projet, les séquences reçues du Génoscope ont été analysées et rendues disponibles aux différentes équipes via un serveur web qui permet les recherches par blast. Les clones sont physiquement disponibles au laboratoire et distribués à la communauté via le Centre de Ressources (cf ci-après).

### *Normalisation par hybridation d'oligonucléotides (Raphaël Thuret, stage de DEA Approches Structurales, Fonctionnelles et Evolutives des Génomes, Université Paris 11)*

Une sélection des séquences correspondant aux 50 espèces de clones les plus abondants dans l'ensemble des deux banques a été utilisée pour définir un jeu d'oligonucléotides (35 mères). Après marquage, ceux-ci ont été hybridés sur les filtres à haute densité de colonies. En deux hybridations utilisant chacune un mélange de 25 oligonucléotides marqués, nous avons identifié 22561 clones positifs soit 20%. Une seconde étape de séquençage a été entreprise sur les clones restants, à concurrence de 30000 clones au total. La liste des clones devant être réordonnés puis séquencés a été transmise au Génoscope fin mai 2002 afin que le projet de séquençage atteigne l'objectif fixé.

### *Analyse des séquences (Ana Carolina Fierro, stage de DEA Approches Mathématiques et Informatiques en Biologie, Université d'Evry)*

En février puis en Août 2003, le Génoscope m'a transmis 24 024 puis 26 598 chromatogrammes correspondant aux séquences des clones des banques Xtbs et Xthr. 22 352 séquences valides issues de 12269 clones de la banque Xtbs ont été

obtenues et 23 621 séquences valides issues de 13 783 clones pour la banque Xthr (Fig. 2).

L'étude de séquences partielles d'ADNc (EST) est un exercice de bioinformatique particulièrement intéressant, dont la bonne réussite est fondamentale dans notre projet. De nombreuses solutions existent aujourd'hui pour le « clustering » des EST, mais peu de comparaisons objectives ont été faites entre les approches utilisées. Les logiciels Stackpack, phrap, cap3, TIGR\_assembler et BioFacet (de la société Gene-IT) ont été comparés au travers de différentes simulations afin de choisir le meilleur outil. Différents critères ont été établis : qualité d'assemblage, discrimination de séquences provenant de familles de gènes à différents taux de ressemblance (de 90 à 99%), cas des transcrits alternatifs. Le logiciel d'assemblage phrap (par ailleurs utilisé dans Stackpack) a été choisi car il a donné les résultats les plus satisfaisants.

Nous avons mis en place une chaîne de traitement des séquences allant du « base calling » à l'assemblage afin de n'utiliser que les séquences informatives. Ce faisant, nous avons établi une compilation des séquences répétées connues chez le xénope mais également assemblé le génome mitochondrial de *Xenopus tropicalis* (à partir de séquences du domaine public).



**Table 3. Statistiques du projet EST de *Xenopus tropicalis***

	xthr	xtbs	xthr and xtbs
Number of sequences reads obtained	30272	27602	57874
Number of clone sequences obtained	16548	14901	31449
Number of valid sequences	26440	22345	48785
Number of clones with valid sequences	15540	12266	27806
Number of clones with 5' and 3' EST	9354	11486	20840
Number of clones with 5' EST only	4831	0	4831
Number of clones with 3' EST only	1355	780	2135
Average trimmed EST length	522	546	534
Number of contigs	4327	4002	8756
Number of contigs groups	497	289	842
Number of contigs grouped	1210	649	2209
Number of unique contigs	3117	3353	6547
Number of clones in contigs	9616	9268	15642
Number of singletons	7958	4262	17018
Number of clones in singletons	5924	2998	12164
Number of putative transcripts	9538	6640	19553
Max. assembled sequence length	3028	3144	3028
Average assembled sequence length	732	782	745
Max. assembled sequence size	147	144	159
Average assembled sequence size	6	5	5
Number of contigs containing			
1	509	97	1152
2	1779	2155	3730
3	452	362	905
4-5	587	545	1159
6-10	505	476	952
11-20	283	207	479
21-30	91	90	174
31-50	80	41	118
50-100	31	26	76
>100	10	3	11

Ce traitement des séquences a été appliqué aux données publiquement accessibles.

Nous avons obtenu 90% de séquences valides ce qui indique la bonne qualité de la préparation des banques d'ADNc et de leur normalisation.

### *Analyse des séquences*

L'assemblage des 45973 séquences par phrap nous a permis de construire 6547 contigs, et 17018 séquences (provenant de 12164 clones ) restent seules. La normalisation a été efficace puisque seuls 34 contigs sont composés de plus de 100 séquences et plus de 5000 contigs sont formés de moins de 10 séquences. Le plus

grand contig (Tubuline alpha) provient de l'assemblage de 322 séquences. L'annotation de ces séquences (Fig. 3) repose sur des comparaisons systématiques avec des banques de données généralistes (swissprot, uniprot, EMBL/Genbank/DDBJ) et des banques plus spécialisées (Séquences de vertébrés, de *X. laevis*, séquences génomiques de *X. tropicalis*). Les ORF sont recherchées avec Framesearch et les séquences protéiques annotées avec InterProscan, ce qui nous permettra d'utiliser le vocabulaire contrôlé de Gene Ontology pour l'annotation. Nous avons pu observer que 46% des contigs contiennent une ORF complète (s'alignant sur toute la longueur avec une séquence protéique de SwissProt).

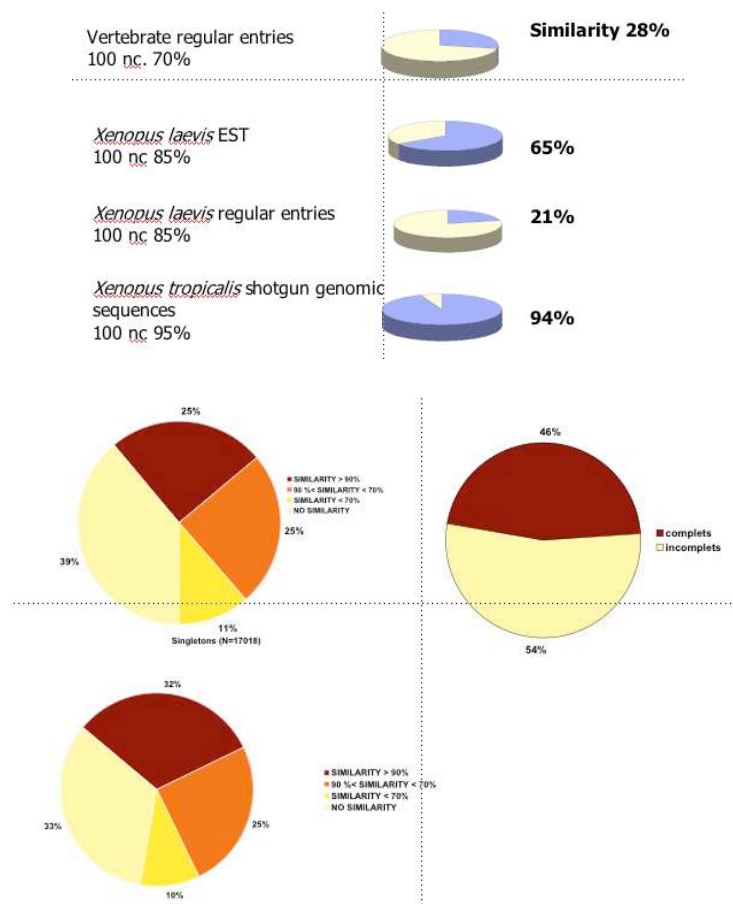


Figure 6. Annotation des séquences obtenues.

### Perspectives

Nous finissons une phase intensive d'annotation des séquences produites. Par ailleurs, nous avons combiné nos données à celles produites au Sanger Institute sur

des banques de stades plus précoces et aux autres données publiques sur *X. tropicalis* afin de générer un répertoire de gènes le plus complet possible. Les séquences ont été rendues disponibles dans les banques EMBL/GenBank/DDBJ (acc CN072222 à CN121006). Un article en cours de rédaction présentera cette ressource:

**Exploring the nervous system transcriptome in the model *Xenopus tropicalis* using EST analysis.**

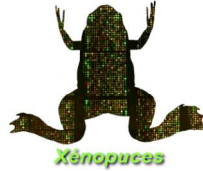
Thuret, R., Fierro, A.C., Coen, L., Perron, M., Demeneix, B., Wegnez, M., Gyapay, G., Weissenbach, J., Wincker, P., Mazabraud, A. and **Pollet, N.**

En parallèle, je collabore avec Mike Gilchrist (Head of Bioinformatics Group, The Wellcome Trust/Cancer Research UK Gurdon Institute, Cambridge) pour mettre au point une méthode originale de clustering d'EST, baptisée « Distiller », utilisée pour les séquences d'ADNc de *Xenopus tropicalis* (<http://informatics.gurdon.cam.ac.uk/online/xt-fl-db.htm>).

Une suite logique à ce projet de séquençage a été retenue par le Génoscope : il s'agit de réaliser le séquençage complet de 5000 clones d'ADNc présumés entiers, soit 10 Mb de séquence.

Le séquençage des ADNc constitue une étape initiale par laquelle nous pourrions traiter notre problématique biologique par une approche de génomique fonctionnelle. Par là, nous entendons identifier les cascades transcriptionnelles en aval des récepteurs aux hormones thyroïdiennes, et caractériser les acteurs moléculaires jouant un rôle de régulateur dans le processus de la métamorphose. Par ailleurs, l'utilisation du modèle *Xenopus tropicalis* nécessite de développer les outils de biologie moléculaire qui permettront de valoriser le séquençage génomique qui est en cours. Ce sont les raisons pour lesquelles je me suis engagé à coordonner deux projets intimement liés, l'un concernant le développement et l'utilisation de puces à ADN, et l'autre visant à produire et utiliser des ressources de cartographie génique et physique.

## ***Xénopuces : Etude globale des transcriptomes au cours du développement chez le modèle xénope à l'aide de puces à oligonucléotides (Thèse de Raphaël Thuret)***



### *Introduction*

Ce projet fédérateur regroupe au sein d'un consortium initial quatre laboratoires de la région parisienne : Laboratoire de Transgenèse et Génétique des Amphibiens, CNRS UMR 8080 Université Paris 11 ; Laboratoire de Biologie Du Développement, CNRS UMR 7622 Université Paris 6; Laboratoire Embryologie Moléculaire et Expérimentale, CNRS UMR 8080 Université Paris 11, Orsay; Laboratoire de Physiologie Générale et Comparée, UMR CNRS 8572, Muséum d'Histoire Naturelle. Ce projet a été financé par le programme Puces à ADN du CNRS.

### *Objectifs*

Nous voulons réaliser des études systématiques de l'expression des gènes à l'aide de puces à ADN en utilisant le modèle xénope afin d'approfondir nos connaissances des réseaux de régulation impliqués dans la voie de signalisation Wnt (De Li Shi), la rétino-genèse (Muriel Perron), la métamorphose et les processus d'apoptose l'accompagnant (Barbara Demeneix, Nicolas Pollet). L'approche envisagée est de produire un **répertoire ordonné de gènes utilisant la technologie des puces à oligonucléotides longs**. Ceux-ci seront choisis à partir des séquences partielles et complètes d'ADNc de xénope pour représenter 1000 à 5000 gènes. Nous envisageons de produire les puces et d'optimiser leur utilisation dans le cadre de nos thématiques spécifiques, puis de générer progressivement les données. L'utilisation de cet outil pourra être complétée par une **analyse systématique des patrons d'expression par hybridation in situ sur embryons entiers ou sectionnés**. Cette approche nous semble tout à fait pertinente tant pour favoriser l'émergence de *X. tropicalis* comme modèle que pour mieux comprendre les réseaux de régulation mis en œuvre au cours du développement.

Cette approche va certainement permettre de définir de nouveaux groupes de synexpression caractéristiques de certaines voies du métabolisme ou de certaines cascades de signalisation en relation avec le développement du système nerveux.

### *Résultats*

Nous avons conduit un projet pilote portant sur l'utilisation d'une puce à oligonucléotides longs (50 mères modifiés en 5' par un groupement NH<sub>2</sub> sur un bras en C6) avec 50 gènes représentés par 60 oligos. Pour 10 gènes, les oligonucléotides ont été conçus soit par MWG biotech soit par nous-même à l'aide du logiciel OligoArray. Nous avons pu mettre au point les conditions d'extraction d'ARN, de marquage de la sonde et d'hybridation.

En parallèle, nous avons établi une liste de 2771 gènes dont 876 sont caractérisés chez le xénope. Nous avons cherché les séquences des gènes orthologues de *X. tropicalis*. Pour ce faire nous avons utilisé toutes les séquences d'ADNc publiquement disponibles (que j'ai assemblées) ainsi que les séquences d'ADN génomique du projet de séquençage en vrac. Après plusieurs étapes de comparaisons de séquences nous avons pu identifier 1400 gènes orthologues et obtenir un fragment de séquence permettant la recherche d'oligonucléotides<sup>4</sup>. Trois autres méthodes de sélection de séquences ont été utilisées pour représenter des gènes :

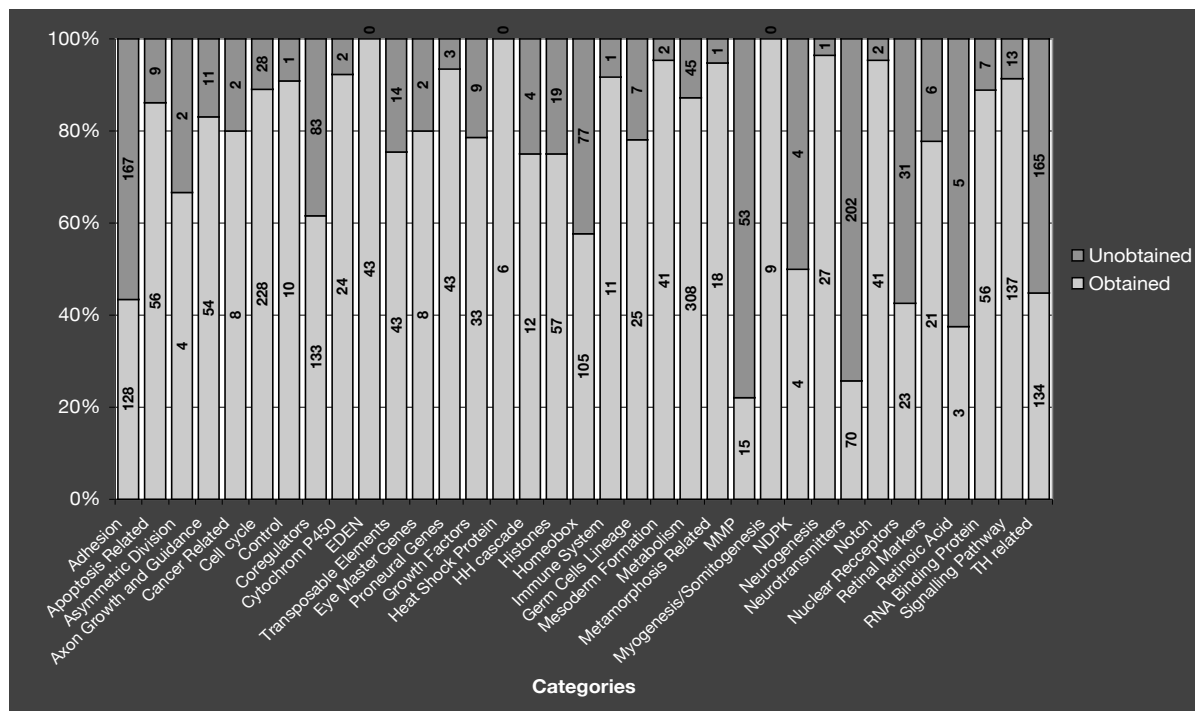
reconstruction de transcrits à partir de séquences génomiques : nous avons développé une procédure bioinformatique largement inspirée de Genotrace (Berezikov et al., 2002)

recherche systématique de facteurs de transcription (identifiés par une recherche de domaines protéiques de fixation à l'ADN)

sélection de transcrits putatifs complets

---

<sup>4</sup> Ces séquences devaient répondre à des critères stricts de qualité pour éviter de souffrir des erreurs de séquençage.

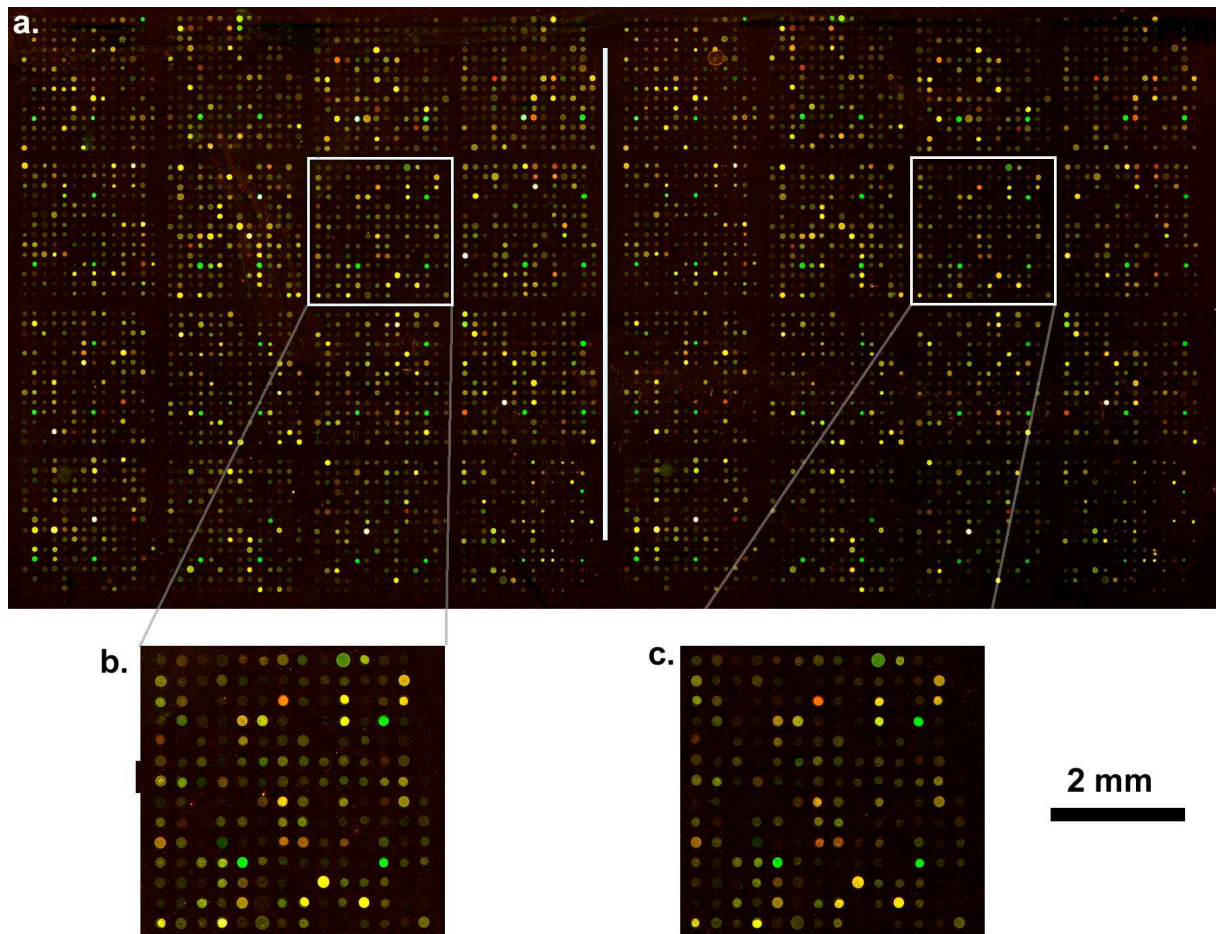


**Figure 7** : Résultats de la recherche de 2771 gènes chez *X. tropicalis*. En clair le nombre de gènes pour lesquels un fragment de séquence de transcrit a été obtenue, en foncé le nombre de cas ou aucune séquence n'a pu être retenue.

Les séquences obtenues ont été ensuite inspectées pour détecter d'éventuels problèmes de chimérisme, pour les orienter correctement vis à vis du brin codant, et éliminer toute redondance. Au final 2898 séquences ont été retenues, deux oligonucléotides devant être conçus pour 102 d'entre elles. Si l'on compare la liste des gènes obtenus à celle dressée au début de ce projet, on constate une inégalité de représentation en fonction des catégories de mots-clefs : par exemple, la quasi-totalité des gènes impliqués dans les voies de signalisation sont représentés alors que très peu de gènes codant des métalloprotéinases de la matrice ont été identifiés (Fig. 4). Ceci reflète l'état de l'art ( tel qu'il était) en ce qui concerne les séquences disponibles pour *X. tropicalis*.

Nous avons sous-traité la conception des oligonucléotides et la production de 360 lames sur lesquelles chaque oligonucléotide est déposé en double à MWG Biotech. Ces lames nous sont parvenues à la fin mars 2004, et les expériences d'hybridation sont en cours. Nous avons purifié des ARN de *X. tropicalis* aux stades ovocytes, neurula, bourgeon caudal, prémétamorphose, climax métamorphique et juvénile afin de constituer un échantillon de référence. En synthétisant des sondes à partir de cette référence, un signal quantifiable est attendu après hybridation pour un maximum d'oligonucléotides. Les échantillons qui seront comparés à cette référence sont des ARNs extraits des différents stades métamorphiques, sur des têtards entiers et sur des organes disséqués (Fig. 8 pour exemple). La qualité des ARN

totaux est inspectée sur puces Agilent, par PCR pour tester la contamination par de l'ADN génomique et par RT-PCR pour tester l'utilisation en transcription inverse. Nous utilisons les ressources de la plate-forme Puces à ADN de Gif-sur-Yvette et celles de la plate-forme de transcriptomique PICT du centre INRA de Jouy-en-Josas pour réaliser ce type d'expériences.



**Figure 8. Exemple de résultats obtenus avec une xénopuce.** Des ARN totaux extraits à partir de têtards de *X. tropicalis* au stade 60 ont été utilisés pour synthétiser une sonde marquée au Cy3 (ici en vert), et un mélange d'ARN de plusieurs stades a été utilisé pour synthétiser une sonde (la référence) au Cy5 (ici en rouge). Ces deux sondes ainsi qu'un oligonucléotide marqué servant de contrôle d'hybridation ont été hybridés entre lame et lamelle dans 35  $\mu$ l de tampon pendant 20h à 42 °C, puis scannés après plusieurs lavages à l'aide d'un scanner GenPix. L'image présentée correspond à la superposition des images obtenues à 532 et 635 nm, avec un contraste optimisé. La ligne blanche en a démarque la zone dupliquée, avec en b et c un agrandissement des deux copies d'un cadran de 14x14 spots d'oligonucléotides 35 mères.

## *Perspectives*

Il y a de nombreuses limites à l'étude du programme transcriptionnel de la métamorphose par une approche de quantification des transcrits à l'aide de puces à ADN.

Tout d'abord, la réponse aux hormones thyroïdiennes est très différente d'un type cellulaire à l'autre, et dans l'idéal il nous faudrait travailler sur des populations cellulaires homogènes. Pour pallier ce problème nous explorerons la possibilité de trier des cellules, par exemple dérivées d'animaux transgéniques exprimant la GFP sous contrôle de promoteurs tissus-spécifiques. Afin également de pouvoir analyser les données d'expression dans le contexte de la régulation par le récepteur  $\beta$  aux hormones thyroïdiennes (TR $\beta$ ) nous produisons des animaux transgéniques pour des formes dominante-négative de ce récepteur (mutants d'activation d'une part et de répression d'autre part, qui ont été développées par Laurent Sachs, UMR CNRS 8572, Muséum d'Histoire Naturelle). Comme ces mutations sont létales, nous utilisons un système conditionnel, en l'occurrence Tet-on (Das et Brown, 2004), afin d'obtenir des animaux en F1 qui pourront être utilisés pour la détermination de profils d'expression, et nous utilisons des protéines de fusion avec la GFP pour visualiser et quantifier le taux de protéine induite.

Par ailleurs, il est nécessaire de pouvoir distinguer les gènes régulés directement par les récepteurs aux hormones thyroïdiennes de ceux régulés indirectement par d'autres facteurs de transcription. Nous utiliserons des cultures de tissus (queue par exemple) permettant l'utilisation de molécules inhibant la traduction (cycloheximide) afin d'explorer les cascades en jeu. Dans un deuxième temps, nous produirons des animaux transgéniques surexprimant les facteurs de transcription identifiés comme étant modulés au cours de la métamorphose, et étudierons les conséquences en termes de profil d'expression par rapport à la situation physiologique.

Finalement, il est souhaitable de pouvoir croiser les résultats obtenus par étude d'expression avec ceux récoltés grâce à une autre approche expérimentale. Avec Laurent Sachs nous débutons une approche d'immunoprécipitation de la chromatine à l'aide d'anticorps contre TR $\beta$  qui nous permettra d'identifier des séquences utilisées comme sites de liaison au complexe transcriptionnel régulé par la T3. Ces séquences permettront via la séquence génomique d'identifier les gènes régulés par TR $\beta$ , et de confirmer ou infirmer les résultats obtenus par l'approche puces à ADN.



## **Centre de Ressources Biologiques pour la cartographie génique et physique du génome de *Xenopus tropicalis***

Ce projet a été retenu dans le cadre de l'appel d'offre « Constitution de Centre de Ressources Biologiques » par le Ministère de l'Enseignement Supérieur et de la Recherche, le CNRS et l'IFREMER. Il vise à constituer un centre de ressources pour les collections d'acides nucléiques que nous aurons développé dans le cadre de nos projets de génomique fonctionnelle.

Il va s'agir de collecter et de centraliser les ressources de banques et de clones d'ADN génomique et d'ADNc de xénope, mais également d'ADN d'espèces différentes et des ADN de constructions utilisées pour les souches transgéniques. L'accès à ces ressources par une base de données assurera l'utilité de cette collection.

a- Collection de clones de grands fragments d'ADN génomique de *X. tropicalis* clonés dans le système des chromosomes artificiels bactériens (BACs) qui seront cartographiés physiquement sur les chromosomes de *X. tropicalis*.

b- Collection de clones ADNc de transcrits du Système Nerveux de *Xenopus tropicalis* partiellement séquencés (5' et 3' ESTs), annotés, et pour lesquels il sera établi un profil d'expression au cours du développement.

c- Collection de banques d'ADNc de xénope

d- Collection des ADN des différentes espèces de xénope

e- Collection des constructions utilisées pour les souches transgéniques

Les éléments a et b de la collection seront développés au laboratoire. Les éléments c, d et e seront constitués par acquisition auprès de différents laboratoires au niveau national et international, ou préparés au laboratoire (c et d).

Le projet scientifique consiste à

- réaliser une étude systématique et à grande échelle de caractérisation de l'expression d'environ 10 000 gènes dans le système nerveux au cours de l'embryogenèse et pendant la métamorphose chez *X. tropicalis*.
- développer une ressource de clones de grands fragments d'ADN génomique dans des BACs
- initier un projet de cartographie physique systématique permettant d'ancrer des séquences génomiques aux chromosomes sur *X. tropicalis*.

### *Résultats*

Nous disposons d'une banque de BAC que nous avons fait construire à façon par Genome Express. Cette banque est la seule existante à être construite à partir d'ADN de femelles (sexe hétérogamétique) *X. tropicalis* de souche Côte d'Ivoire Adiopodoumé. Elle représente 6,9 équivalents génomes et les clones ont des inserts de 127 kpb en moyenne. Pour le criblage, nous disposons de filtres à haute densité de colonies. Nous avons établi une liste de 13 loci pour lesquels il existe au moins une équipe intéressée par l'utilisation de clones BAC pour des études de régions régulatrices. Des marqueurs de type STS (« Sequence Tagged Sites ») ont été définis pour chacun de ces loci, et ont permis de synthétiser des sondes qui seront utilisées lors de criblages par hybridation sur filtres de colonies. Au travers de ces criblages nous pourrions caractériser davantage cette banque de BAC. Ces derniers seront caractérisés par électrophorèse en champs pulsés et étudiés par transgenèse.

### ***Les séquences répétées du génome de *Xenopus tropicalis****

Alors qu'une seconde version de l'assemblage de la séquence génomique de *Xenopus tropicalis* a vu le jour début 2004, on ne connaît ni la nature ni l'importance des séquences répétées qui composent ce génome. Pourtant, il est difficilement concevable de pouvoir annoter correctement les éléments fonctionnels (promoteurs, exons) à partir de la séquence génomique sans que les séquences répétées soient annotées elles-mêmes. De même, l'assemblage des séquences partielles d'ADNc est rendu plus difficile lorsqu'on n'a pas connaissance des séquences répétées, fréquemment retrouvées dans les régions 3' non-traduites des transcrits.

Au cours de ma thèse, j'ai eu le plaisir de découvrir le monde des transposons au travers d'une collaboration avec Yves Bigot (Université de Tours) portant sur la caractérisation (pour la première fois) d'éléments mariner dans le génome humain, et dans celui d'autres mammifères.

**Human and other mammalian genomes contain transposons of the mariner family.**

Auge-Gouillou, C., Bigot, Y., **Pollet, N.**, Hamelin, M.-H., Meunier-Rotival, M. and Periquet, G. (1995)

**FEBS Letters** 368 :541-546.

Internal fragments of the putative transposase gene of mariner-like elements (MLEs) were amplified from human, mouse, rat, chinese hamster, sheep and bovine genomic DNAs by polymerase chain reaction (PCR). The sequences identified in human, ovine and bovine genomes correspond to ancient degenerate transposons. Screening mammalian sequence libraries identified a truncated element in the human ABL gene and the sequence of its 5'-ITR was determined. This ITR sequences were used in PCR experiments with DNA from six mammalian species and detected full-sized and deleted MLEs. The presence of MLE in mammalian genomes demonstrates that they are ubiquitous mobile elements found from fungi to man. This observation strongly raises the possibility that MLE could constitute tools for the modification of eucaryotic genomes.

**Features of the mammal mar1 transposons in the human, sheep, cow, and mouse genomes and implications for their evolution.**

Demattei, M. V., Auge-Gouillou, C., **Pollet, N.**, Hamelin, M. H., Meunier-Rotival, M. and Bigot, Y. (2000).

**Mamm. Genome** 11 :1111-6.

Mariner-like elements (MLE) belong to the Tc1/ mariner superfamily of class II transposons. We have analyzed the mariner related to the cecropia subfamily, and called mammal mar1, in four mammalian genomes, *Bos taurus* (Bovidae), *Homo sapiens* (Primata), *Mus musculus* (Rodentia), and *Ovis aries* (Ovidae). Three kinds of MLE sequences were found in all these species: full-length 1.3-kbp elements, shorter elements 80 bp-1.2 kbp, and single inverted terminal repeats (ITRs). All the 1.3-kbp genomic copies sequenced had an open reading frame encoding a transposase interrupted by stop codons or frame shifts. Phylogenetic analysis of the full-length elements suggested at least two distinct populations of mammal mar1 elements in each species. This was confirmed by using a statistical method that allows defining populations. Finally, the evolutionary origin of the mammal mar1 elements and the paradoxes are discussed.

C'est assez naturellement que je me suis intéressé à la nature des transposons et autres séquences répétées présents dans le génome de *Xenopus tropicalis*. Avec Ludivine Sinzelle, étudiante au thèse au laboratoire, nous avons recherché et caractérisé les famille de transposons à ADN de type „TC1-like elements“ du génome de *X. tropicalis*. Il est remarquable que ces éléments soient abondants et diversifiés en plusieurs familles chez *X. tropicalis*. Il en est de même chez *X. laevis* mais les familles ont une répartition différente, reflétant en ceci la dynamique de ces éléments au cours de la spéciation de ces xénopes. L'existence de transposons complets dont le cadre de lecture de la transposase est intact laisse présumer d'une activité de transposition dans le génome de *X. tropicalis*. Ces transposases ont été clonées pour que leur activité soit testée.

**Discovery and characterization of multiple Tc1 transposon lineages in  
*Xenopus tropicalis* genome.**

Sinzelle, L., **Pollet, N.** and Mazabraud, A.

Submitted to Gene. (voir annexe)

A l'aide de Laurène Meyniel (stage de première année d'IUP Bioinformatique à l'Université d'Evry) j'ai entrepris la compilation des séquences répétées du génome de *X. tropicalis*. De nombreuses applications à ces éléments existent, que ce soit pour la cartographie génétique et physique et les études de polymorphismes ou surtout pour l'annotation correcte du génome de *X. tropicalis*. A titre d'exemple, l'annotation du protéome de *X. tropicalis* défini par le JGI fait apparaitre 1860 protéines contenant un domaine „homeodomain-like“ ([http://supfam.mrc-lmb.cam.ac.uk/SUPERFAMILY/cgi-bin/gen\\_list.cgi?genome=xe](http://supfam.mrc-lmb.cam.ac.uk/SUPERFAMILY/cgi-bin/gen_list.cgi?genome=xe)) alors que seules 289 protéines humaines sont identifiées ainsi. L'inspection de ces résultats fait

apparaître que des transposons de la famille TLE sont annotés comme des gènes nucléaires et sont responsables de cette erreur.

Les travaux en cours visent à circonscrire et annoter les principales familles de séquences répétées et de les annoter sur la séquence génomique de *Xenopus tropicalis* avec nos collègues du JGI.

### **Caractérisation des microARN exprimés au cours de la métamorphose chez le xénope.**

Les microARNs (miRNA) sont une nouvelle classe de petits ARNs (21 à 25 nt) non codants. Ils régulent l'expression génique au niveau post-transcriptionnel en se fixant à des séquences cibles présentes dans les régions 3'UTR des ARNm (Lai, 2003).

Les gènes de miRNA peuvent être groupés en clusters contenant différents assortiments de miRNA, et ont été localisés soit dans des introns, soit dans des régions intergéniques. Les miRNA sont initialement transcrits sous forme de longs précurseurs ARN (les pri-miRNA), mono ou polycistroniques. Dans le noyau, ces pri-miRNA sont clivés en pré-miRNA par la RNAase III Drosha. Ces derniers sont transportés dans le cytoplasme pour être maturés par le complexe RNAase DICER en duplex d'ARN de 22 nt (les pre-miRNA). Seul l'un des brins participe ensuite au complexe ribonucléoprotéique RISC pour fonctionner comme régulateur post-transcriptionnel.

Quelques gènes de miRNA ont été retrouvés conservés entre les insectes et les vertébrés. Peu de processus biologiques ont été associés aux miRNAs. Parmi ceux qui ont été décrits, on trouve le contrôle de la chronologie du développement chez le nématode et de la mort cellulaire programmée chez la drosophile, et chez les végétaux, la prolifération des cellules de feuilles et la différenciation florale .

L'objectif du projet est d'identifier des gènes de miRNAs chez le xénope qui sont exprimés différemment au cours de la métamorphose, d'identifier leur(s) cible(s) et de caractériser la fonction de l'un d'entre eux. Une étude bioinformatique utilisant

ERPIN (Lambert et coll., 2003) réalisée par Mathieu Legendre et Daniel Gautheret (CNRS UMR 6207, Marseille) , avec lesquels nous collaborons, a permis de prédire plus d'une centaine de gènes miRNA à partir des séquences d'ADN génomiques de *X. tropicalis* et servira de point de départ.

Une identification expérimentale est nécessaire car il a été montré que les outils bioinformatiques seuls ne permettent pas encore d'identifier des gènes de miRNA ab initio à partir de la séquence génomique (Lai, 2003). Par ailleurs il a été montré chez la drosophile que les miRNA let-7 and miR-125 ont une expression différentielle au cours de la métamorphose qui est régulée indépendamment du signal ecdysone (Bashirullah et coll., 2003). Ces résultats permettent de poser l'hypothèse selon laquelle il existe un signal chronologique indépendant du contrôle hormonal de la métamorphose chez la drosophile. Nous souhaitons tester cette hypothèse chez le xénope.

## Références

- Adams, M. D., A. R. Kerlavage, et al. (1995). "Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence." *Nature* **377**(6547 Suppl): 3-174.
- Amaya, E. and K. L. Kroll (1999). "A method for generating transgenic frog embryos." *Methods Mol Biol* **97**: 393-414.
- Amaya, E., M. F. Offield, et al. (1998). "Frog genetics: *Xenopus tropicalis* jumps into the future." *Trends Genet* **14**(7): 253-5.
- Auffray, C., G. Behar, et al. (1995). "[IMAGE: molecular integration of the analysis of the human genome and its expression]." *C R Acad Sci III* **318**(2): 263-72.
- Bashirullah, A., A. E. Pasquinelli, et al. (2003). "Coordinate regulation of small temporal RNAs at the onset of *Drosophila* metamorphosis." *Dev Biol* **259**(1): 1-8.
- Beck, C. W. and J. M. Slack (2001). "An amphibian with ambition: a new role for *Xenopus* in the 21st century." *Genome Biol* **2**(10): REVIEWS1029.
- Berezikov, E., R. H. Plasterk, et al. (2002). "GENOTRACE: cDNA-based local GENOME assembly from TRACE archives." *Bioinformatics* **18**(10): 1396-7.
- Bisbee, C. A., M. A. Baker, et al. (1977). "Albumin phylogeny for clawed frogs (*Xenopus*)." *Science* **195**(4280): 785-7.
- Bronchain, O. J., K. O. Hartley, et al. (1999). "A gene trap approach in *Xenopus*." *Curr Biol* **9**(20): 1195-8.
- Burki, E. and M. Fischberg (1985). "Evolution of globin expression in the genus *Xenopus* (Anura: Pipidae)." *Mol Biol Evol* **2**(3): 270-7.
- Chae, J., L. B. Zimmerman, et al. (2002). "Inducible control of tissue-specific transgene expression in *Xenopus tropicalis* transgenic lines." *Mech Dev* **117**(1-2): 235-41.
- Das, B. and D. D. Brown (2004). "Controlling transgene expression to study *Xenopus laevis* metamorphosis." *Proc Natl Acad Sci U S A* **101**(14): 4839-42.
- Dib, C., S. Faure, et al. (1996). "A comprehensive genetic map of the human genome based on 5,264 microsatellites." *Nature* **380**(6570): 152-4.
- Dujon, B. (1998). "EUROFAN and the functional analysis of the *Saccharomyces* genome." *Electrophoresis* **19**: 617-624.
- Evans, B. J., D. B. Kelley, et al. (2004). "A mitochondrial DNA phylogeny of African clawed frogs: phylogeography and implications for polyploid evolution." *Mol Phylogenet Evol* **33**(1): 197-213.
- Gilchrist, M. J., A. M. Zorn, et al. (2004). "Defining a large set of full-length clones from a *Xenopus tropicalis* EST project." *Dev Biol* **271**(2): 498-516.
- Graf, J. D. and M. Fischberg (1986). "Albumin evolution in polyploid species of the genus *Xenopus*." *Biochem Genet* **24**(11-12): 821-37.
- Gyapay, G., J. Morissette, et al. (1994). "The 1993-94 Genethon human genetic linkage map." *Nat Genet* **7**(2 Spec No): 246-339.
- Hartley, K. O., S. L. Nutt, et al. (2002). "Targeted gene expression in transgenic *Xenopus* using the binary Gal4-UAS system." *Proc Natl Acad Sci U S A* **99**(3): 1377-82.
- Hedges, S. B. and S. Kumar (2002). "Genomics. Vertebrate genomes compared." *Science* **297**(5585): 1283-5.
- Hol, F. A., B. C. Hamel, et al. (1995). "Localization of Alagille syndrome to 20p11.2-p12 by linkage analysis of a three-generation family." *Hum Genet* **95**(6): 687-90.
- Houlgatte, R., R. Mariage-Samson, et al. (1995). "The Genexpress Index: a resource for gene discovery and the genic map of the human genome." *Genome Res* **5**(3): 272-304.
- Kenwrick, S., E. Amaya, et al. (2004). "Pilot morpholino screen in *Xenopus tropicalis* identifies a novel gene involved in head development." *Dev Dyn* **229**(2): 289-99.
- Klein, S. L., R. L. Strausberg, et al. (2002). "Genetic and genomic tools for *Xenopus* research: The NIH *Xenopus* initiative." *Dev Dyn* **225**(4): 384-91.
- Ko, M. S. (2001). "Embryogenomics: developmental biology meets genomics." *Trends Biotechnol* **19**(12): 511-8.
- Lahaye, K., S. Kricha, et al. (2002). "XNAP, a conserved ankyrin repeat-containing protein with a role in the Notch pathway during *Xenopus* primary neurogenesis." *Mech Dev* **110**(1-2): 113-24.
- Lai, E. C. (2003). "microRNAs: runts of the genome assert themselves." *Curr Biol* **13**(23): R925-36.

- Lambert, A., J. F. Fontaine, et al. (2004). "The ERPIN server: an interface to profile-based RNA motif identification." *Nucleic Acids Res* **32**(Web Server issue): W160-5.
- Li, L., I. D. Krantz, et al. (1997). "Alagille syndrome is caused by mutations in human Jagged1, which encodes a ligand for Notch1." *Nat Genet* **16**(3): 243-51.
- Marra, M., L. Hillier, et al. (1999). "An encyclopedia of mouse genes." *Nat Genet* **21**(2): 191-4.
- Marsh-Armstrong, N., H. Huang, et al. (1999). "Germ-line transmission of transgenes in *Xenopus laevis*." *Proc Natl Acad Sci U S A* **96**(25): 14389-93.
- Miklos, G. L. and G. M. Rubin (1996). "The role of the genome project in determining gene function: insights from model organisms." *Cell* **86**(4): 521-9.
- Nakayama, T. and R. Graninger (2003). "Generation and Characterization of Developmental Mutations in *Xenopus tropicalis*." *Current Genomics* **4**(8).
- Oda, T., A. G. Elkahloun, et al. (1997). "Identification and cloning of the human homolog (JAG1) of the rat Jagged1 gene from the Alagille syndrome critical region at 20p12." *Genomics* **43**(3): 376-9.
- Onichtchouk, D., Y. G. Chen, et al. (1999). "Silencing of TGF-beta signalling by the pseudoreceptor BAMBI." *Nature* **401**(6752): 480-5.
- Onichtchouk, D., V. Gawantka, et al. (1996). "The Xvent-2 homeobox gene is part of the BMP-4 signalling pathway controlling [correction of controlling] dorsoventral patterning of *Xenopus* mesoderm." *Development* **122**(10): 3045-53.
- Peng, J., B. Riggs, et al. (2003). "Construction of a Set of Full-Length Enriched cDNA Libraries as Genomics Tools for *Xenopus Tropicalis* Research." *Current Genomics* **4**(8): 635-644.
- Richardson, P. and J. Chapman (2003). "The *Xenopus tropicalis* genome project." *Current Genomics* **4**(8): 645-652.
- Stemple, D. L. (2004). "TILLING--a high-throughput harvest for functional genomics." *Nat Rev Genet* **5**(2): 145-50.
- Thiebaud, C. H. and M. Fischberg (1977). "DNA content in the genus *Xenopus*." *Chromosoma* **59**(3): 253-7.
- Tymowska, J. (1973). "Karyotype analysis of *Xenopus tropicalis* Gray, Pipidae." *Cytogenet Cell Genet* **12**(5): 297-304.
- Tymowska, J. and M. Fischberg (1973). "Chromosome complements of the genus *Xenopus*." *Chromosoma* **44**(3): 335-42.
- Tymowska, J. and M. Fischberg (1982). "A comparison of the karyotype, constitutive heterochromatin, and nucleolar organizer regions of the new tetraploid species *Xenopus epitropicalis* Fischberg and Picard with those of *Xenopus tropicalis* Gray (Anura, Pipidae)." *Cytogenet Cell Genet* **34**(1-2): 149-57.
- Zhang, F., J. F. Deleuze, et al. (1990). "Interstitial deletion of the short arm of chromosome 20 in arteriohepatic dysplasia (Alagille syndrome)." *J Pediatr* **116**(1): 73-7.



# CURRICULUM VITAE

**Nom :** Pollet

**Prénom:** Nicolas

**Date et lieu de naissance :** 3 septembre 1968, à Paris 16eme, France.

**Situation de famille :** Marié, deux enfants.

**Adresse personnelle :** 78, route de Limours ; 78470 Saint-Rémy-Lès-Chevreuse

**Adresse professionnelle**

Transgénèse et Génétique des Amphibiens  
CNRS UMR 8080 Développement et Evolution  
IBAIC bat 447, Université Paris-Sud  
91405 Orsay cedex

**Téléphone :** 01 69 15 72 73

**Fax :** 01 69 15 68 16

**mél :** Nicolas.Pollet@ibaic.u-psud .fr

**ouaibe :** <http://indigene.ibaic.u-psud.fr>

**Fonctions actuelles:** Chargé de Recherche 2<sup>ème</sup> classe au CNRS

**Expériences professionnelles :**

2000 – Chargé de Recherche au CNRS (Classé 1<sup>er</sup> en section 23) -  
CNRS ESA 8080 Développement et Evolution, Directeur Pr. Maurice Wegnez.

1999-2000 : "Guest scientist" au Département d'Embryologie  
Moléculaire dirigé par Christof Niehrs, DKFZ, Heidelberg.

1997-1999 : "Marie Curie Research Fellow" au Département  
d'Embryologie Moléculaire dirigé par Christof Niehrs, DKFZ, Heidelberg.

**Etudes universitaires :**

1992-1997 : Université Denis Diderot-Paris 7.

*Thèse de doctorat* en Physiologie du Développement, sous la direction de Michèle Meunier-Rotival à l'unité 347 de l'INSERM. Titre de la thèse: Cartographie physique, génétique et génique de la région p12 du chromosome 20 humain contenant le locus du syndrome d'Alagille. Allocataire de Recherche du Ministère de la Recherche et de l'Espace, et boursier de l'Association Française contre les Myopathies. Mention très honorable avec les félicitations du jury.

1991-1992 : Université Denis Diderot-Paris 7.

*DEA en Physiologie du Développement*, sous la direction de Michèle Meunier-Rotival à l'unité 347 de l'INSERM. Titre du mémoire : Cartographie physique du locus AGS en 20p12 à l'aide de chromosomes artificiels de levure. Mention très bien, classé 1/24.

1990-1991 : Université Pierre et Marie Curie-Paris 6.

*Maîtrise de biologie cellulaire*, options dynamique du développement, aspects cellulaires et moléculaires du développement.

1989-1990 : Université Pierre et Marie Curie-Paris 6.

*Licence de biologie cellulaire et physiologie.*

1986-1989 : Université Pierre et Marie Curie-Paris 6.

*DEUG B*, option biochimie-chimie, embryologie moléculaire.

1986 : Brevet de technicien agricole option générale.

## PRODUCTION SCIENTIFIQUE

### A- Revues internationales avec comité de lecture

1. Deleuze, J.-F., Dhorne, S., Hazan, J., Borghi, E., Raynaud, N., **Pollet, N.**, Meunier-Rotival, M., Deschatrette, J., Alagille, D. and Hadchouel, M. (1994) Deleted chromosome 20 from a patient with Alagille syndrome isolated in a cell hybrid through leucine transport selection : Exclusion of three candidates genes. **Mamm. Genome** 5 :663-669.
  2. **Pollet, N.**, Dhorne-Pollet, S., Deleuze, J.-F., Boccaccio, C., Driancourt C, Raynaud, N., Le Paslier, D., Hadchouel, M. and Meunier-Rotival, M. (1995) Construction of a 3.7-Mb physical map within human chromosome 20p12 ordering 18 markers in the Alagille syndrome locus. **Genomics** 27 :467-474.
  3. Auge-Gouillou, C., Bigot, Y., **Pollet, N.**, Hamelin, M.-H., Meunier-Rotival, M. and Periquet, G. (1995) Human and other mammalian genomes contain transposons of the mariner family. **FEBS Letters** 368 :541-546.
  4. **Pollet, N.**, Boccaccio, C., Dhorne-Pollet, S., Driancourt, C., Raynaud, N., Auffray, C., Hadchouel, M. et Meunier-Rotival, M. (1997) Construction of an integrated physical and gene map of human chromosome 20p12 providing candidate genes for Alagille syndrome. **Genomics** 42 :489-498.
  5. **Pollet, N.**, Gawantka, V. , Delius, H., Vingron, M., Pfister, R., Nitsch, R., Blumenstock, C. and Niehrs C. (1998) Gene expression screening in *Xenopus* identifies molecular pathways, predicts gene function and provides a global view of embryonic patterning. **Mech. Dev.** 77 :95-141.
  6. Crosnier, C., Driancourt., C., Raynaud, N., Dhorne-Pollet, N., **Pollet, N.**, Bernard, O., Hadchouel, M. et Meunier-Rotival, M. (1999) Mutations in JAGGED1 gene are predominantly sporadic in Alagille syndrome. **Gastroenterology**. 116 :1141-1148.
  7. Jen, W.C., Gawantka, V., **Pollet, N.**, Niehrs, C. and Kintner, C. (1999). Periodic repression of Notch pathway genes governs the segmentation of *Xenopus* embryos. **Genes Dev.** 13 :1486-1499.
  8. Niehrs, C. and **Pollet, N.** (1999) Synexpression groups in eukaryotes. **Nature** 402 :483-487.
- Pollet, N.**, Schmidt, H.A., Gawantka, V., Vingron, M. and Niehrs, C. (2000) Axeldb: a *Xenopus laevis* database focusing on gene expression. **Nucleic Acids Res.** 28 :139-140.
9. Chen, Y., **Pollet, N.**, Niehrs, C. and Pieler, T. (2001). Increased XRALDH2 activity has a posteriorizing effect on the central nervous system of *Xenopus* embryos. **Mech. Dev.** 101 :91-103.
  10. Lamar, E., Deblandre, G., Wettstein, D., Gawantka, V., **Pollet, N.**, Niehrs, C. and Kintner, C. (2001). Nrarp is a novel intracellular component of the Notch signaling pathway. **Genes Dev.** 15 :1885-99.
  11. Duncan, P. I., **Pollet, N.**, Niehrs, C. and Nigg, E. A. (2001) Cloning and characterization of Plx2 and Plx3, two additional Polo-like kinases from *Xenopus laevis*. **Exp Cell Res.** 270:78-87.
  12. Li, Y, Fenger, U., Niehrs, C. and **Pollet, N.** (2003) Cyclic expression of *esr9* gene in *Xenopus* presomitic mesoderm. **Differentiation**. 71 :83-89.
  13. Koenig, R., Baldessari, D., **Pollet, N.**, Niehrs, C. and Eils, R. Reliability of gene expression ratios for cDNA microarrays in multiconditional experiments with a reference design. (2004) **Nucleic Acids Res.** 32 :e29.
  14. Böttcher, R., **Pollet, N.**, Delius, H. and Niehrs, C. The transmembrane protein XFLRT3 forms a complex with FGF receptors and promotes FGF signalling. (2004) **Nat Cell Biol** 6 :38-44.
- ### B- Article de Livre
1. Pollet, N. and Niehrs, C. (2001). Expression profiling by systematic high-throughput in situ hybridization to whole-mount embryos. **Methods Mol. Biol.** 175 :309-21.
  2. Pollet, N. Human Chromosome 20. (2002) **In Nature Encyclopedia of the Human Genome.**

### C- Revues nationales

1. Deleuze, J.-F., Dhorne-Pollet, S., **Pollet, N.**, Meunier-Rotival, M. et Hadchouel, M. (1995) Le syndrome d'Alagille en 1995. **Gastroenterol. Clin. Biol.** 19 :587-596.

### D- Conférences internationales avec parution des actes et comité de lecture

1. **Pollet, N.**, Dhorne-Pollet, S., Deleuze, J.-F., Driancourt, C., Raynaud, N., Hadchouel, M. and Meunier-Rotival, M. (1995) Physical mapping of the Alagille syndrome locus on chromosome 20p12. **Med. Genetik** (27th annual meeting of the European Society of Human Genetics, Berlin) 2, G-101.
2. **Pollet, N.**, Dhorne-Pollet, S., Boccaccio, C., Deleuze J-F, Driancourt, C., Raynaud, N., Hadchouel, M. and Meunier-Rotival, M. (1996) Construction of a transcription map of human chromosome 20p12. **Eur. J. Hum. Genet.** (Abstracts of the 28th Annual Meeting of the European Society of Human Genetics, London) 4, S8.005.
3. **Pollet, N.**, Gawantka, V., Delius, H. and Niehrs, C. (2000) Large scale expression screening identifies molecular pathways and predicts gene function. "**Genomics and Proteomics**". Ed. Sandor Suhai, Kluwer Academic/ Plenum Publishers, New-York.
4. **Pollet, N.**, Schmidt, H.A., Gawantka, V., Niehrs, C and Vingron, M. (2000) *In silico* analysis of gene expression patterns during early development of *Xenopus laevis*. **Pac. Symp. Biocomput.** 443-454
5. **Pollet, N.**, Delius, H. and Niehrs, C. Analyse *In Situ* de l'expression des gènes dans les embryons de Xénope. (2003) **Comptes Rendus Biologies.** 326:1011-7.

### E- Conférences sur invitation dans des congrès

*Integration of mapping data for chromosome 20p12.* Second chromosome 20 workshop. The Sanger Center - Wellcome Trust Genome Campus, Hinxton, Cambs, United Kingdom, 26-28 February 1997.

*Large scale expression screening identifies molecular pathways and predicts gene function.* Human Genome Mapping 98, Turin, Italy, 29 mars 1998.

*Large scale gene expression screening identifies molecular pathways and predicts gene function.* International symposium on Genomics and Proteomics, Functional and computational aspects. DKFZ Heidelberg, Germany, 4-7 octobre 1998.

*The role of chromosome 20 editors today.* Third chromosome 20 workshop. The Sanger Center - Wellcome Trust Genome Campus, Hinxton, Cambs, United Kingdom, 15-16 April 1999.

*Axeldb: A Xenopus laevis database.* Workshop on "Large scale embryonic gene expression", DKFZ, Heidelberg, Germany, 16-17 Novembre 1999. Keynote speaker.

*Axeldb: A Xenopus laevis database.* 8th International Xenopus Conference., Estes Park, Colorado, USA, 16-20 August 2000.

*Analysis of gene expression in Xenopus embryos identifies molecular pathways, predicts gene function and provides a global view of embryonic patterning.* Transcriptome 2000, Institut Pasteur, France, 6-9 Novembre 2000. Keynote speaker.

*Axeldb: A Xenopus laevis database focussing on gene expression.* Genome-Wide Approaches for Developmental Biology using *Xenopus*". National Institute for Basic Biology, Okazaki, Japan, 23 Février 2002. Keynote speaker.

*Cyclic expression of esr9 and esr10 genes in Xenopus presomitic mesoderm.* HFSP workshop on Molecular analysis of segmentation mechanisms in vertebrates and arthropods, Cancer Research UK London Research Institute: Lincoln's Inn Fields Laboratories, Londres, United Kingdom, 13-14 Mai 2002.

*Transcriptome studies on the nervous system of Xenopus tropicalis. Xenopus Genomics and Genetics Meeting.* 9th international Xenopus conference, Homerton College, Cambridge, United Kingdom, 21-25 August 2002. Keynote speaker.

*From the functional genomic data to the dynamical analysis of molecular networks.* INSERM Atelier de Formation n°138, La Londe-les-Maures, France, 11-13 septembre 2002. Keynote speaker.

*Expression profiling by systematic in situ hybridization : a spatial look of transcriptomes.* Imaging, Modelling, Manipulating transcriptional regulatory networks. Ambleteuse, 17-22 Octobre 2002, France. Invited speaker.

*A transcriptomic study of metamorphosis in Xenopus tropicalis.* DKFZ Symposium : 10 years of Molecular Embryology at DKFZ. 21 mai 2004, Heidelberg, Germany. Invited speaker.

## **F- Communications orales à des congrès, séminaires invités**

*Construction d'une carte physique localisée en 20p12 contenant le locus associé au syndrome d'Alagille.* Première réunion plénière annuelle du club de clonage positionnel AFM/GREG « le clonage positionnel en 1994-1995 », Gif sur Yvette, 23-24 février 1995.

*Construction of a transcription map of human chromosome 20p12.* Laboratory of Molecular Embryology of DKFZ, Heidelberg, 8 février 1996.

*Construction d'une carte de transcription de la région 20p12.* Centre de Recherches Thérapeutiques en Ophtalmologie, Laboratoire d'Embryologie Humaine, Faculté Necker-Enfants Malades, Paris, 17 avril 1996.

*Clonage positionnel du syndrome d'Alagille.* INSERM U383, Hôpital Necker-Enfants Malades, Paris, 7 mai 1996.

*Criblage d'Internet dans un projet de clonage positionnel.* Laboratoire de Pathologie Moléculaire et Thérapie Génique. Université de Bordeaux II, Bordeaux, 6 septembre 1996.

*Criblage d'après les profils d'expression génique chez Xenopus laevis: identification des voies de signalisation, prédiction de la fonction des gènes et régionalisation de l'embryon.* Séminaire à l'Institut Pasteur, Paris, 7 mai 1998.

*Criblage d'après les profils d'expression génique chez Xenopus laevis: identification des voies de signalisation, prédiction de la fonction des gènes et régionalisation de l'embryon.* Institut d'Embryologie Cellulaire et Moléculaire du CNRS, Nogent-sur-Marne, 13 mai 1998.

*Gene expression screening in Xenopus identifies molecular pathways, predicts gene function and provides a global view of embryonic patterning.* Heidelberg Developmental Biology Club, EMBL, Heidelberg, 6 avril 1998.

*Analyse systématique des profils d'expression génique au cours du développement embryonnaire précoce chez Xenopus laevis.* Réunion du Club amphibien. Montpellier, 15-18 décembre 1999.

*Analyse systématique des profils d'expression génique au cours du développement embryonnaire précoce chez Xenopus laevis.* CNRS ERS 1984 - Génétique moléculaire et biologie du développement, Villejuif, 21 décembre 1999.

*Large scale expression screening identifies molecular pathways and predicts gene function.* BASF/DKFZ workshop. DKFZ, Heidelberg, 29 novembre 1999.

*Analyse in situ des transcriptomes de l'embryon de xénope.* Réunion du groupe « Bioinformatique fonctionnelle des systèmes de régulation génétiques » dans le cadre de l'action IMPG (Informatique, Mathématique, Physique pour la Génomique). ESIL Marseille-Luminy, 30-31 Mars 2001.

*Outils d'analyse des EST de xénope.* Réunion du Club Amphibien. Dourdan, 20-22 Juin 2001.

*Coordinate gene expression in cell differentiation, development and evolution.* Laboratoire de Physiologie Générale et Comparée du Muséum National d'Histoire Naturelle, CNRS UMR 8572, Paris, 14 Février 2002.

*La coordination de l'expression des gènes au cours de la différenciation cellulaire, du développement et de l'évolution.* Institut Jacques Monod, Paris, 27 Mai 2002.

*Cyclic expression of esr9 and esr10 genes in Xenopus presomitic mesoderm.* Laboratoire Régulation post-transcriptionnelle et développement CNRS UMR6061/Université de Rennes 1, Rennes, 31 Janvier 2003.

*Transcriptome studies on the nervous system of Xenopus tropicalis.* Laboratorio de Bioinformática y Expresión Génica, INTA, Santiago, Chili, 7 Novembre 2003.

*Transcriptome studies: from cDNA sequencing to DNA chips.* Laboratorio de Bioinformática y Matemática del Genoma, Centro de Modelamiento Matemático, UMR 2071 UCHILE -CNRS, Universidad de Chile, Santiago, Chili, 14 Novembre 2003.

*Caractérisation des transcriptomes du système nerveux chez Xenopus tropicalis par séquençage*

d'ADNc. Génoscope, 18 février 2004.

*Caractérisation des transcriptomes du système nerveux chez Xenopus tropicalis par séquençage d'ADNc.* Seconde réunion Xénopuces, Orsay 25 février 2004.

*Transcriptome studies on the nervous system of Xenopus tropicalis.* Réunion du projet européen « Jumpy », Orsay, 27-28 février 2004.

*Transcriptome studies on the nervous system of Xenopus tropicalis* Réunion Xenopus genomics. Wellcome Trust/ Cancer Research UK Gurdon Institute Cambridge, 24 Mars 2004.

*Etude du transcriptome du système nerveux chez Xenopus tropicalis par séquençage partiel d'ADNc.* Réunion du Club Amphibien. France, Orsay 6-8 Mai 2004.

## F- Enseignements

*La coordination de l'expression des gènes au cours de la différenciation cellulaire, du développement et de l'évolution.* Module de Génétique *in vitro*, Maîtrise de Génétique. 2 Mai 2002. Université Paris 6.

*Le modèle Xénope en génomique.* DEA Approches Structurales, Fonctionnelles et Comparatives des Génomes. 11 Octobre 2001. Université Paris 11.

*Etude systématique de l'expression des gènes au cours du développement.* Maîtrise des sciences biologiques et médicales, certificat « Mécanismes cellulaires et moléculaires du développement ». 16 Novembre 2001. Université Paris 11.

*Mise en évidence des groupes de synexpression au cours du développement.* DEA Biologie du Développement. 30 Novembre 2001. Université Paris 11.

*Mise en évidence des groupes de synexpression au cours du développement.* DEA Génétique. 30 Novembre 2001. Université Paris 11.

*Le modèle Xénope en génomique.* DEA Approches Structurales, Fonctionnelles et Comparatives des Génomes. 14 Octobre 2002. Université Paris 11.

*Etude systématique des patrons d'expression par hybridation in situ : vers une caractérisation spatiale des transcriptomes.* 25 octobre 2002. INSA Lyon. Filière « Bioinformatique et modélisation ».

*Etude systématique des patrons d'expression par hybridation in situ : vers une caractérisation spatiale des transcriptomes.* 28 octobre 2002. DEA « Génomes et protéines » Ecole Normale Supérieure de Paris.

*Le modèle Xénope en génomique.* DEA Approches Structurales, Fonctionnelles et Comparatives des Génomes. 2003. Université Paris 11.

*Mise en évidence des groupes de synexpression au cours du développement.* DEA Biologie du Développement. 2003. Université Paris 11.

*Etude systématique de l'expression des gènes au cours du développement.* Maîtrise des sciences biologiques et médicales, certificat « Mécanismes cellulaires et moléculaires du développement ». 2003

## Responsabilités dans l'obtention de financements français et européens pour le laboratoire

### **Génoscope – 2000 - *Xenopus tropicalis* - Analyse de banques de cDNA pré et postmétamorphiques du système nerveux.**

J'ai participé à la rédaction de cette demande de projet au Génoscope

### **Xénopuces : Etude globale des transcriptomes au cours du développement chez le modèle xénope à l'aide de puces oligonucléotides. Action Puces à ADN cu CNRS , axe nouvelles applications. 85.000 €.**

J'ai obtenu ce financement, qui profite à une communauté d'équipes : Laboratoire de Biologie Du Développement, CNRS UMR 7622 Université Paris 6, Paris, De-Li Shi ; Laboratoire Embryologie Moléculaire et Expérimentale, UMR 8080 Université Paris 11, Orsay, Maurice Wegnez, Muriel Perron ;

Laboratoire de Transgène et Génétique des Amphibiens, UMR 8080 Université Paris 11, Orsay, André Mazabraud, David Du Pasquier, Raphaël Thuret, Nicolas Pollet; Laboratoire de Physiologie Générale et Comparée, UMR CNRS 8572, Muséum d'Histoire Naturelle, Barbara Demeneix, Laurent Coen

#### **Banque de BAC – CNRS. 48.000 €**

J'ai obtenu pour le laboratoire un financement exceptionnel au département des Sciences de la Vie du CNRS afin de construire une banque d'ADN génomique de *X. tropicalis* dans le système BAC.

#### **Centre de Ressources Biologiques – CNRS – Ministère de la Recherche – IFREMER 120.000 € pour 3 ans. (2002)**

J'ai participé activement à l'élaboration de la demande d'un financement avec pour objectif principal de mettre en place des ressources de génomique pour le xénope (*X. tropicalis* mais également *X. laevis*).

#### **Génoscope – 2003 - Séquençage de clones d'ADNc complets de *Xenopus tropicalis*.**

J'ai obtenu ce projet de séquençage de 1000 clones d'ADNc présumés entiers.

#### **Amphishbird – Projet Intégré – 6ème PCRD**

Même si ce projet n'a finalement pas été retenu pour financement, j'ai participé avec André Mazabraud à cette aventure (!) que constitue le montage et la rédaction d'un projet intégré européen réunissant pas moins de 31 équipes.

#### **X-Omics Action de coordination - 6ème PCRD - "*Xenopus* genomics : coordinating integrated and comparative functional genomics for understanding normal and pathologic development". (2005) 800.000 € pour 3 ans et 11 laboratoires**

J'ai participé avec André Mazabraud à la rédaction de ce projet qui est une émanation du précédent. L'intention est de coordonner les travaux de 11 laboratoires européens pour regrouper les données de génomique fonctionnelle du modèle amphibien (gènes, orthologie avec ceux des autres modèles, expression au cours du développement, données fonctionnelles) dans une base de données "interactive" avec celles de la souris et de l'humain.

#### **Encadrement**

Depuis mon recrutement, j'ai encadré quatre étudiants, dont deux de troisième cycle.

Jennifer Carrière a passé un DEUST à l'Université Paris-Sud et effectué des stages de 6 semaines au laboratoire en alternance avec les cours (11 mois sur 2 ans). Elle a appris les techniques de biologie moléculaire et l'hybridation *in situ* dans le cadre de l'étude de l'expression de gènes anti et pro-apoptotiques au cours de la métamorphose du xénope. Elle travaille en tant que technicienne contractuelle dans une unité INSERM.

La seconde étudiante, Sandrine Franceschi, a effectué son stage dans le cadre de sa seconde année de DEUG à l'Université Paris-Sud. Pendant 8 semaines, elle a travaillé pour mettre au point et optimiser un protocole d'hybridation *in situ* sur les embryons de *X. tropicalis*.

Raphaël Thuret a effectué son DEA « Approches Structurales, Fonctionnelles et Comparatives des Génomes » sous ma responsabilité sur le développement d'outils d'étude du transcriptome du xénope au cours de l'embryogenèse et de la métamorphose. Il dispose d'une bourse du Ministère pour réaliser sa thèse sur le programme transcriptionnel de la métamorphose au laboratoire, sous ma direction.

Ana Carolina Fierro est une stagiaire chilienne du DEA Approches Mathématiques et Informatiques en Biologie de l'Université d'Evry, son sujet a porté sur l'annotation des séquences d'ADNc de *Xenopus tropicalis*.

Qods Ymlahi Ouazzani est une étudiante marocaine que j'encadre pour sa thèse au laboratoire sur la cartographie physique et génique du génome de *Xenopus tropicalis*.

Laurène Meyniel a fait un stage d'un mois au laboratoire dans le cadre de sa formation à l'IUP de bioinformatique de l'Université d'Evry.

### **Autres activités**

En ce qui concerne mes activités liées à l'enseignement, je suis intervenu dans trois DEA : « Approches Structurales, Fonctionnelles et Comparatives des Génomes », « Génomes et protéines » et « Biologie Cellulaire et Moléculaire du Développement ».

J'ai également enseigné au module « Mécanismes cellulaires et moléculaires du Développement » de la Maîtrise de Sciences Biologiques et Médicales (Université Paris Sud) et dans la filière « Bioinformatique et modélisation » de l'INSA Lyon.

Au sein de l'unité et en ce qui concerne les tâches collectives, j'ai suivi la formation d'ACMO et endossé ce rôle. Je suis le webmestre du site <http://indigene.ibaic.u-psud.fr>, du site d'axelddb (<http://www.dkfz.de/abt0135/axelddb.htm>).

Au sein de la communauté internationale des « xénopistes » je fait partie du « *Xenopus tropicalis* Genome Project Advisory Board ».

J'ai été referee pour les revus Genomics, Mutation Research, PLOS Biology, Mechanisms of Development, Nucleic Acids Research. Au niveau national j'ai été membre de jury de recrutement d'ingénieur de recherche au CNRS, rapporteur pour le recrutement d'un Maître de Conférence au Museum National d'Histoire Naturelle, et rapporteur pour un projet ACI IMPBio .

## Annexes