



HAL
open science

**Analyse comparative des génomes d'espèces majeures
pour l'aquaculture par cartographie RH et Identification
des répertoires des récepteurs olfactifs (OR) et TAAR
des cichlides**

Naoual Azzouzi

► **To cite this version:**

Naoual Azzouzi. Analyse comparative des génomes d'espèces majeures pour l'aquaculture par cartographie RH et Identification des répertoires des récepteurs olfactifs (OR) et TAAR des cichlides. Sciences agricoles. Université de Rennes, 2013. Français. NNT : 2013REN1S152 . tel-00989008

HAL Id: tel-00989008

<https://theses.hal.science/tel-00989008>

Submitted on 9 May 2014

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.



THÈSE / UNIVERSITÉ DE RENNES 1
sous le sceau de l'Université Européenne de Bretagne

pour le grade de
DOCTEUR DE L'UNIVERSITÉ DE RENNES 1

Mention : BIOLOGIE

Ecole doctorale VAS

présentée par

NAOUAL AZZOUZI

Préparée à l'unité de recherche IGDR, UMR6290 CNRS
Institut de Génétique et Développement de Rennes
Composante universitaire Sciences de la vie et de l'Environnement

**Analyse Comparative des
génomés d'espèces majeures
pour l'aquaculture par
cartographie RH et
Identification des répertoires
des récepteurs olfactifs
(OR) et TAAR des cichlides**

**Thèse soutenue à Rennes
le 13 Decembre 2013**

devant le jury composé de :

Jean-Nicolas Volff

Professeur Institut de Génomique Fonctionnelle
de Lyon

Ecole Normale Supérieure de Lyon / *rapporteur*

Pascal Favrel

Professeur Université de Caen Basse-Normandie
IBFA, IFR ICORE 146

CNRS INEE FRE 3484 / *rapporteur*

Patrick Prunet

Professeur Université de Rennes1
Laboratoire de physiologie et

Génomique des poissons / *examineur*

Sylvie Mazan

Directeur de recherche au CNRS
Station Biologique de Roscoff

Développement et Évolution
des vertébrés, UMR 7150

/ *examineur*

Francis Galibert

Professeur Université de Rennes 1

IGDR UMR 6290 / *directeur de thèse*

Remerciements

Je tiens dans un premier temps à remercier le Professeur Francis Galibert pour m'avoir confié ce travail de recherches, ainsi que pour son aide et ses précieux conseils au cours de ces années de thèse. Je le remercie pour sa confiance, pour toute l'attention qu'il m'a porté et pour sa disponibilité toutes ces années de thèse.

Je tiens à remercier tous les membres de mon jury de thèse de me faire l'honneur d'assister à ma soutenance.

Je remercie les Professeurs Jean-Nicolas Volff et Pascal Favrel, d'avoir accepté d'être les rapporteurs de ce travail.

Je remercie Madame Sylvie Mazan et le Professeur Patrick Prunet d'avoir accepté de participer à ce jury.

Ce travail n'aurait pu aboutir sans l'aide de nombreuses personnes. Que me pardonnent celles que j'oublie ici, mais j'adresse une pensée particulière à l'ensemble des personnes de l'IGDR UMR 6290 ainsi que Stéphane et Géraldine

J'ai pu travailler dans un cadre particulièrement agréable, grâce à l'ensemble des membres de l'équipe « génétique de chien ». Merci à tous pour votre bonne humeur, pour l'ambiance, et pour toutes les discussions.

Ces remerciements ne seraient pas complets sans une pensée pour Richard et Michaëlle. Merci de m'avoir aidé et encouragé lors de mon arrivée au laboratoire.

Un immense merci à Abdel, Mouhamed, Krisnen, Sylvie et Pascale pour leur lecture de mon manuscrit.

Mes dernières pensées iront vers ma famille, et surtout ma mère et mes deux enfants Abdou et Mouad.

Enfin, j'aimerais adresser un remerciement particulier à mon mari pour son aide et son soutien tout au long de ces années.

À la mémoire de mon Père

RESUMÉ :

La construction de cartes de génomes consiste à baliser les chromosomes par des repères : les marqueurs. Plus une carte est dense en marqueurs régulièrement espacés, plus elle est informative et donc plus les applications ultérieures sont nombreuses. Parmi les différentes stratégies de cartographie, celle exploitant les hybrides d'irradiation dite carte RH, présente de nombreux avantages. Ainsi, des marqueurs polymorphes tels que les microsatellites, utiles pour les analyses de liaisons génétiques et des marqueurs de gènes, polymorphes ou non polymorphes permettant d'établir des cartes comparées avec d'autres génomes et définir des zones de conservation ou de rupture de synténie, peuvent être localisés sur une carte RH. Ces cartes comparées sont utiles non seulement pour l'identification de gènes d'intérêts mais également pour l'étude de l'évolution des génomes. Parmi les nombreux vertébrés d'intérêt, nous nous sommes particulièrement attachés à la construction de cartes RH de poissons et de cichlidés en particulier. Ceux-ci constituent en effet un modèle génétique très intéressant du point de vue économique et évolutif. La réalisation de cartes des génomes de plusieurs poissons devrait aider à l'identification de gènes impliqués dans des traits phénotypiques ou pathologiques voire même des marqueurs liés aux stress et à la reproduction. Mon travail de thèse a consisté à construire une carte du génome du bar, un panel RH de tilapia et la carte RH attenante qui a été utilisée dans la phase finale de l'assemblage des données de séquence du génome de Tilapia. Par ailleurs nous avons construit un panel RH d'esturgeon et un autre d'huitre.

La cartographie du génome de tilapia, réalisée dans le cadre d'un consortium international, nous a donné un accès privilégié aux données de séquences des génomes de cinq cichlidés (*O. niloticus*, *P. nyererei*, *H. burtoni*, *N. brichardi* et *M. zebra*) et nous a permis de participer à l'annotation de ces séquences génomiques en nous intéressant plus particulièrement à l'identification des répertoires des gènes codant pour les récepteurs olfactifs (OR) et les récepteurs connus sous le vocable de TAAR pour 'Trace Amine-Associated Receptors'. C'est ainsi que nous avons identifié et caractérisé 158, 88, 90, 69, 102 gènes OR et 45, 19, 23, 12, 20 gènes TAAR dans les génomes de ces cinq poissons (*O. niloticus*, *P. nyererei*, *H. burtoni*, *N. brichardi* et *M. zebra*)

Abstract :

The construction of genome maps consists in placing tags, called markers, on the chromosomes. The denser in evenly spaced markers is a map, the more it is informative, leading to the development of more future applications. Among the different mapping strategies, those using radiation hybrids (RH) have numerous advantages. Indeed, polymorphic markers such as microsatellites, useful for linkage analyses, as well as gene markers, polymorphic or not and allowing comparative mapping with other genomes and definition of synteny breaks and conservation, can be localized on a RH map. Not only these maps are useful to identify genes of interest but they are also essential tools to study genome evolution. Among numerous vertebrates of interest, we constructed RH maps of fishes and cichlids in particular. Indeed these fishes constitute interesting genetic models from economical and evolution points of view. Having genome maps of several of these fishes would help to identify genes implicated in phenotypical or pathological traits, or even markers linked to stress and reproduction. My thesis work consisted in the construction of a genome map of the sea bass, a tilapia RH panel and its RH map, a sturgeon RH panel as well as an oyster RH panel. The tilapia map was used for the assembly of the sequencing data of the tilapia genome. Thanks to this last work realized with an international consortium we had a privileged access to the sequencing data of five cichlid genomes (*O. niloticus*, *P. nyererei*, *H. burtoni*, *N. brichardi* and *M. zebra*). We then participated to the annotation of these genomic sequences. In particular we have identified and characterized the olfactory receptor gene (OR) repertoires and of the trace amine-associated receptor (TAAR) gene repertoires of these five cichlids. Which are then made of 158, 88, 90, 69, 102 OR genes and 45, 19, 23, 12, 20 TAAR genes respectively.

Liste des Figures

Figure 1 :	<i>Arbre phylogénique des poissons en fonction du temps en millions d'années sur l'axe des abscisses (d'après J-N Volff, 2005).</i>	12
Figure 2 :	<i>Arbre phylogénique des poissons appartenant à la superclasse des Gnathostomes.</i>	16
Figure 3:	<i>Organisation du complexe Hox de drosophile et des 4 complexes (a-d) des mammifères montrant l'impact des cis et trans duplications, suivies de la perte éventuelle de certains gènes absent de l'un ou de l'autre cluster.</i>	19
Figure 4 :	<i>Organisation comparée des complexes hox de divers poissons, mammifère et xenope.</i>	20
Figure 5:	<i>Différentes formes de ploïdie retrouvées chez les différentes espèces d'esturgeons</i>	21
Figure 6 :	<i>Les différents niveaux de résolution des cartes et les techniques associées.</i>	39
Figure 7:	<i>Caryotype humain</i>	40
Figure 8:	<i>Hybridation in situ fluorescente des chromosomes de Tilapia</i>	41
Figure 9 :	<i>Principe et Résultats de cartographie RH</i>	43
Figure 10:	<i>Représentation graphique d'une carte de synténie de trois espèces (homme/souris/chien) réalisé avec le programme AutoGRAPH</i>	45
Figure 11 :	<i>Encéphale d'un poisson</i>	48
Figure 12 :	<i>Schéma montrant l'évolution du système olfactif centrale chez les vertébrés</i>	49
Figure 13:	<i>Arbre de vie et apparition des récepteurs olfactifs.</i>	50
Figure 14 :	<i>Evolution du nombre de récepteurs olfactifs chez les mammifères</i>	51
Figure 15 :	<i>Classification des OR de mammifères (homme, rat et souris) et de poissons (Zebrafish et Tetraodon).</i>	54
Figure 16 :	<i>Représentation schématique d'un récepteur couplé aux protéines G</i>	55
Figure 17 :	<i>Représentation de la séquence en acides aminés d'un récepteur de souris</i>	56
Figure 18 :	<i>Construction d'un panel RH</i>	62
Figure 19 :	<i>Exemples de profils représentatifs obtenus avec deux clones.</i>	100
Figure 20:	<i>Graphe indiquant le % de rétention des cellules hybrides huitre/hamster</i>	101
Figure 21 :	<i>Graphe du ratio d'hybridation des cellules hybride pour le Marqueur AM861117</i>	102
Figure 22 :	<i>Graphe du ratio d'hybridation des cellules hybride pour le Marqueur AM85176</i>	103
Figure 23:	<i>Arbre phylogénique montrant la place des esturgeons (chondrostéens) et celle des téléostéens</i>	104
Figure 24	<i>Les grands lacs de l' Est africain regroupent beaucoup d'espèces de</i>	117

	<i>cichlides</i>	
Figure 25 :	<i>Répartition de l'ensemble des catégories d'OR chez les cinq cichlides</i>	123
Figure 26 :	<i>Arbre phylogénique des OR des 5 cichlides et du poisson Zèbre</i>	125
Figure 27 :	<i>Diagramme du pipeline de calcul des valeurs dN et dS</i>	128
Figure 28 :	<i>Graphe représentant les positions des mutations non silencieuses (en rouge) du couple NYEsplice03/ZEBsplice05.</i>	129
Figure 29 :	<i>Schéma présentant la structure 2D des récepteurs du couple NYEsplice03/ZEBsplice05</i>	130
Figure 30:	<i>Graphe présentant les positions des mutations non silencieuses (en rouge) et silencieuses (en bleu) du couple NYE066/BUR78.</i>	130
Figure 31:	<i>Schéma présentant la structure 2D des récepteurs NYE066/BUR78.</i>	131
Figure 32 :	<i>Les cinq meilleurs motifs caractéristiques des cichlides et de Danio rerio</i>	132
Figure 33 :	<i>Alignement de la séquence du gène BRIspace01 sur son contig</i>	134
Figure 34 :	<i>Alignement de la séquence du gène Tilsplice01 sur son contig</i>	135
Figure 35 :	<i>Alignement de la séquence du gène BURplice01 sur son contig</i>	136
Figure 36 :	<i>Motifs de site donneurs et récepteurs des gènes OR épissés des 5 cichlides</i>	137
Figure 37 :	<i>Répartition de l'ensemble des catégories des TAAR chez les cinq cichlides</i>	155
Figure 38 :	<i>L'arbre phylogénique des TAAR des 5 cichlides et du poisson Zèbre.</i>	157
Figure 39 :	<i>Le Diagramme du pipeline de calcul des valeurs dN et dS pour les gènes TAAR</i>	158
Figure 40:	<i>Graphe présentant les positions des mutations non silencieuses (en rouge) et silencieuses (en bleu) du couple BURspliced13/NYEspliced17</i>	159
Figure 41:	<i>Structure 2D du couple de récepteurs BURspliced13/NYEspliced17</i>	160
Figure 42 :	<i>Graphe présentant les positions des mutations non silencieuses (rouge) et silencieuses (en bleu) du BURt03/ZEBt03</i>	160
Figure 43 :	<i>Structure 2D du couple de récepteurs BURt03/ZEBt03</i>	161
Figure 44 :	<i>Les 5 Motifs protéiques les plus pertinents des TAAR de cichlides et du poisson zèbre.</i>	163
Figure 45 :	<i>Alignement de la séquence du gène NYEspliced17 sur son contig056200</i>	164
Figure 46 :	<i>Motifs de sites donneurs et récepteurs des gènes TAAR épissés des 5 cichlidés</i>	165
Figure 47 :	<i>Présentation avec AutoGRAPH des gènes OR identifiés sur le chromosome LG14 de O. niloticus</i>	172
Figure 48 :	<i>Relation de synténie des gènes OR et TAAR localisés sur le chromosome LG16-21 de Tilapia et de leurs orthologues chez P. nyererei et M.zebra d'une part et H.burtoni et N. brichardi d'autre part.</i>	174

Liste des Tables

Tableau 1 :	<i>La répartition des séquences fournies par le BROAD</i>	119
Tableau 2 :	<i>Résultats des OR pour les 5 cichlides recherchés selon les motifs</i>	121
Tableau 3 :	<i>Résultats des TAAR pour les 5 cichlides recherchés selon les motifs</i>	121
Tableau 4 :	<i>Résultats des OR et des pseudogènes chez les 5 cichlides comparés à ceux des poissons modèles</i>	124
Tableau 5 :	<i>La répartition des familles et sous-familles OR pour chaque poisson</i>	126
Tableau 6 :	<i>Les résultats des moyennes de calcul dN/dS obtenus pour chaque paire de gènes des différentes familles</i>	128
Tableau 7 :	<i>Analyse des valeurs dN et dS sur les différentes parties de la structure 2D pour 14 paires de gènes OR</i>	131
Tableau 8 :	<i>Décompte des gènes formés d'un exon codant et de plusieurs</i>	133
Tableau 9 :	<i>Positions des sites donneurs/accepteurs par rapport à la structure 2D des OR.</i>	138
Tableau 10 :	<i>Résultats des TAAR et des pseudogènes chez les 5 cichlides comparés à ceux des poissons modèles</i>	156
Tableau 11 :	<i>La répartition des famille et sous-familles TAAR pour chaque poisson</i>	158
Tableau 12 :	<i>Les résultats des moyennes de calcul dN/dS obtenus pour les deux familles des gènes TAAR</i>	159
Tableau 13 :	<i>Analyse des valeurs dN et dS sur les différentes parties de la structure 2D pour 14 paires de gènes TAAR</i>	161
Tableau 14 :	<i>Résultats des gènes identifiés formés avec un exon ou plus d'codants</i>	164
Tableau 15 :	<i>Positions des sites d'épissage sur la structure 2D de la protéine TAAR</i>	166
Tableau 16 :	<i>Résultats des données de cartographie du groupe RH3 localisé sur le chromosome LG14 et de l'intégration des données de séquençage.</i>	170
Tableau 17 :	<i>Les gènes OR orthologues identifiés chez les différents cichlides.</i>	171

Abréviations

ADN	Acide DésoxyriboNucléique
ARN	Acide RiboNucléique
EST	Expressed Sequence Tag
SNP	Single Nucleotide Polymorphism
PAML	Phylogenetic Analysis by Maximum Likelihood
Chr	Chromosome
Mb, kb, pb	Megabases, kilobases, paire de bases
PEG	polyéthylène glycol
DMSO	Diméthyle sulfoxyde
DMEM	Dulbecco's Modified Eagle's Medium
CHO	Chinese Hamster Ovary
HPRT	Hypoxanthine-guanine phosphoribosyltransferase
TK	Thymidine Kinase
HAT	hypoxanthine-aminopterin-thymidine medium
OR	Olfactory Receptor
TAAR	Trace Amine-associated Receptors'
ORF	Open Reading Frame
GPCR	G protein coupled receptors
BAC	Bacterial Artificial Chromosome
FISH	Fluorescent <i>In Situ</i> Hybridisation
PCR	Polymerase Chain Reaction
MA	Millions d'Années

TABLE DES MATIÈRES

RESUME	1
Abstract	2
Liste des Figures	3
Liste des Tables	5
Abréviations	6
INTRODUCTION	11
I Phylogénie des Poissons	12
1- L'arbre phylogénique	12
1-1 La famille des Cichlidés	14
1-2 L'importance de la famille des Cichlidés dans les études de l'évolution	16
2- Mécanismes d'évolution des génomes	17
2-1 Evolution de la structure des gènes par brassage d'exons	17
2-2 Remaniements chromosomiques	17
2-3 L'évolution par cis- et trans-duplications	18
2-4 Evolution des génomes des poissons	21
2-5 Séquences répétées et éléments transposables	22
2-6 Pseudogènes et rétropseudogènes	23
2-7 L'évolution par transfert latéral de gènes	23
2-8 Mutations ponctuelles et petites insertions/délétions	24
2-9 La sélection naturelle et les mutations	25
II-Données Générales sur l'aquaculture et son impact sur l'alimentation humaine	27
1- La part des productions aquacoles dans l'alimentation humaine	27
2- La pêche et la surpêche	28
2-1 La pêche	28
2-1-1 La pêche en Europe	28
2-2 La surpêche	29
2-2-1 Les espèces menacées de poisson	29
3 L'aquaculture, la pisciculture	30
3-1 Définition de l'Aquaculture	30
3-2 Historique	31

3-3 Les différents types d'aquaculture.....	32
3-4 L'évolution de la consommation.....	32
3-5 Aliments pour les poissons d'élevage.....	33
3-6 La production aquacole du monde.....	34
3-7 L'aquaculture en France.....	34
3-8 Économie et aquaculture.....	35
3-9 Aquaculture labellisée « biologique ».....	36
4- Génomique et élevage.....	36
4-1-La génomique.....	36
4-2- L'élevage.....	37
III Analyse de la structure des génomes.....	38
1- La cartographie physique.....	39
1-1 Les caryotypes.....	40
1-2 La cartographie cytogénétique.....	40
2 La cartographie de liaison.....	41
2-1 La cartographie génétique.....	41
2-2 La cartographie à l'aide de cellules hybrides d'irradiation (RH)	42
3- La cartographie comparée.....	44
4- Le séquençage.....	46
IV Evolution et développement des familles des récepteurs olfactifs (OR) et des TAAR « Trace Amine-associated Receptors ».....	46
1- L'appareil olfactif	47
2- L'évolution du nombre de gènes codant les récepteurs olfactifs.....	50
3- Les différents récepteurs olfactifs.....	52
4- Aspect génétique et structural.....	53
Objectifs de la thèse.....	57

RESULTATS	59
I- Construction de panels RH et cartographie de génomes d'espèces aquacoles	60
1- Construction d'un panel RH	60
1-1 Les cellules donneuses.....	60
1-2 L'irradiation	61
1-3 Culture cellulaire et clonage cellulaire	61
2- Les étapes de la cartographie d'hybridation.....	62
Articles	
A. Construction du panel d'hybride d'Irradiation du Bar européen (<i>Dicentrarchus labrax</i>) et Cartographie de 1581 marqueurs :	65
Publication n°1 A radiation hybrid map of the European sea bass (<i>Dicentrarchus labrax</i>) based on 1581 markers : Synteny analysis with model fish genomes Richard Guyon et al.	
B. Construction du panel d'hybride d'Irradiation du Tilapia du Nil (<i>Oreochromis niloticus</i>)	80
Publication n°2. A high-resolution map of the Nile tilapia genome : a resource for studying cichlids and other percomorphs. Richard Guyon et al.	
C. Construction de panels d'hybride d'Irradiation d'huitre (<i>Crassostrea gigas</i>) et d'esturgeon (<i>Acipenser baerii</i>)	97
Publication n°3 Construction of Radiation Hybrid panels for tow major aquaculture species: sturgeon and oyster. Naoual Azzouzi et al.	
II- le génome des cichlides	116
1-Annotation des génomes de plusieurs cichlidès.....	119
1-1- Recherche des gènes OR et TAAR par motifs d'acides aminés.....	120
1-2- Recherche des gènes OR et TAAR par BLAST.....	122
Articles	
A- Le répertoire des gènes OR (Publication n° 4 à soumettre)	123
Publication n° 4 Cichlid Olfactory Receptor Genes Naoual Azzouzi, Frédérique Hubler-Barloy, Francis Galibert A soumettre à BMC genomics	
B- Le répertoire des gènes TAAR (Publication n° 5 à soumettre)	155

INTRODUCTION

I- Phylogénie des Poissons

Les poissons sont des vertébrés aquatiques avec des branchies, pourvus de nageoires et dont le corp est le plus souvent couvert d'écaillés. On les trouve en eau douce ou dans les mers. Les poissons jouent un rôle fondamental dans la vie humaine. Ils représentent une source de nourriture non négligeable à coté des végétaux et du bétail, qu'ils soient pêchés ou élevés en pisciculture. Ils sont aussi exploités à des fins récréatives et exposés dans de grands aquariums publics. Ils ont également un rôle important dans l'écosystème en contribuant notamment au cycle du carbone.

1-L'arbre phylogénique

La construction d'arbres évolutifs permet l'identification des ancêtres communs et apporte des informations sur le degré de parenté au niveau des nœuds terminaux. A l'inverse des oiseaux ou des mammifères, les poissons ne forment pas un groupe phylogénique homogène. Ainsi les poissons osseux sont plus proches des tétrapodes que des poissons sans mâchoires et cartilagineux. Ceci indiquerait que les humains et les autres vertébrés terrestres partagent un ancêtre commun poisson de 360-450 millions d'années (figure 1) (J-N Volff, 2005).

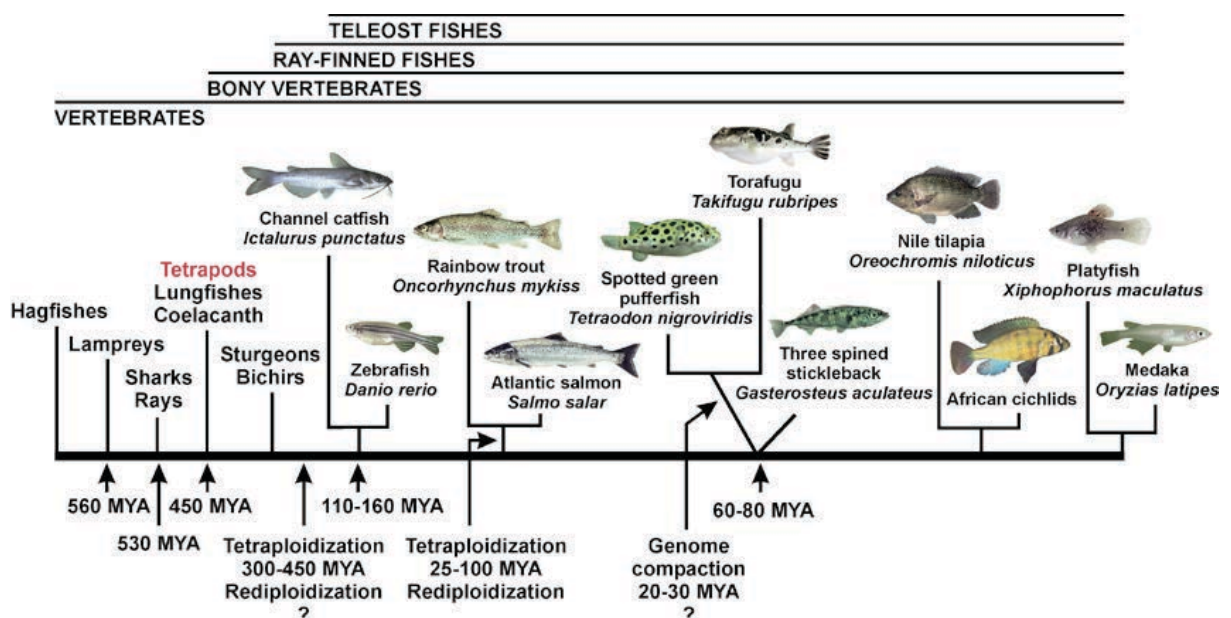


Figure 1 : Arbre phylogénique des poissons en millions d'années sur l'axe des abscisses (J-N Volff, 2005).

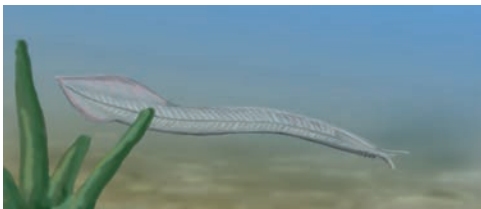
Certains pensent que les poissons ont évolué à partir d'une espèce type ascidie (dont les larves ont des ressemblances avec les poissons primitifs). Les premiers ancêtres des poissons auraient alors conservé leur forme larvaire à l'état adulte, mais l'inverse est aussi possible. Les fossiles candidats au statut de « premier poisson » connus sont : *Haikouichthys*, *Mylokunmingia* et *Pikaia*.



Haikouichthys ercaicunensis est un agnathe primitif apparenté aux lamproies. Il mesure environ 2,5 cm de long. C'est le premier vrai vertébré. Proche de *Pikaia*, il est l'ancêtre de tous les poissons, amphibiens, reptiles, oiseaux et mammifères.



Mylokunmingia fengjiaoa : C'est un poisson agnathe primitif, de petite taille (28 mm de long), dont les fossiles ont été retrouvés près de Kunming en Chine. Il est considéré comme l'un des premiers vertébrés.



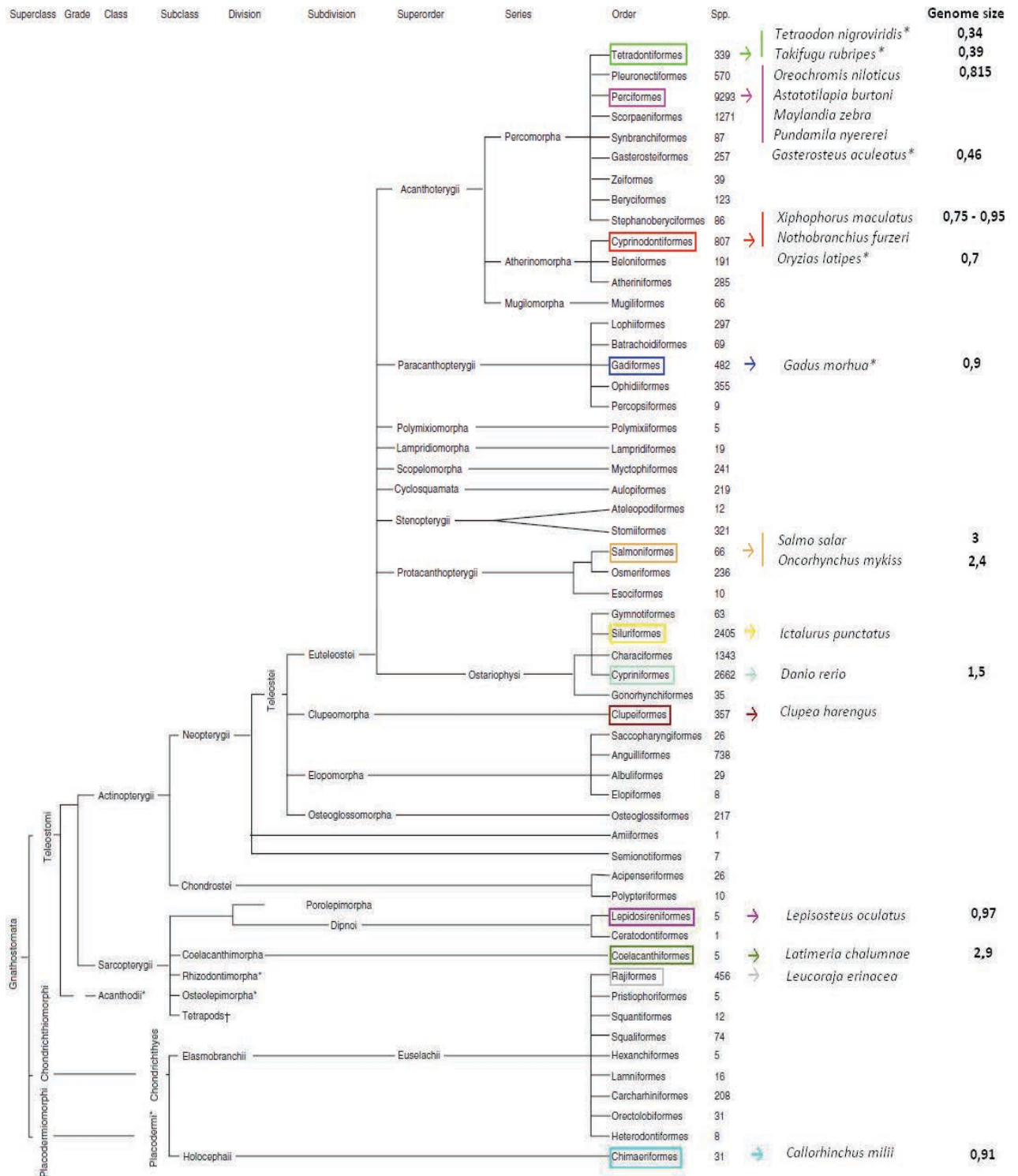
Pikaia gracilens (Walcott CD, 1911) est un céphalochordé fossile. Il avait l'aspect d'une anguille de 5 à 10 cm ; L'amphioxus vivant actuellement, ressemble à *Pikaia*.

Le groupe des poissons représente la première expression biologique proprement dite des vertébrés, dont sont dérivés les *Amphibies*, la première classe de vertébrés en mesure de vivre aussi bien dans l'eau que sur la terre ferme. Dans la classification des espèces, les

poissons appartiennent aux *Cordés*, animaux dotés de “notocorde” (une lamelle cartilagineuse située du côté dorsal de l’animal), sous-type *vertébré*, animaux ayant une colonne vertébrale. Tous les vertébrés sont des Cordés. Contrairement aux cordés inférieurs ou primaires, comme les *Ascidies*, chez qui la “notocorde” permanente chez les vertébrés, est remplacée par une colonne vertébrale osseuse ou cartilagineuse pendant le développement ontogénétique et l’encéphale protégé à l’intérieur d’une capsule crânienne, elle aussi cartilagineuse ou osseuse. Le groupe des poissons est divisé en trois classes, les **Agnathes** ou poissons dépourvus de mâchoires, les *Chondrichthyes* ou poissons cartilagineux et enfin les *Osteichthyes* ou poissons osseux. L’histoire naturelle des poissons a commencé il y a plus de 400 millions d’années, à l’ère primaire dite Paléozoïque (Figure 1). L’apparition d’une mâchoire articulée semble être la raison majeure de la prolifération ultérieure des poissons, car le nombre d’espèces de poissons agnathes devint très faible. Les premières mâchoires ont été trouvées dans les fossiles de placodermes. On ignore si le fait de posséder une mâchoire articulée procure un avantage, par exemple pour la préhension ou la respiration.

1-1-La famille des Cichlidés

Parmi les familles de poissons, la famille des Cichlidés appartient au sous-ordre des Labroidei qui compte aussi la famille des Pomacentridés (Poisson clown) ou la famille des Scaridés (Poisson perroquet). La famille des cichlides est l’une des plus importantes. Elle comporte plus de 200 genres et environ 3000 espèces. La distribution des espèces appartenant à cette famille est mondiale on peut les trouver principalement en Afrique, mais aussi en Amérique Centrale, Amérique du Sud, Madagascar, Syrie, Iran, Sri Lanka ou encore sur les côtes sud des Indes. Par ailleurs de nombreuses espèces sont importées et parfois élevées en Europe à des fins aquariophiles, en raison de leur coloration vive et de leurs mœurs parfois évoluées. Les Cichlidés vivent dans des milieux très différents. Rares sont les familles de poissons qui ont su aussi bien s’adapter aux conditions de l’environnement et aux particularités écologiques que les Cichlidés. Dans les grands lacs africains (Tanganyika, Malawi) les Cichlidés ont tout particulièrement réussi à occuper différentes niches écologiques et à s’imposer face à la concurrence.



1-2-L'importance de la famille des Cichlidés dans les études de l'évolution

La famille des cichlides appartient à l'ordre des Perciformes. Cette famille très importante présente un niveau élevé de biodiversité affectant la morphologie, le comportement ainsi que de nombreux autres aspects de leur biologie (Barlow GW, 2008). Cette grande variabilité rend les poissons en général et les cichlidés en particulier extrêmement intéressants pour l'étude de nombreuses questions biologiques, en particulier celles liées à l'évolution. Des données de séquençage ont récemment révélé plusieurs caractéristiques particulières des génomes de poissons qui pourraient avoir eu un rôle dans l'évolution des poissons et leur spéciation. Il y a maintenant des preuves substantielles d'un cycle de tétraploïdisation / rediploïdisation qui a eu lieu au cours de l'évolution précoce de la lignée des poissons à nageoires rayonnées et que des centaines de gènes dupliqués générés par ce mécanisme ont été maintenus sur des centaines de millions d'années d'évolution (J-N Volff, 2005). La perte de fonction des gènes dupliqués pourrait être impliquée dans la génération de la variabilité du poisson. Comme chez les mammifères, les génomes de poissons téléostéens ont aussi plusieurs éléments transposables actifs, qui pourraient avoir eu un rôle dans la spéciation en affectant la stérilité des hybrides et leur viabilité. La diversité phénoménale des systèmes de détermination du sexe et la plasticité des chromosomes sexuels observées chez les téléostéens ont été impliquées dans l'isolement reproducteur (mécanisme empêchant ou limitant fortement l'hybridation de deux espèces habitant la même région) à la fois avant et après l'accouplement. La comparaison des données générées par les projets actuels et futurs d'analyse de génomes des poissons ainsi que d'autres espèces permettront de mieux comprendre les mécanismes moléculaires et par conséquent découvrir le mécanisme d'évolution et la diversité des poissons. Ceci va certainement contribuer de manière significative à notre compréhension de la fonction et de l'évolution des gènes chez les humains et les autres vertébrés (J-N Volff, 2005).

Les origines de la diversité du monde vivant se trouvent au niveau moléculaire. Chacun des individus, appartenant à une espèce ou une sous-espèce, est en effet légèrement différent des autres sur le plan génétique et cette diversité génétique est la condition nécessaire pour créer la diversité biologique qui permettra aux espèces de mettre en œuvre des stratégies alternatives, lesquelles sont leurs réponses adaptatives aux changements de l'environnement. La variabilité intra espèce ne se limite pas en effet à la composante génétique, mais se traduit aussi par une gamme de phénotypes différents.

2. Mécanismes d'évolution des génomes

La molécule d'ADN est le support biologique de l'information héréditaire et l'histoire de l'évolution est enregistrée dans les génomes des organismes actuels. Ainsi, le séquençage à grande échelle des génomes peut fournir une image des forces évolutives qui ont façonné la structure des différentes espèces.

2-1 Evolution de la structure des gènes par brassage d'exons

En 1970 W. Gilbert a défini le brassage d'exons comme l'un des mécanismes permettant de créer des nouveautés génétiques. Il s'agit essentiellement d'assemblages et remaniements d'exons provenant de gènes différents. Ce phénomène serait fréquent chez les espèces eucaryotes et on dénombre aujourd'hui une centaine de familles de gènes, principalement de mammifères ainsi constituées (Gilbert, 1970). Elles répondent en particulier aux fonctions de signalisation entre cellules de l'organisme (propres aux métazoaires) (Pathy, 2003). Le brassage d'exons est le plus couramment mis en avant, mais il existe également, par exemple, des processus de brassage de promoteurs (Luchetta *et al.*, 2005). Le brassage d'exons est très favorable à l'acquisition de nouvelles fonctions protéiques voire même de nouvelles combinaisons de fonctions protéiques.

2-2 Remaniements chromosomiques

En terme de modifications génomiques, on distingue les remaniements qui affectent la taille du génome par amplification ou réduction et ceux qui ne la changent pas. On parle alors de réarrangements équilibrés. L'observation des caryotypes d'espèces plus ou moins proches révèle parfois l'existence de différences cytogénétiques importantes. Ils témoignent des réarrangements génomiques de grands fragments d'ADN qui sont relativement fréquents chez les mammifères, jusque dans leur histoire évolutive récente (Jones *et al.*, 1992). Parmi les principaux facteurs de l'amplification génomique qui constituent de véritables duplications sub-génomiques, on trouve le phénomène des crossing-overs inégaux lors de la méiose, échanges inégaux entre chromatides sœurs, translocations chromosomiques et transpositions de multiples copies d'ADN appelées aussi duplications. Ces processus peuvent cependant conduire à une perte de matériel génétique. D'autres remaniements chromosomiques ne modifient ni la taille ni le contenu nucléotidique du génome. Il s'agit alors de phénomène de translocation équilibrée, de transposition de simple copie d'ADN, d'inversion chromosomique ou encore de fusion ou fission de chromosomes. Nadeau et Taylor ont suggéré par l'étude des relations

synténiques, c'est-à-dire des bornes de ces remaniements, que les points de coupures (breakpoints) entre segments homologues devaient être répartis de façon uniforme et aléatoire sur le génome (Eichler and Sankoff, 2003; Nadeau and Taylor, 1984). Néanmoins, les régions centromériques et télomériques semblent être davantage concernées, probablement en raison de leur richesse en séquences répétées qui favoriseraient les événements de recombinaison homologues (Eichler and Sankoff, 2003). Le taux de réarrangements chromosomiques est variable d'un génome à un autre. En effet, il a été mis en évidence un taux variant de 0,2 à 1 ou 2 remaniements par million d'années chez les vertébrés (Burt *et al.*, 1999). Cependant, l'ensemble de ces réarrangements chromosomiques est ensuite transmis à la descendance avec une fréquence faible, en raison de l'infertilité qui en découle le plus souvent.

2-3 L'évolution par cis- et trans-duplications

La duplication de fragments d'ADN constitue une forme de remaniement chromosomique que l'on peut observer dans la plupart des génomes. Elle concerne des fragments de quelques kilobases à plusieurs centaines de kilobases. Ces duplications sont dites segmentaires par opposition aux duplications de génome entier appelées polyploïdisation.

Les duplications partielles ou totales des génomes ont profondément modifiées la structure de ces derniers et sont pour une large part responsables de la diversité biologique. Trois évènements d'amplification totale ont été proposés pour expliquer l'évolution ancienne des vertébrés : deux à l'origine du groupe et une troisième spécifique des poissons (Meyer et Schartl 1999; Meyer et Van de Peer 2005). Alors que la controverse persiste quant aux deux premiers évènements, le troisième affectant les poissons est sans ambiguïté, comme le révèle le positionnement dans le génome de tétraodon de 1078 paires de gènes paralog, ainsi que l'étonnant appariement de gènes orthologues identifiés et positionnés sur les chromosomes de tétraodon et de l'homme (Jaillon *et al.* 2004). Ceci fait des poissons un modèle unique pour l'analyse de ces phénomènes du moins chez les vertébrés.

Le devenir des gènes dupliqués peut être variable : disparaître, accumuler des mutations faux sens et non sens et devenir des pseudogènes comme nombre d'OR ou encore accumuler des mutations faux sens et gagner une nouvelle fonction, éventuellement complémentaire de celle du gène ancestral (Balakirev and Ayala, 2003).

Une bonne illustration de ce phénomène de duplication/polyploïdisation est donnée par l'analyse des gènes Hox.

Les gènes Hox codent des facteurs de transcription. Leur expression gouverne la

différenciation cellulaire et la formation des différentes parties du corps selon l'axe antéro-postérieur. Chez amphioxus, ils constituent un ensemble de 14 paralogues et chez la drosophile on en dénombre 9 regroupés sur un seul chromosome (García-Fernández J. 2005). Chez les tétrapodes d'une façon générale suite à deux tours de polydiploïdisation, on trouve 4 clusters (Amemiya CT *et al.*, 2008). Dans le génome humain (et de tous les mammifères) ils constituent une petite famille de 39 gènes. Leur organisation dans tous les génomes illustre parfaitement leur genèse. En effet plusieurs cis-duplications rendent compte de la présence d'une dizaine de paralogues regroupés en un locus donné d'une centaine de kilobases et de trans duplication ou polypléïdisation avec l'existence de 4 groupes de gènes dénommés a, b, c, d sur les chromosomes humains 6, 11, 15, 2 respectivement (Lemons D et McGinnis W. 2006; Duboule D. 2007). L'expression de ces gènes suit une règle de colinéarité dans laquelle les gènes situés en amont du complexe s'expriment les premiers et gouvernent la différenciation de la partie antérieure du corps alors que les gènes situés à l'autre extrémité s'expriment plus tardivement et sont impliqués dans la différenciation des régions postérieures.

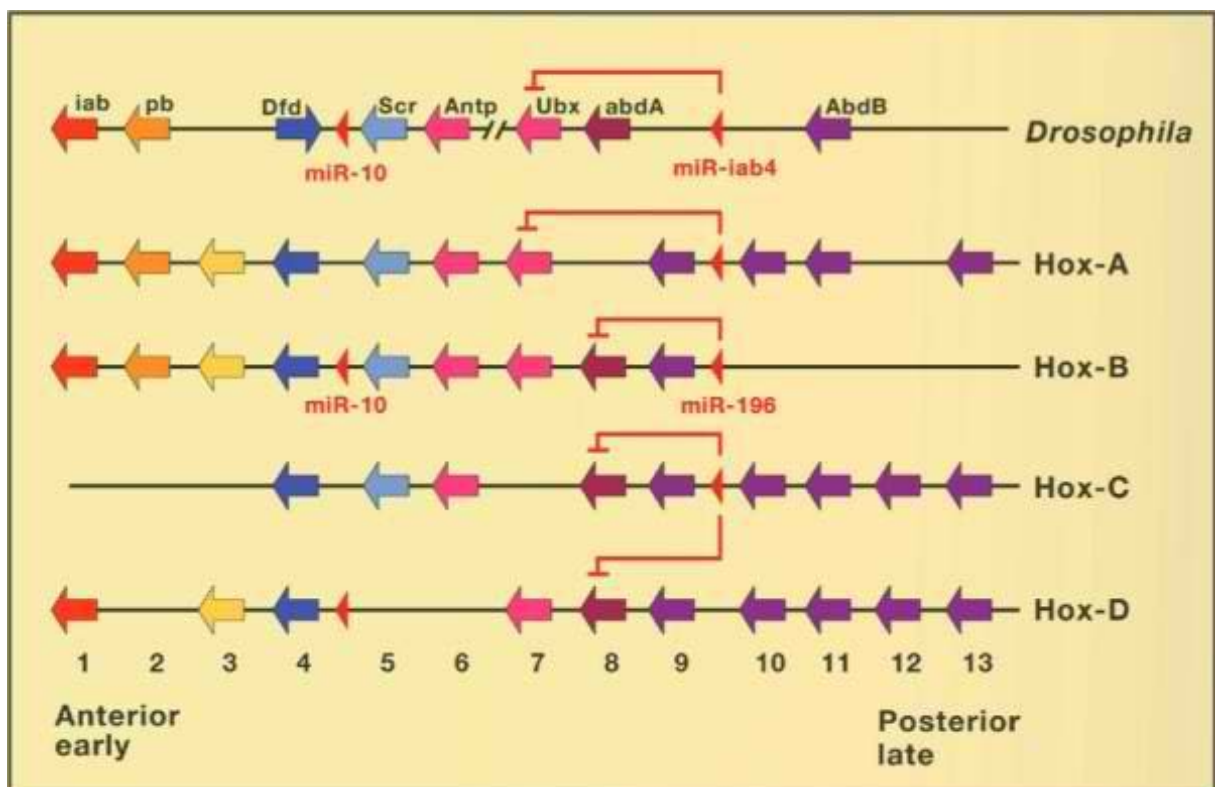


Figure 3 : Organisation du complexe Hox de drosophile et des 4 complexes (a-d) des mammifères montrant l'impact des cis et trans duplications, suivies de la perte éventuelle de certains gènes absents de l'un ou de l'autre cluster. (De Robertis 2008)

Comme cela a été rappelé plus haut, les poissons ont subi un troisième cycle de polydiploïdisation qui se reflète parfaitement dans l'organisation des différents complexes Hox au nombre de 8 dénommés Aa, Ab, Ba, Bb, Ca, Cb, Da et Db et présentant la même structure d'ensemble que celle observée chez les mammifères.

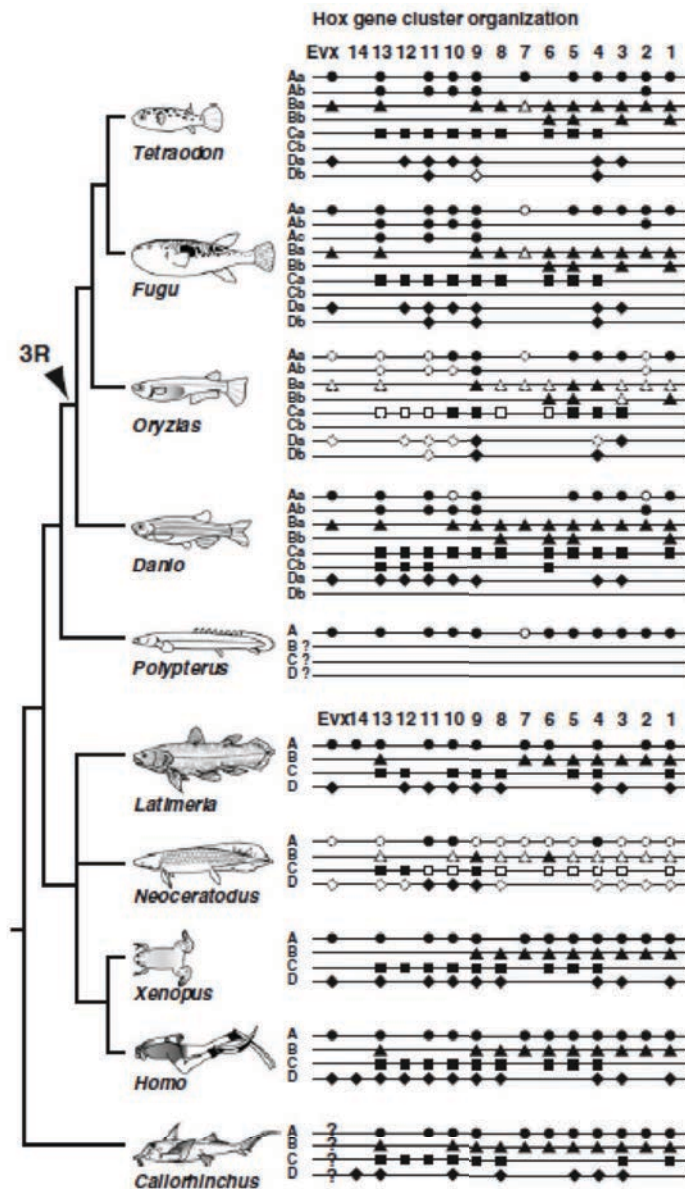


Figure 4 : Organisation comparée des complexes hox de divers poissons, mammifères et xenope. (Yukuto S et al., 2010)

Au delà de ces modifications ayant affecté l'ensemble des lignées, on peut également noter quelques cas particuliers d'amplification comme celles observées chez les esturgeons chez lesquels on trouve même des espèces hexaploïdes comme *Acipender persicus* qui possède 360 paires de chromosomes

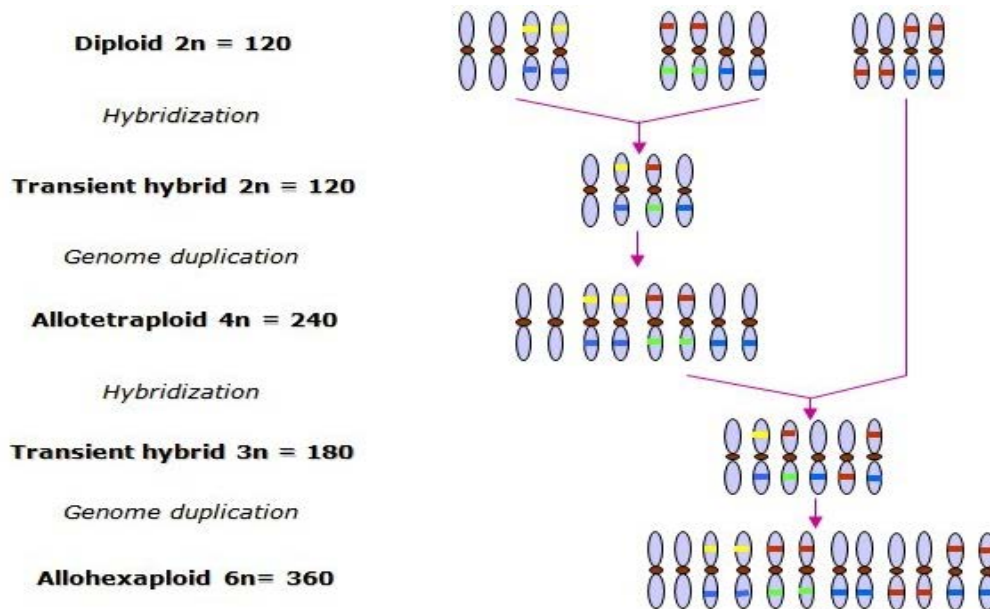


Figure 5 : Différentes formes de ploïdie retrouvées chez les différentes espèces d'esturgeons sont présentées ici (*A. persicus* avec $6n = 360$) (Biology, Conservation and Sustainable Development of Sturgeons. Springer Science B.V. 2009. Chapter 3.

2-4 Evolution des génomes des poissons

Chez les poissons, l'évolution du nombre et de la forme des chromosomes est particulièrement spectaculaire dans la famille des Cyprinodontidae pour laquelle ils existent des espèces très proches morphologiquement, alors que les caryotypes sont différents et qu'elles ne peuvent donc s'hybrider. Chez les *Aphyosemion*, le nombre haploïde varie entre $n = 24$ chez *A. rectoogoense* et $n = 9$ chez *A. chrystyi* (Scheel 1968). Il existe un exemple particulièrement intéressant de spéciation avec réduction du nombre de chromosomes de 15 à 13, 11, 10 et 9 paires, respectivement, dans la série *Aphyosemion* : *melanopteron-cognatum-schoutedeni-elegans-chrystyi*. Les phénotypes de ces espèces sont très semblables et elles ne peuvent être identifiées par la seule utilisation de caractères métriques ou méristiques. Elles se distinguent seulement par des détails de coloration de la livrée des mâles adultes. Ces différences avaient d'abord été attribuées à du polymorphisme intra-espèce, jusqu'à ce que l'étude des caryotypes et des expériences d'hybridation aient montré qu'il s'agissait d'espèces isolées sur le plan reproducteur. La polyploïdie, qui se manifeste par un accroissement important du nombre de chromosomes, n'est pas rare chez

les poissons, en comparaison avec les autres vertébrés. Ce phénomène paraît en particulier jouer un rôle majeur dans l'évolution des caryotypes de la famille des Cyprinidae et l'on a pu montrer que certaines espèces de grands *Barbus* africains étaient hexaploïdes et possédaient 148-150 chromosomes. Il est également possible d'induire expérimentalement la polyploïdisation chez certaines espèces de poissons (Lévêque C et Didier P, 1999). Plusieurs espèces de poissons téléostéens sont particulièrement étudiées aux niveaux génétiques et génomiques; certains d'entre eux ont fait l'objet de projets de séquençage de leurs génomes, comme le saumon de l'Atlantique *Salmo salar*, la truite arc-en ciel *Oncorhynchus mykiss* et le tilapia du Nil *Oreochromis niloticus*. Ces analyses et les programmes de génomique en général ont par ailleurs une motivation économique. Ces programmes visent essentiellement à identifier des caractères qualitatifs et quantitatifs au niveau moléculaire (comme les QTL) afin de pouvoir maîtriser la croissance, la reproduction, la pigmentation, la tolérance à l'environnement ou la résistance aux maladies, qui sont des traits très pertinents pour l'aquaculture.

C'est ainsi que notre laboratoire s'est intéressé tout d'abord à la daurade *sparus aurata* (programme européen Bridge-Map No. Q5RS-2001-01797), puis successivement au bar *dicentrarchus labrax* (programme européen : Marine Genomics Europe (contract GOCE-CT-2004-505403 et project Aquafirst), au tilapia *oreochromis niloticus* (ANR-07-GANI-005) et à l'esturgeon ; divers programmes auxquels j'ai participé.

2-5 Séquences répétées et éléments transposables

De nombreuses familles de séquences hautement répétées sont trouvées dans les génomes des eucaryotes. On distingue les séquences répétées en tandem et l'ADN répétitif dispersé, impliquant les transposons. Les répétitions en tandem se divisent en trois sous-groupes d'ADN non codant : ADN satellite, ADN minisatellite et ADN microsatellite. L'ADN satellite, se compose de répétitions d'une séquence plus ou moins complexe et constitue l'essentiel de l'hétérochromatine, alors que l'ADN minisatellite comprend une variété de répétitions en tandem, dispersées sur le génome et couvrant des fragments de taille moyenne (quelques kilobases). Ces éléments hypervariables sont particulièrement présents aux télomères et ne sont que rarement transcrits. On considère qu'ils constituent ainsi des « points chauds » (hotspots) de recombinaison homologue. Au niveau des télomères, ces éléments protègent l'extrémité des chromosomes de la dégradation et de la perte de matériel génétique. L'ADN microsatellite se définit par des séries d'unités simples répétées et dispersées dans le génome. L'unité de base de l'ADN microsatellite ne compte

que quelques nucléotides voire quelques dizaines (Luchetta *et al.*, 2005). Le caractère répétitif de l'ensemble de ces séquences en fait des éléments favorisant les remaniements chromosomiques et géniques par recombinaison homologue.

2-6 Pseudogènes et rétroseudogènes

Les pseudogènes sont des gènes non fonctionnels qui présentent des similitudes avec un ou plusieurs gènes fonctionnels. L'absence de fonction est due soit à la perte de séquences régulatrices soit à des mutations qui perturbent la maturation des pré-messagers en altérant les signaux de maturation des pré-messagers soit encore par la présence de mutations dans la partie codante introduisant des codons stop ou des changements de phases de lecture du messager. Selon les modalités de leur mise en place on distingue les pseudogènes proprement dit qui sont des gènes ayant subi des mutations délétères et les rétrogènes.

Les rétroseudogènes existent chez les Métazoaires et proviennent d'une rétroposition ou transcription inverse d'un ARN messager suivie d'une intégration dans l'ADN génomique. Récemment il a été montré que le phénotype de chondrodysplasie des chiens était dû à un rétrogène (Parker *et al.*, 2009).

Le rétroseudogène ressemble donc au messager dont il est issu et ne comporte ni intron ni promoteur. Son intégration s'accompagne de la formation de petites régions répétées directes à chaque extrémité ainsi que d'une région polyadénylée en aval, provenant du polyA de l'ARN messager (Rogers, 1985). L'évolution des rétroseudogènes dépend en partie du lieu où ils s'insèrent. S'ils sont insérés dans une région qui possède fortuitement un promoteur, ils peuvent redevenir actifs, soit rester « en attente » d'une nouvelle fonction s'ils sont dans une région qui permettra plus tard de recréer un promoteur (Long, 2001). Les séquences flanquantes peuvent en effet, dans certains cas, acquérir un rôle de promoteur après des mutations ponctuelles. De même, des pieds d'introns/exons peuvent apparaître de façon aléatoire complexifiant ainsi la structure du rétroseudogène qui sera co-transcrit avec ces nouveaux exons (Long, 2001).

2-7 L'évolution par le transfert latéral de gènes

Le transfert latéral de gènes définit un échange de matériel génétique entre espèces distinctes et favorise l'acquisition de fonctions de l'organisme donneur, systématiquement procaryote, par l'organisme receveur qui peut être eucaryote. Appelé aussi transfert horizontal de gène, ce sont des événements rares, mis en évidence par la présence de gènes similaires dans des espèces très distantes (Genereux and Logsdon, 2003). Le matériel

génétique étranger doit d'abord pénétrer dans la cellule receveuse, sous forme d'ADN nu ou au sein de la cellule donneuse. Ce phénomène se produit généralement entre un procaryote et un organisme phagotrophe, le gène est ensuite incorporé dans le noyau receveur. Pour être maintenu dans le génome, le gène transféré doit également conférer des propriétés sélectionnées dans la population (Andersson, 2005).

2-8 Mutations ponctuelles et petites insertions/délétions

Les mutations résultent d'erreurs de réplication et de lacunes du système de réparation de l'ADN. Les mutations peuvent être induites par exposition à différents types de mutagènes présents dans notre environnement ou produites par l'environnement des cellules. Elles proviennent également d'erreurs spontanées de la réplication et de la réparation de l'ADN. Si ces dernières affectent les cellules de la lignée germinales, elles seront transmises à la descendance et susceptibles d'affecter l'espèce. Ce qui n'est pas le cas des mutations affectant les cellules somatiques qui peuvent affecter l'individu mais pas l'espèce. Parmi les mutations on distingue les transitions qui sont des substitutions entre purines (adénine ou guanine) ou entre pyrimidines (thymine ou cytosine) et les transversions qui sont des substitutions entre purines et pyrimidines. Il y a plus de transitions que de transversions, ce qui s'explique probablement par une moindre différence d'encombrement stérique et un moindre impact fonctionnel des transitions. On distingue aussi les insertions et délétions de quelques nucléotides, qui lorsqu'elles surviennent dans une région codante peuvent provoquer un décalage du cadre de lecture et une modification conséquente de la séquence protéique. Les mutations affectent en principe aussi bien les régions codantes que non-codantes des génomes, mais en raison de la perte de fonction qu'elles peuvent entraîner selon la région, elles ne sont pas toutes conservées et on observe des taux différents entre les exons, introns et séquences intergéniques. Les insertions et délétions sont très rares dans les régions codant des protéines en raison de la perte de fonction qui en découle. Les mutations sont dites non-synonymes ou synonymes, lorsqu'elles se situent dans une séquence codant une protéine. Les mutations synonymes, plus fréquentes, sont également dites silencieuses car la séquence protéique n'est pas modifiée, en raison de la redondance du code génétique. Les mutations non-synonymes peuvent induire l'arrêt prématuré de la synthèse protéique par insertion d'un codon stop (mutation non-sens) ou en modifier la séquence (mutation faux-sens). L'évolution des génomes par mutation est extrêmement lente, ce qui impose de comparer les séquences d'organismes différents pour pouvoir l'observer. En outre, les substitutions, qui apparaissent de façon aléatoire, ne peuvent

envahir la population qu'à deux conditions : affecter le génome de la lignée germinale et résister à la sélection naturelle.

2-9 La sélection naturelle et les mutations

La sélection naturelle est l'un des mécanismes de l'évolution des séquences d'ADN.

L'évolution peut donc être considérée comme étant sous le contrôle combinatoire de deux types de pression : les mutations qui se produisent aléatoirement dans le génome et une pression sélective qui peut éliminer les mutations ou, au contraire, augmenter leur fréquence dans la population. La sélection purificatrice dite négative permet l'élimination progressive des individus porteurs de mutations défavorables et donc la non dissémination de ces mutations dans la population, alors que la sélection darwinienne dite positive, permet la fixation de mutations favorables. Les effets de la pression de sélection négative sont observés essentiellement au cours de l'élimination des mutations dans des régions codantes et qui aboutissent à des changements phénotypiques significatifs. C'est la raison pour laquelle les régions codantes sont exceptionnellement bien conservées entre espèces, même distantes.

Kimura a développé la théorie neutraliste qui met en avant l'importance d'événements aléatoires, de mutations et de dérive, dans l'évolution moléculaire (Kimura 1968). Dans la théorie neutraliste de l'évolution, les variations des séquences sont neutres d'un point de vue évolutif et les fréquences des différents allèles dans la population varient de manière aléatoire ; c'est la dérive génétique. La dérive génétique est la principale cause des différences entre les populations qui aboutissent à la création d'espèces séparées, tandis que les pressions de sélection jouent un rôle dans l'adaptation des populations et des espèces, mais peu dans leur définition et leur différenciation.

Le calcul des taux de substitutions synonymes et non-synonymes permet de savoir quel type de sélection s'exerce sur une séquence codante. Ces valeurs sont calculées entre des séquences orthologues dans deux espèces différentes selon différentes formules. Ces formules sont établies d'après la méthode de Jukes et Cantor qui considère que les taux de transitions et transversions sont égaux.

MS et MA sont les nombres de substitutions, respectivement synonymes et non-synonymes, et NS et NA les nombres de sites synonymes et non synonymes, selon la dégénérescence du code génétique au locus analysé. On considère généralement qu'un rapport dN/dS (ou $dN = MA/NA$ et $dS = MS/NS$) voisin de 1 indique l'absence de sélection, $dN/dS < 1$ une sélection négative et $dN/dS > 1$ une sélection positive. En effet, un

rapport $dN/dS > 1$ indique que des substitutions changeant des acides aminés s'accumulent davantage que les mutations neutres, suggérant qu'elles sont positivement sélectionnées.

On utilise soit des méthodes basées sur des modèles Markoviens pour évaluer la sélection des différents codons (Robinson, *et al.*, 2003; Ouyang and Liang, 2007), soit des méthodes de maximum de vraisemblance pour calculer le nombre de site ayant un $dN/dS \ll 1$, ~ 1 et > 1 (Li, *et al.*, 1985; Goldman and Yang, 1994) à partir des alignements multiples des séquences. Un test de rapport des vraisemblances ou Likelihood Ratio Test « LRT » (Yang Z and Nielsen, 1998; Zhang Jianzhi, *et al.*, 2005b; Yang Z and Dos Reis, 2011) est utilisé pour tester la présence de sélection positive dans un gène donné. L'identification des cibles de la sélection positive au sein d'une espèce par approche dN/dS a pour objectif d'identifier les gènes et les réseaux de gènes impliqués au cours de l'évolution naturelle de l'espèce considérée.

L'analyse du rapport dN/dS permet donc d'obtenir des indications sur les pressions de sélection s'exerçant sur les séquences codantes. En particulier, des phénomènes tels que la conversion génique sont susceptibles de réduire ou d'accentuer les effets d'une éventuelle pression de sélection. Makalowski et Bogulski ont calculé une valeur moyenne de 0,11 du rapport dN/dS sur un ensemble de 1880 gènes orthologues entre l'homme et les rongeurs (Makalowski et Bogulski, 1998), indiquant que la vaste majorité des gènes sont soumis à une sélection négative. Pour autant une sélection positive a pu être confirmée, aussi bien chez les procaryotes que chez les eucaryotes. Les gènes pour lesquels une sélection positive a été mise en évidence sont essentiellement impliqués dans les fonctions des systèmes de défense et immunité, des systèmes de la reproduction et des systèmes olfactifs (Fitch et al. 1991 ; Yu Fet al., 2005 ; Chen M et al.,2010).

Pour notre part nous nous sommes particulièrement intéressés aux familles des gènes des récepteurs olfactifs des cichlidés, qui s'imposent comme un bon modèle d'étude de l'évolution.

II- Données Générales de l'aquaculture et son impact sur l'alimentation humaine

1- La part des productions aquacoles dans l'alimentation humaine

En dehors des végétaux et des animaux terrestres, l'alimentation humaine comporte une bonne part de poissons de mer ou de rivière dont l'aquaculture est la principale source. Ayant une forte dépendance vis-à-vis des ressources halieutiques, en tout cas pour l'élevage de certaines espèces, les recherches actuelles portent beaucoup sur l'assurance d'un développement durable de la filière tant sur le plan européen que sur le plan mondial. Du point de vue de la nutrition humaine, les poissons constituent une source de protéines à valeur biologique élevée, couvrant près de 20% de l'apport protéique. Ils sont aussi des sources riches en micronutriments (oligo-éléments, vitamines ou pro-vitamines). Les produits aquatiques, notamment les poissons, sont d'excellentes sources d'acides gras longs poly- insaturés (AGLPI) de la série 3, comme l'acide eicosapentanoïque (EPA) ou le docosahexaénoïque (DHA). Plusieurs études ont bien démontré l'effet bénéfique d'un apport régulier en AGLPI de la série 3 sur la santé humaine (Gissi, 1999 ; Bucher *et al.*, 2002). Les AGLPI de la série 3 comme l'EPA et le DHA sont impliqués dans diverses actions connues : amélioration de la fluidité membranaire, diminution de l'agrégation plaquettaire et, par conséquent, diminution des maladies cardiovasculaires, augmentation de la résistance immunitaire et à la cancérogenèse (Simopoulos, 2001). Ces acides gras sont importants aussi pour la vision et le développement cérébral. On considère même qu'ils ont joué un rôle dès l'origine de l'humanité (Broadhurst *et al.*, 1998 ; Crawford *et al.*, 1999). En revanche, un rapport nutritionnel $\omega 6/\omega 3$ trop élevé (de l'ordre de 15), tel qu'il est observé dans les pays les plus développés est inquiétant et pourrait avoir des conséquences sur la santé (Simopoulos, 2002). La diminution de ce rapport est donc considérée comme un enjeu majeur pour la santé humaine et la consommation régulière de poissons semble être une solution particulièrement intéressante dans ce contexte. Bien que ne disposant pas de chiffres précis sur l'apport nutritionnel conseillé, les recommandations actuelles sont de l'ordre de 0,4 à 1 g d'AGLPI3 par jour, variable selon l'âge, le sexe ou l'état physiologique (Martin, 2001). Au plan mondial, la consommation de poisson dans l'alimentation humaine et animale augmente régulièrement. Elle pourrait atteindre 179 millions de tonnes en 2015 selon l'agence alimentaire de l'Organisation des Nations Unies

(FAO), soit 36 % de plus qu'en 2002. Il faut retenir aussi que les acides gras à longue chaîne comme l'oméga-3 se trouvent dans les graines de chia, le lin, la noix, la cameline, le colza, le soja et les algues. Face à la raréfaction de certaines ressources halieutiques, la nécessité d'une pêche durable et responsable apparaît plus évidente d'autant que la pêche seule ne pourra satisfaire les besoins des populations en produits aquacoles. Aujourd'hui, près de la moitié du poisson consommé dans le monde provient déjà de l'aquaculture et non plus de la pêche.

2- La pêche et la surpêche

2-1- La pêche

La pêche est une activité consistant à capturer des animaux aquatiques (poissons, crustacés, céphalopodes...) dans leur milieu naturel (océans, mers, cours d'eau, étangs, lacs, mares). Cette activité est pratiquée par les pêcheurs, comme loisir ou profession. Les techniques utilisées pour la pêche sont nombreuses. Elles dépendent de l'espèce recherchée, du milieu, voire des navires utilisés et du niveau de professionnalisation. La pêche est le plus souvent encadrée par une réglementation de plus en plus sévère, afin de protéger au mieux la biodiversité, l'environnement et les ressources halieutiques.

2-1-1- La pêche dans le monde et en Europe

Près de 90% des poissons capturés à l'état sauvage dans le monde viennent des océans et des mers. Plus de 40 % des océans sont très endommagés et très peu d'eaux marines restent vierges. Selon le secrétariat de la Convention sur la diversité biologique, qui dépend de l'ONU, les zones mortes côtières ont pratiquement doublé chaque décennie depuis 1960. Les zones les plus touchées sont la Mer du Nord, le sud et l'est de la Mer de Chine, le bassin des Caraïbes, la côte est de l'Amérique du Nord, la Méditerranée, la Mer Rouge, le Golfe Persique, la Mer de Béring et plusieurs régions occidentales du Pacifique.

La puissance de pêche en Europe augmente de plus en plus. Grâce à des améliorations techniques permanentes sur les bateaux ou en matériels de détection électronique, la flotte de pêche de l'Union européenne est toujours plus puissante. Elle gagne 3% par an en « puissance réelle de pêche », ce gain se fait malgré la diminution du nombre de navires de pêche européens de 105.000 à 80.000 navires entre 1992 et 2009. Celle-ci ne fournit que 4

millions de tonnes de prises, soit seulement 38% des 10,7 millions de tonnes de poissons consommés annuellement par les Européens.

Selon la FAO l'organisation des Nations Unies pour l'agriculture, 7 des 10 plus importantes espèces de poissons sont au bord du dépeuplement total. Chaque année dans le monde 130 millions de tonnes de poissons sont capturées; cette quantité n'augmente plus depuis plusieurs années et seule la pisciculture permettra de compenser la baisse annoncée des quantités pêchées. Les écosystèmes océaniques les mieux préservés sont ceux des régions polaires. Selon la FAO en 2010, 2/3 des espèces sont surexploitées dans le monde. On estime qu'une espèce s'effondre quand les prises de pêche ont diminué de 90%, ce qui était le cas de 29% des espèces en 2003. Directement ou indirectement, la vie de plus de 500 millions de personnes dans les pays en voie de développement dépend des pêcheries et de l'aquaculture.

FAO : The State of World Fisheries and Aquaculture.

<http://www.fao.org/docrep/013/i1820e/i1820e00htm> 2010.

2-2-La surpêche

La surpêche est la pêche excessive (légale ou illégale) pratiquée par l'homme sur certains poissons, crustacés ou mollusques. Cette surexploitation observée dans presque toutes les pêcheries de la planète préoccupe la FAO et l'ONU car elle menace la sécurité alimentaire et les équilibres écologiques marins. La surpêche a causé une réduction de certains stocks halieutiques commerciaux de plus de 90%. Elle frappe actuellement de nombreux stocks et a pour effet de diminuer la biomasse et d'augmenter la capture de juvéniles non commercialisables et d'espèces non ciblées. Il est donc urgent de réduire la surpêche et de mettre en œuvre des mesures appropriées pour développer la biomasse des stocks reproducteurs. Grâce à la recherche et aux progrès technologiques, les engins de pêche devraient être perfectionnés afin de réduire les prises accessoires.

2-2-1- Les espèces menacées de poissons

Depuis 1950, 7% des espèces marines ont disparu. Globalement, une espèce de poissons pêchée sur trois est menacée d'extinction et la moitié parvient tout juste à se renouveler.

Dans le monde 29% des 600 espèces pêchées sont en voie d'extinction totale. C'est le cas au Canada de la morue, du saumon, du marsouin commun, du fouille-roche gris, de la raie tachetée, du colin, de l'églefin entre autres, selon l'Union Internationale pour la Conservation de la Nature (IUCN).

La NOAA « National Oceanic and Atmospheric Administration » est une administration américaine qui fournit et assure un accès à des données environnementales mondiales et des services d'information à partir de satellites et d'autres sources dans le but de promouvoir, de protéger et d'améliorer l'économie, la sécurité, l'environnement et la qualité de vie de la nation. Mais elle a aussi un rôle de surveillant sur les eaux territoriales américaines. En 2010, la NOAA avait interdit les sorties de chalutiers dans certaines zones du golfe du Mexique, suite à la marée noire de la plate-forme Deepwater Horizon.

Le rapport de janvier 2010 publié par La NOAA insiste sur le respect de la légalité en matière de pêche et pointe du doigt les pays "pirates" que sont la Colombie, l'Equateur, l'Italie, le Panama, le Portugal et le Venezuela. En effet, les chalutiers de ces pays ont été pris en flagrant délit de pêche illégale en 2009 et 2010.

3-L'aquaculture, la pisciculture

3-1-Définition

Le mot aquaculture, qui s'est progressivement substitué à celui d'aquiculture ou halieuculture, désigne l'ensemble des activités qui ont pour objet la production d'espèces aquatiques animales ou végétales par l'homme. Par définition, on parle d'aquaculture dès lors qu'il existe au moins une intervention humaine au cours du cycle de vie des organismes élevés, intervention distincte de l'opération de capture ou de récolte et donc antérieure à la mort de l'organisme.

On distingue de nombreuses formes d'aquaculture :

- La transplantation qui consiste à déplacer des organismes au sein de leur aire de répartition géographique naturelle,
- L'acclimatation qui correspond à l'introduction des organismes dans un lieu situé en dehors de leur aire de répartition,

- L'élevage intensif avec distribution de nourriture artificielle dans des bassins aménagés ou des cages à fort renouvellement d'eau. En ce sens, l'aquaculture peut être définie comme la gestion d'un écosystème aquatique en vue d'y favoriser la production d'une ou de plusieurs espèces d'intérêt commercial.

Il existe également une définition de l'aquaculture d'ordre juridique : c'est une activité d'élevage d'espèces aquatiques dont la propriété du stock appartient à une personne juridique identifiée tout au long du cycle d'élevage. Cette définition est adoptée par exemple par la F.A.O. Les activités de transformation, conditionnement et commercialisation des produits de l'aquaculture, propriété d'une personne juridique, comportent en effet des particularités qui les distinguent de la pêche.

L'aquaculture concerne aussi bien les eaux douces que les eaux saumâtres et marines. À la différence de la pêche, dont 90 % des apports mondiaux proviennent des océans, les produits de l'aquaculture (animaux et végétaux) sont issus à peu près à parts égales entre eaux douces et salées.

3-2-historique

L'élevage des animaux aquatiques est une idée très ancienne. La pisciculture d'eau douce très simple s'est développée 1 500 ans avant notre ère en Chine et en Inde. Le premier traité de pisciculture connu a été écrit par Fan-Li en 475 avant J.-C. A la même époque, la pisciculture extensive en lagune a été développée par les Étrusques et les Romains sur les rives de la Méditerranée, alors que le grossissement d'huîtres est entrepris en Grèce, puis en Italie. C'est à partir de l'observation des animaux aquatiques conservés en viviers et avec des techniques de pêche archaïques que les premières formes d'élevage se sont développées. Ainsi, les juvéniles de nombreuses espèces nées en mer pénètrent naturellement dans les lagunes qui communiquent avec la mer. Leur capture après une saison de grossissement en lagune, lors de leur retour en mer à l'aide de grilles, de claies ou de filets barrant les passes a constitué une forme primitive d'aquaculture. Cette technique, à mi-chemin entre la pêche et l'élevage, a évolué au cours de la Renaissance vers un véritable engraissement. En Europe centrale, la pisciculture d'eau douce de la carpe prend naissance au Moyen Âge et la maîtrise complète du cycle biologique conduit rapidement à

une véritable domestication de cette espèce. L'élevage de la moule remonte à la même époque, avec l'introduction en baie de l'Aiguillon des bouchots (gros pieux de bois enfoncés dans le sol permettant la fixation des jeunes moules et leur croissance). C'est aux XVIII^e et XIX^e siècles que débute, au Japon, puis en Europe et notamment en France, l'élevage des huîtres à partir du captage de naissains sauvages sur des supports appropriés, les collecteurs. À la fin du XIX^e siècle, l'élevage d'huîtres était très intense.

3-3-Les différents types d'aquaculture

Il existe plusieurs types d'aquaculture selon l'organisme aquatique :

- La pisciculture, c'est l'élevage de poissons ;
- La conchyliculture, c'est l'élevage de coquillages. Les types les plus courants de conchyliculture sont :
 - o l'ostréiculture (élevage des huîtres),
 - o l'halioticulture (élevage des ormeaux),
 - o la mytiliculture (élevage des moules),
 - o la pectiniculture (élevage de coquilles Saint-Jacques ou de pétoncles) ;
- L'élevage de crustacés se décompose en différents volets selon les espèces :
 - o L'astaciculture est l'élevage des écrevisses,
 - o La pénéculture ou élevage des crevettes
- L'algoculture, ou culture d'algues.

3-4-L'évolution de la consommation

L'homme consomme chaque année 130 millions de tonnes de produits aquatiques depuis le début du XXI^e siècle. Cette consommation représente 7 fois la consommation de 1950. Les produits aquatiques constituent la principale source de protéines animales pour de nombreux pays d'Asie, d'Afrique et d'Amérique latine. Ils proviennent pour 65% de la pêche et pour 35% de l'aquaculture. La production de la pêche mondiale étant stabilisée depuis 1990, aux environs de 90 à 95 millions de tonnes par an (dont une partie est utilisée

pour la fabrication de farines afin de satisfaire le développement de l'aquaculture) l'aquaculture est appelée à accroître encore sa production pour répondre aux besoins des consommateurs. En effet, indépendamment de l'accroissement démographique, la consommation mondiale moyenne de poissons et d'invertébrés marins a doublé en plus de quarante ans, passant de 8 à 16 kilogrammes par personne et par an entre 1960 et 2004. La consommation augmente en particulier dans les des pays émergents, au premier rang desquels on trouve la Chine et l'Inde. La progression de la population mondiale attendue à 7,5 milliards d'habitants d'ici 2020 va accroître significativement la demande. Selon les économistes cette demande devrait atteindre 30 kilogrammes par personne et par an pour l'année 2020. Si la pêche reste stable en raison des difficultés d'accès à la ressource et son nécessaire contrôle, la production aquacole annuelle mondiale devrait pour satisfaire cette nouvelle demande passer de quelque 40 millions de tonnes actuellement à plus de 150 millions de tonnes d'ici à 2020. Un tel objectif n'est pas irréaliste, la principale limitation étant certainement celle de la disponibilité des espaces nécessaires et des ressources alimentaires.

3-5-Aliments pour les poissons d'élevage

La farine de poisson est utilisée pour l'alimentation des animaux aussi bien terrestres qu'aquatiques. Une forte concurrence existe entre les deux secteurs, le secteur terrestre en utilisant 43% contre 57% pour les poissons et crustacés. En aquaculture, la farine de poisson est utilisée pour diverses espèces carnivores et omnivores. Les crevettes, les poissons exclusivement marins et les salmonidés sont les principaux consommateurs. Concernant l'huile de poisson, l'aquaculture en consomme 87% dont la moitié est utilisée par les salmonidés. L'origine de la fabrication de la farine de poisson a été un moyen de tirer parti des surplus de la pêche dédiée à la consommation humaine. Une pêche spécialisée a été par la suite créée et prit le nom de pêche minotière. Face à la demande croissante de l'aquaculture et à la stagnation prévisible de la production des farines et huiles, issues de la pêche, il a été nécessaire de réduire la proportion de farine de poisson dans les aliments des élevages aquacoles. La recherche s'est donc orientée vers d'autres sources protéiques comme la matière première végétale tout en cherchant à conserver les qualités nutritionnelles et organoleptiques des poissons d'aquaculture.

3-6-La production aquacole du monde

La pisciculture est la branche la plus importante de l'aquaculture. Elle produit un peu plus de 28 Mt, dont 23,8 Mt de poissons d'eau douce, 2,8 Mt de poissons amphibiotiques (appelés également diadromes) et plus de 1,4 Mt de poissons marins. En mer, c'est l'élevage de coquillages ou conchyliculture qui domine avec 13,24 Mt, l'élevage de crustacés atteignant 3,67 Mt dont 2,47 Mt pour les seules crevettes. Chez les végétaux, ce sont les algues marines, principalement les algues brunes (phéophycées) et les algues rouges (rhodophycées) dont la consommation est largement répandue dans les pays du Sud-Est asiatique, qui représentent plus des trois quarts de la production mondiale de l'aquaculture végétale. Du point de vue de la production, l'aquaculture l'emporte très largement sur la pêche pour les poissons d'eau douce en 2004 avec 23,8 Mt contre 7,8 Mt, plus faiblement pour les poissons amphibiotiques avec 2,8 Mt contre 1,5 Mt. Mais pour les poissons marins, qui constituent l'essentiel de la pêche il y a 1,4 Mt de production aquacole contre 71,5 Mt de prises, ainsi que pour les crustacés 3,67 Mt contre 6,19 Mt. Alors que pour les mollusques, l'aquaculture l'emporte à nouveau (13,24 Mt contre 7,31 Mt) et surtout pour les algues et les autres végétaux non identifiés (13,92 Mt contre 1,43 Mt). Sur le plan géographique, on constate que l'Asie vient largement en tête avec 40,5 Mt de production, devant l'Europe avec 2,2 Mt, l'Amérique du Sud 1,1 Mt, l'Amérique du Nord 0,9 Mt et l'Afrique avec seulement 0,6 Mt.

3-7-L'aquaculture en France

Depuis plus de 1000 ans, la France a une tradition de pisciculture en étangs dans le Limousin et les Dombes par exemple. De nombreux viviers avaient été créés par les moines avec une utilisation intensive des retenues de moulins dont les vers de farine et les déchets de meunerie alimentaient les truites et d'autres poissons ainsi sédentarisés. Au début du XX^e siècle environ 6 000 exploitants d'étangs étaient déclarés, surtout localisés en Région Centre, Rhône-Alpes et Lorraine pour une production de 12 000 tonnes de carpes, gardons, brochets et tanches, représentant un chiffre d'affaires d'environ 16 millions d'euros. De cette production, 80 % étaient réservés à la consommation, 12 % servaient au repeuplement pour la pêche de loisir et 8 % au repeuplement des rivières. Les pratiques les plus récentes sont la salmoniculture en rivière puis la pisciculture marine. Au début des années 2000, 60 000 tonnes de poissons étaient produites par an, pour environ 222 millions d'euros de chiffres d'affaires. La salmoniculture a permis de produire environ 41 000

tonnes de truites arc-en-ciel (Bretagne et Aquitaine surtout), 52 producteurs en mer ont livré 3 000 tonnes de bar, 1 200 tonnes de dorades royales et 910 tonnes de turbots.

La conchyliculture (huîtres, moules et coquillages) s'est fortement développée sur la façade atlantique. Ce type d'aquaculture produit 90 300 tonnes d'huîtres, 4 100 tonnes d'autres coquillages, par 52 600 concessions sur 18 100 hectares du domaine public et 1 570 km de littoral.

Des statistiques mondiales de l'aquaculture sont mises à disposition par l'Ifremer et des stratégies d'innovation développées, incluant des projets de domestication de nouvelles espèces de poisson d'eau douce en France, sont étudiées par l'INRA et le CNRS.

3-8-Economie et aquaculture

En 2008 l'aquaculture mondiale a produit 68,3 millions de tonnes de poissons et plantes aquatiques dont 28,8 millions de tonnes de poissons, alors qu'au début des années 1950, cette production ne dépassait pas le million de tonnes. En 1970 la part de l'aquaculture dans la production totale de poissons était de 4%, en 1980 de 8%, alors qu'en 2002 la production est arrivée à 30% et en 2008 à 36,9 %. Un peu plus de la moitié des élevages (57,7 %) sont en eau douce, et les espèces les plus fréquemment élevées sont des cyprinidés (carpes) et des cichlidés (tilapia). Dans les élevages marins, on élève surtout des saumons, du thon, des daurades, des bars, des mollusques dont les huîtres et les moules, ainsi que des crustacés comme les crevettes. La production des plantes aquatiques a été de 13,9 millions de tonnes en 2004, principalement représentée par la laminaire japonaise avec 4,5 millions de tonnes de production, le wakamé avec 2,5 millions de tonnes et le nori avec 1,3 million de tonnes. Les principaux pays producteurs sont la Chine (10,7 millions de tonnes), les Philippines (1,2 million de tonnes), la Corée du Sud (0,55 million de tonnes) et le Japon (0,48 million de tonnes). Depuis 1970, la production a connu une croissance annuelle moyenne de 8,8 %. Il s'agit de la plus grande expansion dans le secteur de la production alimentaire. Au cours de la même période, la pêche a progressé de 1,2 % et l'élevage terrestre de 2,8 %. Les pays qui ont enregistré la plus forte croissance annuelle moyenne entre 2000 et 2008, hors algoculture, sont la Birmanie (27,1 %), le Vietnam (22,1 %), le Chili (10,1 %), l'Indonésie (10 %) et l'Égypte (9,3 %).

En 2008, l'aquaculture employait environ 10 800 000 personnes dans le monde. Comparativement à l'élevage d'animaux terrestres, l'aquaculture présente l'avantage de nécessiter moins de nourriture et donc d'avoir un impact environnemental inférieur. En effet, les poissons étant des animaux à sang froid et vivant dans l'eau, ils utilisent moins d'énergie pour garder leur corps à température ou pour constituer une ossature.

3-9- Aquaculture labellisée «biologique»

L'aquaculture biologique est la forme de production aquacole qui répond à un cahier des charges imposant certains critères de production dits « *biologiques* ». Il s'agit en quelque sorte de l'équivalent aquacole de l'agriculture biologique. Cette aquaculture labellisée croît rapidement, même si elle ne concerne qu'une très petite niche dans le marché aquacole lui-même en très forte croissance. En 2005 l'aquaculture labellisée représentait moins de 0,1% de l'aquaculture mondiale, mais la demande et l'offre se développent dans de nombreux pays. L'aquaculture biologique concerne quelques espèces de poissons, quelques crustacés et de petites productions de mollusques.

4- Génomique et Elevage

4-1-La génomique

Les avancées les plus remarquables des biotechnologies agricoles sont dues aux recherches sur la structure des génomes et sur les mécanismes génétiques qui impactent fortement sur le plan économique. La génomique est une discipline qui progresse à grands pas et elle fournit des informations sur l'identité, l'emplacement, l'impact et la fonction des gènes qui ont une incidence sur les productions animales ou végétales. La génomique a permis le développement de disciplines nouvelles comme la protéomique et la métabolomique qui livrent des connaissances sur les protéines ainsi que sur leurs fonctions et interactions. Ces disciplines visent une compréhension au niveau moléculaire de la biologie des organismes.

Une gamme très vaste de technologies et de matériels nouveaux sont élaborés à un rythme rapide pour générer et traiter les informations sur la structure et la fonction des systèmes biologiques. Par ailleurs la bioinformatique aide non seulement l'acquisition et le stockage de données en grand nombre mais de plus elle aide la prédiction de la fonction des gènes.

Ainsi la comparaison des cartes physiques et génétiques et des séquences d'ADN de différents espèces de poissons permet de réduire sensiblement le temps requis pour identifier et sélectionner les gènes potentiellement intéressants.

4-2-L'élevage

La sélection est un phénomène naturel par lequel les individus les mieux adaptés se reproduisent en plus grand nombre alors que les autres ont tendance à disparaître. L'homme a sélectionné de façon artificielle les animaux qui l'entouraient soit de façon à éliminer les caractères négatifs en éliminant les individus malades ou faibles, soit de façon positive, en gardant les individus produisant plus de viande ou de lait par exemple. Dans le monde aquacole, cette sélection artificielle est devenue primordiale. En général, la sélection assistée par marqueurs (SAM) permet d'améliorer la précision et la vitesse de la sélection de l'espèce d'intérêt. Son efficacité par rapport à la sélection classique est d'autant plus grande que la liaison physique du caractère d'intérêt et des marqueurs de sélection est forte. L'utilisation optimale de la sélection assistée par marqueurs nécessiterait l'identification des gènes à effets quantitatifs et qualitatifs. Les avantages potentiels de la sélection assistée par marqueurs sont plus importants pour les caractéristiques contrôlées par de nombreux gènes, comme la résistance aux maladies. Les marqueurs peuvent aussi être utilisés pour accélérer ou améliorer l'introduction de nouveaux gènes d'une population à l'autre. L'utilisation de marqueurs génétiques polymorphes en amélioration génétique des poissons est discutée dans deux situations : dans une population issue du croisement entre deux espèces bien différenciées ou dans le cadre d'une sélection intrafamiliale. En général, il y aurait peu de gain à espérer par rapport à la sélection sur performances pour des caractères d'héritabilité assez élevée et faciles à mesurer. En revanche cette forme de sélection semble indiquée dans les cas suivants : sélection précoce des candidats à la reproduction, optimisation de l'introgession d'un gène marqueur, contrôle de caractères faiblement héréditaires ou difficiles à mesurer. Dans le cas d'une sélection intrafamiliale il semble que la combinaison d'une sélection sur performances avec une sélection sur marqueurs ne soit efficace que si un très petit nombre de marqueurs (un à trois) sont simultanément pris en compte.

Dans tous ces cas, des cartes de liaison génétique et/ou des cartes RH plus faciles à obtenir avec un grand nombre de marqueurs sont indispensables pour localiser et sélectionner les gènes qui affectent les caractéristiques d'importance économique chez les plantes ou les animaux.

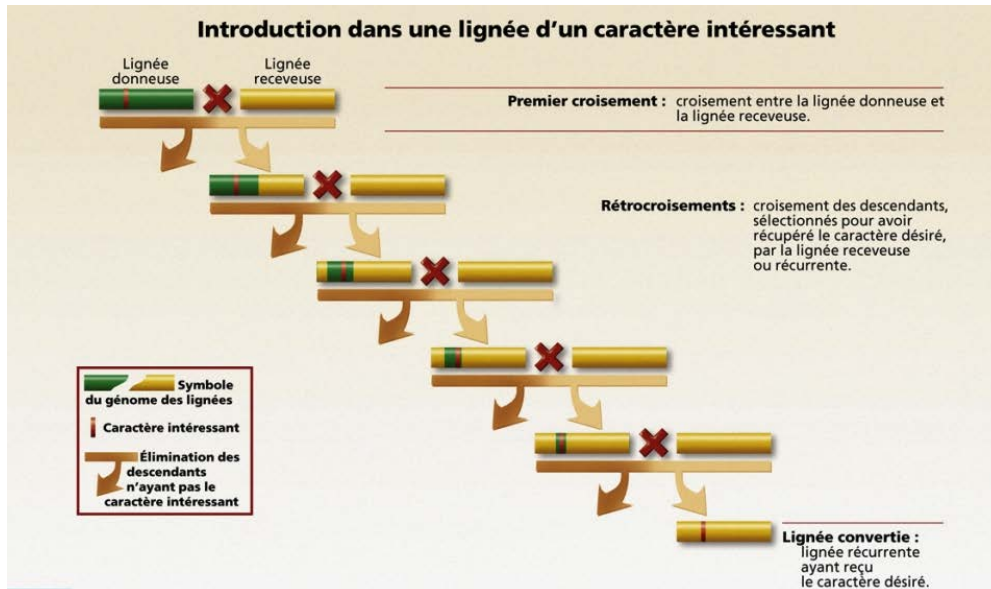


Schéma illustrant la méthodologie SAM

III- Analyse de la structure des génomes

Pendant de nombreuses années l'analyse de la structure des génomes ou génomique structurale s'est limitée à la cartographie de gènes et de marqueurs polymorphes avant d'évoluer vers le séquençage complet des génomes. L'action de cartographier consiste à déterminer les positions relatives d'éléments du génome pouvant être caractérisés de façon non ambiguë. Ces éléments, appelés marqueurs, sont très divers. Il peut s'agir de simples fragments d'ADN sans caractérisation fonctionnelle (segments anonymes), d'éléments répétés (microsatellites en particulier), de SNP « single nucleotide polymorphism » ou de gènes. La méthode de positionnement des marqueurs et la définition de la distance qui les sépare ont conduit à l'existence de plusieurs types de cartes que l'on peut schématiquement présenter en deux groupes : La cartographie de liaison (les cartes génétiques et les cartes d'hybridation irradiation RH) et la cartographie physique (les cartes de restriction et les cartes cytogénétiques).

Plusieurs méthodes sont utilisées pour cartographier les génomes, avec un niveau de

résolution variable, du chromosome jusqu'à la séquence nucléotidique complète du génome.

1- La cartographie physique

Il existe plusieurs types de cartes physiques en fonction de la nature des marqueurs utilisés. Celles consistant à positionner de grands clones chevauchants recouvrant l'intégralité d'un

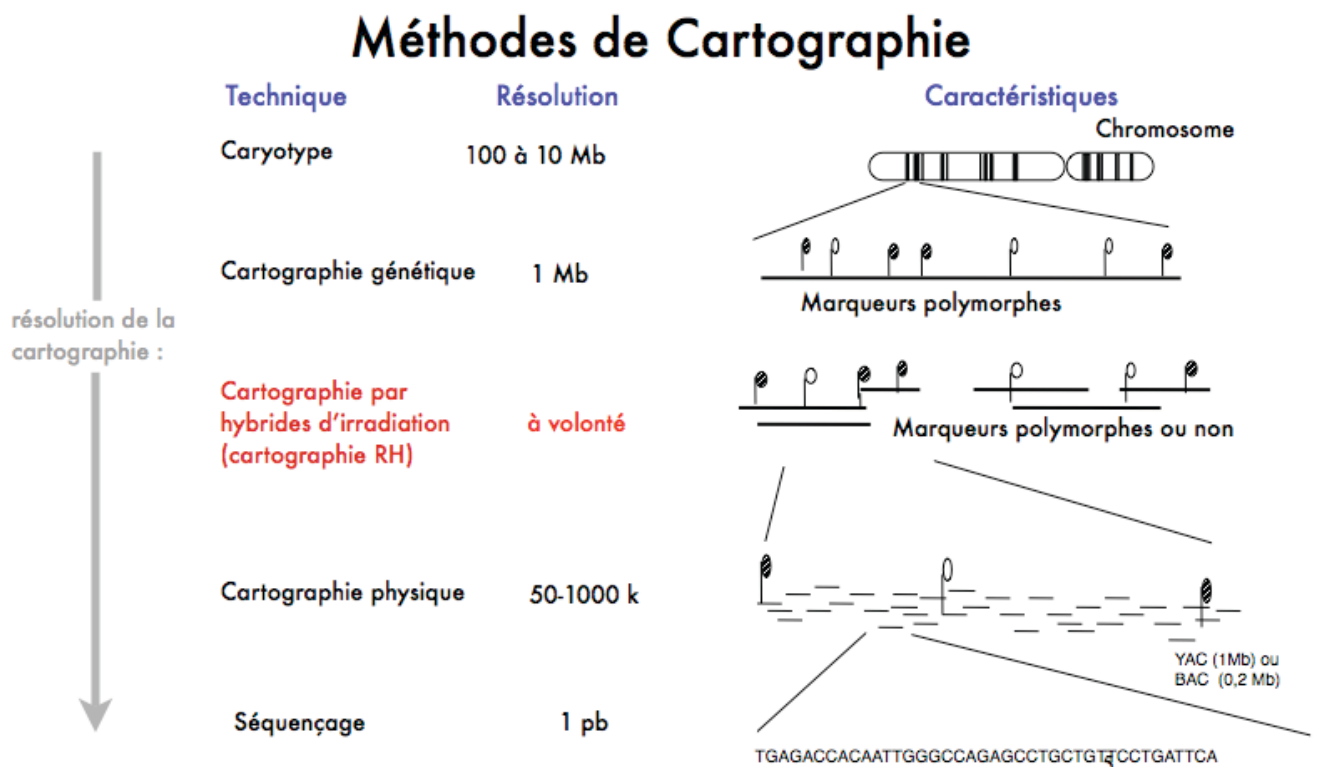


Figure 6 : Les différents niveaux de résolution des cartes et les techniques associées.

génomique ont constitué pendant longtemps la dernière étape avant le séquençage complet des génomes des espèces modèles comme l'homme (International Human Genome Sequencing Consortium, 2001), la souris (Mouse Genome Sequencing Consortium, 2002), le rat (Rat Genome Sequencing Project Consortium, 2004) et récemment des poissons modèles comme zébrafish (Zebrafish Genome Sequence handed over to Genome Reference Consortium 2010). En revanche, la stratégie de séquençage complet frontal dite en « shot gun » inventée par Frederick Sanger et beaucoup plus tard popularisée et adaptée aux grands génomes par Craig Venter a supplanté l'approche par clones ordonnés nécessitant la construction de carte préalable.

1-1- Les Caryotypes

Le caryotype est un inventaire ordonné de l'ensemble des chromosomes d'une espèce. Les chromosomes sont classés par ordre décroissant de taille en fonction de la position du centromère qui délimite les deux bras du chromosome. Les chromosomes sont découpés en régions délimitées par des bandes « repères » et notées de 1 à n sur chacun des bras chromosomiques, depuis le centromère jusqu'au télomère. Pour réaliser un caryotype une culture cellulaire (souvent de lymphocytes) de l'espèce d'étude est mise en présence de colchicine qui perturbe les fuseaux mitotiques et bloque les cellules en métaphase de la mitose. Ces cellules récoltées et incubées dans un milieu hypotonique sont ensuite fixées et étalées sur des lames. Cette préparation colorée, le plus souvent au Giemsa, entraîne l'apparition de bandes sur les chromosomes analysables au microscope et caractéristiques des différentes paires de chromosomes.

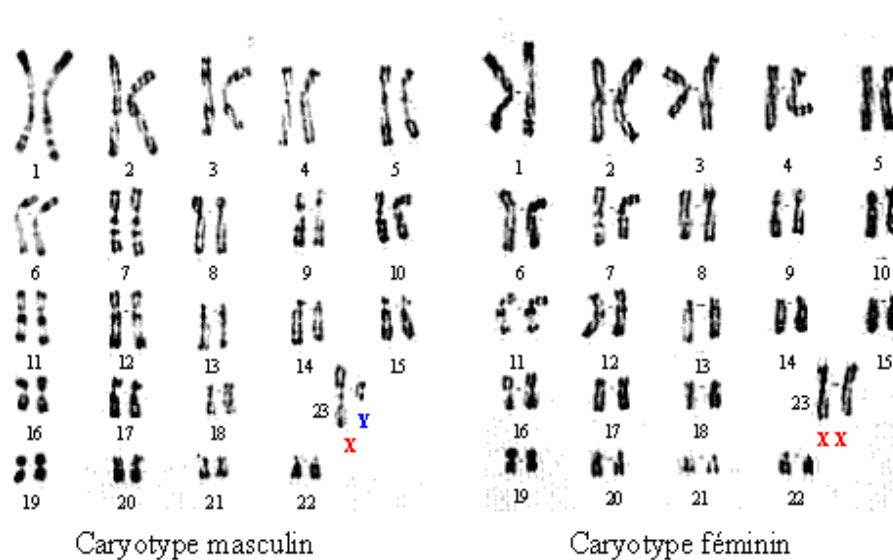


Figure 7 : *Caryotype humain (Caryotype masculin le chromosome 23 est XY, caryotype féminin le chromosome 23 est XX).* (Schrock, E et al 1996)

1-2-La Cartographie Cytogénétique

La cytogénétique ou l'hybridation moléculaire *in situ* permet une localisation de sondes généralement d'ADN sur des préparations de chromosomes métaphasiques. Ces hybridations permettent d'assigner des sondes marquées par radioactivité ou coloration à des chromosomes particuliers voire à préciser leur localisation au niveau des bandes chromosomiques, tandis que les hybrides somatiques ne permettent qu'une assignation chromosomique générale. Le niveau de résolution de l'hybridation moléculaire *in situ* se situe entre celui des hybrides somatiques et celui des hybrides d'irradiation.

Le niveau de résolution est celui de la bande chromosomique qui varie de 3 à 20 Mb. La technique d'hybridation *in situ* fluorescente ou FISH « fluorescent *in situ* hybridisation » a été mise au point par Viegas-Péquignot et collaborateurs (1989) et Lemieux et collaborateurs (1992). Elle a été légèrement modifiée par Hayes *et al.* (1992).

L'hybridation en parallèle de plusieurs sondes marquées avec des fluorophores différents a grandement augmenté les potentialités de cette technique.

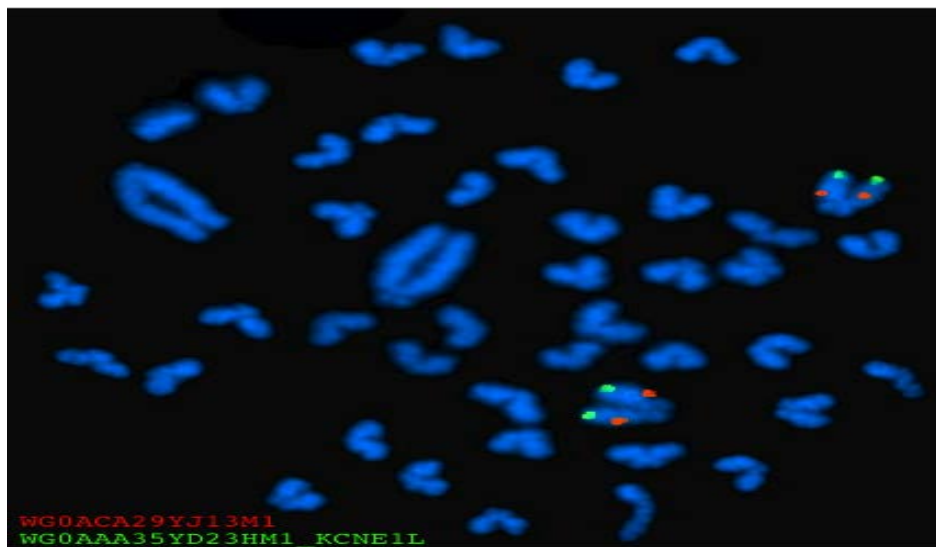


Figure 8 : hybridation *in situ* fluorescente des chromosomes de *Tilapia* (Guyon R *et al.*, 2012)

- marquage d'un BAC situé au niveau du centromère
- marquage d'un BAC situé au niveau du télomère

2- La cartographie de liaison

2-1-La Cartographie Génétique

À partir d'expériences de croisement de pois, Gregor Mendel (1822 -1884) a publié en 1866 ses célèbres lois de transmission des caractères héréditaires, la loi d'uniformité de la F1 et la loi de pureté des gamètes qui traduisent la ségrégation indépendante des allèles d'un gène. En 1902, Walter S Sutton (1876-1916) note la très forte ressemblance entre la transmission des gènes et celle des homologues de chaque paire de chromosomes et construit l'hypothèse selon laquelle les chromosomes constituent le support morphologique de l'hérédité. La notion de carte génétique remonte à 1913 avec les travaux sur la drosophile de Thomas H. Morgan (1866-1945) et Alfred H. Sturtevant (1891-1970). Ces travaux montrent d'une part qu'il existe une liaison de certains caractères avec la ségrégation des chromosomes sexuels et d'autre part que deux marqueurs sur le même chromosome sont en général transmis ensemble.

La carte génétique est basée sur l'ordonnement de marqueurs grâce à l'analyse statistique de leur ségrégation au cours de la méiose. Le centimorgan (cM) est l'unité de mesure des distances entre deux locus chromosomiques pour lesquels un produit méiotique sur cent est recombinant (1 crossing-over pour 100 méioses = 1cM). Les crossing-overs se produisent aléatoirement durant la méiose le long du chromosome. Plus la distance séparant deux locus portés par le même chromosome est importante, plus la probabilité de recombinaison entre ces locus est élevée. Le niveau de résolution d'une carte génétique dépend de sa densité en marqueurs et de la taille et de la structure de la population analysée. La cartographie génétique est souvent assez lourde à établir de sorte que les cartes génétiques ne comportent qu'un nombre limité de marqueurs et qu'il faut analyser plusieurs générations pour le même caractère. Par ailleurs les distances exprimées en cM ne correspondent que partiellement aux distances physiques.

2-2-La cartographie des cellules Hybrides d'irradiation (RH) :

Dans les années 1960, les premières cellules hybrides ont été observées avec la mise en évidence des phénomènes de fusion cellulaire spontanée (Barski *et al.*, 1960). Différentes techniques de fusion cellulaire ont été mises au point un peu plus tard, en utilisant le virus de Sendai inactivé par irradiation aux rayons UV (Schneeberger et Harris, 1966 et Yerganian et Nell, 1966) ou du polyéthylène glycol (PEG) (Pontecorvo, 1975). La formation de cellules hybrides capables de se multiplier indéfiniment et dont les noyaux contiennent des chromosomes des deux génomes a bénéficié de nombreuses avancées techniques (Harris *et al.*, 1965). L'hybridation de fragments de gènes marqués avec des sondes fluorescentes sur des chromosomes métacentriques d'hybrides somatiques permet de localiser ces séquences au niveau du chromosome entier ou au niveau d'une région large du chromosome dans le meilleur des cas. Dans les années 80 cette technique a permis la localisation de plusieurs gènes sur des chromosomes humains particuliers. En revanche elle ne permet que difficilement de déterminer l'ordre des marqueurs et ne permet pas d'envisager le positionnement d'un grand nombre de marqueurs. Le nombre d'hybrides somatiques est en général de moins de 50 lignées.

En 1975 il a été montré que des cellules diploïdes humaines irradiées à une dose létale pouvaient être « sauvées » par fusion avec des cellules receveuses de hamster (Goss et Harris 1975). Les hybrides irradiés sont obtenus de la même façon que les hybrides somatiques mais la cellule donneuse est soumise préalablement à une irradiation aux rayons X destinée à casser les chromosomes avant la fusion. Les clones de cellules

hybrides irradiées perdent aléatoirement, au cours de leur prolifération en culture, des fragments de chromosomes de la lignée donneuse, alors que les chromosomes de la lignée receveuse sont conservés. A terme on collectionne des clones de cellules hybrides contenant chacun une combinaison différente de fragments de chromosomes de la lignée donneuse. C'est en 1994 que l'idée de construire des cartes appelées d'hybrides irradiés à été proposé (Walter *et al.* 1994) et la première carte RH du génome humain a été réalisée l'aide d'un panel de 168 hybrides (Gyapay *et al.* 1996). A la différence de ce que l'on fait avec les hybrides somatiques, l'ADN de chacun des clones est extrait et l'ensemble des ADN constitue la collection d'hybrides. Une étape de caractérisation de la collection est ensuite nécessaire pour ne conserver que les échantillons les plus pertinents pour la cartographie. Cette caractérisation se fait grâce à la technique d'amplification PCR « polymerase chain reaction », avec des amorces spécifiques de marqueurs.

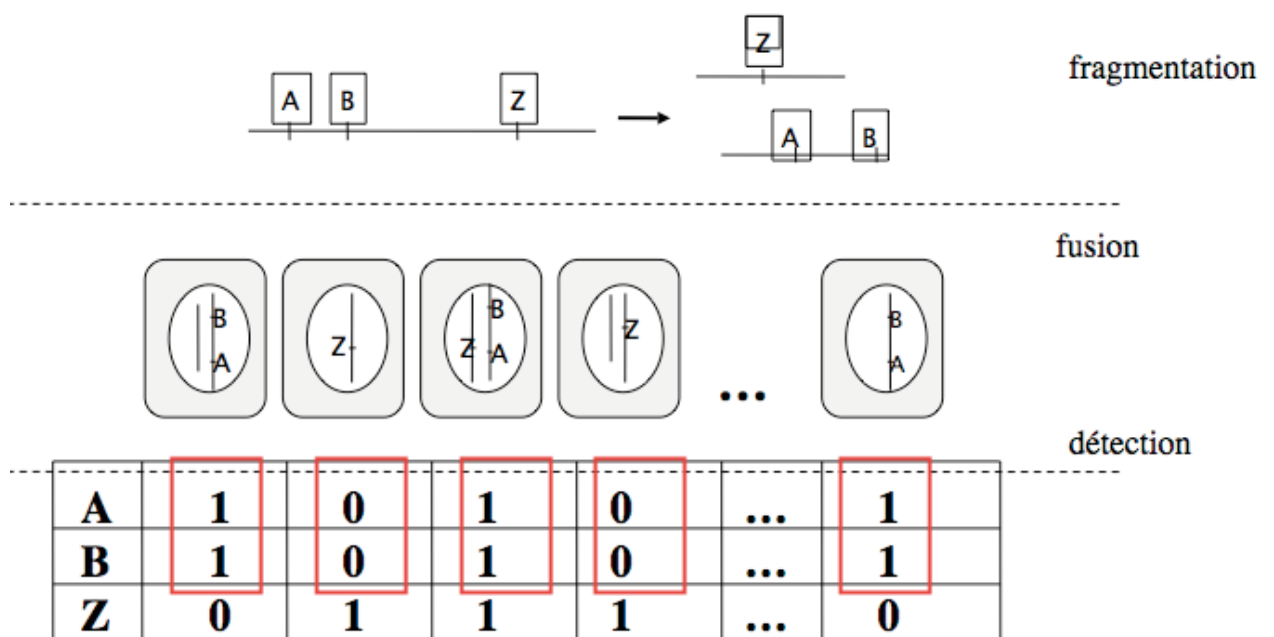


Figure 9 : Principe et Résultats de cartographie RH :

Soient trois marqueurs A, B et Z (A et B étant proches l'un de l'autre, alors que z est éloigné) : 1 : indique la présence du marqueur ; 0 : indique l'absence du marqueur dans la lignée considérée. La fréquence d'apparition conjointe d'une paire de marqueurs est inversement proportionnelle à la distance séparant les deux partenaires.

Les cartes d'hybrides d'irradiation ou RH « radiation hybrids » reposent sur l'analyse de la

fréquence des cassures chromosomiques induites artificiellement. Les cellules d'hybrides d'irradiation possèdent des fragments chromosomiques, d'autant plus petits que la dose d'irradiation de départ est élevée (3000 à 12 000 Rad). Ainsi, plus les fragments chromosomiques obtenus sont petits et plus la résolution de la cartographie sera grande. L'unité pour le calcul des distances est le centiRay (cR), avec 1 cR (N rad) qui correspond à une fréquence de cassure de 1 % entre deux marqueurs dans une collection construite avec une dose d'irradiation de N Rad de rayons X (Cribs et Schiller, 2000). Le rapport entre cR et distance physique réelle n'est pas stricte, mais beaucoup moins aléatoire que ne l'est le rapport cM/ distance physique. La valeur de ce rapport dépend de l'intensité d'irradiation et du nombre de lignées. Un panel d'hybrides d'irradiation est en général composé pour des questions pratiques d'analyse d'une petite centaine de lignées. Les cartes d'hybrides d'irradiation sont basées sur l'apparition de cassures dues aux rayons X, tout comme les cartes génétiques sont basées sur l'apparition de cassures dues aux recombinaisons lors de la méiose et sur la recherche de liaisons entre locus, mais l'avantage de la carte RH est que l'on peut positionner des marqueurs polymorphes ou non polymorphes et en beaucoup plus grand nombre compte tenu de la plus grande facilité d'analyse.

La présence des marqueurs dans les différentes lignées est le plus souvent déterminée par PCR sur les ADN extraits des lignées cellulaires, à l'aide d'amorces spécifiques. Les résultats sont ensuite analysés statistiquement à l'aide de divers programmes de bioinformatique.

3 - La cartographie comparée

La comparaison de cartes d'espèces différentes permet l'analyse des relations de synténie qui peuvent exister entre des espèces plus ou moins phylogénétiquement éloignées. La cartographie comparée consiste à déterminer les chromosomes ou segments chromosomiques porteurs des mêmes marqueurs et donc conservés dans plusieurs espèces. L'identification de ces régions conservées permet de transposer les données acquises dans des espèces modèles très étudiées (souris, rat, ou des poissons comme le poisson zèbre, medaka ou stickelback) à d'autres espèces moins étudiées. Les résultats de la cartographie comparée représentent un atout à l'identification de gènes d'intérêt et apportent des connaissances concernant l'évolution des génomes. Tous les types de cartes se prêtent à ces analyses mais les plus instructives sont celles obtenues par FISH ou par RH avec des marqueurs de gènes orthologues qui permettent de définir des régions de conservation de

synténie (Nadeau et Sankoff, 1998). Une région où la synténie est conservée entre deux espèces correspond à un segment contenant deux ou plusieurs gènes orthologues sur le même chromosome dans chacune des deux espèces, quelque soit l'ordre des gènes. En revanche, un segment d'ordre conservé est un groupe de synténie dont les gènes sont disposés dans le même ordre dans les deux espèces (O'Brien *et al.*, 1993). La conservation de synténie a été démontrée par la cartographie des gènes orthologues dans plusieurs espèces, même dans le cas d'espèces éloignées (Andersson *et al.*, 1996 ; Nadeau et Taylor, 1984). En 1991 une première carte de synténie a été publiée synthétisant l'information connue dans 28 espèces (O'Brien, 1991). Mais la résolution de cette carte était faible, car le nombre de gènes localisés en commun dans ces espèces était réduit. Deux autres études ont permis d'affiner la première carte contenant respectivement, 321 et 410 locus d'ancrage (O'Brien *et al.*, 1993 et Lyons *et al.*, 1997). Le séquençage complet de génomes permet également des comparaisons de synténie mettant en évidence non seulement la conservation de blocs de synténie importants en taille mais également la conservation de l'ordre des gènes orthologues.

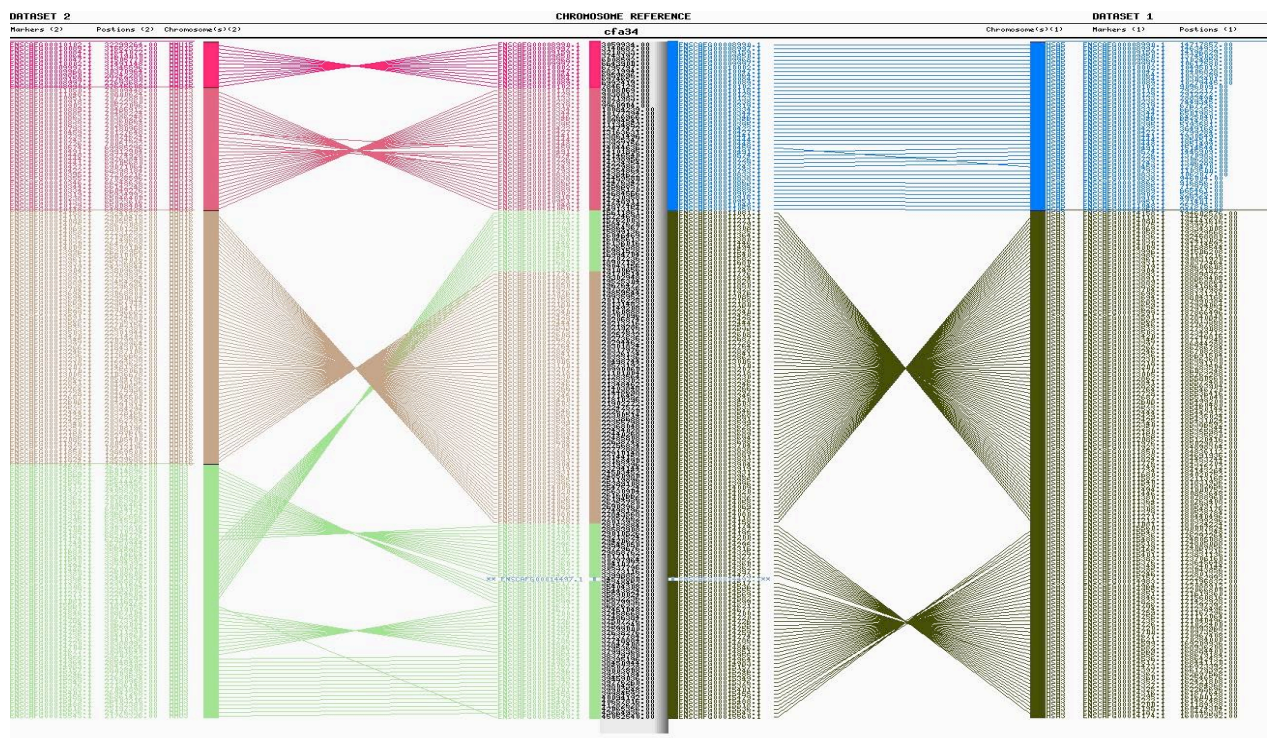


Figure 10 : Représentation graphique d'une carte de synténie de trois espèces (homme/souris/chien) réalisée avec le programme AutoGRAPH (Derrien T *et al.* 2007). Une fraction du chromosome 34 de chien, ici l'espèce de référence, est présentée au milieu

de la figure. A droite est présentée la région correspondante humaine et à gauche celle de la souris. Pour chaque espèce, les marqueurs sont identifiés par leurs coordonnées génomiques. Les lignes obliques et colorées reliant les génomes des trois espèces montrent la conservation de l'ordre des gènes.

4-Séquençage

Le développement récent des méthodes de séquençage à haut débit de l'ADN permet d'envisager le séquençage du génome d'un grand nombre d'espèces. Toutefois ces nouvelles approches présentent le défaut de fournir des longueurs de lecture faibles qui rendent très difficile l'assemblage *de novo* des données brutes. Par ailleurs même si le coût du séquençage a fortement diminué, il ne permet pas d'envisager un séquençage complet pour toutes les espèces d'intérêt, d'où la production de séquençage léger (4X pour des données de type Sanger ou 20/ 30 X pour des données NGS). Si un tel séquençage permet d'avoir une très bonne idée du niveau d'analogie entre espèces, il ne fournit aucune ou peu d'information sur la structure des génomes car il exclut tout assemblage complet. Ainsi loin d'avoir rendu obsolète la construction de cartes, les méthodes de séquençage à haut débit demandent pour être pleinement efficaces de disposer de cartes denses (Hitte *et al.*, 2005 ; Lewin HA *et al.*, 2009).

Au cours de ces dernières années, plusieurs programmes de cartographie et de séquençage de génomes de poissons ont été menés. L'un des derniers a concerné les génomes de cinq cichlidés : *O. niloticus*, *P. nyererei*, *H. Burton*, *N. brichardi* et *M. zebra*. L'intérêt de ce groupe de poissons est double. D'une part le tilapia est l'une des espèces les plus importantes en aquaculture et d'autre part le groupe des cichlidés de par sa grande diversité phénotypique et la rapidité de sa diversification constitue un modèle de choix pour l'étude de l'évolution. Nous avons participé à ce programme en construisant une carte RH dense du génome de tilapia qui a servi à l'assemblage des données de séquence. Par ailleurs nous avons été responsables de l'annotation et de l'identification des gènes codant pour les récepteurs olfactifs (OR) et des « trace amine-associated receptors » (TAAR).

VI - Evolution et développement des familles des récepteurs olfactifs (OR) et des TAAR « Trace Amine-Associated Receptors »

Toutes les espèces animales possèdent des capacités sensorielles permettant d'assurer leur survie en interagissant avec leur environnement immédiat. Ces sens leur permettent de détecter et de caractériser leur milieu, notamment en assurant des fonctions essentielles,

telles la recherche de nourriture, la détection de prédateurs, l'établissement de relations sociales ou encore la reproduction. Ainsi lorsqu' un poisson d'un banc se fait attaquer, il secrète des phéromones qui sont détectées par les autres poissons du banc.

1-L'appareil olfactif

L'olfaction est un sens prédominant chez la plupart des animaux qui s'appuient très largement sur celle-ci pour toutes leurs fonctions vitales. Chez les animaux aquatiques la détection s'effectue en milieu liquide. Les molécules odorantes doivent être solubles et dissoutes dans l'eau à la différence de ce qui se passe en milieu aérien où les molécules odorantes sont volatiles et secondairement dissoutes dans le mucus. Les molécules odorantes sont détectées et analysées lorsque l'eau circule dans les cavités nasales. Les poissons, dans leur ensemble, auraient un odorat de sensibilité comparable à celui de nombreux mammifères, en revanche la variété des molécules détectées est vraisemblablement plus faible.

Les poissons ne respirent pas de la même façon que les mammifères. Ils possèdent des narines mais celles-ci leurs servent uniquement pour l'odorat. Elles sont au nombre de deux et parfois quatre chez certaines espèces. Les OR sont au niveau d'un épithélium olfactif qui tapisse les parois internes des narines. Le courant d'eau entre par un orifice antérieur et sort par un orifice postérieur mettant en contact l'épithélium olfactif avec le milieu extérieur (Laberge F et Hara TJ, 2001).

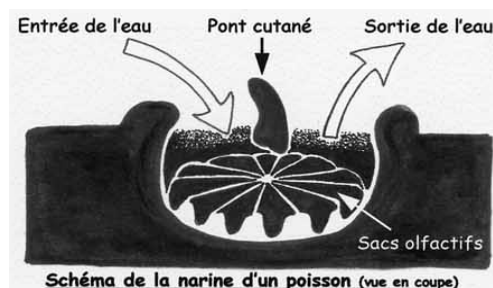


Image partielle de la tête d'un poisson dans laquelle la narine est entourée (à gauche) et schéma représentant le flux d'entrée et sortie de l'eau baignant l'épithélium olfactif (à droite).

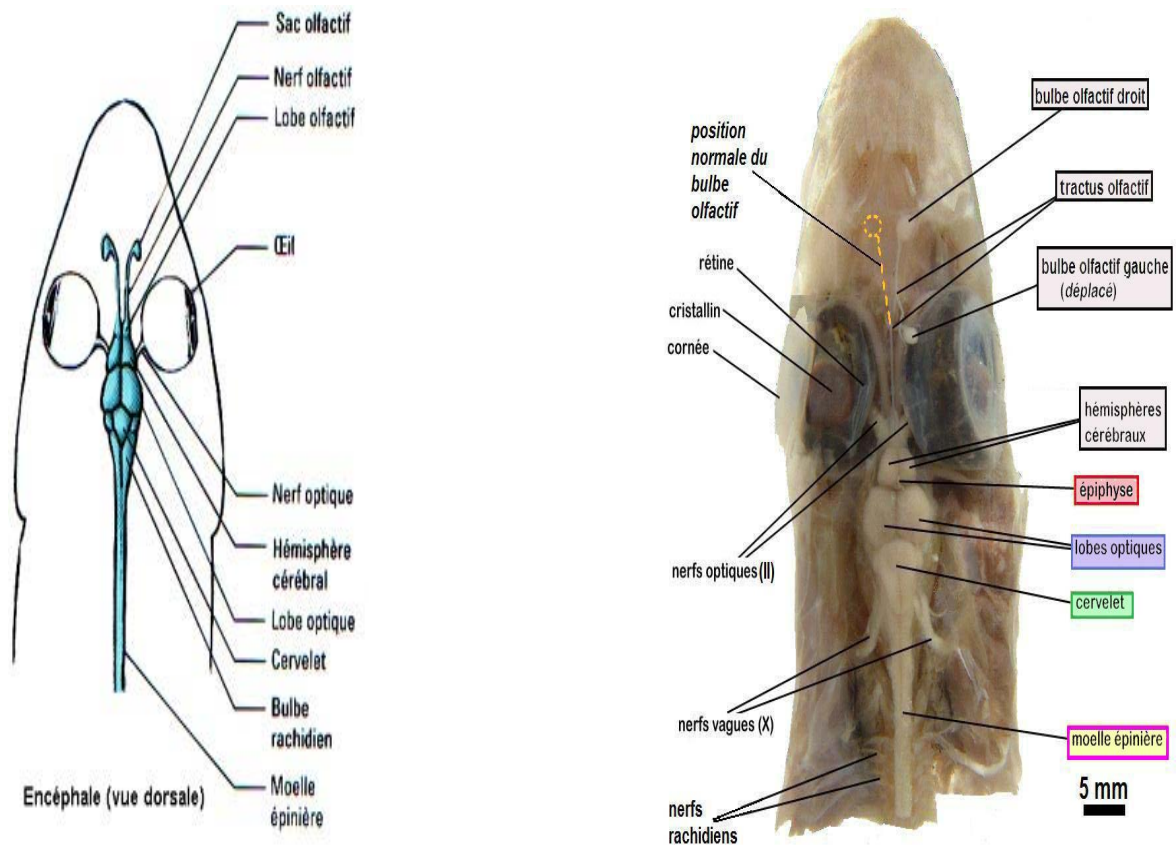


Figure 11 : Encéphale d'un poisson

(Spécimen UNSCIBA.Z. 000673 de la collection pédagogique de l'Université de Nantes J. Baudet).

Dans les années 1990 on pensait que la muqueuse olfactive des poissons était constituée d'une seule unité fonctionnelle (Ngaib *et al.*, 1993), mais par la suite des analyses quantitatives plus rigoureuses, utilisant des sondes spécifiques permettant de préciser la position des neurones, ont révélé la présence d'au moins trois zones distinctes (Wete *et al.*, 1996).

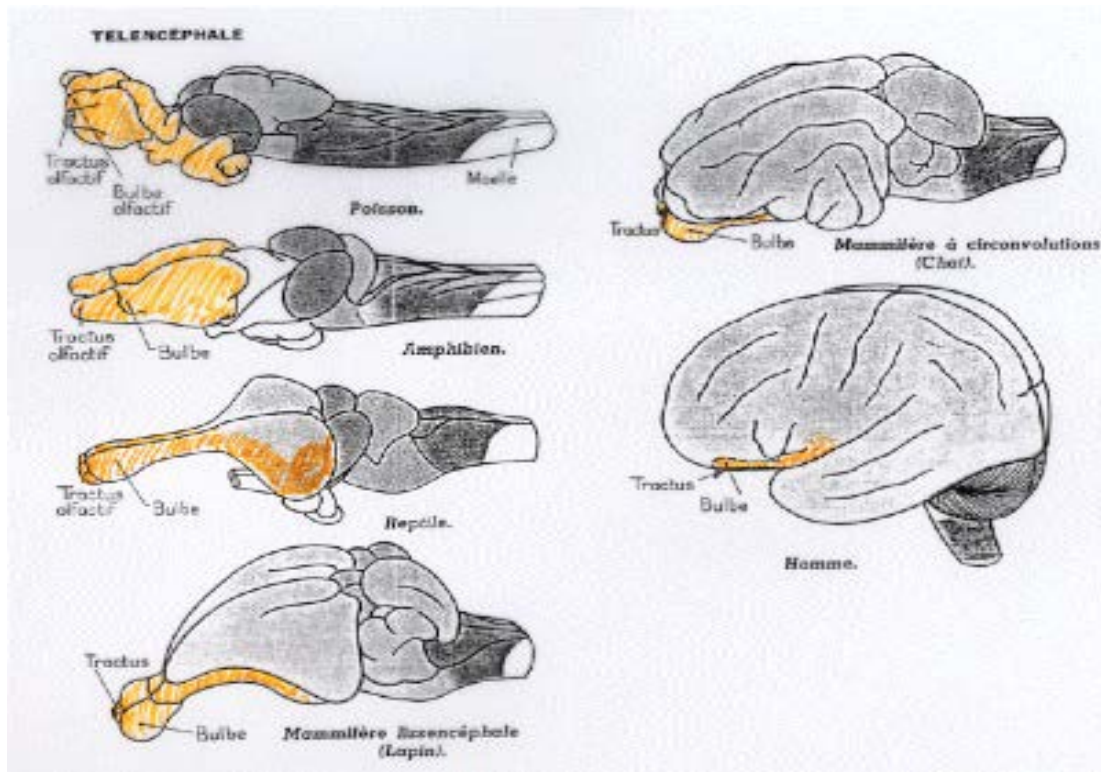


Figure 12 : schéma montrant l'évolution du système olfactif central chez les vertébrés.

De près de 50% de la masse totale du cerveau chez les poissons, on arrive à 3% chez le chat et à moins de 0,5% chez l'homme. (*Généralités sur la sensorialité*, R. Pujol; 2005)

Le cortex moderne des mammifères dérive du rhinencéphale des poissons. Le fait que l'ancienne structure olfactive des poissons ait donné naissance aux hémisphères cérébraux reflète l'importance qu'avait l'olfaction chez les premiers mammifères et qu'elle possède encore aujourd'hui pour beaucoup d'entre eux. Une hypothèse informelle suggère que cette évolution pourrait être due à l'absence d'information portée par les sensations olfactives. Alors que les signaux auditifs peuvent par eux-mêmes fournir des informations sur la taille, la position et le mouvement des autres animaux, ce n'est pas le cas des signaux olfactifs. Pour être utiles, ils doivent être associés à des traces mnésiques qui les mettent en relation avec des souvenirs visuels ou auditifs, d'autant que les premiers mammifères étaient vraisemblablement des animaux nocturnes de la taille d'une souris. Les hémisphères cérébraux se sont donc développés et ont reçu des afférences visuelles, auditives et somatiques en vue d'intégrer toutes ces informations avec les signaux olfactifs. Ceci n'est qu'une hypothèse et il n'existe aucun moyen de la vérifier, l'évolution permettant de constater la succession des étapes, mais pas d'en donner les raisons, surtout pour un organe fossilisant aussi mal que le cerveau.

2-Evolution du nombre des gènes codant les récepteurs olfactifs

Les récepteurs olfactifs constituent la plus grande famille multigénique mise en évidence chez les mammifères. Les gènes qui les codent représentent plus de 1 % du génome chez l'homme, soit 906 gènes pour 322 protéines fonctionnelles (Glusman *et al.*, 2001), 1094 gènes chez le chien et 1493 gènes chez le rat (Quignon *et al.*, 2005). Les récepteurs olfactifs sont très anciens du point de vue de l'évolution. En effet, on les observe chez de nombreux organismes comme les insectes, les nématodes, les poissons ou les vertébrés. (Voir l'arbre de vie)

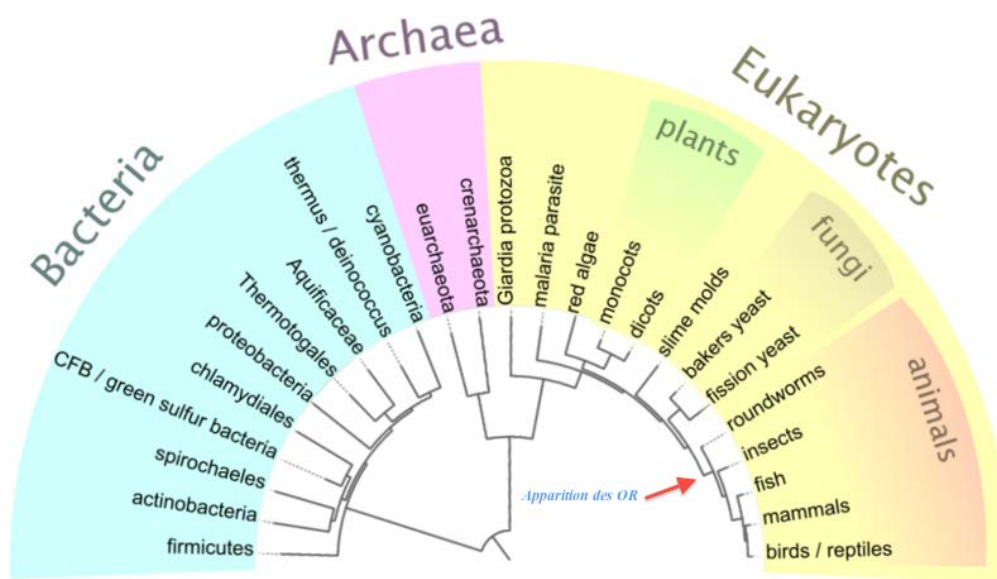


Figure 13 : Arbre de vie et apparition des récepteurs olfactifs. (Woese *et al.* 1990 modifié)
La flèche rouge positionne l'apparition des OR dans l'arbre de vie

L'évolution du nombre de récepteurs olfactifs chez les vertébrés montre une phase d'expansion avec l'apparition de la vie terrestre, puis une régression du nombre de gènes fonctionnels, concomitante de la station verticale des primates et de l'envol des oiseaux. Cependant, l'homme a compensé cette perte relative par un développement du cortex et d'autres capacités sensorielles. L'évolution de ces répertoires marque la plupart des grandes transitions évolutives des vertébrés, comme par exemple la transition du milieu marin au milieu terrestre (Niimura and Nei 2005), le retour au milieu aquatique de certains vertébrés (Kishida *et al.*, 2007), la diversification des rongeurs (Emes *et al.*, 2004; Grus and Zhang 2004) ou encore la spécialisation du mode de vie humain (Gilad *et al.*, 2003).

L'évolution rapide de ces répertoires de récepteurs par alternance de phase d'expansion et de contraction ainsi que la sélection positive dont ils sont l'objet témoignent du rôle clé que ces gènes jouent dans l'adaptation à de nouveaux milieux, de nouveaux modes de vie

ou de conditions environnementales (Niimura and Nei 2006). Chez l'homme on les trouve rassemblés en groupes plus ou moins importants sur tous les chromosomes exceptés les chromosomes 20 et Y (Rouquier *et al.*, 2000) alors qu'ils ne se trouvent que sur 24 des 39 chromosomes canins. Pour autant cette moindre dispersion n'est qu'apparente. Elle est en fait due au fractionnement du caryotype canin (39 chromosomes autosomiques). En effet on observe une grande conservation de la synténie (Quignon *et al.*, 2005). Des pseudogènes en nombre variable selon les espèces accompagnent les gènes OR. Ainsi chez l'homme l'ensemble du répertoire comporte près de 70% de pseudogènes, chez les singes de l'ancien monde (les plus évolués) ce pourcentage est d'environ 30% et chez les singes du nouveau monde il varie entre 15 et 20%, ce qui est similaire à ce qui est trouvé chez la souris (Rouquier *et al.*, 2000 ; Gilad *et al.*, 2004 ; Rouquier et Giorgi, 2007) le rat ou le chien (Quignon *et al.*, 2005). Il a été observé que le singe hurleur, seul singe du nouveau monde à posséder comme l'homme et les singes de l'ancien monde une vision en couleur (Gilad *et al.*, 2004), possédait un nombre plus important de pseudogènes (environ 30%). L'acquisition de la vision en couleur aurait selon certains auteurs fortement abaissé la pression de sélection sur le système olfactif et favorisé une pseudogénéisation importante (Rouquier et Giorgi, 2007). On peut aussi noter que l'apparition de la station debout de l'homme aurait défavorisé le système olfactif en éloignant l'organe olfactif des sources odorantes, le plus souvent proches du sol (Shepherd GM., 2004).

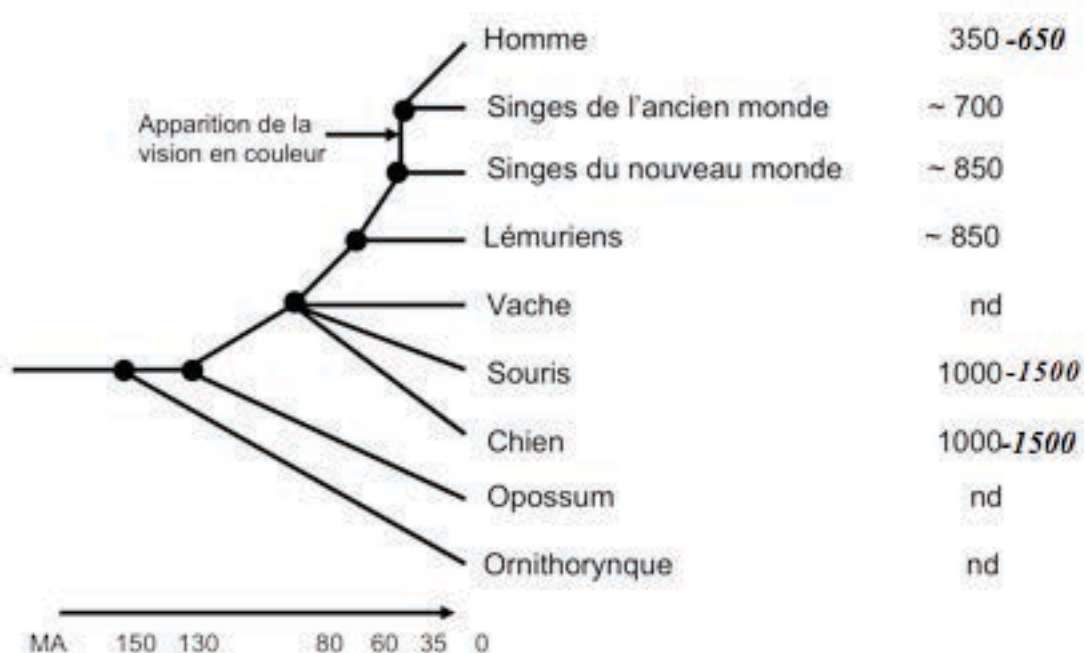


Figure 14 : Evolution du nombre de récepteurs olfactifs chez les mammifères.

Représentation d'un arbre phylogénétique indiquant l'échelle de temps pour la séparation des différents groupes en millions d'années (MA). L'apparition de la vision en couleur chez les vertébrés est indiquée par une flèche. A droite, comparaison du nombre de récepteurs olfactifs pour chaque espèce (Rouquier et Giorgi, 2007).

3-Les différents récepteurs olfactifs et leur évolution

A coté des gènes OR, acteurs majeurs, il existe trois autres entités ayant un rôle chimiosensoriel : les récepteurs voméronasaux (VR) (Dulac et Axel 1995), les TAAR (trace amine-associated receptors) (Liberles et Buck 2006) et les FPR (formyl peptide receptor like proteins (Liberles *et al.*, 2009). Les récepteurs voméronasaux (VR) ont un rôle majeur dans la médiation des réponses sociales et de défense aux signaux chimiques sexuels et spécifiques (phéromones) de la plupart des vertébrés terrestres (Keverne 2004; Brennan and Binns 2005).

Il existe deux classes de VR : les V1R découverts en 1995 (Dulac et Axel 1995) et une deuxième classe V2R découverte dans l'organe voméronasal du rat et de souris (Herrada et Dulac 1997). Les gènes V1R et V2R sont organisés en groupes séparés sur plusieurs chromosomes. Chez les rongeurs le répertoire des gènes VR est plus compact et il est distribué sur six chromosomes chez le rat et huit chez la souris. Les poissons possèdent peu de gènes V1R et une trentaine de V2R. Les gènes VR y sont exprimés au niveau de l'épithélium olfactif car les poissons ne possèdent pas d'organe voméronasal. L'évolution de la communication olfactive, qu'elle soit sexuelle, sociale ou interspécifique passe probablement par l'évolution des répertoires des deux classes de récepteurs voméronasaux. Par exemple, ces répertoires sont marqués par de récentes expansions chez les rongeurs (Lane *et al.*, 2002; Zhang *et al.*, 2004; Yang et al. 2005; Young *et al.*, 2005) chez lesquels la communication olfactive est centrale dans la plupart des interactions. Par ailleurs, ces répertoires sont souvent caractéristiques de chaque espèce et divergent entre espèces proches (Grus and Zhang 2004; Lane *et al.*, 2004). Au delà de leur rôle dans l'adaptation, il est donc possible que les récepteurs voméronasaux participent à l'évolution de barrières reproductives et donc à la spéciation.

Les TAAR forment une famille spécifique de GPCR chez les vertébrés. Les TAAR étaient initialement considérés comme des récepteurs de neurotransmetteurs, mais en 2006 il a été montré que les TAAR de souris fonctionnaient comme des récepteurs chimiosensoriels dans l'épithélium olfactif. Chez les vertébrés les dynamiques évolutives de la famille des gènes TAAR ont été clarifiées, par l'identification des répertoires des gènes et des

pseudogènes TAAR chez des poissons téléostéens (Zebrafish, stickleback, fugu et médaka), les lamproies (H. Yasuyuki and N. Mutsumi 2007) et les mammifères [humains, souris et Rat (Gloriam D.E. *et al.*, 2005)]. Des recherches ont révélé que les poissons avaient de nombreux gènes TAAR (13-109 gènes), alors que le nombre de gènes TAAR est relativement faible chez les tétrapodes (3-22 gènes). L'analyse phylogénétique de ces gènes a indiqué que la famille des gènes TAAR se divise en sous-familles qui ont divergé avant la divergence des poissons à nageoires rayonnées et les tétrapodes. Chez les tétrapodes, pratiquement tous les gènes TAAR sont localisés en une seule région spécifique de leurs génomes et forment un « cluster » de gènes, mais chez les poissons, les gènes TAAR sont dispersés sur plusieurs régions génomiques. Cela est peut-être à mettre en relation avec une duplication du génome entier qui a eu lieu chez l'ancêtre commun des poissons à nageoires rayonnées. L'analyse de l'expression des gènes TAAR de zebrafish et stickleback a révélé que ces gènes s'expriment dans l'organe olfactif, ce qui suggère l'importance relativement élevée des TAAR en tant que récepteurs chimiosensoriels chez les poissons (H. Yasuyuki and N. Mutsumi, 2007).

4-Aspect génétique et structural

Les GPCR ont été regroupés en trois classes sur des critères d'homologie de séquence. Les membres d'une même classe comportent au moins 25 % d'identité protéique. La première classe, la plus abondante comprenant les récepteurs olfactifs a été séparée en trois familles (A, B et C). Les récepteurs de la famille A comprenant la rhodopsine et les récepteurs olfactifs sont caractérisée par un domaine N terminal extracellulaire très court (20 à 30 acides aminés). De nombreuses études ont approché la compréhension de la structure des GPCR de façon indirecte jusqu'à la publication de la structure tridimensionnelle de la rhodopsine obtenue par cristallisation (Palczewski *et al.*, 2000). Cette structure a permis de confirmer la présence d'un ectodomaine N-terminal extracellulaire et de 7 domaines transmembranaires formés d'enchaînement d'acides aminés hydrophobes. Les 7 domaines transmembranaires repliés en hélice α forment la structure commune à tous les GPCR.

Les gènes de récepteurs olfactifs ont été identifiés par des techniques de clonage moléculaire à partir de l'épithélium olfactif du rat (Buck & Axel 1971). A la suite de cette étude princeps de nombreux répertoires de gènes OR ont été déterminés (Kieffer *et al.*, 1996 ; Zhao *et al.*, 1998 ; Hatt *et al.*, 2001 ; Spher *et al.*, 2003 ; Quignon *et al.*, 2005) .

Les récepteurs olfactifs sont organisés en deux classes I et II, selon leur niveaux d'analogie mais également de l'histoire de leur mise en évidence. On a longtemps pensé que la classe

I des OR était spécifique des substances odorantes solubles et donc des poissons chez lesquels on ne trouve pas d'OR de classe II et que les OR de classe II étaient spécifiques des odorants volatils. La découverte de ces deux classes chez un amphibien a renforcé cette théorie (Freitag *et al.*, 1998). Mais aucune expérience n'a confirmé cette hypothèse qui par ailleurs est infirmée par la découverte d'OR de classe I chez plusieurs espèces comme l'homme, le rat, la souris ou le chien.

La comparaison des caractéristiques structurales des deux classes de récepteurs de différentes espèces révèle une différence au niveau des boucles extracellulaires, ce qui peut contribuer à la spécificité du ligand. La comparaison du nombre et de la diversité des gènes des récepteurs olfactifs chez différentes espèces permet de mieux comprendre l'origine et l'évolution de cette famille de gènes (Freitag *et al.*, 1998).

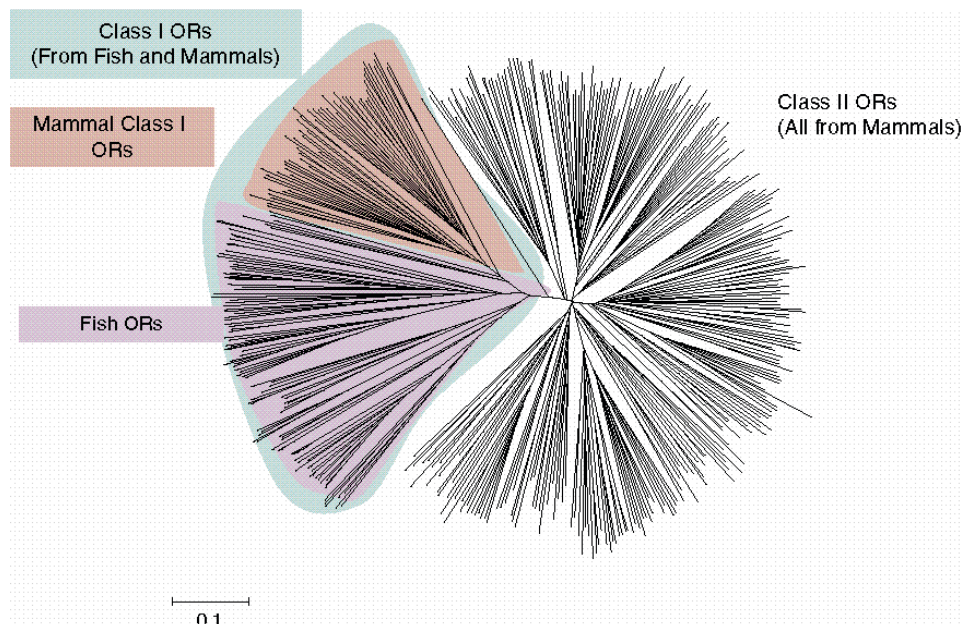


Figure 15 : Classification des OR de mammifères (homme, rat et souris) et de poissons (Zebrafish et Tetraodon) (Zhang and Firestein 2009).

Il est important de noter que les classes I d'OR de poissons et de mammifères constituent une branche à part mais qu'à l'intérieur de cette branche les gènes de poissons et de mammifères se séparent en deux sous-branches.

Les OR appartiennent à la superfamille de récepteurs couplés aux protéines G (GPCR) (Buck et Axel, 1991). C'est une superfamille très large, dont les membres sont capables de reconnaître des signaux extracellulaires aussi variés que des photons, des odorants, des

phéromones, des oses ou des macromolécules. Exprimés principalement au niveau de l'épithélium olfactif, les OR sont le premier élément dans une longue cascade de réactions transformant un signal odorant chimique en un signal électrique transmis à différentes aires cérébrales où il est analysé. Les OR sont organisés en familles et sous-familles de gènes codant des récepteurs présentant respectivement plus de 40% et plus de 60% d'identité alors que les membres d'une même classe comportent au moins 25 % d'identité protéique (Glusman *et al.*, 2000).

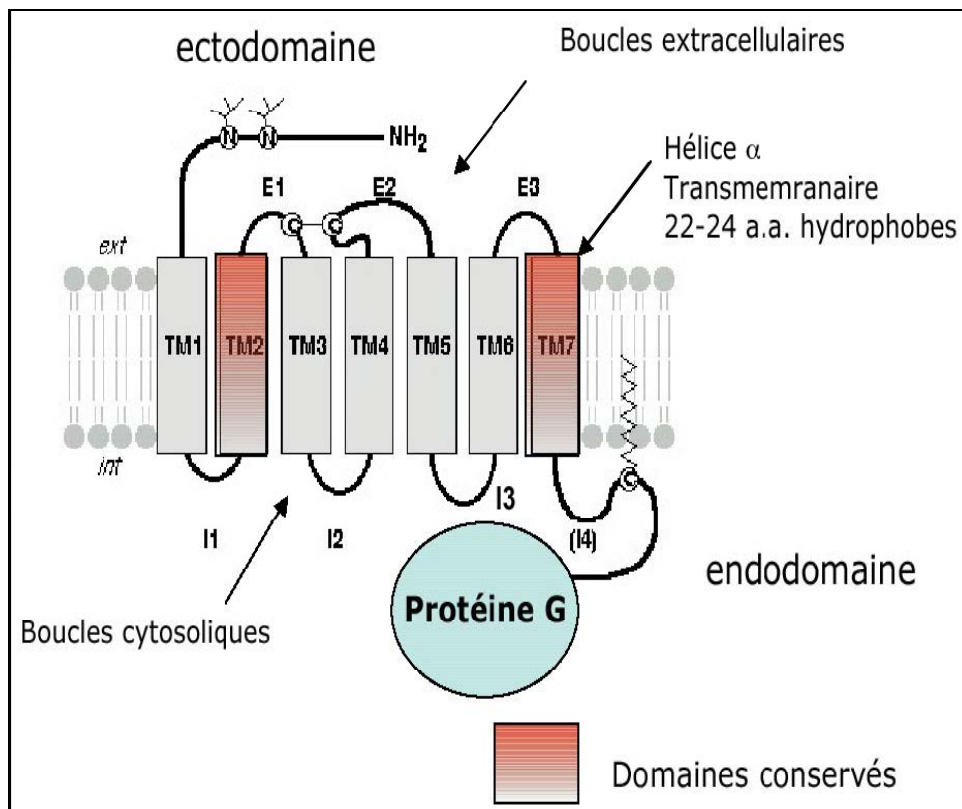


Figure 16 : Représentation schématique d'un récepteur couplé aux protéines G.

Les 7 domaines transmembranaires sont notés TM1-7. Les boucles extracellulaires et intracellulaires sont notées E1-3 et I1-3. Les boucles intracellulaires et la partie C-terminale interagissent avec la sous-unité α de la protéine G (appelée Golf dans le cas des récepteurs olfactifs). Pour les OR, la boucle E2 est particulièrement longue alors que l'extrémité N-terminale est courte.

Les OR fixent des ligands de petite taille au sein d'une cavité située entre les domaines transmembranaires 3 et 6. Ils ont une séquence de 300 à 350 acides aminés, sans peptide signal défini expliquant leur adressage à la membrane cellulaire. Les OR possèdent en

particulier une extension extracellulaire de 38-40 acides aminés, contenant une paire de cystéines très conservées (Zhao et Kirstein, 1999). Les gènes sont formés de deux exons dont seul le second est codant. La séquence exonique codante continue des OR est une particularité qui a facilité l'identification de ces gènes dans les génomes d'espèces séquencés. Outre des portions très conservées, en particulier dans les domaines transmembranaires 2, 6 et 7, ils possèdent des zones hypervariables au sein des domaines transmembranaires 3, 4 et 5 (Zozulya *et al.*, 2001). Il a été montré par mutagenèse que ces zones hypervariables étaient impliquées dans l'interaction avec les odorants (Kanada *et al.*, 2005, Abaffy *et al.*, 2007 ; Rouquier et Giorgi, 2007). Par ailleurs, les séquences conservées interagiraient avec la machinerie de transduction, commune à tous les récepteurs olfactifs.

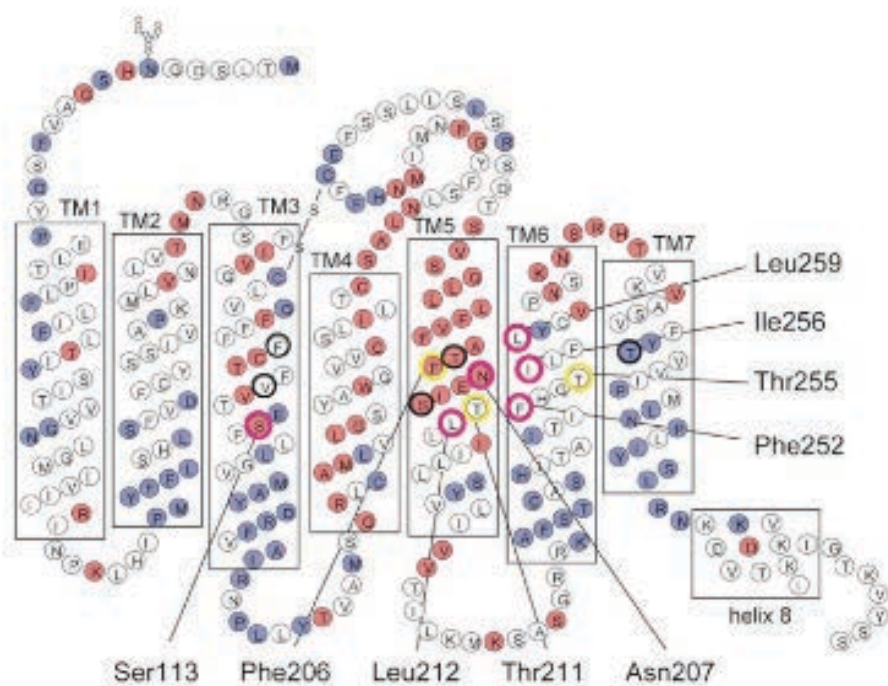


Figure 17 : Représentation de la séquence en acides aminés d'un récepteur de souris

- Les acides aminés situés dans des zones hypervariables
- Les acides aminés conservés dans la famille des GPCR.
- Les acides aminés entourés en rose, en jaune et en noir représentent ceux dont la mutation affecte fortement, faiblement ou pas du tout les capacités de fixation du récepteur (extrait de Kanada *et al.*, 2005).

Objectifs de la thèse

L'objectif de cette thèse était dans un premier temps de construire des panels RH et la carte RH de plusieurs espèces aquacoles et dans toute la mesure du possible de généraliser le protocole afin qu'il soit applicable à d'autres espèces aquacoles d'intérêt économique. Dans un deuxième temps, mon travail de thèse a consisté à utiliser plusieurs programmes bioinformatiques pour identifier l'ensemble des gènes OR et TAAR de cinq poissons de la famille des cichlidés dont les génomes ont été séquencés par le Broad Institute dans le cadre d'un programme international auquel mon laboratoire participait. J'ai ensuite fait des comparaisons inter espèces de ces gènes replacés dans le cadre évolutif de ces espèces. L'ensemble de ce travail présenté dans ce manuscrit décrit les connaissances actuelles sur l'évolution des génomes de poissons et l'olfaction. Les résultats obtenus sont présentés sous forme de publications avec au préalable une description détaillée. Une partie discussion sur l'ensemble des travaux expose les limites, les conclusions et les perspectives de ce projet de thèse.

RÉSULTATS

I- Construction de panels RH et cartographie de génomes d'espèces aquatiques

La cartographie à l'aide d'hybrides d'irradiation est une approche puissante pour produire des cartes dont la densité peut être facilement adaptée à la problématique, de quelques centaines de marqueurs à plusieurs milliers voire dizaines de milliers. Les cartes RH sont des ressources importantes pour les études génétiques. Elles constituent une aide appréciable sinon indispensable à l'assemblage des données de séquençage, ainsi qu'une approche aisée à la comparaison des génomes d'espèces voisines et détection des zones de conservation de synténie (Priat *et al.*, 1998; Kwitek *et al.*, 2004; Hitte *et al.*, 2005; Lindblad-Toh *et al.*, 2005; Lewin *et al.*, 2009).

Les cartes RH permettent le positionnement de marqueurs de toute nature, qu'ils soient polymorphes ou non polymorphes et ainsi constituent un lien entre cartes génétiques et cartes physiques.

La construction d'une carte RH se décline en plusieurs étapes principales :

1-Construction d'un panel de clones hybrides cellulaires

1-1-Les cellules donneuses

Dans le cadre des premiers panels construits, les cellules donneuses étaient le plus souvent sinon toujours obtenues à partir de cellules en culture. Bien que commode cette pratique est contestable dans la mesure où les cellules des lignées cellulaires sont l'objet de remaniements chromosomiques altérant de ce fait la structure d'ensemble du génome à étudier. L'utilisation de fibroblastes obtenus par des cultures primaires ou mieux des cellules isolées fraîches comme les cellules des lignées sanguines blanches est de beaucoup préférable. La rate quand l'animal peut être sacrifié est également une source intéressante de cellules, dans la mesure où sa dilacération dans un milieu isotonique avec un broyeur « potter » ou « dounce » est très facile et efficace. Dans la recherche d'une source de cellules donneuses, la seule question que généralement l'expérimentateur se pose est celle de l'accessibilité des cellules. Une deuxième question pourtant cruciale est celle de la capacité des cellules donneuses de fusionner avec les cellules receveuses. Cette capacité peu étudiée et mal comprise dépend apparemment des tailles relatives des cellules donneuses et receveuses qui ne doivent pas être trop différentes et surtout peut-être la composition lipidique de leurs membranes respectives. Si ce type de questionnement ne se pose guère pour la construction de panels de cellules de mammifères pour lesquels la cellule réceptrice est généralement une cellule de hamster, le problème

devient aigu lorsque la cellule donneuse s'éloigne beaucoup phylogénétiquement de la cellule receveuse. Dans notre expérience ce fut le cas avec les cellules d'huitre en particulier.

1-2-Irradiation

L'irradiation aux rayons X se pratique aisément avec un irradiateur de cellules sanguines. La dose d'irradiation appliquée va conditionner la résolution de la carte. Des doses élevées vont induire des cassures chromosomiques fréquentes. Elles sont utilisées pour ordonner des marqueurs proches et donc permettre de produire des cartes à haute résolution. A l'inverse, de faibles doses permettront d'ordonner des marqueurs plus éloignés et de produire des cartes à résolution faible. Mais dans tous les cas, la dose d'irradiation doit être suffisante pour tuer toutes les cellules donneuses et empêcher un envahissement ultérieur des cultures par des cellules donneuses intactes. Dans les conditions opératoires habituelles on délivre quelques 3000 Rad et au maximum 50.000 rad, doses à partir de laquelle les chromosomes des cellules donneuses sont si fragmentés que le nombre de marqueurs à positionner devient ingérable sauf à s'intéresser à une région particulière d'un génome. Ce fut le cas d'un panel RH humain qui irradié à 50.000 rads n'a guère été utilisé.

1-3-Culture cellulaire et clonage cellulaire

C'est l'étape la plus critique. Au cours de cette étape, on souhaite qu'après fusion cellulaire, la protéine codée par le gène de sélection muté dans la cellule receveuse puisse être fonctionnellement compensée par la protéine codée par le gène orthologue de la cellule donneuse. Or dans le même temps on ne souhaite pas que beaucoup de protéines de la cellule donneuse soient fonctionnelles dans leur nouvel environnement, au risque de perturber le développement de la cellule hybride. Par ailleurs le nombre de gènes de sélection actuellement disponible est très faible, il se résume aux gènes de la thymidine kinase (TK) et à celui de l'hypoxanthine-guanine phosphoribosyl transférase (HPRT).

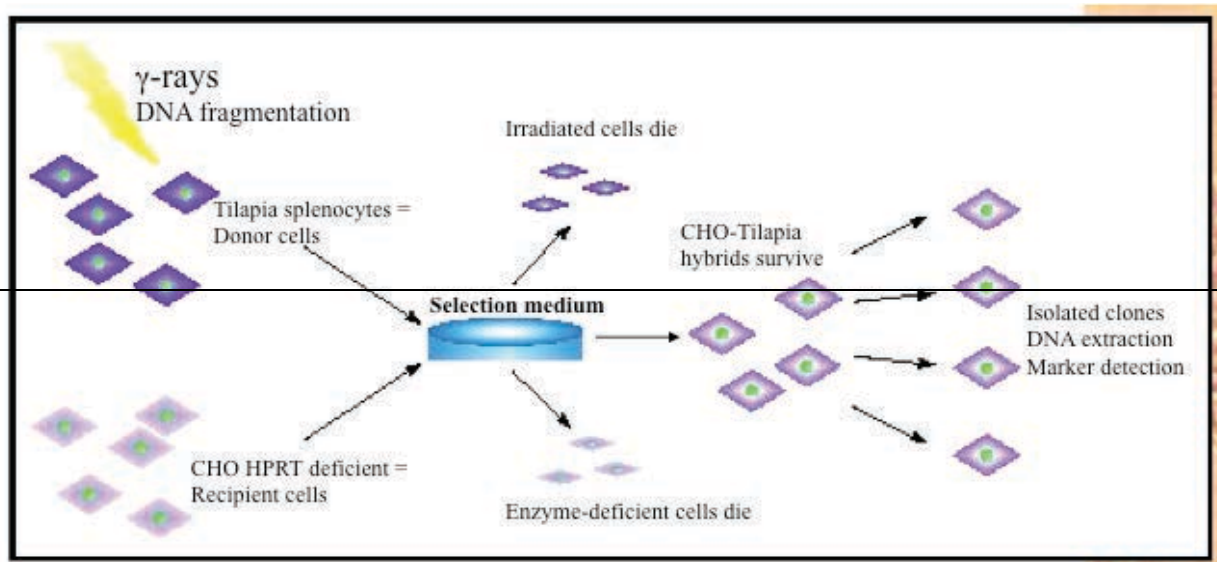


Figure 18 : Construction d'un panel RH

Ce schéma illustre les trois étapes clés de la construction d'un panel RH comprenant la fusion des cellules donneuses et receveuses, la sélection des hybrides cellulaires au cours de la culture en milieu HAT et enfin l'extraction de l'ADN des différents clones.

2-Les étapes de la cartographie d'hybridation :

Différentes méthodes permettent d'interroger chaque hybride cellulaire quant à la présence des différents marqueurs que l'on souhaite positionner sur la carte. Le plus souvent il s'agit d'analyse par PCR mais on peut également envisager d'autres approches quantitativement plus efficaces comme l'hybridation sur des puces sur lesquelles des milliers d'oligonucléotides peuvent être déposés. Au terme de cette analyse la présence d'un marqueur dans une lignée est notée 1 son absence 0. Ainsi pour chaque marqueur on dispose d'une suite de 0 et de 1 égale au nombre de lignées et appelée vecteur. On procède alors à l'analyse comparative des vecteurs avec des programmes dédiés tels que Ultime et Cartagène. Celle-ci procède en deux temps. On procède d'abord à une analyse dite deux points, au cours de laquelle on considère les marqueurs deux à deux. Au cours de cette étape on forme des groupes de marqueurs liés avec un support statistique exprimé en Lod score [Logarithm of odds : logarithme décimal du rapport de la probabilité d'observation d'une liaison de 2 locus à d'une distance Q (ou Θ , souvent utilisé) à la probabilité d'observation d'indépendance de ces 2 locus ($Q = 0.50$)] dont on fera varier la valeur expérimentalement en se guidant sur les résultats obtenus. Dans une deuxième analyse dite multipoint on ordonnera dans chaque groupe les marqueurs les uns par rapport aux autres, en calculant la distance entre eux.

La construction d'un panel RH pour des poissons, a la réputation d'être difficile, voire impossible à faire et il est vrai que de nombreux laboratoires ont tenté de produire des panels RH de poissons et ont abandonné.

Du point de vue expérimentale, la cartographie RH consiste à établir une collection de clones cellulaires hybrides contenant le génome des cellules receveuses (hamster ou souris) et une fraction aléatoire du génome des cellules donneuses (dans le cas présent des espèces aquacoles) dont le contenu est différent d'un hybride à l'autre. L'ensemble des hybrides cellulaires constitue un panel RH.

Le principe de la cartographie par hybrides cellulaires est le suivant. Si deux marqueurs sont proches sur le génome à analyser, la probabilité qu'ils soient séparés par une cassure liée à l'irradiation sera faible et les deux marqueurs auront tendance à se retrouver dans les mêmes hybrides et ce d'autant plus qu'ils seront plus proches. En revanche s'ils sont éloignés ou mieux sur deux chromosomes différents ils auront une distribution aléatoire l'un par rapport à l'autre.

Pour les mammifères, la cartographie par hybride d'irradiation ne présente plus de difficultés techniques, mais lorsque l'on s'écarte du point de vue phylogénique des mammifères et en particulier avec les espèces marines à chaque étape on est confronté à des difficultés majeures pour la construction des panels. Deux cartes RH seulement de poisson avaient été réalisées lorsque nous nous sommes intéressés à cette problématique et encore celles-ci avaient été construites avec des panels issus de la fusion de cellules de hamster avec des cellules de poisson zèbre en lignées continues.

La construction d'un panel de cellules RH se subdivise en plusieurs sous thèmes qui chacun posent des problèmes spécifiques dépendant de l'espèce considérée. La première étape consiste à trouver une source adéquate de cellules donneuses, ce qui s'est révélé délicat sachant que les cellules de poissons se cultivent difficilement et que par ailleurs les lignées continues sont à proscrire en raison des remaniements chromosomiques qui ne manquent pas de se produire au cours des différents passages.

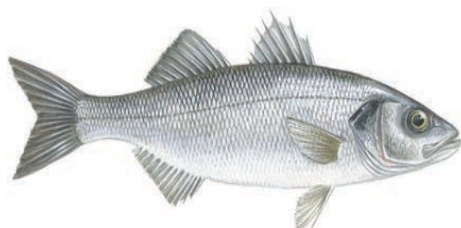
Notre équipe travaille depuis plus de 10 ans sur les panels et cartographie RH des mammifères et depuis quelques années, nous nous intéressons à la construction des panels RH et la cartographie des espèces aquacoles. Le premier panel et carte RH de ces espèces réalisés avant mon arrivée au laboratoire étaient ceux de la dorade (Senger *et al.*, 2005).

Cette construction a été très difficile et a pris beaucoup de temps, en fait plusieurs années. La source des cellules donneuses choisie était des fibroblastes obtenus à partir des nageoires dorsales. Les difficultés essentielles furent dues à l'obtention des fibroblastes en grande

quantité de l'ordre de plusieurs dizaines de millions de cellules et de la croissance des clones hybrides qui s'est révélée extrêmement laborieuse. Les différentes difficultés rencontrées lors de cette première construction se sont estompées avec la construction des autres panels qui ont bénéficié de l'expérience acquise.

Les séquences de poissons qui sont disponibles concernent les poissons dits modèles comme le Tetraodon, le Zebrafish, le Melaka et le Stickleback dont aucun n'intéresse l'aquaculture. Aussi mon projet de thèse a concerné la construction de panels RH et de cartes de plusieurs espèces majeures pour l'aquaculture et qui présentent un intérêt économique à savoir le bar, le tilapia, l'esturgeon et l'huître.

A. Construction d'un panel d'hybrides d'Irradiation du Bar européen (*Dicentrarchus labrax*) et Cartographie de 1581 marqueurs : (Publication n°1)



Bar européen (*Dicentrarchus labrax*)

Le bar est un poisson Téléostéen, il appartient à la famille des Moronidés de l'ordre des Perciformes. C'est l'ordre le plus diversifié des vertébrés. Le terme de « bar » ou *loup* désigne principalement deux espèces : le bar commun (*Dicentrarchus labrax*) et le bar tacheté (*Dicentrarchus punctatus*). Ce poisson de littoral aime particulièrement les rochers, mais on le trouve aussi dans les estuaires et les étangs saumâtres. Le bar vit en bancs, les jeunes plutôt en surface et les adultes en profondeur. On les trouve le long des côtes atlantiques, de la Norvège au Maroc, mais aussi en Méditerranée et dans la mer Noire.

Ce sont des poissons très résistants, capables de supporter des variations de salinité et de température très élevées. Le loup est constamment aux aguets et attend le passage d'un banc de poissons sur lequel il bondit dès qu'il l'aperçoit. Il affectionne particulièrement les zones de forts courants comme le raz de Sein en France dans lesquelles il chasse ses proies. Le bar fait depuis longtemps l'objet d'une aquaculture importante, sa production annuelle est de 138 000 tonnes pour l'année 2008 selon la FAO.

(http://www.aquamedia.org/production/species/seabasses/seabassprod_en.asp)

Son expansion est une conséquence de la mise au point de techniques permettant le contrôle de la reproduction et de la croissance dans les années 70 ainsi que de l'efficacité alimentaire. Cependant, l'élevage du bar souffre encore de certaines contraintes majeures. Il y a un accès limité pour la sélection des lignés qui peut en partie être attribué à un mauvais contrôle de l'accouplement. Par ailleurs les conditions de domestication sont encore difficilement comprises. C'est pour ces raisons qu'un programme subventionné par l'Union Européenne

(FP6 Network of Excellence Marine Génomiques Europe - contract GOCE-CT-2004-505403 et the FP6 project Aquatiras - contract STREP-2004-513692) a été lancé en 2004 visant à renforcer les bases scientifiques d'une sélection assistée par marqueur. C'est dans ce cadre que notre laboratoire a construit un panel RH et une carte de 1581 marqueurs.

Le panel RH du Bar est le premier panel construit avec des splénocytes comme cellules donneuses. L'intérêt de cet organe, aisément assimilable à un simple sac rempli de splénocytes est qu'un broyage doux de ce dernier, simplement émincé aux ciseaux, dans un broyeur de type « dounce ou potter » permet l'obtention de millions de splénocytes directement utilisables pour une fusion cellulaire. Après une longue expérimentation visant à définir plusieurs paramètres, trois fusions ont été réalisées entre des cellules d'hamster déficientes pour le gène HPRT et des splénocytes (cellules de rate) du bar européen (*Dicentrarchus labrax*) irradiés à 3000 rad. Ces trois fusions ont fourni 290 clones individuels qui ont été repiqués et cultivés indépendamment. Après plusieurs passages les cellules ont été récoltées, les ADN extraits puis quantifiés. Chacun de ces 290 ADN a été caractérisé par le typage de 96 marqueurs. A partir des résultats de ce typage, nous avons sélectionné 93 ADN présentant pour la plupart la plus forte rétention de marqueurs.

C'est sur ce panel de 93 ADN sélectionnés sur des critères quantitatifs et qualitatifs qu'ont été cartographiés 1581 marqueurs par la méthode PCR classique.

La carte RH du génome du Bar ainsi construite a une densité moyenne de 1 marqueur pour 480 Kb. La plus grande part des marqueurs a été dessinée à partir d'EST et ARNm déchargés à partir de la base de données GenBank (Avril 2009). Afin de construire une carte riche en marqueurs orthologues de marqueurs présents sur les cartes de poissons modèles, les séquences déchargées ont été alignées avec les séquences génomiques des espèces modèles. Seules les séquences ayant un score supérieur à 250 et une longueur comprise entre 80 et 300 pb ont été choisies pour dessiner des paires d'oligonucléotides propices à amplifier par PCR des marqueurs de 60 à 150 pb de long. Le génotypage de 141 microsatellites de la carte génétique nous a par ailleurs permis de comparer la carte génétique et la carte RH du bar. La comparaison de la carte RH avec les séquences des génomes de stickleback, tetraodon, medaka et zebrafish a été réalisée pour établir les relations de synténie existant entre le génome du bar et celles des quatre génomes modèles. Cette comparaison révèle en particulier qu'aucune rupture de synténie n'existe entre les génomes du bar et de la dorade ou du medaka, alors que quelques points de ruptures existent avec ceux du stickleback et du tetraodon. En raison d'un assemblage de séquence génomique insuffisant ce genre d'analyse n'a pas été concluant avec le génome du zebrafish.

Cette carte RH facilitera les études de liaison et l'identification de gènes candidats ainsi que des QTL (Quantitative Trait Loci) dont l'utilisation pour la sélection assistée par marqueurs est importante pour soutenir l'expansion de l'aquaculture. Le bar représente par ailleurs un modèle génétique de l'évolution des Perciformes.

Publication n°1

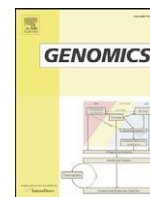
A radiation hybrid map of the European sea bass (*Dicentrarchus labrax*) based on 1581 markers: Synteny analysis with model fish genomes

Richard Guyon ^a, Fabrice Senger ^b, Michelle Rakotomanga ^a, Naoual Sadequi ^a, Filip A.M. Volckaert ^c, Christophe Hitte ^a, Francis Galibert ^{a,*}

Genomics 2010, 96, 228-238

Contents lists available at [ScienceDirect](http://www.sciencedirect.com)

Genomics

journal homepage: www.elsevier.com/locate/ygeno

A radiation hybrid map of the European sea bass (*Dicentrarchus labrax*) based on 1581 markers: Synteny analysis with model fish genomes

Richard Guyon ^a, Fabrice Senger ^b, Michaëlle Rakotomanga ^a, Naoual Sadequi ^a, Filip A.M. Volckaert ^c, Christophe Hitte ^a, Francis Galibert ^{a,*}

^a CNRS UMR 6061 Institut de Génétique et Développement, Université de Rennes 1, 2 Avenue du Pr Léon Bernard, F-35043 Rennes Cedex, France

^b CNRS UMR 6026 Interactions Cellulaires et Moléculaires, Université de Rennes 1, Campus de Beaulieu, Bâtiment 13, F-35042 Rennes Cedex, France

^c Laboratory of Animal Diversity and Systematics, Katholieke Universiteit Leuven, Charles Deberiotstraat 32, B-3000 Leuven, Belgium

article info

Article history:

Received 21 April 2010

Accepted 20 July 2010

Available online 24 July 2010

Keywords:

Sea bass

RH-panel

Mapping

Synteny

Comparative genomics

abstract

The selective breeding of fish for aquaculture purposes requires the understanding of the genetic basis of traits such as growth, behaviour, resistance to pathogens and sex determinism. Access to well-developed genomic resources is a prerequisite to improve the knowledge of these traits. Shaving this aim in mind, a radiation hybrid (RH) panel of European sea bass (*Dicentrarchus labrax*) was constructed from splenocytes irradiated at 3000 rad, allowing the construction of a 1581 marker RH map. A total of 1440 gene markers providing ~4400 anchors with the genomes of three-spined stickleback, medaka, pufferfish and zebrafish, helped establish synteny relationships with these model species. The identification of Conserved Segments Ordered (CSO) between sea bass and model species allows the anticipation of the position of any sea bass gene from its location in model genomes. Synteny relationships between sea bass and gilthead seabream were addressed by mapping 37 orthologous markers. The sea bass genetic linkage map was integrated in the RH map through the mapping of 141 microsatellites. We are thus able to present the first complete gene map of sea bass. It will facilitate linkage studies and the identification of candidate genes and Quantitative Trait Loci (QTL). The RH map further positions sea bass as a genetic and evolutionary model of Perciformes and supports their ongoing aquaculture expansion.

© 2010 Elsevier Inc. All rights reserved.

Introduction

Teleost fishes represent the most numerous and diversified group of vertebrates. Within this group, the order of Perciformes is by far the most numerous and highest evolved. European sea bass (*Dicentrarchus labrax*; family of Moronidae) is one of its biologically well-documented members and one of the more valuable marine finfish. The European aquaculture production amounted to 138,156 tons in 2008 (http://www.aquamedia.org/production/species/seabasses/seabassprod_en.asp?). Its expansion is a consequence of the development of techniques allowing the control of reproduction and growth from larva to portion size in the seventies and eighties [1].

However today's sea bass industry still suffers from some major constraints. There is limited access to selected lines, which may be attributed to poor acceptance of controlled mating. Sea bass breeders have essentially a wild genetic profile; they are experiencing poorly managed domestication selection and relaxation from natural selection pressure under aquaculture conditions [2]. Most appreciated by Producer and consumers are traits related to growth, feeding efficiency, gender, disease resistance and immunity. Fortunately the heritability of many aquaculture traits is moderate to high [3], making the potential for selection high. It has resulted in single generation gains of 23% in growth in European sea bass [2]. Several QTLs involved in body weight, a multigenic trait, and shape have been identified in European sea bass [4]. But production losses in aquaculture remain huge due to infection by viruses and bacteria, and to a lesser extent by parasites. Enhancement of innate and acquired immunity through breeding, husbandry and vaccination is an acknowledged strategy to address this issue (see [5] for review). Also, considerable environmental and behavioural stresses are induced by the atypical aquaculture environment. Since stress has a genetic component, selection should be feasible [4,6]. Hile sex determination is polygenic in sea bass [7], sex differentiation is growth-dependent and environmentally tuned [8]. Henné gender control remains a big issue in sea bass breeding. Another challenge is the induction of sterility to optimise the production of portion-sized fish and to limit the environmental impact from escapees [9]. Of a more recent interest is the substitution of current fish food with plant ingredients for a more sustainable aquaculture [10]. This strategy may directly affect quality traits related to growth and metabolism, whose heritability is under investigation in sea bass [11].

Hence, understanding the genetic basis of traits of interest will directly benefit the identification of the best breeders and the management of the broodstock. But the feasibility of these studies relies on the availability of well-developed genomic resources. The more established fish models such as zebrafish *Danio rerio*, spotted green pufferfish *Tetraodon nigroviridis*, fugu *Takifugu rubripes*, three-spined stickleback *Gasterosteus aculeatus* and medaka *Oryzias latipes* are of no aquaculture interest. However, a growing group of economically important fishes (e.g. Atlantic salmon *Salmo salar*, Nile tilapia *Oreochromis niloticus*, carp *Cyprinus carpio*, rainbow trout *Oncorhynchus mykiss*, channel catfish *Ictalurus punctatus*, gilthead seabream *Sparus aurata* and European sea bass) is being supported by genomic resources driven by important technological developments in molecular biology.

The sea bass genome recently underwent a 2× WGS sequencing combined with the alignment of both ends of ~45,000 BAC clones on the stickleback sequence [12]. Establishing dense and accurate maps for these species along with comparative data with other species makes it very attractive to extrapolate genome architecture, gene function, evolution and structure. Genetic linkage maps of the sea bass genome, consisting of 174 and 368 polymorphic markers, were constructed by Chistiakov et al. in 2005 [13] and 2008 [14], respectively. These maps provided a first characterisation of the genome and identified consistent low resolution synteny relationships with model fish species. The reason is that linkage maps are usually made with relatively few gene markers, which limit interspecific comparisons. By contrast, radiation hybrid (RH) mapping is suited to map all types of markers, it does not require locus-specific polymorphism, and it offers higher density mapping. RH mapping thus allows to construct comparative maps through the mapping of genes and the localisation of their orthologs in other species [15]. In fish genomics this was illustrated by the construction of two gene-based RH maps of the gilthead seabream genome and their correspondence with the spotted green pufferfish genomic sequence [16,17]. However synteny comparisons between seabream and sea bass, and between seabream and tilapia were made indirectly via comparisons with stickleback. Global synteny relationships between farmed Perciformes (seabream, sea bass and Nile tilapia) and model species (three-spined stickleback, medaka and spotted green pufferfish) [18] fit with the generally accepted phylogeny based on the mitogenome [19] and 42 orthologous genes [20].

High resolution physical maps such as RH maps are also of great help in the assembly of sequencing data in shotgun projects [21]. Indeed, in many cases “finished” genome sequences from shotgun projects contain large sequence gaps that imply inconsistencies in the positioning of scaffolds [22]. In addition, low-depth sequences lack long-range continuity and provide only a fragmented view of a genome. This is exemplified by the fugu genome sequence made of 7213 unconnected scaffolds without any chromosome assignment [23].

Still, the construction of RH panels of fish species has turned out to be more difficult than in mammals, where most RH panels have been developed (see [15] for review). To date, only three fish RH panels have been reported: two RH panels of zebrafish derived from stable cell lines [24,25] and one RH panel of seabream derived from primary fibroblasts [16]. Here we report for the first time the production of fish radiation hybrid lines from splenocytes. The sea bass RH panel was used to construct a 1581 marker RH map of the sea bass genome. The mapping of 1440 gene markers designed from ESTs and complete mRNA allowed to establish gene order along the 24 sea bass chromosomes and to identify synteny relationships with other fish model genomes. Moreover, the mapping of 141 microsatellites made it possible to integrate the sea bass genetic linkage map in the RH map. It is the first dense RH map of the sea bass genome obtained from a high resolution panel with good performance. It is also the first complete gene map of European sea bass.

Methods

Selection of radiation hybrid cell lines

Juvenile sea bass of average weight 50 g were kindly provided by Ifremer (Brest, France) and kept at the INRA, SCRIBE (Rennes, France) fish husbandry facility until needed for fusion experiments. When proceeding with the fusion experiment, fish were euthanised with an overdose of 2-phenoxyethanol and rinsed briefly in 70% ethanol. The spleen was rapidly dissected and rinsed several times in washing medium (Leibovitz L-15; 1× Penicillin/Streptomycin/Fungizone). The spleen was cut in small pieces and fragmented between the frosted edges of glass slides. The obtained cell suspension was ready for irradiation. A panel of radiation hybrid cell lines was constructed using the methodology described previously [26]. Briefly, the sea bass spleen cell suspension was irradiated by a 3000-rad γ -ray exposure. Splenocytes were fused with HPRT⁻ derivative CHO cells in a 5:1 ratio in the presence of polyethylene glycol 1500 (Roche, Mannheim, Germany). Following selection in HAT medium and minimum cell culture expansion (3 to 4 weeks), DNA was extracted from individual clones (QIAmp DNA kit, protocol for cultured cells; Qiagen, Hilden, Germany).

DNA amplification

Once hybrid cell lines were selected on quantitative and qualitative criteria, their DNA was expanded by Whole Genome Amplification (WGA) using ϕ 29 DNA polymerase of V2 and HY GenomiPhi kits (GE healthcare, Fairfield CT, USA). DNA concentration was first estimated with a Nanodrop (Thermo Scientific, Wilmington DE, USA) for every cell line. A first step consisting of two separate WGA was performed from 2× 10 ng of DNA with the V2 GenomiPhi kit. WGA products were pooled to constitute the stock panel. A second step consisting of two separate WGA was performed with the HY GenomiPhi kit from 2× 10 ng of DNA from the stock panel. Products were pooled to constitute a working panel allowing to genotype several hundreds of markers. When needed, additional material was prepared by repeating the second step. The WGA reliability was previously demonstrated by Senger et al. [16].

Marker definition

All sea bass ESTs and mRNAs available at the time of the study were downloaded from the GenBank database (April 2009). Sequences were masked for simple repeats with the RepeatMasker web server (<http://www.repeatmasker.org/>) and aligned together with the CAP3 software [27] in order to build contigs of overlapping sequences. Default parameters of the CAP3 software were used for the assembly step.

In order to identify orthologous sequences between sea bass and four model fishes (three-spined stickleback, medaka, pufferfish and zebrafish), the contigs and singlets obtained from the assembly step were aligned to the respective genome sequences using the Exonerate v1.4.0 software [28]. Sea bass sequences having a hit with a minimal score of 250 and an alignment size of 80 to 300 bp with model species genomic sequences were selected and further used to design markers. PCR markers of 60 to 150 bp long with primers of 23 nt average length were designed using the Primer3 v0.4.0 software [29]. For each of the sea bass sequences the coordinates of the best hit on each of the model fish genomes were considered as the location of the putative orthologs. Orthologous genes with gilthead seabream were also identified by alignment of sea bass sequences to the seabream sequences mapped by Sarropoulou et al. [17]. Hits between sea bass and seabream sequences were selected using the same criteria as above, except that no maximal alignment size was imposed, as sequences of both species consisted of ESTs.

Orthologs 1:1:1:1 between sea bass, stickleback, medaka, pufferfish and zebrafish were used to estimate gene sequence conservation

between these species. All hits localised between coordinates of these orthologs as annotated by Ensembl were identified and a conservation coefficient between sea bass and each species was calculated as follows: $\text{coeff} = \frac{\sum(\text{identity_}\% \times \text{identity_length})}{\sum(\text{identity_length})}$. The Ensembl gene coordinates for each species were downloaded from the Biomart webserver (<http://www.biomart.org>).

Genotyping

PCR were performed on 50 ng of DNA in a final volume of 10 μ l and at a final concentration of 0.3 μ M of each primer, 250 μ M of each dNTP, 1.5 or 2 mM MgCl_2 , 50 mM KCl, 10 mM Tris-HCl, plus 0.05 U of AmpliTaq Gold polymerase (Applied Biosystems, Foster city, CA). Reactions were carried out in MJ (MJ Research, Cambridge, MA) or GeneAmp 9700 (Applied Biosystems) thermocyclers with the following PCR program: 1 inducing step of 95 °C for 8 min; 20 cycles of 94 °C for 30 s, 61 °C (or 57 °C) for 30 s (−0.5 °C/cycle), 72 °C for 1 min; 15 cycles of 94 °C for 30 s, 51 °C (or 47 °C) for 30 s, 72 °C for 1 min; 1 final extension step of 72 °C for 2 min.

PCR products were analysed by electrophoresis in a 2% agarose gel in 0.5× TBE buffer. Gel images were recorded with a high resolution CCD camera (Vilber Lourmat, Torcy, France). Results were scored as present, absent or ambiguous in a semi-automated fashion. Microsatellite markers mapped on the first generation linkage map [13] were also genotyped.

Data computation

A two-point analysis was performed using the Multimap v2.0 software [30], starting at a lod score of 3.5. The multipoint analysis was performed with the CarthaGène v1.0 software [31]. RH groups that harboured obvious aberrations were re-analysed at higher two-point lod scores of up to 4.5, before performing the multipoint analysis again. Distances between markers were expressed in centirays (cR_{3000}). Coordinates of the putative orthologous genes retrieved from the Exonerate analysis of sea bass sequences with the four model genomes were aligned with the corresponding sea bass markers on the graphic representation. Ordered RH groups were tentatively oriented according to two-point lod scores between their end-markers. The microsatellite order on the linkage map and the synteny relationships with the four model species were compared with the RH map. Conserved Segments (CS) and Conserved Segments Ordered (CSO) between sea bass and the model species were identified using the AutoGRAPH web server [32]. A default adjacency penalty of 5 was used for the counting of CSO.

Results and discussion

Radiation hybrid panel

Sea bass splenocytes were γ -irradiated with a 3000 rad dose, fused with HPRT[−] derivative CHO host cells and selected on HAT medium as described in the Methods section. A total of 290 hybrid cell lines was obtained through three fusion experiments. The retention frequency was estimated for every clone by typing a set of 96 microsatellite markers taken from the genetic map [13] and by scoring their presence/absence pattern. The retention frequency of the 96 microsatellite set in every clone is shown in Fig. 1. The average retention frequency of the 290 clones was 19.8%.

A subset of 93 hybrid cell lines was selected to constitute the RH panel. Hybrid cell lines with high retention frequency were selected with the highest priority. However hybrid cell lines exceeding 50% retention were not selected in order to normalise the retention frequency of the hybrid lines of the RH panel. These cell lines were replaced by other cell lines, taking into consideration the retention

value of each individual marker to get a better representation of the whole genome.

The average retention frequency of the RH panel is 30.6%. This value corresponds to the average amount of the sea bass genome retained in each clone as well as to the chance that a given marker is present in a given clone. Retention frequency of the clones ranged from 17% to 49% with 73% of the clones ranging from 25 to 40% retention. Finally, 94 of the 96 microsatellites are present in more than 5% of the hybrid cell lines.

A WGA with ϕ 29 DNA polymerase was subsequently performed from a DNA aliquot of each cell line of the panel. To our knowledge this is the first time that a RH panel is constructed from fresh cells requiring no primary culture. Moreover this cell type is convenient to use as the dissociation of the spleen is performed in a relatively short time preceding the irradiation. By doing so, the chromosomal rearrangements that frequently occur in long-term cultured cell lines were avoided. The WGA reliability and its assets upon cell culture expansion to generate a large amount of DNA were discussed previously in Senger et al. [16].

Marker identification

Sequences of 44,358 ESTs and 399 mRNAs were downloaded from GenBank (April 2009), excluding mitochondrial sequences. Once simple repeats were masked, the assembly with the CAP3 software [27] with default parameters resulted in 7495 contigs and 13,371 singlets for a total of 20,866 unique sequences. Alignments with the Exonerate software [28] with the criteria described above revealed that 12,882 (~62%) of these sequences had a hit in the stickleback genome, 10,488 (~50%) in the medaka genome, 10,402 (~50%) in the pufferfish genome, and 7093 (~35%) in the zebrafish genome. A Venn diagram recapitulates how hits are assigned to the four model genomes, pointing to 6305 sequences shared by all of them (Fig. 2). Coordinates of the corresponding genes in the genomic sequences were retrieved based on the Ensembl gene annotation for each model species. All hits localised within these coordinates and putatively corresponding to exons of the orthologous genes were identified. Global sequence conservation between sea bass and the model species was assessed as described in the Methods section. The identity coefficient was estimated to 83.6% with stickleback, 81.1% with medaka, 79.4% with green spotted pufferfish and 76.1% with zebrafish. According to these values the stickleback is evolutionarily the closest to the sea bass, and the furthest from the zebrafish while distances with medaka and pufferfish are intermediate. Although the values may be biased as the sequencing completion of the model genomes is unequal, these results are consistent with phylogenies previously observed by Kuhl et al. [12] when aligning BAC-end sequences on the four model fish genomes. The decreasing proximity of sea bass with stickleback, medaka, pufferfish and zebrafish is also consistent with the phylogenetic gene trees constructed from the genes SOX17 [33] and IFN [34]. Trees made with seabream and the model species show a similar phylogeny [18,35].

EST-based markers were designed in priority from sequences conserved among stickleback, pufferfish and medaka. Based on sequence alignment scores, 1440 gene markers were selected, of which 1373 (95%) were conserved in stickleback, 1144 (79%) in medaka, 1132 (79%) in pufferfish and 758 (53%) in zebrafish. The Ensembl identifiers of orthologous genes in each species are listed in Supplemental Table 2 (S2). Finally, 1420 markers (99%) were conserved in at least one fish model species. Of those 1420 markers, 702 (49%) allowed anchorage between sea bass and the four species, 341 (24%) with three of them, 199 (14%) with two and 178 (14%) with only one model species, accounting for a total of 4407 anchors. However not all of those orthologous markers provided true anchors with the genomes of the fish model species, as the availability of coordinates depends on the assembly status of the genome sequence.

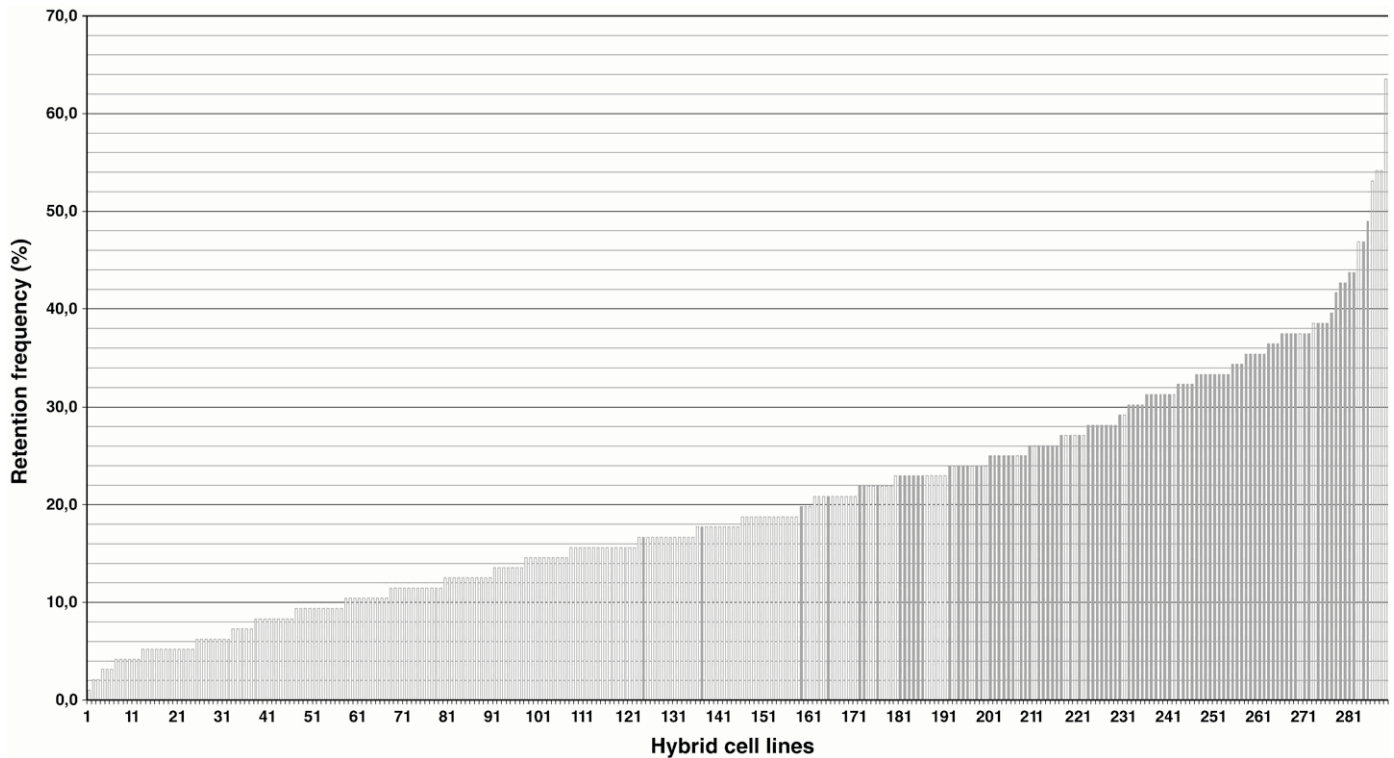


Fig. 1. Retention frequency of the sea bass hybrid cell lines. Hybrid cell lines are numbered from 1 to 290 on the X axis. Their retention frequency, estimated by PCR determination of the presence/absence of 96 microsatellite markers spread all over the sea bass genome [13], is represented on the Y axis. The 93 hybrid cell lines selected on qualitative and quantitative criteria that constitute the sea bass RH panel are figured in dark grey.

Thus 66 gene markers had orthologs identified on unlocalised scaffolds (“chromosome unknown”) of the stickleback genome sequence, 351 markers aligned with the “chromosome unknown” of the pufferfish genome sequence, 135 markers aligned with the “chromosome unknown” of the medaka genome sequence and 55 markers aligned with unlocalised scaffolds of the zebrafish genome sequence. These proportions are in accordance with the proportion of the sequence assemblies assigned to chromosomes for the model species [36,37].

A score of 250 with the Exonerate software corresponds to a minimal alignment size of 50 bp, which is consistent with usual recommendations for mapping orthologous ESTs [38]. The maximal

alignment size of 300 bp was imposed to avoid hits actually corresponding to retro-pseudogenes that may be mistaken for true orthologs. This value was based on the average size of exons estimated to be ~180 bp in the pufferfish [36]. The synteny conservation with the stickleback in the assigned RH groups for 933 markers out of 974 proved that these criteria were consistent, even though they are less efficient for identifying true orthologs in phylogenetically distant species like zebrafish.

Characteristics of the RH map

The sea bass RH map contains a total of 1581 markers consisting of 1440 genes and 141 microsatellites. Assuming the 763 Mb size estimate of the sea bass genome [39] is correct, this represents a density of one marker per 480 kb. The typing data were analysed using the Multimap and CarthaGène softwares in an approach that minimised the number of RH groups and maximised their size while avoiding illegitimate fusions of RH groups.

The two-point analysis performed at a lod score of 3.5 resulted in 46 RH groups. As some of them harboured obvious aberrations, increasing lod scores were applied in a step-wise fashion up to the breaking point. Markers responsible for inappropriate fusions were also removed. Finally, 91 RH groups were obtained from two-point analyses performed at lod scores of 3.5 to 4.5. Characteristics of RH groups in terms of size, number of markers and associated orthology are shown in Table 1.

Data processing revealed 1552 markers (98%) with unique positions and 25 positions with two to four co-localised markers. These 91 RH groups with more than three markers represented 1342 markers, while 239 markers remained unlinked or in small groups of just two markers. Sixty-five of the RH groups containing three to 74 markers were assigned to the 24 sea bass chromosomes with the help of the microsatellite positions on the genetic map and the two-point lod scores between end-markers. It amounts to an average of 2.7 RH

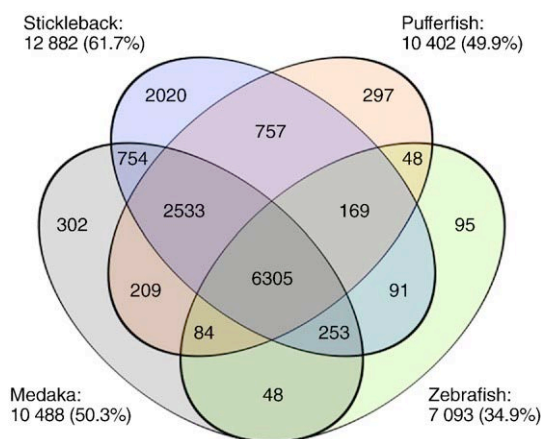


Fig. 2. Distribution of hits between sea bass and stickleback/medaka/pufferfish/zebrafish represented as a Venn diagram. Each model species is represented by an ellipse. The number of hits shared by two species or more are indicated in every intersection. For each model species, the number of sea bass sequences having one hit or more and their percentage of the total number of sequences analysed are indicated.

Table 1
Characteristics of RH groups.

RH group	No. of markers ^a	No. of positions	No. of genes ^a	Size (cR _{anno})	Lod score	Stickleback		Pufferfish		Medaka		Zebrafish	
						chr ^c	No. of anchors ^b	chr ^c	No. of anchors ^b	chr ^c	No. of anchors ^b	chr ^c	No. of anchors ^b
LOD3.5-RH13	19	18	17	723	3.5	16	7	13	7	9	14 (2)	10	21, 14, 10, 9, 12
LOD3.5-RH14	18	18	13	670	3.5	13	18	8	14	9	24	5	20, 23
LOD3.5-RH16	16	16	15	533	3.5	15	7 (1)	9	7	13	14 (18, 13)	6	14, 5, 24, 18, 15
LOD3.5-RH19	8	8	8	257	3.5	7	8 (14, 1)	1	(1)	2	(4, 12)	5	6, 21, 2
LOD3.5-RH21	7	7	6	205	3.5	6	18	5	14	6	24	4	20
LOD3.5-RH5	48	48	45	1879	3.5	43	11 (5, 14, 13, 12)	24	3 (18, 2, 9)	40	8 (9, 7, 12, 10, 1)	26	3, 8, 6, 24, 22, 1
LOD3.5-RH10	34	33	25	1420	3.5	21	5	17	2 (15, 3)	17	19 (16)	15	12, 3, 8, 24, 20, 17
LOD3.5-RH15	16	16	16	627	3.5	14	18 (11)	14	14, 3	14	24 (22)	9	20, 3, 17, 13
LOD3.5-RH17	15	15	13	631	3.5	12	11 (15)	2	(11, 10)	8	8 (7)	8	3
LOD3.5-RH18	9	9	8	329	3.5	7	10	0		5	11	3	24, 19, 17
LOD3.5-RH20	8	8	8	237	3.5	8	19	8	13	6	6	6	18, 25, 2
LOD3.5-RH22	6	6	6	248	3.5	5	18 (4)	2	14	2	24	0	
LOD3.5-RH23	6	6	6	197	3.5	3	7	4	7	3	14	4	21
LOD3.5-RH24	5	5	5	236	3.5	5	19	4	13	5	6 (3)	3	4, 25, 18
LOD3.5-RH25	5	5	4	221	3.5	4	3 (7)	0		2	17	2	2
LOD3.5-RH26	4	4	4	161	3.5	3	12	3	9	2	7	2	8
LOD3.5-RH27	4	4	3	98	3.5	0		2	18	0		2	22
LOD3.5-RH28	4	4	4	122	3.5	4	6	0		3	(15, 10, 1)	2	7, 1
LOD3.5-RH29	4	4	4	124	3.5	3	13 (14)	2	12	1	(9)	2	8, 5
LOD3.5-RH3	74	72	67	2899	3.5	61	8 (15)	36	1 (8, 13)	48	4 (1)	25	22, 20, 2, 8, 11, 10, 6, 21, 17, 16
LOD3.5-RH31	4	4	4	146	3.5	4	4	1	(15)	4	23 (22, 17)	2	4, 11
LOD3.5-RH32	4	4	4	166	3.5	3	16	0		4	21	4	9, 6
LOD3.5-RH33	4	4	3	175	3.5	2	13	2	12	2	(9, 8)	2	8, 3
LOD3.5-RH34	4	4	4	112	3.5	4	18	3	14	3	24 (15)	3	20, 12
LOD3.5-RH35	4	4	4	171	3.5	3	1	3	3	0		3	1, 25
LOD3.5-RH36	3	3	3	142	3.5	3	14	0		1	(12)	2	5
LOD3.5-RH37	3	3	3	127	3.5	3	19	2	13	3	6 (17)	3	7, 4, 1
LOD3.5-RH38	3	3	3	107	3.5	2	19	2	13	1	(6)	2	25, 18
LOD3.5-RH39	3	3	3	95	3.5	3	18	3	14	0		2	20
LOD3.5-RH40	4	3	3	13	3.5	2	19	2	13	2	6	1	25
LOD3.5-RH41	3	3	3	63	3.5	3	1 (4)	0		2	(18, 13)	1	15
LOD3.5-RH42	3	3	3	140	3.5	2	13	0		2	9	0	
LOD3.5-RH43	3	3	2	88	3.5	0		1	(11)	1	(5)	1	6
LOD3.5-RH44	3	3	3	93	3.5	0		0		0		2	8
LOD3.5-RH45	3	3	2	94	3.5	2	15	0		0		0	
LOD3.5-RH46	3	3	3	43	3.5	2	7	1	(15)	3	18	1	5
LOD3.5-RH9	34	34	30	1338	3.5	27	7	3	20 (18)	22	18 (17)	10	7, 1
LOD3.6-RH28	7	7	7	209	3.6	7	9	1	(16)	2	1	3	1
LOD3.6-RH6	43	42	42	2025	3.6	34	15 (7, 18)	32	10 (14)	32	22 (11, 24, 7, 14, 1)	26	17, 2, 20, 19, 6, 24, 13
LOD3.6-RH10	25	25	21	1071	3.6	20	7 (1)	10	7	14	14 (13, 18)	12	10, 5, 15, 24, 21, 18
LOD3.6-RH12	23	22	21	1006	3.6	19	21 (20, 1)	11	6 (3)	20	20 (6, 2)	19	24, 2, 6
LOD3.6-RH13	19	19	18	905	3.6	17	19	11	13	11	6	10	25, 7, 18
LOD3.6-RH21	10	10	8	417	3.6	8	19	6	13	7	6 (11)	4	18, 17
LOD3.6-RH25	8	8	8	456	3.6	7	19 (20)	2	(5, 13)	3	6 (16)	3	25, 18
LOD3.6-RH29	6	6	6	273	3.6	6	4	2	(9, 13)	5	23	4	4, 7, 13
LOD3.6-RH38	4	4	3	126	3.6	0	0			2	20	2	24
LOD3.6-RH45	4	3	3	82	3.6	2	19	2	13	2	6	0	
LOD3.7-RH18	18	18	16	789	3.7	15	5	9	2 (9)	12	19 (7, 5)	9	12, 24, 20, 13
LOD3.7-RH52	3	3	2	49	3.7	1	(3)	0		1	17	0	
LOD3.7-RH11	25	25	24	1191	3.7	22	10 (9)	16	21 (20, 18)	15	11 (16, 1)	11	19, 16, 8
LOD3.7-RH47	4	4	4	121	3.7	2	(20, 10)	1	(17)	3	11	3	19
LOD3.8-RH15	30	30	30	946	3.8	28	20	20	8 (21)	24	16	15	16, 2
LOD3.8-RH27	17	17	17	882	3.8	17	20 (3)	13	8 (15)	10	16 (2, 17)	10	16, 20, 19, 1
LOD3.8-RH29	16	16	13	686	3.8	12	1	8	10, 16	9	13	7	15, 16, 1
LOD3.8-RH31	14	14	14	618	3.8	14	3	6	15	9	17	6	2, 24
LOD3.8-RH13	33	32	30	1327	3.8	29	4 (13)	19	1	21	10 (9)	19	14, 9, 8
LOD3.8-RH20	21	21	17	899	3.8	16	20	1	(8)	12	16	6	16
LOD3.8-RH22	20	20	20	724	3.8	20	15	14	10	17	22	9	17, 13, 21
LOD3.8-RH3	65	64	59	2761	3.8	54	14 (9, 4, 11)	23	4 (3)	46	12 (9, 8, 1)	33	21, 10, 5, 8, 3, 2, 19, 16, 12
LOD3.8-RH30	15	15	15	723	3.8	15	13 (5)	9	12	8	9	5	5, 21
LOD3.8-RH35	13	12	13	497	3.8	10	1	10	16	9	13	5	15, 5
LOD3.8-RH39	8	8	8	267	3.8	8	3 (6)	8	15 (9, 17)	7	17 (7)	7	2, 18, 24, 13
LOD3.8-RH57	5	4	5	108	3.8	4	1 (10)	4	16 (21)	2	13	1	15

Table 1 (continued)

Stickleback						Pufferfish			Medaka			Zebrafish		
						No. of anchors ^b	chr ^c	No. of anchors ^b	chr ^c	No. of anchors ^b	chr ^c	No. of anchors ^b	chr ^c	
LOD3.8-RH6	43		36	2074	3.8	30	2 (4)	19	5	31	3	20	7, 18, 17, 6, 5, 25, 1	
LOD3.8-RH67	3	3		86	3.8	3	5	2	2	2	19	1	14	
LOD3.8-RH9	41		35	1689	3.8	33	3 (8, 10, 1)	23	15	33	17 (4, 13)	17	2, 20, 8, 4, 23, 22, 11	
LOD3.9-RH3	53		49	2478	3.9	42	17 (15)	26	11 (9, 10)	38	5 (7, 22, 16)	30	6, 11, 23, 22, 8, 5, 4, 25, 21, 2, 19, 17, 1	
LOD3.9-RH43	8	8	7	348	3.9	7	6	1	(17)	5	15 (18)	3	17, 12	
LOD3.9-RH17	30	25		991	3.9	18	9	15	18	16	1 (12)	8	1, 6, 22, 21, 19	
LOD3.9-RH47	7	7	7	223	3.9	7	16 (5, 1)	4	2 (3)	7	21 (2)	1	11	
LOD3.9-RH50	6	6	6	205	3.9	6	19	5	13	6	6	3	4, 25	
LOD3.9-RH52	6	6	6	258	3.9	6	20	5	8	4	16	3	16, 19	
LOD3.9-RH73	3	3	3	82	3.9	2	12	1	(9)	2	7	1	23	
LOD4-RH41	12	12		371	4	9	9 (2)	8	18 (9)	7	1	6	1, 14	
LOD4-RH15	27	27		1413	4	24	16	19	2 (1, 17)	10	21 (7)	12	9, 6, 22, 16, 1	
LOD4-RH49	8	8	8	217	4	1	(17)	6	11	3	5	6	6, 4, 12, 11	
LOD4-RH52	7	7	7	309	4	7	13	3	12	5	9 (12)	1		
LOD4.1-RH14	27	26		1142	4.1	26	12	2	(9, 13)	14	7 (4)	18	23, 8, 7, 6, 4	
LOD4.1-RH53	7	7	6	333	4.1	6	12 (11)	1	(19)	4	7 (8)	3	8, 23, 16	
LOD4.1-RH59	6	6	6	275	4.1	6	4	3	19	4	23	1	4	
LOD4.1-RH7	34	34		1406	4.1	32	6 (10)	21	17 (18)	24	15	17	13, 17, 20, 19	
LOD4.1-RH82	4	4	4	79	4.1	4	1	4	3	3	2	2	9, 17	
LOD4.2-RH11	30	29		1390	4.2	25	1 (21, 16)	16	3	15	2 (20)	12	1, 7, 6, 22, 9, 23	
LOD4.2-RH22	20	20		814	4.2	20	10 (1)	7	21 (8, 16)	12	11 (17, 13)	11	19, 15, 5, 20, 1	
LOD4.2-RH37	13	13		482	4.2	10	12 (19)	10	9 (19)	12	7 (23)	9	23, 6, 22, 11	
LOD4.2-RH87	4	4	3	141	4.2	3	13	2	12	2	9	2	8	
LOD4.4-RH3	40	39		1552	4.4	34	13 (15)	31	12 (4, 10)	26	9 (22)	22	8, 5, 21, 23, 13	
LOD4.4-RH45	11	10		322	4.4	10	13	9	12 (4)	8	9	4	5, 17, 12	
LOD4.4-RH54	7	7	6	233	4.4	2	16	6	2	5	21	4	1, 9	
LOD4.5-RH21	17	17		702	4.5	15	12	9	9	12	7	8	8, 23, 6	
LOD4.5-RH31	14	14		675	4.5	14	4 (3)	9	19	11	23	7	4	
		1320	1237	54373		1104		682		869		639		

No distinction is made between conserved segments and singletons in the case of zebrafish.

^a Including the co-localised markers. Note that co-localised markers are not taken into the account of anchors to model genomes.

^b Only anchors connecting assigned scaffolds are taken into account.

^c Number in parentheses correspond to singleton localization.

groups per chromosome. Chromosome RH maps were named after the linkage groups from the genetic maps [13,14]. Good correspondence was found between RH and linkage maps for most chromosomes. We found that linkage groups LG18 and LG21 were assigned to the same chromosome as that observed in the second generation linkage map. Linkage groups LG22 and LG25 were also found assigned to a unique chromosome. The main discrepancy between the genetic and RH maps concerned linkage group LG1 which we found was split in two chromosomes named LG1a and LG1b. One RH chromosome map contained no microsatellite from the genetic maps (LGx). The assigned RH groups represented 1222 markers (91% of the linked subset) and ranged in size from 49 to 2899 cR. Twenty-six of the RH groups containing three to 15 markers remained unassigned. These unassigned RH groups represented 120 markers (9% of the linked subset) and ranged in size from 13 to 723 cR. Given that the 91 RH groups represent 1342 markers out of a total of 1581, the probability to map a marker of interest in one of the existing RH groups can be estimated at ~85%. The RH map of sea bass chromosome LG1a is shown in Fig. 3. The whole RH map is available in Supplemental Fig. S1. Further cytogenetic studies would help confirm the assignment of the RH groups to the 24 chromosomes. Most of all, such studies would help standardise the sea bass karyotype.

Map construction has revealed an unexpectedly high resolution of our sea bass RH panel leading to a high number of RH groups ($n = 91$). By contrast, the seabream RH panel constructed at the same irradiation dose resulted in far fewer RH groups ($n = 28$) starting with less than 500 markers. Moreover the total size of the sea bass RH map amounts to 54,373 cR, which corresponds to 14 kb/cR (763 Mb/54,373 cR), compared to 140 kb/cR in the seabream RH map. This

resolution power of the sea bass RH panel irradiated at 3000 rad may be a consequence of the cell type used in the panel construction. Maybe the generated DNA fragments were smaller than expected. As the sea bass RH map consists mostly of gene markers, some gaps between RH groups could be due to a weak marker density corresponding to gene-poor regions such as centromeres.

Stickleback orthologous genes corresponding to sea bass markers were identified based on the Ensembl gene annotation (Supplemental Table S2). Among the 1236 gene markers positioned on the map, 74 markers happened to correspond to 33 orthologous genes in the stickleback, i.e. a redundancy of two to five markers per gene. These 74 markers were designed from independent ESTs or contigs and therefore treated as distinct sequences after the CAP3 assembly step. For 29 out of the 33 genes, markers that tagged the same gene were co-localised or mapped to adjacent positions. Only in three cases were the two duplicates mapped to two separate positions. The blind mapping of these duplicate markers thus provides a quality control for map construction.

Comparative genomics

Synteny relationships were established from the gene markers mapped in the assigned RH groups and having an ortholog identified with location information in at least one of the model species. Out of 1083 markers with an orthologous relationship, 323 (30%) allowed anchorage between sea bass and the four model species, 347 (32%) with three of them, 245 (23%) with two and 168 (16%) with just one model species for a total of ~3000 anchors. Synteny relationships identified by two or more consecutive markers defined a conserved

C1808_N	0	cR	CHR17	2473223	CHR.UN	-	CHR5	32002587	CHR11	25202907
C4176_N	49	cR	CHR17	1890689	CHR.UN	-	CHR5	33240856	CHR11	12435946
CX660539	104	cR	CHR17	1114552	CHR.UN	-	CHR5	25385095	CHR6	37046202
DQ149510	180	cR	CHR17	379872	CHR9	6164737	CHR7	16538149	CHR22	32765872
DLA0164_LG1	239	cR	-	-	-	-	-	-	-	-
DT044786	324	cR	CHR17	988141	CHR.UN	-	CHR.UN	-	CHR6	48602243
FM007504_N	376	cR	CHR17	1562977	CHR.UN	-	CHR.UN	-	CHR23	17882512
C41	412	cR	CHR17	3405990	CHR11	7809846	CHR5	11783081	CHR6	36601545
FM004843_N3	457	cR	-	-	-	-	-	-	-	-
FM004843_N	467	cR	CHR17	2923940	-	-	-	-	-	-
DN832034	505	cR	CHR17	3508640	CHR11	7730458	-	-	-	-
DLA122_LG1	534	cR	-	-	-	-	-	-	-	-
C116	574	cR	-	-	-	-	-	-	-	-
CX660764	588	cR	CHR17	3913886	CHR11	7386844	CHR5	12831699	CHR6	36718414
DLA0130_LG1	600	cR	-	-	-	-	-	-	-	-
CX660608	630	cR	CHR17	4110070	CHR11	7232437	CHR5	13114251	CHR22	5331932
CX660843	694	cR	CHR17	4641662	CHR11	3270788	CHR5	10514970	-	-
DT044699	719	cR	CHR17	4971829	CHR11	2997610	CHR5	9939522	CHR6	58954036
FM028815_N	763	cR	CHR17	5335290	CHR11	2693123	CHR5	9303556	CHR2	39693249
C3793_N	801	cR	CHR17	5659342	CHR11	3457921	CHR5	8766425	CHR11	1612290
C269	894	cR	CHR.UN	-	CHR11	3865742	CHR5	7752906	-	-
DT044559	946	cR	CHR17	6021169	-	-	-	-	-	-
DQ821115	1016	cR	CHR.UN	-	CHR11	8922901	CHR5	28577419	-	-
CX535546	1064	cR	CHR.UN	-	CHR11	8992745	CHR5	28410928	ZV7_NA2995	-
C4869_N	1136	cR	CHR17	6721502	CHR11	8332385	CHR5	29784323	CHR25	29792538
DT044625	1196	cR	CHR17	7114061	CHR.UN	-	CHR5	30314887	CHR23	7989490
AM040728	1228	cR	CHR17	7063546	CHR.UN	-	CHR5	30257347	CHR6	25936540
C93	1321	cR	CHR17	7945544	CHR.UN	-	CHR5	17195547	CHR5	992795
C271	1363	cR	CHR17	8810538	CHR11	1762964	CHR5	15807477	CHR1	5642510
FM022019_N2	1390	cR	CHR17	8854681	CHR11	1731178	CHR5	15747611	CHR11	16305128
DT044855	1440	cR	CHR17	8659507	CHR11	908385	CHR5	16111821	CHR21	7969958
C267	1475	cR	CHR17	8571415	-	-	CHR5	16180942	-	-
DT044753	1528	cR	-	-	-	-	CHR5	16326562	-	-
AF224281	1580	cR	CHR17	8627589	CHR11	887111	CHR5	16161202	-	-
C205	1661	cR	CHR17	9828279	CHR11	1029737	CHR5	31875994	CHR8	49266092
C4526_N2	1723	cR	CHR17	9586450	CHR9	8169808	CHR5	14667930	CHR11	14578899
DV217037	1762	cR	CHR17	9850930	CHR11	1015796	CHR5	31835005	-	-
DV217211	1795	cR	CHR17	10329978	CHR11	9063279	CHR5	26644383	CHR11	17138712
C718_N	1830	cR	CHR17	10535884	CHR.UN	-	CHR5	26897984	-	-
FM016504_N	1889	cR	CHR17	14117796	CHR11	10259575	CHR7	20770329	CHR23	850922
C4767_N	1964	cR	CHR17	13779517	CHR11	9999585	CHR5	21209758	CHR6	31408317
DN832245	2046	cR	CHR17	10940037	CHR.UN	-	CHR5	22518674	ZV7_NA1329	-
FM011326_N	2100	cR	CHR15	10554863	CHR10	10493251	CHR22	4135576	CHR17	36415257
FM006508_N	2154	cR	CHR17	11257130	CHR11	10494906	CHR5	24329841	CHR4	17640390
FM016415_N	2208	cR	CHR17	11708893	CHR11	10768865	CHR5	25915339	CHR11	23949791
FM001339_N2	2280	cR	CHR17	12386667	CHR.UN	-	CHR16	12429815	CHR19	853522
C45	2314	cR	CHR17	12538824	-	-	CHR5	19298401	-	-
DV217156	2345	cR	CHR17	12652640	-	-	-	-	-	-
DN832155	2362	cR	CHR17	12665854	CHR.UN	-	CHR5	19033489	-	-
FM010382_N2	2428	cR	CHR17	13104781	-	-	-	-	-	-
C1189_N2	2447	cR	CHR17	13134465	-	-	-	-	-	-
DLA0016_LG1	2478	cR	-	-	-	-	-	-	-	-

Fig. 3. Radiation hybrid map of the sea bass chromosome LG1a. Marker names are listed in column 1. All names correspond to gene-based markers except names in bold that correspond to microsatellites. Column 2 represents cumulative distances expressed in centiRays (cR). Following columns correspond to comparative data with, from left to right, stickleback, pufferfish, medaka and zebrafish. For every marker, chromosome number and coordinates of the putative orthologs in the genome sequences of the four model species are displayed. Boxes represent CSO between sea bass and model species except zebrafish.

segment (CS) while a single marker identified a singleton. Oxford grids shown in Fig. 4 recapitulate the conserved segments (CS) between sea bass and stickleback (Fig. 4A), medaka (Fig. 4B) and pufferfish (Fig. 4C) identified from the assigned RH groups. The Oxford grid between sea bass and zebrafish (Fig. 4D) is shown only as a rough guide as zebrafish does not belong to the Percomorpha and is phylogenetically too distant from sea bass to establish a pattern of chromosomal events. Synteny relationships between sea bass and gilthead seabream were also indirectly investigated with the help of sea bass–medaka relationships established from this work and medaka–seabream relationships described by Sarropoulou et al. [18] (Fig. 4E).

Synteny relationships are best described with stickleback through 1014 anchors to the sea bass assigned RH groups. The RH map revealed 24 CS between these two species. Every sea bass chromosome consists of a unique CS with the stickleback while three stickleback chromosomes (GAC1, 4, 7) each consist of two CS with sea bass. This implied three inter-chromosomal rearrangements between these two species while 18 chromosomes remained without inter-chromosomal changes. Two inter-chromosomal rearrangements (GAC04-LG2/LGx and GAC07-LG3/LG14) were also observed by Kuhl et al. [12]. The third one (GAC01-LG13/LG24) was only observed by us. An additional break (GAC20-LG16/LG18) identified by a singleton in the RH map was observed by Kuhl et al. [12].

Synteny relationships between sea bass and medaka were established from 806 true anchors. The map revealed 24 CS between

the two species. Every sea bass chromosome consists of a unique CS with medaka and conversely, which implies no synteny breakpoint between these two species. Synteny relationships between sea bass and pufferfish were established from 625 anchors revealing 26 CS between the two species. Of these 26 CS, 16 correspond to entire chromosomes with no major rearrangement. Four pufferfish chromosomes and one sea bass chromosome are composed of two CS implying five inter-chromosomal rearrangements since lineage split. The CS between sea bass LG3 and pufferfish TNI20 was identified from only two orthologous genes while most other pufferfish orthologs for this RH group were located in the “chromosome unknown”. Moreover no other synteny was identified with TNI20 because of the assembly status for this chromosome. Finally, 13 chromosomes did not undergo any rearrangement since the evolutionary radiation of Percomorpha some 200 million years ago (MYA) [40,41]. These chromosomes are written in bold in Figs. 4A, B, C. The analysis of synteny relationships between sea bass and zebrafish revealed 47 CS. However many anchors between the sea bass and the zebrafish genomes were singletons putatively identifying additional CS. Due to the phylogenetic distance between the two species, genomes must have been intensively rearranged since lineage split, creating a high number of short CS which could possibly be identified with a higher density of orthologous markers.

On an intra-chromosomal scale, Conserved Segments Ordered (CSO) between two species are regions in which the order of orthologous genes is perfectly conserved due to the absence of

rearrangement after lineage split [42,43]. The identification of CSO allows to anticipate the localisation of all genes in a specific region given that the synteny has been established with the genome of a model species.

A

Our sea bass RH map aimed at identifying these CSO with stickleback, medaka, pufferfish and zebrafish in order to benefit from the comprehensive sequencing of these genomes and to hypothesise the location on the sea bass genome of most if not all sea bass genes having a clear orthologous relationship with genes of the model species. Moreover the comparison with several species allowed to simultaneously ascertain the extent and boundaries of shared CSO, reinforcing the probability of the hypothesis, while being able to reveal the break points that arose in some lineages. CSO with stickleback, medaka and pufferfish identified using the AutoGRAPH web server [32] are indicated in Table 2.

A

	GAC17	GAC05	GAC04	GAC08	GAC02	GAC19	GAC09	GAC11	GAC10	GAC03	GAC06	GAC15	GAC01	GAC07	GAC18	GAC20	GAC18	GAC21	GAC14	GAC13	GAC12	CHR.UN
LG1a	41											1										3
LG1b		36																				2
LG2			28																	1		1
LG3				65								1	1	27								1
LG4					29														1			1
LG5		1				40																2
LG6							33															4
LG7					1			50					1								1	1
LG8		1							47				1	1					1	1	1	1
LG9				1			1			55	1		1	1								1
LG10											38											3
LG11												52		1				1				2
LG12													22									3
LG13														3	51							2
LG14															31							3
LG15		1								1						66						4
LG16																	35					3
LG17								2										17				3
LG18-21													1									3
LG19			1				1	1											54			2
LG20											1									46		2
LG22-25						1		1													58	2
LG24												23			1							2
LGx			29							1												2

B

	OLA05	OLA19	OLA10	OLA18	OLA04	OLA03	OLA06	OLA01	OLA08	OLA11	OLA17	OLA15	OLA22	OLA13	OLA14	OLA21	OLA16	OLA24	OLA20	OLA12	OLA09	OLA07	OLA02	OLA23	CHR.UN
LG1a	34																1					2			2
LG1b	1	26															1					1			5
LG2			20																		1				3
LG3				21							1														1
LG4					46																1				7
LG5						31																			2
LG6							29				1														6
LG7								24													1				5
LG8			1						42												1	1	2		1
LG9										28	1														6
LG10											48												1		4
LG11			1	2								28													1
LG12										1	3														2
LG13														18											1
LG14														3	33										6
LG15			2													20						1			7
LG16												1											1		6
LG17													1												1
LG18-21																									1
LG19																									5
LG20																									4
LG22-25						1																			6
LG24																									7
LGx																									22

C

	TNI11	TNI20	TNI01	TNI05	TNI13	TNI18	TNI03	TNI21	TNI15	TNI17	TNI10	TNI16	TNI07	TNI2	TNI08	TNI14	TNI06	TNI04	TNI12	TNI09	TNI19	CHR.UN	
LG1a	23																				2		12
LG1b							1		2						22						1		8
LG2				19																			5
LG3		2				1																	19
LG4			35													1							20
LG5				19																			11
LG6					31																		5
LG7						24						1										1	9
LG8	1						3	18				1											17
LG9		1							19				1										16
LG10										35												1	15
LG11											21												7
LG12												45											6
LG13													4	14									4
LG14															36								9
LG15			1													26							9
LG16									1	1	1						37						18
LG17																							3
LG18-21																							11
LG19																							26
LG20																							1
LG22-25						1																	24
LG24																							4
LGx							16				1												8

Fig. 4. Oxford grids between sea bass and (A) stickleback, (B) medaka, (C) pufferfish, (D) zebrafish, and (E) seabream. Lines correspond to sea bass chromosome maps named after the genetic map. Columns correspond to chromosomes of model fishes named as follows: DLA: sea bass chromosomes; GAC: stickleback chromosomes; OLA: medaka chromosomes; TNI: pufferfish chromosomes; DRE: zebrafish chromosomes; SAU: seabream chromosomes. The last column corresponds to the unlocalised part of the genome sequence (chromosome unknown). Conserved segments are figured in grey squares containing the number of orthologous markers that identified them. Singletons are figured in white squares. In the case of the sea bass–zebrafish Oxford grid, orthologous markers were figured indistinctly whether they consist of conserved segments or singletons because of the complexity of synteny relationships. In the case of the sea bass–seabream Oxford grid, conserved segments were deduced from the seabass–medaka and the seabream–medaka synteny relationships.

the genetic map. Columns correspond to chromosomes of model fishes named as follows: DLA: sea bass chromosomes; GAC: stickleback chromosomes; OLA: medaka chromosomes; TNI: pufferfish chromosomes; DRE: zebrafish chromosomes; SAU: seabream chromosomes. The last column corresponds to the unlocalised part of the genome sequence (chromosome unknown). Conserved segments are figured in grey squares containing the number of orthologous markers that identified them. Singletons are figured in white squares. In the case of the sea bass–zebrafish Oxford grid, orthologous markers were figured indistinctly whether they consist of conserved segments or singletons because of the complexity of synteny relationships. In the case of the sea bass–seabream Oxford grid, conserved segments were deduced from the seabass–medaka and the seabream–medaka synteny relationships.

D

	DRE01	DRE02	DRE03	DRE04	DRE05	DRE06	DRE07	DRE08	DRE09	DRE10	DRE11	DRE12	DRE13	DRE14	DRE15	DRE16	DRE17	DRE18	DRE19	DRE20	DRE21	DRE22	DRE23	DRE24	DRE25	CHR.UN
LG1a	1	1		1	1	7		1		7		13	1			1		1		2	1	2	3		1	2
LG1b			2					1	1								1			2			2			2
LG2														14												3
LG3	2						6																			2
LG4		4				4		2		2	2					1	1			6	2	5				1
LG5	1				1	1	11									2	2		2						1	1
LG6		1		1			2									1	8								9	1
LG7	12					1							1					1			1	3				
LG8	1		26			1		1													1			1		3
LG9	1				1			1							2	2	1		13	1				1	2	2
LG10		20		1				1			1	1					3	2		2		1	1	2		
LG11												1	13				3		1	1						1
LG12		4				1						4				13		2	3	1				1		6
LG13	1				1										9	1										
LG14					4				1	6	1	1		4	3			2			9			2		
LG15	3					1			8		1				1							1				2
LG16	1	1													26			2	1							1
LG17			3	1							1	1					1		2	15			1			3
LG18-21						1																		14		1
LG19		1	1			7			4		7		1						1		8					4
LG20						9			10			1	1	1										1		1
LG22-25				1			3	1	7			1				1					4			1	23	2
LG24	3					2			1													2	1			1
LGx				11			1				1		1													

E

	SAU21	SAU11	SAU22	SAU15	SAU2	SAU18	SAU12	SAU17	SAU24	SAU19-20	SAU6	SAU7	SAU8	SAU1	SAU9	SAU10	SAU14	SAU13	SAU5	SAU3	SAU16	SAU23	SAU25	SAU4	
LG1a	2																								
LG1b		1																							
LG2			1																						
LG3				1																					
LG4					4																				
LG5						1																			
LG6							1																		
LG7								1																	
LG8									1																
LG9										2															
LG10											5														
LG11												2													
LG12													1												
LG13														1											
LG14															1										
LG15																2									
LG16																	2								
LG17																		2							
LG18-21																									
LG19																									
LG20																									
LG22-25																									
LG24																									
LGx																									
	OLA05	OLA19	OLA10	OLA18	OLA04	OLA03	OLA06	OLA01	OLA08	OLA11	OLA17	OLA15	OLA22	OLA13	OLA14	OLA21	OLA16	OLA24	OLA20	OLA12	OLA09	OLA07	OLA02	OLA23	

Fig. 4 (continued).

A total of 112 CSO were identified between sea bass and stickleback. Some chromosomes, such as LG3 and LG18–21, are remarkable; they did not undergo any intra-chromosomal rearrangement and consist of a perfect CSO. In contrast, chromosomes LG8, LG10, LG12 are the most rearranged. However the synteny relationships for these chromosomes are too complex to identify the rearrangement events that led to the synteny pattern. All other chromosomes underwent local insertions or inversions leading to three to seven CSO. By aligning both ends of 45,000 BAC clones on the stickleback sequence, Kuhl et al. [12] identified 139 rearrangements leading to a greater number of CSO than was found in the present study. Although the density of anchors between the two genomes is lower in our approach, its resolution is comparable and could be improved by mapping additional markers. The next step will consist in verifying CSO that differ between both studies.

A total of 92 CSO were counted between sea bass and medaka. AutoGRAPH detected one CSO in chromosomes LG2, LG18–21 and LG19 while chromosomes LG10 and LG12 were the most rearranged with eight CSO. All other chromosomes harboured two to six CSO. A total of 70 CSO were identified between sea bass and pufferfish. As a relatively large fraction of the genome sequence is unassembled, this value likely represents an underestimate and does not mean fewer recombination events in the pufferfish lineage compared to stickleback and medaka lineages. Hidden CSO may lie in the region conserved with the pufferfish “unknown chromosome” for instance in LG3, LG4, LG9 or LG10. The sea bass map provides elements for assembling the unknown set of genome sequences of pufferfish, such as contigs containing orthologs localised in these regions.

When comparing sea bass, stickleback, medaka and pufferfish with each other, particular regions are highlighted such as chromosomes LG18–21, LG3, LG19 and LG13 that underwent very few rearrangements since lineage split. This may reveal structural stability linked with evolutionary constraints for these chromosomes. Conversely chromosomes LG12, LG18 and LG20

were intensively rearranged. However in some cases the quality of the genome sequence assembly of the model species as well as a mapping analysis bias may be responsible for some heretical CSO.

Forty-one singletons were identified between sea bass and stickleback, 61 between sea bass and medaka and 45 between sea bass and pufferfish. These singletons will have to be investigated individually to establish if they are real or not. Indeed the orthologous location of a given sequence in a model species was defined as the highest hit on the genome of that species, but it may sometimes not correspond to the true ortholog.

Correspondence between the sea bass and seabream RH maps was established from their medaka synteny relationships and with the help of 37 orthologous markers spread over 19 chromosomes (Supplemental Table S3). No inter-chromosomal rearrangement was observed between the sea bass and the medaka nor between the seabream and the medaka [18], which implies no inter-chromosomal rearrangement between the sea bass and the seabream. Additional markers will have to be found to confirm the correspondence between the five remaining chromosomes. A dense comparative map will also allow the identification of intra-chromosomal rearrangements between the two species.

Table 2

Seabass	Stickleback	Medaka	Pufferfish	Zebrafish			
LG1a	1	3	1	6	1	4	2
LG1b	1	4	1	4	1	2	1
LG2	1	6	1	1	1	2	1
LG3	1	1	1	3	1	1	2
LG4	1	7	1	5	1	1	3
LG5	1	3	1	4	1	1	1
LG6	1	5	1	3	1	5	2
LG7	1	3	1	5	1	6	2
LG8	1	11	1	4	1	1	1
LG9	1	3	1	6	1	2	3
LG10	1	8	1	8	1	2	2
LG11	1	4	1	4	1	3	2
LG12	1	8	1	8	1	4	3
LG13	1	2	1	3	2	4	1
LG14	1	4	1	2	1	5	4
LG15	1	4	1	4	1	3	1
LG16	1	7	1	4	1	4	1
LG17	1	5	1	5	1	5	1
LG18–21	1	1	1	1	1	1	2
LG19	1	3	1	1	1	3	4
LG20	1	6	1	3	1	7	3
LG22–25	1	7	1	3	1	1	2
LG24	1	4	1	2	1	2	2
LGx	1	3	1	3	1	1	1
All chr	24	112	24	92	25	70	47

CS: Conserved segments.

CSO: Conserved segment ordered.

Conclusion

For the first time a complete gene map of all 24 chromosomes of European sea bass is presented. Our RH map represents a new resource to facilitate genome navigation, to explore evolutionary history, to understand adaptation and to support marker-assisted selection of traits of aquaculture interest. The RH map complements the assembly of the draft resulting from the ongoing sequencing of the sea bass genome. Knowledge of all breakpoints and CSO with model species maximises the chances to map in silico any gene of interest knowing its location in the model genomes. Chromosomes that underwent few rearrangements during evolution have been identified and point at possible constraints to genome architecture. Based on these synteny relationships, the RH map should improve the assembly of sequenced model genomes. The ongoing efforts to establish synteny relationships with other Perciformes of aquaculture interest, such as the gilthead seabream and Nile tilapia, and fundamental models, such as the three-spined stickleback and haplochromids, will allow the quick transposition of research results to conspecifics. Access to functional information on genes and gene clusters in other taxa will directly benefit the interpretation of QTL in sea bass. Access to a dense genome map is essential as depth of coverage and length of reads often limit the assembly of next generation sequences. Therefore, our simplified method for the construction of a RH panel for a lower vertebrate from spleen is promising given the wealth of sequence data already appearing from next generation sequencing [44].

Acknowledgments

Research has been funded by the FP6 Network of Excellence Marine Genomics Europe (contract GOCE-CT-2004-505403) and the FP6 project Aquafirst (contract STREP-2004-513692). The authors

would like to thank Ifremer for providing fish. The authors also thank Dr. Emeric Sevin for assistance with the English language.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ygeno.2010.07.007.

References

- [1] G. Barnabé, F. René, Reproduction contrôlée du loup *Dicentrarchus labrax* (Linné) et production en masse d'alevins, CR Acad. Sci. Paris D 275 (1972) 2741–2744.
- [2] M. Vandeputte, J.F. Baroiller, P. Haffray, E. Quillet, Amélioration génétique des poissons: quelles réalisations et quels défis pour demain? Cah. Agricoles 18 (2009) 262–269.
- [3] T. Gjedrem, Selection and Breeding Programs in Aquaculture, Springer, Dordrecht, The Netherlands, 2005.
- [4] C. Massault, B. Hellemans, B. Louro, C. Batargias, J.K. Van Houdt, A. Canario, F.A. Volckaert, H. Bovenhuis, C. Haley, D.J. de Koning, QTL for body weight, morphometric traits and stress response in European sea bass *Dicentrarchus labrax*, Anim. Genet. (2009).
- [5] D.A. Chistiakov, B. Hellemans, F.A. Volckaert, Review on the immunology of European sea bass *Dicentrarchus labrax*, Vet. Immunol. Immunopathol. 117 (2007) 1–16.
- [6] S. Millot, M. Bégout, B. Chatain, Exploration behaviour and flight response toward a stimulus in three sea bass strains (*Dicentrarchus labrax* L.), Appl. Anim. Behav. Sci. 119 (2009) 108–114.
- [7] M. Vandeputte, M. Dupont-Nivet, H. Chavanne, B. Chatain, A polygenic hypothesis for sex determination in the European sea bass *Dicentrarchus labrax*, Genetics 176 (2007) 1049–1057.
- [8] N. Ospina-Alvarez, F. Piferrer, Temperature-dependent sex determination in fish revisited: prevalence, a single sex ratio response pattern, and possible effects of climate change, PLoS ONE 3 (2008) e2837.
- [9] T. Svásand, D. Crosetti, E. García-Vázquez, E. Verspoor, Genimpact — evaluation of genetic impact of aquaculture activities on native populations, Final Scientific Report, 2007.
- [10] S.J. Kaushik, D. Coves, G. Dutto, D. Blanc, Almost total replacement of fish meal by plant protein sources in the diet of a marine teleost, the European seabass, *Dicentrarchus labrax*, Aquaculture 230 (2004) 391–404.
- [11] E. Saillant, M. Dupont-Nivet, M. Sabourault, P. Haffray, S. Laureau, M.-O. Vidal, B. Chatain, Genetic variation for carcass quality traits in cultured sea bass (*Dicentrarchus labrax*), Aquat. Living Resour. 22 (2009) 105–112.
- [12] H. Kuhl, A. Beck, G. Wozniak, A.V. Canario, F.A. Volckaert, R. Reinhardt, The European sea bass *Dicentrarchus labrax* genome puzzle: comparative BAC-mapping and low coverage shotgun sequencing, BMC Genomics 11 (2010) 68.
- [13] D.A. Chistiakov, B. Hellemans, C.S. Haley, A.S. Law, C.S. Tsigonopoulos, G. Kotoulas, D. Bertotto, A. Libertini, F.A. Volckaert, A microsatellite linkage map of the European sea bass *Dicentrarchus labrax* L., Genetics 170 (2005) 1821–1826.
- [14] D.A. Chistiakov, C.S. Tsigonopoulos, J. Lagnel, Y.M. Guo, B. Hellemans, C.S. Haley, F.A. Volckaert, G. Kotoulas, A combined AFLP and microsatellite linkage map and pilot comparative genomic analysis of European sea bass *Dicentrarchus labrax* L., Anim. Genet. 39 (2008) 623–634.
- [15] T. Faraut, S. de Givry, C. Hitte, Y. Lahbib-Mansais, M. Morisson, D. Milan, T. Schiex, B. Servin, A. Vignal, F. Galibert, M. Yerie, Contribution of radiation hybrids to genome mapping in domestic animals, Cytogenet. Genome Res. 126 (2009) 21–33.
- [16] F. Senger, C. Priat, C. Hitte, E. Sarropoulou, R. Franch, R. Geisler, L. Bargelloni, D. Power, F. Galibert, The first radiation hybrid map of a perch-like fish: the gilthead seabream (*Sparus aurata* L.), Genomics 87 (2006) 793–800.
- [17] E. Sarropoulou, R. Franch, B. Louro, D.M. Power, L. Bargelloni, A. Magoulas, F. Senger, M. Tsalavouta, T. Patamello, F. Galibert, G. Kotoulas, R. Geisler, A gene-based radiation hybrid map of the gilthead sea bream *Sparus aurata* refines and exploits conserved synteny with *Tetraodon nigroviridis*, BMC Genomics 8 (2007) 44.
- [18] E. Sarropoulou, D. Nousdili, A. Magoulas, G. Kotoulas, Linking the genomes of nonmodel teleosts through comparative genomics, Mar. Biotechnol. (NY) 10 (2008) 227–233.
- [19] M. Miya, H. Takeshima, H. Endo, N.B. Ishiguro, J.G. Inoue, T. Mukai, T.P. Satoh, M. Yamaguchi, A. Kawaguchi, K. Mabuchi, S.M. Shirai, M. Nishida, Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial DNA sequences, Mol. Phylogenet. Evol. 26 (2003) 121–138.
- [20] D. Steinke, W. Salzburger, A. Meyer, Novel relationships among ten fish model species revealed based on a phylogenomic analysis using ESTs, J. Mol. Evol. 62 (2006) 772–784.
- [21] C. Hitte, J. Madeoy, E.F. Kirkness, C. Priat, T.D. Lorentzen, F. Senger, D. Thomas, T. Derrien, C. Ramirez, C. Scott, G. Evanno, B. Pullar, E. Cadieu, V. Oza, K. Lourgant, D.B. Jaffe, S. Tacher, S. Dreano, N. Berkova, C. Andre, P. Deloukas, C. Fraser, K. Lindblad-Toh, E.A. Ostrander, F. Galibert, Facilitating genome navigation: survey sequencing and dense radiation-hybrid gene mapping, Nat. Rev. Genet. 6 (2005) 643–648.
- [22] H.A. Lewin, D.M. Larkin, J. Pontius, S.J. O'Brien, Every genome sequence needs a good map, Genome Res. 19 (2009) 1925–1928.
- [23] S. Aparicio, J. Chapman, E. Stupka, N. Putnam, J.M. Chia, P. Dehal, A. Christoffels, S. Rash, S. Hoon, A. Smit, M.D. Gelpke, J. Roach, T. Oh, I.Y. Ho, M. Wong, C. Dettler, F.

- Verhoef, P. Predki, A. Tay, S. Lucas, P. Richardson, S.F. Smith, M.S. Clark, Y.J. Edwards, N. Doggett, A. Zharkikh, S.V. Tavtigian, D. Pruss, M. Barnstead, C. Evans, H. Baden, J. Powell, G. Glusman, L. Rowen, L. Hood, Y.H. Tan, G. Elgar, T. Hawkins, B. Venkatesh, D. Rokhsar, S. Brenner, Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*, *Science* 297 (2002) 1301–1310.
- [24] N.A. Hukriede, L. Joly, M. Tsang, J. Miles, P. Tellis, J.A. Epstein, W.B. Barbazuk, F.N. Li, B. Paw, J.H. Postlethwait, T.J. Hudson, L.I. Zon, J.D. McPherson, M. Chevrette, I.B. Dawid, S.L. Johnson, M. Ekker, Radiation hybrid mapping of the zebrafish genome, *Proc. Natl. Acad. Sci. USA* 96 (1999) 9745–9750.
- [25] C. Kwok, R. Critcher, K. Schmitt, Construction and characterization of zebrafish whole genome radiation hybrids, *Meth. Cell Biol.* 60 (1999) 287–302.
- [26] F. Vignaux, C. Hitte, C. Priat, J.C. Chuat, C. Andre, F. Galibert, Construction and optimization of a dog whole-genome radiation hybrid panel, *Mamm. Genome* 10 (1999) 888–894.
- [27] X. Huang, A. Madan, CAP3: a DNA sequence assembly program, *Genome Res.* 9 (1999) 868–877.
- [28] G.S. Slater, E. Birney, Automated generation of heuristics for biological sequence comparison, *BMC Bioinform.* 6 (2005) 31.
- [29] S. Rozen, H. Skaletsky, Primer3 on the WWW for general users and for biologist programmers, *Meth. Mol. Biol.* 132 (2000) 365–386.
- [30] T.C. Matise, M. Perlin, A. Chakravarti, Automated construction of genetic linkage maps using an expert system (MultiMap): a human genome linkage map, *Nat. Genet.* 6 (1994) 384–390.
- [31] T. Schiex, C. Gaspin, CARTHAGENE: constructing and joining maximum likelihood genetic maps, *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 5 (1997) 258–267.
- [32] T. Derrien, C. Andre, F. Galibert, C. Hitte, AutoGRAPH: an interactive web server for automating and visualizing comparative genome maps, *Bioinformatics* 23 (2007) 498–499.
- [33] L. Navarro-Martin, M. Galay-Burgos, G. Sweeney, F. Piferrer, Different sox17 transcripts during sex differentiation in sea bass, *Dicentrarchus labrax*, *Mol. Cell. Endocrinol.* 299 (2009) 240–251.
- [34] D. Casani, E. Randelli, S. Costantini, A.M. Facchiano, J. Zou, S. Martin, C.J. Secombes, G. Scapigliati, F. Buonocore, Molecular characterisation and structural analysis of an interferon homologue in sea bass (*Dicentrarchus labrax* L.), *Mol. Immunol.* 46 (2009) 943–952.
- [35] M. Minghetti, M.J. Leaver, E. Carpena, S.G. George, Copper transporter 1, metallothionein and glutathione reductase genes are differentially expressed in tissues of sea bream (*Sparus aurata*) after exposure to dietary or waterborne copper, *Comp. Biochem. Physiol. C Toxicol. Pharmacol.* 147 (2008) 450–459.
- [36] O. Jaillon, J.M. Aury, F. Brunet, J.L. Petit, N. Stange-Thomann, E. Mauceli, L. Bouneau, C. Fischer, C. Ozouf-Costaz, A. Bernot, S. Nicaud, D. Jaffe, S. Fisher, G. Lutfalla, C. Dossat, B. Segurens, C. Dasilva, M. Salanoubat, M. Levy, N. Boudet, S. Castellano, V. Anthouard, C. Jubin, V. Castelli, M. Katinka, B. Vacherie, C. Biemont, Z. Skalli, L. Cattolico, J. Poulain, V. De Berardinis, C. Cruaud, S. Duprat, P. Brottier, J.P. Coutanceau, J. Gouzy, G. Parra, G. Lardier, C. Chapple, K.J. McKernan, P. McEwan, S. Bosak, M. Kellis, J.N. Voff, R. Guigo, M.C. Zody, J. Mesirov, K. Lindblad-Toh, B. Birren, C. Nusbaum, D. Kahn, M. Robinson-Rechavi, V. Laudet, V. Schachter, F. Quetier, W. Saurin, C. Scarpelli, P. Wincker, E.S. Lander, J. Weissenbach, H. Roest Crolius, Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype, *Nature* 431 (2004) 946–957.
- [37] M. Kasahara, K. Naruse, S. Sasaki, Y. Nakatani, W. Qu, B. Ahsan, T. Yamada, Y. Nagayasu, K. Doi, Y. Kasai, T. Jindo, D. Kobayashi, A. Shimada, A. Toyoda, Y. Kuroki, A. Fujiyama, T. Sasaki, A. Shimizu, S. Asakawa, N. Shimizu, S. Hashimoto, J. Yang, Y. Lee, K. Matsushima, S. Sugano, M. Sakaizumi, T. Narita, K. Ohishi, S. Haga, F. Ohta, H. Nomoto, K. Nogata, T. Morishita, T. Endo, I.T. Shin, H. Takeda, S. Morishita, Y. Kohara, The medaka draft genome and insights into vertebrate genome evolution, *Nature* 447 (2007) 714–719.
- [38] W.J. Kent, BLAT—the BLAST-like alignment tool, *Genome Res.* 12 (2002) 656–664. [
- [39] S. Peruzzi, B. Chatain, B. Menu, Flow cytometric determination of genome size in European seabass (*Dicentrarchus labrax*), gilthead seabream (*Sparus aurata*), thinlip mullet (*Liza ramada*), and European eel (*Anguilla anguilla*), *Aquat. Living Resour.* 18 (2005) 77–81.
- [40] Y. Yamanoue, M. Miya, J.G. Inoue, K. Matsuura, M. Nishida, The mitochondrial genome of spotted green pufferfish *Tetraodon nigroviridis* (Teleostei: Tetraodon-tiformes) and divergence time estimation among model organisms in fishes, *Genes Genet. Syst.* 81 (2006) 29–39.
- [41] Y. Azuma, Y. Kumazawa, M. Miya, K. Mabuchi, M. Nishida, Mitogenomic evaluation of the historical biogeography of cichlids toward reliable dating of teleostean divergences, *BMC Evol. Biol.* 8 (2008) 215.
- [42] S.J. O'Brien, J.E. Womack, L.A. Lyons, K.J. Moore, N.A. Jenkins, N.G. Copeland, Anchored reference loci for comparative genome mapping in mammals, *Nat. Genet.* 3 (1993) 103–112.
- [43] W.J. Murphy, S. Sun, Z. Chen, N. Yuhki, D. Hirschmann, M. Menotti-Raymond, S.J. O'Brien, A radiation hybrid map of the cat genome: implications for comparative mapping, *Genome Res.* 10 (2000) 691–702.
- [44] P.A. Hohenlohe, S. Bassham, P.D. Etter, N. Stiffler, E.A. Johnson, W.A. Cresko, Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags, *PLoS Genet.* 6 (2010) e1000862.

B. Construction d'un panel d'hybrides d'irradiation du Tilapia du Nil (*Oreochromis niloticus*) : (Publication n°2)



Le Tilapia, *Oreochromis niloticus* appartient à la famille des cichlides de l'ordre des Perciformes. Parmi les 3000 espèces de cichlides, il existe près d'une cinquantaine d'espèces de Tilapia qui sont très diversifiées grâce à leur morphologie et leur pigmentation. Ils sont proches du bar et de la dorade.

Le Tilapia est une espèce aquacole de grande importance et constitue également un modèle d'étude des « species flocks » de *Cichlidae*. La distribution est mondiale, on peut les trouver en Amérique Centrale et du Sud, dans les grands lacs Africains, à Madagascar et en Inde.

Connu par son court intervalle de génération, le Tilapia peut servir de modèle d'étude à l'amélioration d'espèces françaises et européennes proches comme le bar, la dorade, le pagre qui ont une maturité sexuelle tardive, ce qui ne facilite pas leur étude génétique.

C'est la deuxième production aquacole au monde après la carpe. Son élevage est facile. C'est un poisson herbivore qui figure en bas de la chaîne alimentaire. Il se nourrit d'algues, plancton, déchets végétaux, contrairement aux autres poissons élevés en aquaculture qui nécessitent des quantités importantes de poissons, de farine et d'huile de poisson.

Les Tilapias ont développé une capacité exceptionnelle de résistance aux stresses, tolèrent bien les variations de pH, salinité, température, de concentration en oxygène. Ils se sont facilement adaptés à différentes niches écologiques. A cause de leur diversité exceptionnelle c'est le principal groupe utilisé pour l'étude des phénomènes de spéciation (naissance d'une nouvelle espèce). Pour toutes ces raisons, plusieurs laboratoires européens et américains se sont regroupés pour construire une carte RH du génome de tilapia *oreochromis niloticus* dans le cadre d'un projet soutenu par une subvention ANR. Secondairement, ce groupe s'est associé à un grand consortium international qui s'est organisé pour le séquençage des génomes de 5 cichlidés sous l'égide du NIH.

Grâce à l'acquis expérimental du laboratoire la construction du panel de tilapia s'est faite sans aucune difficulté. Trois fusions ont été réalisées entre des cellules de hamster déficientes pour le gène HPRT et des splénocytes de tilapia (*Oreochromis niloticus*). Elles ont fourni plus de 500 clones cellulaires exploitables qui ont été cultivés indépendamment pour avoir une quantité suffisante d'ADN sur lesquels 48 marqueurs ont été génotypés par PCR. A partir de ces résultats, nous avons sélectionné 190 lignées présentant la meilleure rétention quantitative et qualitative de marqueurs. Ces 190 lignées ont constitué le panel RH sur lequel ont été typés 1358 marqueurs par une méthode originale dérivée de la méthode Illumina pour le génotypage de SNP.

La comparaison de la carte RH du génome de Tilapia avec les séquences des génomes de stickleback, tetraodon, medaka et zebrafish a été réalisée pour établir les relations de synténie existant entre tous ces poissons. Par ailleurs avec nos collègues Parisiens, nous avons procédé à des hybridations *in situ* sur préparations chromosomiques (FISH) d'un certain nombre de clones BACs dont les extrémités avaient été positionnées sur la carte RH. La localisation cytogénétique de ces marqueurs a permis d'orienter la carte RH le long des chromosomes, d'attribuer chacun des groupes RH à un chromosome et de fixer le caryotype *d'oreochromis niloticus*. Le typage de microsatellites sur le panel RH a par ailleurs permis l'intégration des cartes génétiques et RH.

Malgré l'existence d'une séquence complète du génome de tilapia (Federica di Palma *et al.*, publication N° 6 à soumettre), la carte RH d'un maniement beaucoup plus léger reste un outil appréciable, totalement complémentaire. Elle reste très utile pour le positionnement et l'analyse des déterminants divers et de la différenciation sexuelle, déjà très étudiés du point de vue biologique chez cette espèce et faire ainsi du Tilapia un modèle de référence pour plusieurs espèces de l'ordre perciforme d'intérêts aquacoles majeurs. Par ailleurs la carte RH a permis de faire des études de comparaison de synténie avec d'autres perciformes, comme le bar et la dorade, qui présentent également un grand intérêt en aquaculture nationale et européenne (Sarropoulou *et al.*, 2008). Enfin cette carte a été utilisée par le BROAD pour l'assemblage des données de séquences génomiques du tilapia.

Publication n°2

A high-resolution map of the Nile tilapia genome : a resource for standing cichlids and other percomorphs

Richard Guyon,^{#1} Michelle Rakotomanga,^{#1} Naoual Azzouzi,¹ Jean Pierre Coutanceau,³ Celine Bonillo,⁴ Helena D'Cotta,⁵ Elodie Pepey,³ Lucile Soler,³ Marguerite Rodier-Goud,⁴ Angelique D'Hont,⁴ Matthew A Conte,⁵ Nikkie EM van Bers,⁶ David J Penman,⁷ Christophe Hitte,¹ Richard PMA Crooijmans,⁶ Thomas D Kocher,⁵ Catherine Ozouf-Costaz,² Jean Francois Baroiller,³ and Francis Galibert^{1,8}

BMC Genomics. 2012, 13:222

A high-resolution map of the Nile tilapia genome: a resource for studying cichlids and other percomorphs

Richard Guyon^{1†}, Michaëlle Rakotomanga^{1†}, Naoual Azzouzi¹, Jean Pierre Coutanceau³, Céline Bonillo⁴, Helena D'Cotta⁵, Elodie Pepey³, Lucile Soler³, Marguerite Rodier-Goud⁴, Angélique D'Hont⁴, Matthew A Conte⁵, Nikkie E.M van Bers⁶, David J Penman⁷, Christophe Hitte¹, Richard P.M.A Croijmans⁶, Thomas D Kocher⁵, Catherine Ozouf-Costaz², Jean François Baroiller³ and Francis Galibert^{1,8*}

Background: The Nile tilapia (*Oreochromis niloticus*) is the second most farmed fish species worldwide. It is also an important model for studies of fish physiology, particularly because of its broad tolerance to an array of environments. It is a good model to study evolutionary mechanisms in vertebrates, because of its close relationship to haplochromine cichlids, which have undergone rapid speciation in East Africa. The existing genomic resources

for Nile tilapia include a genetic map, BAC end sequences and ESTs, but comparative genome analysis and maps of quantitative trait loci (QTL) are still limited.

Results: We have constructed a high-resolution radiation hybrid (RH) panel for the Nile tilapia and genotyped 1358 markers consisting of 850 genes, 82 markers corresponding to BAC end sequences, 154 microsatellites and 272 single nucleotide polymorphisms (SNPs). From these, 1296 markers could be associated in 81 RH groups, while 62 were not linked. The total size of the RH map is 34,084 cR₃₅₀₀ and 937,310 kb. It covers 88% of the entire genome with an estimated inter-marker distance of 742 Kb. Mapping of microsatellites enabled integration to the genetic map. We have merged LG8 and LG24 into a single linkage group, and confirmed that LG16-LG21 are also merged. The orientation and association of RH groups to each chromosome and LG was confirmed by chromosomal in situ hybridizations (FISH) of 55 BACs. Fifty RH groups were localized on the 22 chromosomes while 31 remained small orphan groups. Synteny relationships were determined between Nile tilapia, stickleback, medaka and pufferfish.

Conclusion: The RH map and associated FISH map provide a valuable gene-ordered resource for gene mapping and QTL studies. All genetic linkage groups with their corresponding RH groups now have a corresponding chromosome which can be identified in the karyotype. Placement of conserved segments indicated that multiple inter-chromosomal rearrangements have occurred between Nile tilapia and the other model fishes. These maps represent a valuable resource for organizing the forthcoming genome sequence of Nile tilapia, and provide a foundation for evolutionary studies of East African cichlid fishes.

* Correspondence: galibert@univ-ennes1.fr

[†]Equal contributors

¹Institut Génétique et Développement (UMR 6061) CNRS/Université de Rennes 1, Rennes, France

⁸UMR CNRS 6061, Faculté de Médecine, Université de Rennes 1, Rennes, France

Full list of author information is available at the end of the article

Background

Tilapia is a common name for a large number of species belonging to the order Perciformes which accounts for one fourth of all vertebrate species. They are members of the family *Cichlidae* which consists of more than 3000 species distributed across tropical and subtropical regions. Tilapia are currently the second most farmed fish in the world with an annual production exceeding 2.8 million tons in 2010 [1]. Tilapia are a valuable source of protein for developing and emerging countries, but it is now also a prime fish commodity in developed countries. Apart from their domestic importance, there is a wealth of studies on different aspects of tilapia biology, e.g. on their physiology, endocrinology, immunology, toxicology and genetics. Tilapia have a short generation time, are sufficiently large in size for physiological studies and can be easily reared making them a perfect model system. They exhibit a versatile adaptability to different environmental conditions to match the vast array of their ecological habitats. They can tolerate incredible variations in temperature (12 to 43°C), pH (6 to 10), salinity (0 to 135 g/L), and oxygen levels (0.3 to 1.5 mg/L [2-5]). Therefore, they constitute exquisite models for environmental genomics, to analyse the interactions between the genome and the environment, and the adaptive responses to environmental stresses [6]. Because tilapia are closely related to the cichlid fishes in the Great Lakes of East Africa, which have undergone a spectacular radiation, they will contribute to our understanding of evolutionary mechanisms. The 2000 cichlid species in these lakes represent a collection of natural mutants that may provide insight into the genetic mechanisms of speciation and adaptation [7]. These unique biological features have motivated the development of a range of genomic tools for the Nile tilapia, *Oreochromis niloticus*, one of the most farmed tilapia species. An extensive collection of ESTs was recently constructed to aid the annotation of the forthcoming Nile tilapia genome and for gene expression studies [8]. Likewise the analysis of 106,259 BAC end sequences and their alignment on the genome sequence of four model fish species (stickleback, medaka, pufferfish and zebrafish) provides a valuable intermediate resource for the mapping of genes in cichlids [9]. The culmination of these efforts is the whole genome sequence currently being assembled by the Broad Institute (Cambridge, USA).

There are several economic traits in tilapia such as growth [10] and sex-ratio [11] that need improvement and require genetic markers for their selection. Likewise, identification of QTLs (Quantitative Trait Loci) for other economic traits are being performed in tilapia [12] as well as for immune responses [13]. Two Nile tilapia genetic maps were constructed, for QTL mapping and for selection purposes, of which the latest contains 538 microsatellites and 21 gene markers [14,15]. These genetic maps established 24 linkage groups although the tilapia karyotype is composed of just 22 chromosome pairs [16]. These genetic maps provided a first characterization of the tilapia genomes. Because only a few gene-based markers were mapped, synteny relationships with model fish species were only possible at low resolution. The map has been updated with a few more gene-based markers but the number of comparative markers remains limited [17,18].

Radiation hybrid (RH) mapping is suited to mapping all types of markers including gene-based markers, and can order them at high resolution. RH maps can integrate genetic maps through the mapping of polymorphic markers, as well as construct comparative maps through the mapping of non polymorphic markers (orthologous genes). In zebrafish the RH map allowed much higher gene-marker coverage of the genome and permitted comparisons with other vertebrates [19,20]. These advantages are particularly evident for fish species lacking genome tools, such as sea bass, where a gene-based RH map enabled comparisons with the genomes of stickleback, pufferfish, medaka and zebrafish [21]. Likewise, the two RH maps constructed for the seabream allowed comparisons with the pufferfish genome sequence [22,23]. Global synteny relationships were also established between three farmed Perciformes (seabream, European seabass and Nile tilapia) and with the model species (stickleback, medaka and pufferfish) [24].

High-resolution RH maps are also of great help in scaffolding genome sequences developed in shotgun projects [25,26]. In many cases draft and even “finished” genome sequences from shotgun projects contain large sequence gaps that imply inconsistencies in the placement of scaffolds. In addition, low in-depth sequences lack long-range continuity and provide only a fragmented view of a genome. This was precisely the case for the fugu genome sequence, which consists of 7213 unconnected scaffolds without any chromosome assignment [27]. The construction of RH panels for fish species has not been an easy task. To date, only four fish RH panels have been reported. Two zebrafish RH panels were derived from permanent cell lines [20,28]. More recently, a seabream RH panel was constructed from primary fibroblasts [22] and the European seabass RH panel was derived from splenocytes [21], thus avoiding the problems of genome rearrangements that arise in cell lines.

Here we describe the construction of a Nile tilapia RH panel derived from fresh splenocytes and a gene-rich RH map of 1358 markers. The RH map was integrated with the Nile tilapia karyotype by FISH analysis. This allowed us to assign the RH groups to the 22 chromosomes as well as to identify their orientation with respect to the centromere. The mapping of 154 microsatellites permitted the anchoring of the genetic map to the RH map. Amongst the different markers selected for the Nile tilapia RH map there was a large proportion related to growth and reproduction. In addition, a large number of SNPs identified in individuals from the 10th generation of the widely cultured GIFT (Genetically Improved Farmed Tilapia) strain were included in the RH map (van Bers *et al.*, submitted to Molecular Ecology Resources).

Results and Discussion

RH panel

We fused *Hprt*⁻ derivative CHO host cells with Nile tilapia splenocytes γ -irradiated at 3500 rad. Hybrid cells maintaining Nile tilapia chromosome fragments were selected based on their growth in HAT medium. A total of 381 hybrid cell lines were obtained through 3 fusion experiments. The retention frequency (i.e. the estimated percent of markers per clone) was determined for every clone by typing a set of 48 microsatellites selected from the genetic map [15]. We selected 190 hybrid cell lines on the basis of their retention frequency and their genome representation (Figure (Figure1).1). Further typing of 56 additional markers on this 190 cell line panel led to a cumulative retention frequency of 11.7 %. This rather low retention value is compensated by the unusually high number of hybrid cell lines, which allowed us to substantially increase the number of genotyping data. The Nile tilapia RH panel was therefore constructed from fresh live cells that required no primary culture, an important condition to avoid genome rearrangements typical of permanent cell lines. Splenocytes are convenient to use, as they are abundant, their dissociation is easy and can be performed in a relatively short time preceding the irradiation step.

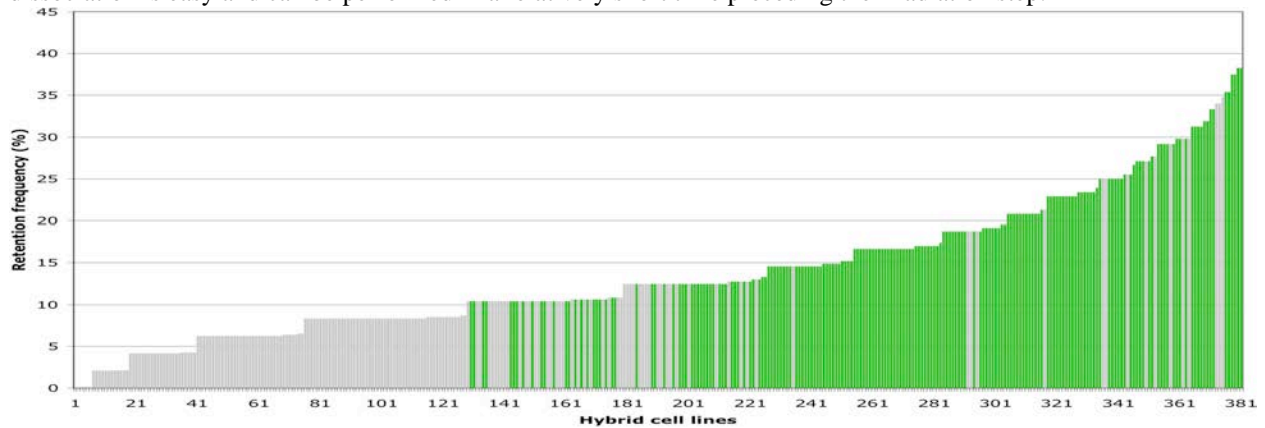


Figure 1 Retention frequency of the Nile tilapia hybrid cell lines. The hybrid cell lines are numbered from 1 to 381 on the X axis. Presence/absence of 48 microsatellite markers spread all over the tilapia genome [15] was estimated by PCR determination. Their retention frequency per clone is presented on the Y axis. The 190 hybrid cell lines selected on quantitative and qualitative criteria that constitutes the tilapia RH panel are in green.

Marker selection

A total of 16,195 Nile tilapia expressed sequence tags (ESTs) were collected, consisting of 5161 sequences from the CIRAD, 3537 sequences from the NCBI nucleotide database and 7497 sequences from the RBEST database (October 2008). A proportion of 3.5 % of these sequences were identified as simple repeats and masked by the RepeatMasker program. Sequence alignment with the CAP3 software resulted in 1476 contigs and 5692 singlets i.e. 7168 unique sequences putatively corresponding to as many genes (Additional file 1: Data S1). These 7168 sequences were aligned onto the stickleback, pufferfish, medaka and zebrafish genome sequences using the Exonerate software. A minimum score of 250, corresponding to a minimal alignment length of 50 bp, was applied following the usual recommendations to map orthologous ESTs [29]. A maximal alignment size of 300 bp was imposed to avoid hits that may in fact correspond to retrogenes rather than orthologs (retrogenes as opposed to pseudogenes which are characterized by non-sense or frame shift mutations). According to these criteria, 2475 of the unique sequences had a hit with at least one of the model genomes: 1920 (77.6%) had a hit with stickleback, 1836 (74.3%) with medaka, 1715 (69.3%) with pufferfish and 1304 (52.7%) with zebrafish (Figure (Figure2).2). A total of 942 Nile tilapia sequences were conserved across all four model species, while 224 Nile tilapia sequences were conserved exclusively with stickleback, 157 with medaka, 117 with pufferfish and 90 with zebrafish.

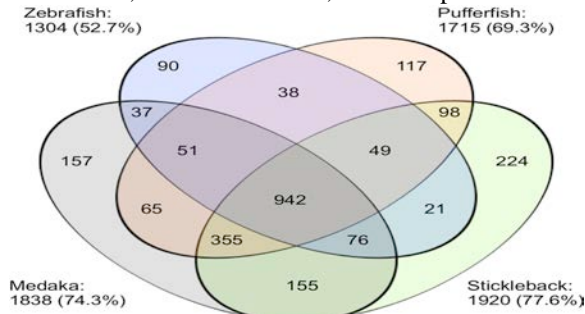


Figure 2 Venn diagram representing the distribution of markers shared by Nile tilapia and stickleback/medaka/ pufferfish/zebrafish. Each model species is represented by an ellipse. Number of markers shared by two species or more are indicated in every intersection. For each model species, the number of markers and the percentage of the 2475 Nile tilapia markers are indicated.

We preferentially designed gene markers from Nile tilapia ESTs having the highest conservation with the stickleback. Although Nile tilapia is phylogenetically closer to medaka [30], the stickleback genome appears to be the best reference sequence because it is assembled with highest confidence [9]. Each marker was designed from the aligned region having the best homology (highest score) with the reference genome. This strategy minimized the possibility of an intron lying between the oligonucleotides. The Nile tilapia Illumina BeadArray contained 1536 markers consisting

of 1300 genes, 97 BACs and 139 microsatellites. Sequence analysis of a Nile tilapia reduced representation library (RRL) resulted in the detection of 3569 SNPs. Of these, a subset of 384 SNPs was genotyped on the RH panel.

Genotyping

Out of the initial 1536 selected markers and the subsequent 384 SNP markers genotyped on the RH panel, 1026 and 272 markers respectively showed an exploitable profile and were selected to construct the RH map. Roughly one third of the markers had to be removed from the initial set, as their calls could not be separated in two distinct clusters of presence and absence on the graphic representation. This ratio of failure is high, but is in the range of what was observed for our previous RH map constructions using PCR genotyping (dog, seabream and seabass). With this strategy a similar proportion of designed pairs of oligonucleotides either do not properly amplify the test DNA or amplify both the test DNA and the carrier hamster cell DNA and thus are not useful for RH map construction. The current Nile tilapia map was constructed with a final set of 1358 markers consisting of 850 genes, 154 microsatellites, 82 BACs and 272 SNP, including 60 markers genotyped by PCR (see the vector suite in Additional file 2: Data S2).

RH group characteristics

The two-point analysis of the 1358 markers using the Multimap software started at a lod score of 4.0, which was then increased in a step-wise fashion up to a threshold of lod 7.0. The final map consists of 81 RH groups containing between 2 to 89 markers each for a total of 1296 markers. Sixty-two markers remained unlinked. The 1296 markers are spread in 1255 positions of which 1220 positions consist of a single marker, 32 positions contain two co-localized markers and three positions contain 3 to 6 co-localized markers. Multipoint analysis was carried out with CarthaGène software that ordered markers within each RH group and determined their distances expressed in centirays (cR_{3500}). RH groups ranged in size from 5 to 1906 cR_{3500} . Inter-marker distances vary between 1 and 164 cR_{3500} with an average of 27 cR_{3500} . Characteristics of the RH groups in terms of size and marker content are presented in Additional file 3: S3. Assuming a Nile tilapia genome size of 1060 Mb [31] the mapping of 1358 markers corresponds to a density of 1.28 marker per megabase or one marker per 780 kb, when considering an even distribution of the markers.

The relationship between cR_{3500} and kilobase (kb) can be estimated from the ratio of RH to genetic distances (cR/cM) knowing the ratio between cM and kb. To this end, we identified 82 pairs of microsatellites separated by a known distance measured in genetic and RH units. The cumulated distance of these 82 couples is 18,446 cR_{3500} on the RH map and 604 cM on the genetic map i.e. a ratio of 30.5 cR_{3500}/cM . The size of the Nile genome being 1060 Mb [31] and the size of the genetic map being 1311 cM the ratio was estimated to 840 kb/cM [15]. Therefore, the relation of physical unit to RH unit is estimated to be 27.5 kb/ cR_{3500} . The calculated size in kb of the RH map is 937,310 kb ($34,084 cR \times 27.5$) and thus corresponds to a coverage of 88% of the entire genome size. Considering that ~96% of the markers (1298 markers out of 1358) were mapped in the 81 RH groups, one can estimate the probability of mapping a novel marker of interest in one of the existing RH group to be 96 %. This figure can most likely be considered as a better estimate of the coverage of the Nile RH map.

Because most of the markers designed for the Nile tilapia RH map were from ESTs, the genome regions that remain uncovered may correspond to gene-poor regions such as heterochromatin or regions containing genes that are poorly expressed.

Integration of RH map and FISH data

The FISH mapping of BAC clones analyzed two by two allowed us (i) to assign RH groups to specific chromosomes with higher confidence, (ii) to orient them relative to each other and (iii) to localize centromeric and telomeric ends on the chromosome maps. They also served as a validation of linkage group assignment. In addition to ordering based on FISH mapping, the RH groups were tentatively ordered and orientated on the basis of the two-point analyses between markers close to RH group extremities and on the basis of the genetic map. The chromosome map of LG7 featuring RH and FISH maps along with the genetic map is presented in Figure Figure3.3. All chromosome maps are available in Additional file 4: Data S4 and online (<http://www.BouillaBase.org>). Table Table11 presents chromosome characteristics in terms of number of markers and RH groups. We first selected 42 BAC clones that contained markers assigned to 33 RH. A second set of 48 BAC clones was selected from regions of interest based on synteny with reference species, obtained from the BouillaBase server.

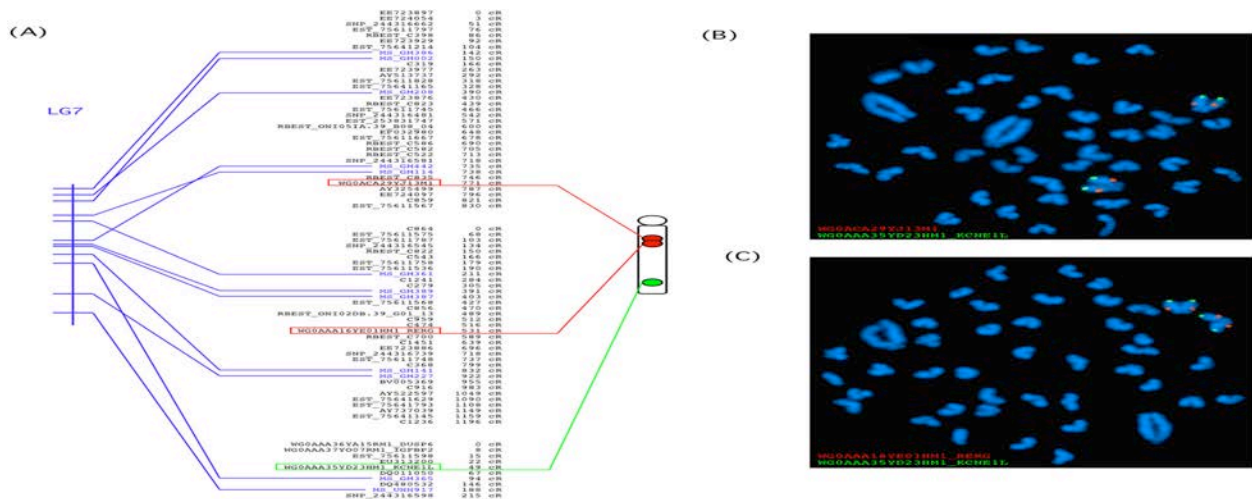


Figure 3 (A) Integrated genetic-RH-FISH map of the tilapia chromosome LG7. The RH map on the middle consists of three RH groups containing ordered markers whose coordinates are indicated in cR₃₅₀₀. Microsatellites (in blue) allowed the anchorage of the RH map to the genetic map [15] figured by a vertical bar on the left. Double-FISH of BAC clones highlighted by a red or green frame indicate the relative position of the RH groups on the chromosome symbolized on the right side. The chromosome is orientated with its centromere up. (B) Results of Double-FISH experiment of BAC clone WG0AAA35YD23HM1 revealed with FITC (green) and BAC clone WG0AAA16YE01HM1 revealed with Rhodamin (red) on a chromosome preparation. (C) Results of Double-FISH experiment of BAC clone WG0AAA35YD23HM1 revealed with FITC (green) and BAC clone WG0ACA29YJ13M1 revealed with Rhodamin (red) on a chromosome preparation.

Table 1: Characteristics of Nile tilapia chromosome maps

	No of RH groups	Size (cR3500)	No. of positions	No of co-localized markers	No. of markers	No. of genes	No. of BAC	No. of microsat	No. of SNP	No. of anchors	
										Stickleback	Pufferfish
LG1	2	1145	45		45	30	4	4	7	30	21
LG2	1	1038	44		44	34	3	2	5	34	24
LG3	5	1167	35	1	36	13	2	6	15	16	2
LG4	2	1296	55	2	57	45	2	3	7	42	23
LG5	3	1862	61	2	63	40	5	7	11	30	28
LG6	3	1622	60		60	43	4	8	5	44	24
LG7	3	2241	80		80	57	5	12	6	51	29
LG8-24	1	1677	61	2	63	46	2	5	10	42	26
LG9	2	1177	37		37	20	2	10	5	20	15
LG10	3	468	23		23	17	2	4	0	17	14
LG11	1	1465	48	3	51	29	5	7	10	27	19
LG12	1	1906	80	9	89	55	2	11	21	53	37
LG13	3	1349	48	2	50	31	2	4	13	34	29
LG14	1	1508	53	1	54	34	2	10	8	31	18
LG15	2	1269	47	3	50	36	2	5	7	29	20
LG16-21	4	1624	55		55	27	5	10	13	26	21
LG17	2	1489	51	3	54	30	3	8	13	25	16
LG18	2	1325	54	1	55	38	2	3	12	37	16
LG19	2	1462	53		53	36	3	6	8	36	27
LG20	3	1373	55	2	57	41	3	3	10	44	24
LG22	3	1095	40	4	44	25	2	4	13	24	8
LG23	1	1011	38	3	41	22	2	5	12	21	8
Sub-Total	50	30,569	1123	38	1161	749	64	137	211	713	449
Orphan groups		3515	133	4	137	71	13	8	45	72	38
Unlinked					60	31	5	8	16		
Total		34,084			1358	851	82	153	272		

Table 1 : Characteristics of Nile tilapia chromosome maps

Of the 90 initially selected BAC clones, FISH data from 45 BACs were not taken into account. They correspond to BACs producing many signals on different chromosome pairs (chimeric BACs or BACs that hybridize to duplicated regions), and/or to the observation of strong background signals in spite of the use of competitor and carrier DNA. They were deleted from the analysis and in many circumstances replaced by other BACs from the same linkage group. Successful hybridization results were those in which the two probes for the same LG group gave a clear signal and could be repeatedly observed in at least 10 metaphase spreads. Thus a total of 55 BAC markers (Table (Table2)2)

were successfully hybridized and mapped in 39 RH groups. This allowed all of the main RH groups ($n \geq 30$ markers) to have at least one BAC mapped by RH and FISH. For 17 chromosomes, two to four BACs were used to assemble several RH groups onto a single chromosome (Table (Table2).2). This was the case for instance, of LG7, constituted by three RH groups (Figure (Figure3).3). Each of the 22 chromosomes can now be identified with the help of one to four fluorescent probes. This is a particularly important result because, with the exception of the largest chromosome pair (Chr1/LG3), which is three times larger than any other Nile tilapia chromosomes and of the second chromosome pair (Chr2/LG7), none of the other chromosomes can be easily distinguished using classical cytogenetic techniques, due to similarity in size and fluctuations in the chromatin condensation [32,33]. These BAC chromosome markers can also be used to identify orthologous chromosomal regions among closely related species within the Tilapia group (such as other *Oreochromis* spp., or *Sarotherodon*). They can also be used to provide insights on the evolution of chromosome regions that have taken place since the divergence of tilapia and other cichlids from their ancestors. The list of clones, their reference and chromosome assignation is presented in Table Table22.

BAC Markers positioned by RH mapping and FISH analysis

All the BAC probes hybridized to the long arm of the chromosomes (such as LG7, shown in Figure Figure3)3) with the exception of LG15. This LG is composed of two RH groups and the BAC probes taken from each RH group both hybridized to the short arm of the chromosome. Chromosome LG15 is a small submetacentric chromosome in which the small arm is often clearly visible. It very probably corresponds to the chromosome 6 as defined by Ferreira et al. [34].

Integration of RH and genetic maps

The published genetic map of Nile tilapia was constructed with 545 microsatellite markers and 20 gene markers present on 24 linkage groups (LG) [15]. The integration of the RH map onto this genetic map was established using the 132 microsatellites present in both maps. Hence, fifty of the 81 RH groups were connected to the 24 genetic linkage groups, placing them onto 22 chromosome maps which contained on average 2.3 RH groups per chromosome. These 50 RH groups totalize 1123 map positions containing 1161 markers, which represent 89.4% of the markers located on the map. The remaining 31 small RH groups containing two to nine markers totalize 137 markers, which correspond to 10.6% of the markers on the map. They cannot be assigned to any chromosomes presently. These groups ranged in size from 5 to 345 cR₃₅₀₀.

In the RH map we were able to associate two small LGs, LG8 and LG24 into a single chromosome by RH mapping of four microsatellite markers (GM027 and UNH129 from LG8, GM104 and GM173 from LG24). We also merged and confirmed by FISH analysis the previously grouped LG16 and LG21 into a unique chromosome [16]. LG2, LG11, LG12, LG14 and LG23 correspond each to a single RH group whereas 15 LGs are made of two to five RH groups (see Table Table1).1). Consequently we have been able to locate the 24 linkage groups and placed them onto 22 chromosome pairs. For simplicity, we named the chromosomes maps after the genetic linkage groups (LG) of the Nile tilapia genome [15].

Overall the microsatellites in LG11, LG12, LG14 or LG23 are in the same order in the RH and genetic maps except for small local inversions that may be due to vector quality (see the data computation chapter) in one of the mapping methodologies. However, a larger discrepancy was observed in the lower part of LG14. This RH group was tentatively broken at higher lod scores (up to 7.0) but no reordering of the resulting groups was consistent with the genetic map. Given the high lod score to which this group stayed unbroken we believe that the correct order is that of the RH map.

Comparative genomics

Synteny relationships were established from the markers of the assigned RH groups having localized orthologous genes in the sequences of model species. Orthologs localized in the “chromosome unknown” of model species were not taken into account. Of the 1123 mapped positions in the assigned RH groups, 277 markers allowed anchorage of the Nile tilapia genome with all four model fish species, 268 with three species, 165 with two species and 78 with one species representing a total of 788 orthologous markers providing 2320 anchors. Synteny relationships identified by two or more consecutive conserved markers defined a conserved segment (CS) while a single marker identified a singleton [35][36].

The Oxford grid shown in Figure Figure4A4A recapitulates the CS found between Nile tilapia and stickleback. The Nile tilapia RH map and the stickleback genome sequence were connected by 713 anchors defining 23 CS. Seventeen CS correspond to entire chromosomes in which synteny is perfectly conserved between the two species. The Nile tilapia chromosome LG7 is made of two CS corresponding to stickleback chromosomes GAC14 and GAC19. Conversely, Nile tilapia chromosomes LG2 and LG17 consist of one CS each that are fused in stickleback to form chromosome GAC04. Furthermore, Nile tilapia chromosomes LG3 and LG10 both correspond to stickleback chromosome GAC07. This pattern of synteny would imply at least three inter-chromosomal rearrangements between the two lineages. The presence of two interstitial telomeric signals in Nile tilapia LG3 [32,33] suggests that this chromosome arose by two fusions. It has been suggested that these occurred within the cichlid lineage [34] but the stickleback-tilapia synteny (LG3 – GAC07) may suggest that these are older.

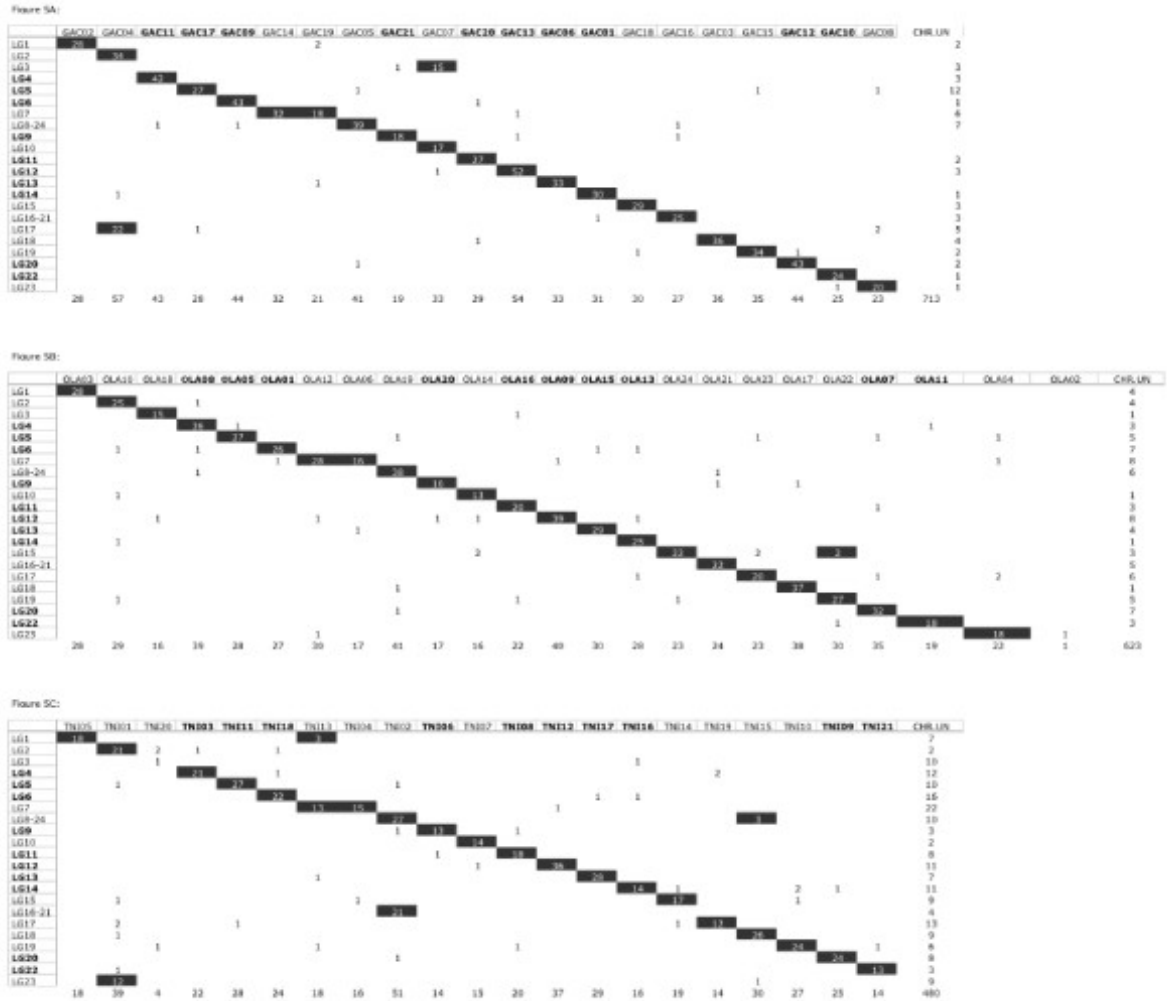


Figure 4
Oxford grids between Nile tilapia and (A) stickleback, (B) medaka, (C) pufferfish. Chromosomes are named as follows : LG : Nile tilapia chromosomes; GAC : stickleback chromosomes; OLA : medaka chromosomes; TNI : pufferfish chromosomes. Conserved chromosomes ...

A total of 623 anchors identified 24 CS connecting the Nile tilapia RH map and the medaka genome sequence (Figure (Figure4B).4B). Synteny is entirely conserved between 20 chromosomes of the two species. As with stickleback, the Nile tilapia chromosome LG7 is made of two CS corresponding to medaka chromosomes OLA06 and OLA12. The Nile tilapia chromosome LG15 is also made of two CS, a large one corresponding to medaka chromosome OLA24 and a small one corresponding to medaka chromosome OLA22. This pattern implies two inter-chromosomal rearrangements that would have occurred in one or the other lineage. Medaka chromosome OLA02 is the only chromosome with no Nile tilapia chromosome counterpart in the Oxford grid (Figure (Figure4B).4B). However three arguments suggest that LG23 is the missing counterpart of medaka chromosome OLA02 in the Oxford grid: (a) a two-point analysis between end-markers links orphan group RH36 to LG23 (see Methods section) (b) RH36 and medaka chromosome OLA02 share five ortholog sequences (see Additional file 1: data S1) and (c) tilapia marker AF116240 located on LG23 has an ortholog sequence on medaka chromosome OLA02. The grouping of RH36 and LG23 creates an additional synteny breakpoint with medaka as well as with stickleback and pufferfish.

We identified 449 anchors connecting the Nile tilapia RH map and the pufferfish genome (Figure (Figure4C)4C) defining 24 CS between these two species. Synteny appeared totally conserved between 14 chromosomes of the two species. Chromosome LG1 corresponds to two CS, a large one that corresponds to chromosome TNI05 and a small CS made of two markers, which corresponds to chromosome TNI13. The chromosome LG7 consists of two CS, one with chromosome TNI13 and one with chromosome TNI04. Pufferfish chromosome TNI01 is made of two CS corresponding to LG2 and LG23 respectively. Chromosome TNI02 is also made of two CS with Nile tilapia chromosomes LG8-24 and LG16-21. A small additional segment of chromosome TNI15 is conserved with Nile tilapia chromosome LG8-24. This pattern of conservation implies four inter-chromosomal events. Assignment of the orphan group RH36 to LG23 as discussed above would create an additional CS and would imply another inter-chromosomal event. Additional CS exist but have not been identified yet such as in chromosome LG3 for which most of the pufferfish orthologs are located in the “chromosome unknown” file of the pufferfish assembly.

Ten chromosomes show no synteny breakage across the four species (bolded in Figure Figure4A,4A, 4B, 4C). Twenty five singletons were identified between Nile tilapia and stickleback, 44 between Nile tilapia and medaka and 38 between Nile tilapia and pufferfish. These singletons suggest putative new CS but they also may be artefacts. Indeed the orthologous location of a given gene in a model species was defined as the best hit on the genome sequence of that species. However the best hit may in some instances have corresponded to a paralog especially when the true ortholog has been lost as hypothesized by Soler et al. [9] to explain a possibly overestimated number of breakpoints. Consequently each singleton will have to be established as a new CS by the mapping of additional and informative markers.

The number of CS appeared similar between Nile tilapia and each of the three reference models investigated in this work. This is in concordance with what was previously observed in the comparative map of the sea bass genome and the same fish models [21]. Finally, Nile tilapia and zebrafish were considered too distant phylogenetically to establish a pattern of chromosomal conservation despite the fact that 531 anchors were identified between the two species.

On an intra-chromosomal scale, Conserved Segments Ordered (CSO) are regions in which the order of orthologous genes is perfectly conserved [35,36]. The simultaneous comparison of several species allowed us to ascertain the extent and boundaries of shared CSO while also revealing the breakpoints that arose in some lineages. The Nile tilapia RH map aimed at identifying these CSO with stickleback, medaka and pufferfish in order to benefit from the comprehensive sequencing of these model genomes. Thus the location on the Nile tilapia genome of unmapped genes having a clear orthologous relationship with genes of the model species could be hypothesized with high confidence. CSO between stickleback, medaka, pufferfish and Nile tilapia were identified using the AutoGRAPH web server and are presented in Table Table3.3. Detailed CSO of LG7 are shown in Figure Figure5.5. Comparative maps of each of the 22 chromosomes are presented in Additional file 5: Data S5.

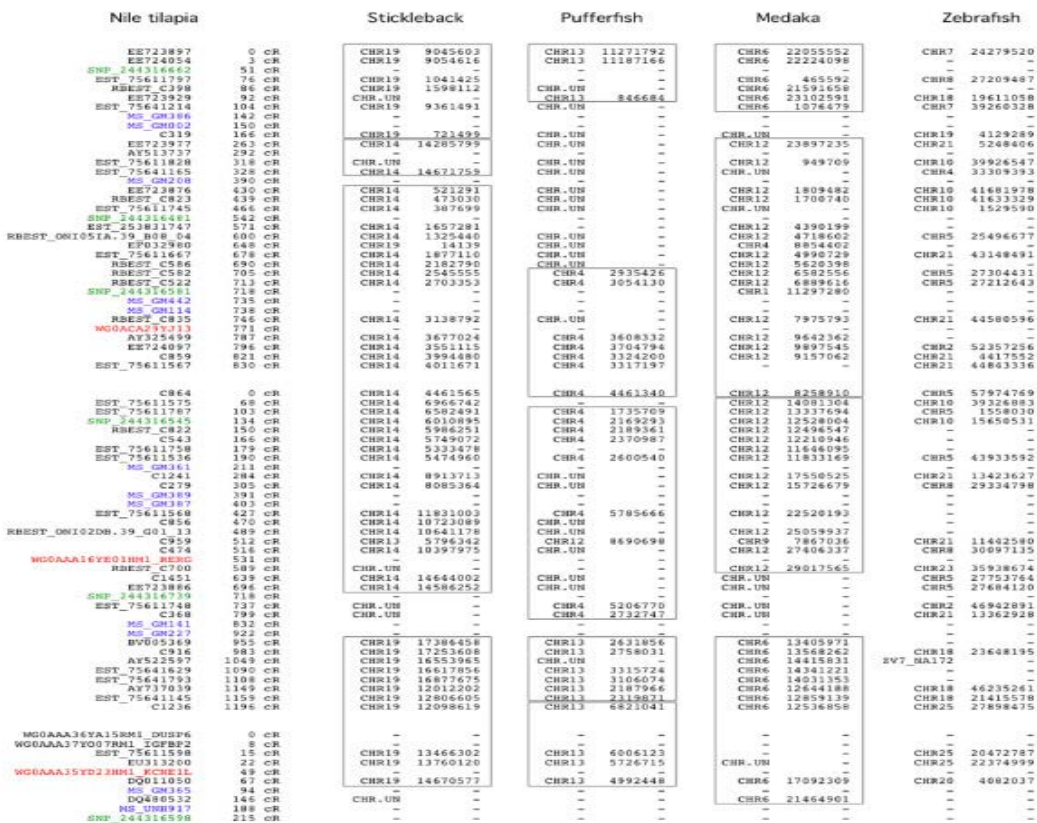


Figure 5 Comparative map of the Nile tilapia chromosome LG7. Column 1 corresponds to marker names. All markers are gene-based markers except (a) those with prefix “MS” which correspond to microsatellites (in blue) taken from Lee et al. (2005), (b) those with prefix “WG0” which are BAC end markers (in red) and (c) those with prefix “SNP” which correspond to SNP-based markers (in green). Column 2 corresponds to marker coordinates expressed in centiRays (cR₃₅₀₀). Following columns correspond to comparative data with, from left to right, stickleback, pufferfish, medaka, zebrafish. For every marker, chromosome numbers and coordinates of the putative orthologs in the genome sequences of the four model species are displayed. CSO between Nile tilapia and stickleback/medaka/pufferfish are figured in boxes.

We identified 90 CSO between Nile tilapia and stickleback. Chromosomes LG14, LG9, LG10 and LG13 underwent few rearrangements with one to two CSO only (1, 2, 2, 2 respectively) while for LG4, LG18 and LG20, seven, six and seven CSO respectively were identified, showing evidence of considerable rearrangement. A total of 79 CSO were identified between Nile tilapia and medaka. One CSO was detected in chromosomes LG2 and LG10 while LG15 and LG18 were the most rearranged. Only 57 CSO were detected between Nile tilapia and pufferfish (Table Table33).

A higher number of CSO was identified with stickleback than with medaka and pufferfish. Considering that the phylogenetic position based on various parameters between Nile tilapia and the other model species indicates a closer

relationship between Nile tilapia and medaka one would rather have expected the contrary [30,37]. However, in terms of sequence similarity Nile tilapia is actually closer to stickleback than to medaka and pufferfish. This apparent discrepancy, if confirmed, would indicate that forces shaping the overall genome structure are different from those affecting gene function and evolution. Finally, from a technical standpoint, it is important to recall that the number of observed CSO shared by two species depends in part on the number of anchors used to establish the respective comparative maps. Indeed the higher the number of orthologous genes is, the more resolution the comparative map will have. Ideally in this study, the comparative maps should be based solely on 1:1:1 orthologs between Nile tilapia, stickleback, medaka and pufferfish. However only half of the markers satisfy this condition. Comparative maps based on this smaller number of markers have too little resolution to reach any conclusions.

The sequence assembly status in some genomic regions of model species also prevents the identification of CSO with Nile tilapia. For example, assemblies of stickleback GAC17 (syntenic of Nile tilapia LG5), medaka OLA09 (syntenic of LG12), pufferfish TNI04 (syntenic of LG7) or TNI19 (syntenic of LG17) are incomplete and lead to an underestimation of the number of CSO. Assuming a conserved gene order with the Nile tilapia the RH map would provide an opportunity to locate the unassigned contigs of model species. In this way the Nile tilapia RH map can be seen as a tool for improving the sequence assemblies of other fish species.

Chromosomes LG10, LG14, LG9 and LG13 appear to be the least rearranged between Nile tilapia and the three model genomes. Conversely LG5, LG6 and LG18 were the most rearranged. This observation suggests that the genome plasticity and the underlying evolutionary constraints are not evenly distributed across the genome.

Go to:

Conclusions

Through a spectacular decrease in cost and with the capability to generate more than hundred gigabases per week, the New Sequencing Technologies (NGS) have revolutionized the field of genomics over the last few years. It is now possible to obtain deep knowledge of the genomes of many more species than we could have dreamed of even ten years ago. However the main drawback of NGS is the short length of their reads. Although steadily increasing, sequence reads are still very short (~100 nucleotides). This is not a problem when the goal is to re-sequence individuals and align the reads to a reference sequence. However this short size, even with a pair ends sequencing approach, renders the problem of *de novo* sequencing of large genomes difficult. Many of the novel assemblies produced with this approach are composed of a very large number of scaffolds [38]. This discontinuity does not affect gene discovery, polymorphism analysis and sequence comparison between closely related species but it greatly limits the study of the genome structure and evolution. RH mapping and FISH mapping of markers present in different contigs and scaffolds allow to link them and deduce larger super scaffolds.

Here we report the construction of a high-resolution RH map of Nile tilapia containing ESTs, genes, microsatellites and SNPs. The RH map has an estimated density of one marker every 780 kb. Fifty RH groups, which contained the vast majority of the markers (1161 out of 1358), were assigned to the 24 previously known LGs, which in turn were located and oriented on the 22 Nile tilapia chromosomes through BAC multicolor FISH mapping on metaphase chromosomes. Already this RH map allows us to locate a large number of physiologically important genes. For example, group RH17 located on chromosome LG15 contains the estrogen receptor gene together with *gata4* and the follicle-stimulating hormone gene (*fshb*). This last gene has been shown to regulate the activity of Gata4, a transcription factor involved in ovarian function, by regulating the aromatase *cyp19* gene [39]. Likewise, the growth hormone receptor gene *ghr1* (marker C456) was mapped in the group RH3 assigned to chromosome LG12 and the growth hormone receptor gene *ghr2* (marker C474) was mapped in the group RH9 assigned to chromosome LG7.

The RH map associated to the FISH data also offers a detailed synteny analysis with three of the four model species (stickleback, medaka, pufferfish). Due to the great evolutionary distance separating Nile tilapia from zebrafish (>300MY), it was not possible to reach definitive conclusions about synteny between these species. Furthermore, it provides a foundation for studying karyotypic evolution in the flocks of haplochromine cichlids in East Africa, including the evolution of sex chromosomes [40-42] and the origins of B chromosomes [43]. By contributing to the construction of a golden path for the Nile tilapia genome assembly, these maps will enable QTL and association mapping of adaptive traits in each of the haplochromine species flocks.

The mapping of a number of SNPs derived from 20 individuals of the 10th generation of the widely cultured GIFT strain are included. To the best of our knowledge these SNPs are the first set of genome wide SNPs publicly available for Nile tilapia. SNPs are gaining popularity for use in e.g. parental assignment [44] and for the estimation of genetic parameters in tilapia breeding. The 272 SNP markers were shown to allow the discrimination between different strains and species of tilapia (van Bers *at al.*, submitted Molecular Ecology Resources), and will be used in the near future to assess the genetic diversity of natural populations of Nile tilapia (Richard Crooijmans, personal comm.). The physical mapping position of these SNPs determined in this study allows the selection of unlinked SNPs for these future applications. Finally the map will help to place and orientate on the Nile tilapia karyotype many of the scaffolds identified in the forthcoming tilapia genome sequence determined with the Illumina technology and assembled by the BROAD Institute.

Go to:

Methods

Construction of a Nile tilapia radiation hybrid panel

The RH map was constructed from a fully-inbred homozygous clonal line of *O. niloticus* consisting of all-female fish, generated at the University of Stirling (Scotland, UK). These fish were derived by gynogenesis from a strain originating from Lake Manzala (Egypt) [45]. A panel of radiation hybrid cell lines was constructed using the methodology described previously [22,46]. Briefly for each fusion, a splenocyte suspension was prepared using one clonal fish as described in Guyon et al. [21]. The suspension was γ -irradiated at 3500 rad. Splenocytes were fused with *Hprt*⁻ derivative CHO cells in a 5: 1 ratio (Splenocyte/CHO) in the presence of polyethylene glycol 1500 (Roche, Mannheim, Germany). Cells were seeded in 6-well microplates at a total concentration of 150,000 cells per well and cultivated with HAT medium for 3 to 4 weeks until hybrid clone appearance. Each clone was recovered and further cultivated under HAT selection approximately one week in a 60 mm diameter Petri dish. After trypsinisation DNA was extracted from individual clones using the NucleoSpin Tissue kit (Macherey-Nagel, Düren, Germany). DNA concentration was estimated by fluorescence quantitation using Quant-iT Picogreen assay kit and a Qubit measuring device (Invitrogen, Carlsbad NM, USA). DNA extracts of hybrid cell lines were amplified by a Whole Genome Amplification (WGA) procedure when additional material was needed. In these cases, two separate WGA were performed with 10 ng of DNA each using V2 GenomiPhi kits (GE healthcare, Fairfield CT, USA). WGA products were pooled providing ~10 μ g of material for subsequent genotyping. The reliability of WGA was previously demonstrated by Senger et al. [22]. Fishes were anesthetized with the addition of phenoxybarbital in the water tank (3ml per 10 liter) and then sacrificed following the rules established by the ethical local committee.

Marker selection

Nile tilapia expressed sequences (ESTs and complete mRNA sequences) were either downloaded from NCBI nucleotide and RBEST (<http://reprobio.nibb.ac.jp/>) databases or provided by the CIRAD (Montpellier, France). After masking simple repeats using the RepeatMasker program [47] sequences were aligned together using the CAP3 software [48] with default parameters. The resulting unique sequences were aligned on stickleback (v1.0 assembly), pufferfish (v8 assembly), medaka (v1.0 assembly) and zebrafish (Zv7 assembly) genome sequences using the Exonerate software [49]. Orthologous sequences were searched using a minimal score of 250 and an alignment size of 80 to 300 nt. Sequences which showed conservation with the highest number of species were selected to design markers. For every marker, the coordinate of the best hit on each of the model genomes was considered as the location of the putative ortholog in the model species. In addition, BAC end sequences from the CIRAD and the University of Maryland (USA) [9] containing genes of interest were selected for the RH mapping. Microsatellites of the genetic linkage map of Nile tilapia (second generation) [15] were also selected in order to anchor the genetic map to the RH map.

In addition SNP markers were identified in a pool of 20 individuals obtained from the 10th generation of the GIFT population (WorldFish Center, Malaysia). The development of the SNP markers is described in detail elsewhere (van Bers et al., in prep). In brief, an RRL was prepared and sequenced using the Illumina GAI sequencing technology. Pooled DNA was digested with the *RsaI* restriction enzyme and fragments of 3.5-4 kb were isolated by electrophoresis. The fragments were sheared and used for high throughput sequencing. Sequence reads stringently filtered for quality were first assembled using SSAKE [50] to constitute a reference draft sequence. Less stringently filtered reads were subsequently mapped onto the reference draft using MAQ [51], allowing the detection of SNPs. The Minor Allele Frequency (MAF) was calculated based on how many times a SNP was observed in the sequence data. Only SNPs showing the minor allele at least three times were considered as true SNPs. Illumina type II SNPs with a design score >0.75 and a MAF >0.16 were selected for genotyping.

Marker sequences used to construct the map are given as supplemental Additional file 6: data S6 and have been submitted to NCBI: accession numbers, ss 244316446–244316740 (SNP markers) and 253831740–253831804 and 75611463–75642120 (EST markers). BAC markers are deposited in <http://www.BouillaBase.org>.

Genotyping

Genes, microsatellites and BAC end markers were typed on the Nile tilapia RH panel using the 1536-marker Illumina BeadArrays system. Amplification of 45–55 bp loci was performed using oligonucleotides 20–23 nt in length complementary to the Nile tilapia sequences and designed using the Illumina proprietary design program. For this, the program was adapted to design oligonucleotides flanking non-polymorphic sites. Oligonucleotides were synthesized and spotted onto two 96-sample array matrices by the Illumina company (San Diego, CA, USA). The genotyping was carried out using the Illumina GoldenGate technology. SNP markers were typed on the RH panel with a 384 SNP multiplex genotyping assay using the GoldenGate Assay. Oligonucleotides were designed flanking a SNP according to the Illumina design program specifications. The assay was deployed on a BeadXpress platform using the Veracode technology.

The Illumina Genome Studio software was used to visualize typing results and score the presence/absence of the markers in the hybrid cell lines by a method adapted from McKay et al. [52]. In our experiment all markers including the SNPs behaved as homozygous markers because the hybrid lines were constructed from a homozygous clonal Nile tilapia line implying homozygosity at all loci. Consequently a single allele-specific oligonucleotide was used in the GoldenGate extension step and only the axis of the corresponding fluorochrome was taken into account for a given marker. Clones located above a threshold of 0.3 on this axis on the cartesian plot were scored as present regardless to the value on the other axis. Clones located under the threshold were scored as absent. The threshold was adjusted depending on the overall distribution of dots on the typing profile. Clones located close to the threshold were scored

as ambiguous (Figure (Figure6).6). Microsatellites used to characterize hybrid cell lines and additional BAC end markers were typed by PCR and scored as described in Guyon et al. [21].

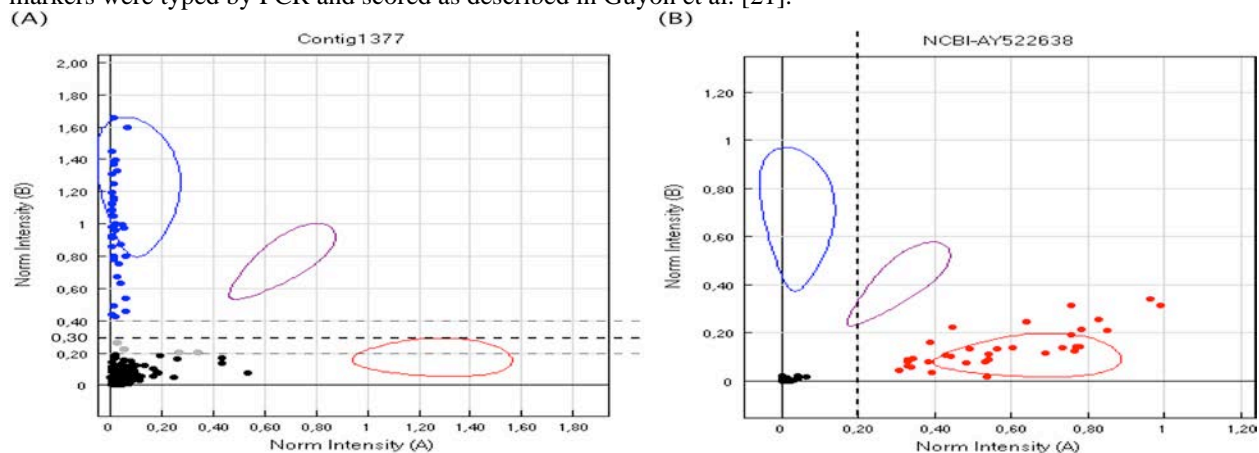


Figure 6 Cartesian plots of radiation hybrid typing by the GoldenGate technology. (A) Dots located above a threshold of 0.30 on the y-axis corresponded to positive clones scored "1", dots located under the threshold corresponded to negative clones scored "0". Dots located close to the threshold were considered as ambiguous results scored "2" (grey dots). (B) According to the overall repartition of dots on the profile of typing the threshold was lowered to 0.20 on the x-axis.

Data computation

All vectors were integrated in a single file and a two-point analysis was performed using the Multimap v2.0 software [36] starting at a lod score of 4.0. The multipoint analysis was performed with the CarthaGène v1.0 software [53]. RH groups that harboured obvious aberrations were re-analysed at higher two-point lod scores of up to 7.0 before performing the multipoint analysis again. Distances between markers were expressed in centirays (cR_{3500}). Coordinates of the putative orthologous genes in the four model genomes were aligned with the corresponding Nile tilapia markers on the graphic representation. Ordered RH groups were tentatively oriented according to two-point lod scores between their end-markers. Similarly, comparison of the lod values obtained between the markers at the extremities of two RH groups, which are for other reasons supposed to be close to one another on the genome allowed to confirm or rule out this hypothesis. The microsatellite order on the RH map was compared with that of the linkage map [15]. Conserved Segments (CS) and Conserved Segments Ordered (CSO) [35,36] between Nile tilapia and the model species were identified using the AutoGRAPH web server [54].

CMap construction

A comparative map viewer was constructed using the web-based tool CMap [55] in order to visualize and compare the RH map with the genetic map of *O. niloticus* [14,15]. Tab-delimited map and correspondence files were created between the two maps based on marker names and loaded into the CMap database. In addition, comparative maps were created between each of the RH map, the *O. niloticus* genetic map, as well as maps of two haplochromine cichlid lineages: *Astatotilapia burtoni* [56] and *Metriaclima zebra/Labeotropheus fuelleborni* [57]. The comparative maps can be viewed through CMap at <http://cichlid.umd.edu/cgi-bin/cmap/viewer>.

Fluorescent In Situ Hybridization (FISH)

In order to validate the anchoring of the RH groups to a particular linkage group and to orient correctly the groups on the corresponding chromosome, physical mapping was performed by fluorescent in situ hybridization (FISH) using as probes at least two BACs per linkage group/chromosome.

Chromosome preparations

Chromosomes were prepared by direct in vivo methods from spleen and head-kidney cell suspensions as described in Fischer et al. [58], with a hypotonic treatment performed in a 28°C water bath for only 20 min. All chromosomes preparations were made from the XX genotype of the sequenced strain.

BAC clones preparation and purification

BAC DNA were purified from two Nile tilapia BAC libraries, T3 library (mean insert size 145 kb) and T4 library (mean insert size 194 kb) [59]. Individual clones were cultured in 100 ml 2YT broth with 12.5 µg/ml chloramphenicol at 37°C for 24 hours. BAC DNA was isolated using the plasmid midi kit (Qiagen) following the manufacturer's protocol, obtaining between 20 to 50 µg yields. The BAC DNA was then validated and its quality verified before FISH by PCR using specific primers.

Fluorescence in situ hybridization

BAC probes were prepared from 2 µg of BAC preparation fragmented by heating at 98°C for 30 min and subsequently labelled with either DIG- or biotin using a High Prime DNA labelling kit (Roche Applied Science), according to the supplier's protocol. To facilitate double BAC FISH experiments, we first prepared stock solutions for each component which were stored separately at -20°C. The BAC probe pellet was resuspended in the hybridization buffer (50% formamide, 2xSSC, 10% dextran sulphate and 50 mM of sodium phosphate) at a concentration of 16 ng/µl and incubated overnight at 37°C under constant agitation. The competitor consisted of sonicated *O. niloticus* DNA and the carrier was sonicated bovine DNA. Both were resuspended in hybridization buffer at 8 µg/µl and 10 µg/µl, respectively.

For the FISH, 2.5 µl of BAC probe, 1 µl of competitor and 4 µl of carrier were preheated at 45°C in a water bath before mixing. The BAC probe mixture was then denatured for 5 min at 85°C and pre-hybridized in a water bath at 37°C for 90 min to eliminate non-specific signals generated by small abundant repetitive sequences (essentially microsatellites) present in BAC inserts or generated by the BAC vector. For the double FISH, both BAC probes were pooled, mixed just *prior* to the hybridization. Chromosomes on slide preparations were denatured for 10 seconds in 70 % formamide/2x SSC at 72°C, followed by a dehydration in 70 %, 80 %, 98 % ethanol bath series. After quickly air drying the slides, the reannealed probe mixture was loaded onto the slides, covered with 22 x 22 mm plastic coverslips, and hybridized at 37°C in a moist chamber during 48 h. For the post-hybridization, the coverslip was removed and the slides were washed in 0,4x SSC, 0,3% Tween 20 (v/v) at 60°C for 2 min and 2x SSC, 0.1% Tween 20 (v/v) at room temperature for 1 min. The hybridized probes were detected with 30 µl of a dual colour solution of anti-dig Rhodamin/streptavidin-FITC (Roche Diagnostics) placed under a 24 x 40 mm coverslip, during 5 min in the dark. Slides were then washed three times in 4x SSC, 0.1% Tween 20, 2 min each followed by dehydration in a series of increasing ethanol percentages.

For the FISH observations, the slides were mounted in DAPI/antifade and analysed with a fluorescent microscope Zeiss Axio imager M1 equipped with a CoolSNAP camera (Photometrics) and the animal karyotyping/FISH imaging software *Genus* (Genetix).

Go to:

Competing interest

The authors declare no competing interest

Go to:

Authors' contributions

FG, JFB, COC, HDC and TDK conceived the experiments. DJP selected the clonal line of *O. niloticus*. RG selected the gene markers, constructed the RH map. MR and NA made the RH panel and genotyped the gene markers. NEMVB and RPMC identified and genotyped the SNP markers. LS and MC compared the genetic and RH maps. HDC, CB and EP selected and prepared the BACs. JPC, CB, ADH, MRG and COC made the FISH experiments. FG and RG wrote the manuscript. All authors read and edited the manuscript.

Go to:

Acknowledgments

This work was supported by grant to COC, JFB and CH (ANR-07-GANI-005) from ANR (Agence Nationale de la Recherche, France).

Go to:

References

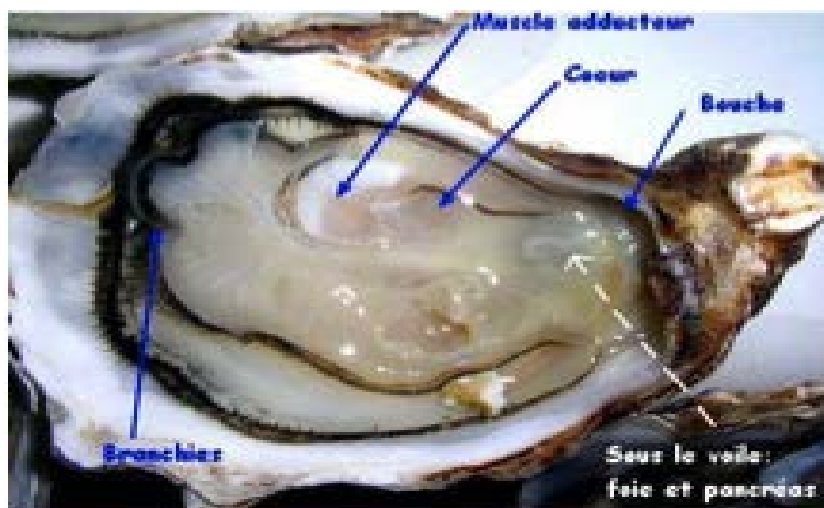
1. FAO. The State of World Fisheries and Aquaculture. 2010. <http://www.fao.org/docrep/013/i1820e/i1820e00.htm>.
2. Randall DJ, Wood CM, Perry SF, Bergman H, Maloiy GM, Mommsen TP, Wright PA. Urea excretion as a strategy for survival in a fish living in a very alkaline environment. *Nature*. 1989;337(6203):165–166. [PubMed]
3. Tine M, de Lorgeril J, D'Cotta H, Pepey E, Bonhomme F, Baroiller JF, Durand JD. Transcriptional responses of the black-chinned tilapia *Sarotherodon melanotheron* to salinity extremes. *Marine Genomics*. 2008;1:37–46. [PubMed]
4. Baroiller JF, D'Cotta H, Bezault E, Wessels S, Hoerstgen-Schwark G. Tilapia sex determination: Where temperature and genetics meet. *Comparative biochemistry and physiology*. 2009;153(1):30–38. [PubMed]
5. Trewavas E. Tilapiine fishes of the genera *Sarotherodon*, *Oreochromis* and *Danakilia*, British Museum Natural History, London, UK; 1983.
6. Cossins AR, Crawford DL. Fish as models for environmental genomics. *Nature reviews*. 2005;6(4):324–333. [PubMed]
7. Kocher TD. Adaptive evolution and explosive speciation: the cichlid fish model. *Nature reviews*. 2004;5(4):288–298. [PubMed]
8. Lee BY, Howe AE, Conte MA, D'Cotta H, Pepey E, Baroiller JF, di Palma F, Carleton KL, Kocher TD. An EST resource for tilapia based on 17 normalized libraries and assembly of 116,899 sequence tags. *BMC genomics*. 2010;11:278. [PMC free article] [PubMed]
9. Soler L, Conte MA, Katagiri T, Howe AE, Lee BY, Amemiya C, Stuart A, Dossat C, Poulain J, Johnson J, et al. Comparative physical maps derived from BAC end sequences of tilapia (*Oreochromis niloticus*) *BMC genomics*. 2010;11:636. [PMC free article] [PubMed]
10. Rezk MA, Ponzoni R, Khaw HL, Kamel E, Dawood T, John G. Selective breeding for increased body weight in a synthetic breed of Egyptian Nile tilapia, *Oreochromis niloticus*: Response to selection and genetic parameters. *Aquaculture*. 2009;293(3–4):187–194.
11. Wessels S, Hörstgen-Schwark G. Selection experiments to increase the proportion of males in Nile tilapia (*Oreochromis niloticus*) by means of temperature treatment. *Aquaculture*. 2007;272:S80–S87.
12. Korol A, Frenkel Z, Cohen L, Lipkin E, Soller M. Fractioned DNA pooling: a new cost-effective strategy for fine mapping of quantitative trait loci. *Genetics*. 2007;176(4):2611–2623. [PMC free article] [PubMed]
13. Cnaani A, Zilberman N, Tinman S, Hulata G, Ron M. Genome-scan analysis for quantitative trait loci in an F2 tilapia hybrid. *Mol Genet Genomics*. 2004;272(2):162–172. [PubMed]
14. Kocher TD, Lee WJ, Sobolewska H, Penman D, McAndrew B. A genetic linkage map of a cichlid fish, the tilapia (*Oreochromis niloticus*) *Genetics*. 1998;148(3):1225–1232. [PMC free article] [PubMed]
15. Lee BY, Lee WJ, Streebman JT, Carleton KL, Howe AE, Hulata G, Slettan A, Stern JE, Terai Y, Kocher TD. A second-generation genetic linkage map of tilapia (*Oreochromis* spp.). *Genetics*. 2005;170(1):237–244. [PMC free article] [PubMed]
16. Majumdar KC, McAndrew BJ. Relative DNA content of somatic nuclei and chromosomal studies in three genera: Tilapia, *Sarotherodon* and *Oreochromis* of the tribe Tilapiini. *Genetica*. 1986;68:165–168.
17. Shirak A, Seroussi E, Cnaani A, Howe AE, Domokhovskiy R, Zilberman N, Kocher TD, Hulata G, Ron M. Amh and Dmrta2 genes map to tilapia (*Oreochromis* spp.) linkage group 23 within quantitative trait locus regions for sex determination. *Genetics*. 2006;174(3):1573–1581. [PMC free article] [PubMed]
18. Cnaani A, Lee BY, Ozouf-Costaz C, Bonillo C, Baroiller JF, D'Cotta H, Kocher T. Mapping of sox2 and sox14 in tilapia (*Oreochromis* spp.) *Sex Dev*. 2007;1(3):207–210. [PubMed]
19. Geisler R, Rauch GJ, Baier H, van Bebber F, Bross L, Dekens MP, Finger K, Fricke C, Gates MA, Geiger H, et al. A radiation hybrid map of the zebrafish genome. *Nature genetics*. 1999;23(1):86–89. [PubMed]

20. Hukriede NA, Joly L, Tsang M, Miles J, Tellis P, Epstein JA, Barbazuk WB, Li FN, Paw B, Postlethwait JH. et al. Radiation hybrid mapping of the zebrafish genome. *Proceedings of the National Academy of Sciences of the United States of America*. 1999;96(17):9745–9750. [PMC free article] [PubMed]
21. Guyon R, Senger F, Rakotomanga M, Sadequi N, Volckaert FA, Hitte C, Galibert F. A radiation hybrid map of the European sea bass (*Dicentrarchus labrax*) based on 1581 markers: Synteny analysis with model fish genomes. *Genomics*. 2010;96(4):228–238. [PubMed]
22. Senger F, Priat C, Hitte C, Sarropoulou E, Franch R, Geisler R, Bargelloni L, Power D, Galibert F. The first radiation hybrid map of a perch-like fish: the gilthead seabream (*Sparus aurata* L) *Genomics*. 2006;87(6):793–800. [PubMed]
23. Sarropoulou E, Franch R, Louro B, Power DM, Bargelloni L, Magoulas A, Senger F, Tsalavouta M, Patarnello T, Galibert F. et al. A gene-based radiation hybrid map of the gilthead sea bream *Sparus aurata* refines and exploits conserved synteny with *Tetraodon nigroviridis*. *BMC genomics*. 2007;8:44. [PMC free article] [PubMed]
24. Sarropoulou E, Nousdili D, Magoulas A, Kotoulas G. Linking the genomes of nonmodel teleosts through comparative genomics. *Mar Biotechnol (NY)* 2008;10(3):227–233. [PubMed]
25. Hitte C, Madeoy J, Kirkness EF, Priat C, Lorentzen TD, Senger F, Thomas D, Derrien T, Ramirez C, Scott C. et al. Facilitating genome navigation: survey sequencing and dense radiation-hybrid gene mapping. *Nature reviews*. 2005;6(8):643–648. [PubMed]
26. Lewin HA, Larkin DM, Pontius J, O'Brien SJ. Every genome sequence needs a good map. *Genome research*. 2009;19(11):1925–1928. [PMC free article] [PubMed]
27. Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Dehal P, Christoffels A, Rash S, Hoon S, Smit A. et al. Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science (New York, NY)*. 2002;29(5585):1301–1310. [PubMed]
28. Kwok C, Critcher R, Schmitt K. Construction and characterization of zebrafish whole genome radiation hybrids. *Methods in cell biology*. 1999;60:287–302. [PubMed]
29. Kent WJ. BLAT—the BLAST-like alignment tool. *Genome research*. 2002;12(4):656–664. [PMC free article] [PubMed]
30. Chen W, Mayden R. A phylogenomic perspective on the new era of ichthyology. *BioScience*. 2010;60:421–432.
31. Gregory TR. Animal Genome Size Database. 2011. <http://www.genomesize.com>.
32. Chew JSK, Oliveira C, Wright JM, Dobson MJ. Molecular and cytogenetic analysis of the telomeric (TTAGGG)_n repetitive sequences in the Nile tilapia, *Oreochromis niloticus* (Teleostei: Cichlidae) *Chromosoma*. 2002;111(1):45–52. [PubMed]
33. Mota-Velasco JC, Ferreira IA, Cioffi MB, Ocalewicz K, Campos-Ramos R, Shirak A, Lee BY, Martins C, Penman DJ. Characterisation of the chromosome fusions in *Oreochromis karongae*. *Chromosome Res*. 2010;18(5):575–586. [PubMed]
34. Ferreira IA, Poletto AB, Kocher TD, Mota-Velasco JC, Penman DJ, Martins C. Chromosome evolution in African cichlid fish: contributions from the physical mapping of repeated DNAs. *Cytogenetic and genome research*. 2010;129(4):314–322. [PMC free article] [PubMed]
35. O'Brien SJ, Womack JE, Lyons LA, Moore KJ, Jenkins NA, Copeland NG. Anchored reference loci for comparative genome mapping in mammals. *Nature genetics*. 1993;3(2):103–112. [PubMed]
36. Murphy WJ, Sun S, Chen Z, Yuhki N, Hirschmann D, Menotti-Raymond M, O'Brien SJ. A radiation hybrid map of the cat genome: implications for comparative mapping. *Genome research*. 2000;10(5):691–702. [PMC free article] [PubMed]
37. Volff JN. Genome evolution and biodiversity in teleost fish. *Heredity*. 2005;94(3):280–294. [PubMed]
38. Li R, Fan W, Tian G, Zhu H, He L, Cai J, Huang Q, Cai Q, Li B, Bai Y. et al. The sequence and de novo assembly of the giant panda genome. *Nature*. 2010;463(7279):311–317. [PubMed]
39. Kwintkiewicz J, Cai Z, Stocco C. Follicle-stimulating hormone-induced activation of Gata4 contributes in the up-regulation of Cyp19 expression in rat granulosa cells. *Molecular endocrinology (Baltimore, Md)*. 2007;21(4):933–947. [PubMed]
40. Cnaani A, Lee BY, Zilberman N, Ozouf-Costaz C, Hulata G, Ron M, D'Hont A, Baroiller JF, D'Cotta H, Penman DJ. et al. Genetics of sex determination in tilapiine species. *Sex Dev*. 2008;2(1):43–54. [PubMed]
41. Roberts RB, Ser JR, Kocher TD. Sexual conflict resolved by invasion of a novel sex determiner in Lake Malawi cichlid fishes. *Science (New York, NY)*. 2009;326(5955):998–1001. [PMC free article] [PubMed]
42. Ser JR, Roberts RB, Kocher TD. Multiple interacting loci control sex determination in lake Malawi cichlid fish. *Evolution; international journal of organic evolution*. 2010;64(2):486–501. [PMC free article] [PubMed]
43. Yoshida K, Terai Y, Mizoiri S, Aibara M, Nishihara H, Watanabe M, Kuroiwa A, Hirai H, Hirai Y, Matsuda Y. et al. B chromosomes have a functional effect on female sex determination in lake victoria cichlid fishes. *PLoS genetics*. 2011;7(8):e1002203. [PMC free article] [PubMed]
44. Hauser L, Baird M, Hilborn R, Seeb LW, Seeb JE. An empirical comparison of SNPs and microsatellites for parentage and kinship assignment in a wild sockeye salmon (*Oncorhynchus nerka*) population. *Molecular ecology resources*. 2011;11(Suppl 1):150–161. [PubMed]
45. Sarder MR, Penman DJ, Myers JM, McAndrew BJ. Production and propagation of fully inbred clonal lines in the Nile tilapia (*Oreochromis niloticus* L.) *The Journal of experimental zoology*. 1999;284(6):675–685. [PubMed]
46. Vignaux F, Hitte C, Priat C, Chauat JC, Andre C, Galibert F. Construction and optimization of a dog whole-genome radiation hybrid panel. *Mamm Genome*. 1999;10(9):888–894. [PubMed]
47. Smit AFA, Hubley R, Green P. RepeatMasker Open-3.0. 1996–2010. <http://www.repeatmasker.org>.
48. Huang X, Madan A. CAP3: A DNA sequence assembly program. *Genome research*. 1999;9(9):868–877. [PMC free article] [PubMed]
49. Slater GS, Birney E. Automated generation of heuristics for biological sequence comparison. *BMC bioinformatics*. 2005;6:31. [PMC free article] [PubMed]
50. Warren RL, Sutton GG, Jones SJ, Holt RA. Assembling millions of short DNA sequences using SSAKE. *Bioinformatics (Oxford, England)* 2007;23(4):500–501. [PubMed]
51. Li H, Ruan J, Durbin R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome research*. 2008;18(11):1851–1858. [PMC free article] [PubMed]
52. McKay SD, Schnabel RD, Murdoch BM, Aerts J, Gill CA, Gao C, Li C, Matukumalli LK, Stothard P, Wang Z. et al. Construction of bovine whole-genome radiation hybrid and linkage maps using high-throughput genotyping. *Animal genetics*. 2007;38(2):120–125. [PMC free article] [PubMed]
53. Schiex T, Gaspin C. CARTHAGENE: constructing and joining maximum likelihood genetic maps. *Proceedings/International Conference on Intelligent Systems for Molecular Biology; ISMB*. 1997;5:258–267. [PubMed]
54. Derrien T, Andre C, Galibert F, Hitte C: AutoGRAPH. an interactive web server for automating and visualizing comparative genome maps. *Bioinformatics (Oxford, England)* 2007;23(4):498–499. [PubMed]
55. Youens-Clark K, Faga B, Yap IV, Stein L, Ware D. CMap 1.01: a comparative mapping application for the Internet. *Bioinformatics (Oxford, England)* 2009;25(22):3040–304. [PMC free article] [PubMed]
56. Sanetra M, Henning F, Fukamachi S, Meyer A. A microsatellite-based genetic linkage map of the cichlid fish, *Astatotilapia burtoni* (Teleostei): a comparison of genomic architectures among rapidly speciating cichlids. *Genetics*. 2009;182(1):387–397. [PMC free article] [PubMed]

57. Albertson RC, Streelman JT, Kocher TD. Directional selection has shaped the oral jaws of Lake Malawi cichlid fishes. *Proceedings of the National Academy of Sciences of the United States of America*. 2003;100(9):5252–5257. [PMC free article] [PubMed]
58. Fischer C, Ozouf-Costaz C, Roest Crollius H, Dasilva C, Jaillon O, Bouneau L, Bonillo C, Weissenbach J, Bernot A. Karyotype and chromosome location of characteristic tandem repeats in the pufferfish *Tetraodon nigroviridis*. *Cytogenetics and cell genetics* 2000, 8. 2000;88(1–2):50–55. [PubMed]
59. Katagiri T, Kidd C, Tomasino E, Davis JT, Wishon C, Stern JE, Carleton KL, Howe AE, Kocher TD. A BAC-based physical map of the Nile tilapia genome. *BMC genomics*. 2005;6:89. [PMC free article] [PubMed]

C- Construction de panels d'hybrides d'Irradiation d'huître (*Crassostrea gigas*) et d'esturgeon (*Acipenser baerii*)

C-1 Huître (*Crassostrea gigas*)



L'huître creuse *Crassostrea gigas* ou huître japonaise est un mollusque bivalve appartenant au group des Lophotrochozoaires. Elle est originaire du pacifique nord-ouest et a été introduite dans de nombreuses régions du monde à des fins commerciales.

Aujourd'hui c'est l'espèce la plus cultivée au monde avec une production annuelle d'environ 4,5 millions de tonnes pour une valeur marchande d'environ 3 milliards d'euros.

Connue pour sa haute fécondité et le haut degré de polymorphisme de séquence de son ADN, elle se situe à une position phylogénétique clé, représentant des organismes pour lesquels relativement peu de données moléculaires sont actuellement disponibles. C'est une espèce sentinelle de choix pour comprendre l'évolution des écosystèmes côtiers et estuariens.

Pour toutes ces données, l'huître creuse constitue un modèle animal intéressant pour la biologie et la génétique des populations.

Malheureusement l'ostréiculture connaît des épisodes récurrents de mortalité importante dite estivale (température optimale...) ou liée à des infections (herpès....) qui peuvent atteindre jusqu'à 60% de mortalité la première année. Ceci renforce l'intérêt des études génétiques de cette espèce afin de mieux comprendre ce phénomène en repérant les traits de sensibilité et de résistance à la mortalité et envisager une amélioration de la reproduction. Il semble en effet exister un lien entre les mortalités estivales et le processus de reproduction. C'est dans ce contexte qu'un programme ANR (projet Gametogenes 08-GENM-041-03) regroupant plusieurs équipes universitaires dont la

nôtre et de l'Ifremer a été développé. Notre responsabilité dans ce programme était de construire une carte RH du génome de l'huitre.

Pour réaliser le panel de cellules hybrides nous nous sommes heurtés à plusieurs difficultés méthodologiques. La première difficulté a été comme toujours dans le choix d'une source de cellules donneuses.

Notre première source de cellules donneuses a été les ovocytes et les spermatozoïdes, mais après plusieurs tentatives de fusions qui n'ont pas donné de clone, on a pensé qu'il pouvait s'agir d'un problème de taille de cellules et plus particulièrement de différence de taille entre les ovocytes beaucoup plus gros que les cellules de hamster receveuses ou entre les spermatozoïdes beaucoup plus petits que ceux-ci.

Nous nous sommes orientés alors vers d'autres sources tissulaires et avons tenté des prélèvements de cellules cardiaques. Les rendements étant beaucoup trop faible nous nous sommes intéressés au muscle. Ce dernier tout comme la rate de poissons nous a fourni par simple écrasement dans un broyeur « douce » des cellules en grand nombre, de taille adéquate. Malheureusement les premières fusions réalisées avec des cellules CHO TK- ou HPRT- n'ont strictement rien donné. Nous avons alors pensé qu'il pouvait s'agir d'un tout autre problème, d'un problème d'incompatibilité environnementale entre les cellules d'huitres et les cellules de hamster. De fait nous avons constaté que des cellules d'huitres seules mises dans les conditions de milieu et de température propres à la fusion mourraient instantanément. A ce stade nous avons émis l'hypothèse qu'il s'agissait d'un problème d'isotonicité. En effet les huitres sont des organismes dont le système circulatoire est ouvert ou lacunaire. Ils sont osmoconformes, c'est à dire qu'ils ne régulent que de manière très modeste la pression osmotique de leur milieu intérieur qui reflète ainsi celle du milieu environnant. Ainsi, des cellules d'huitre préparées comme précédemment mais mises dans du milieu MEM enrichi en sel pour égaler la pression osmotique de l'eau de mer survivent parfaitement. Inversement des cellules de hamster dans un milieu où les cellules d'huitres sont parfaitement viables meurent instantanément.

Un changement des conditions de fusion a donc été recherché, en ajoutant 50% d'eau de mer et 50% DMEM, à différents moments et sous différentes conditions pendant ou après la fusion, en présence ou non des cellules de hamster HPRT- ou TK-, mais sans le moindre résultat. Une acclimatation des huitres à des conditions de salinité moindre a alors été tentée, en se fondant sur l'observation que l'affinage de certaines huitres se fait dans des claires qui sont des bassins alimentés avec une partie d'eau douce et que par ailleurs les huitres aiment bien coloniser les estuaires de rivières où la concentration saline est moindre.

Des huitres achetées sur le marché ont été immergées dans des bassines remplies d'eau de mer. Au bout d'une semaine d'acclimatation, de l'eau douce correspondant à 10% du volume initial a été ajoutée. Le processus d'ajout de 10 % d'eau douce a été poursuivi toutes les semaines et ce pendant deux mois. A l'issue de cette période la concentration saline dans la bassine était équivalente à celle du DMEM. Alors que quelques huitres étaient mortes, la très grande majorité d'entre elles étaient vivantes.



Images des huitres acclimatées au laboratoire

Des tests de viabilité réalisés avec des cellules musculaires préparées avec des huitres acclimatées ont montré la bonne tolérance de celles-ci au DMEM selon le tableau suivant :

	Huître normale	Huître acclimatée
Culots cellulaires + DMEM	0% cellules vivantes	65% cellules vivantes

De nouveaux essais de fusion ont été alors réalisés. Des cellules musculaires issues d'huitres acclimatées ont été préparées dans de l'eau d'acclimatation stérile puis irradiées à 3500 rads et fusionnées avec des cellules de hamster réceptrices HPRT^r ou TK^r. On a alors pu observer des cellules hybrides résistantes à la mise en culture dans du milieu de sélection HAT. Malheureusement ces cellules hybrides ne se multipliaient pas. On en a déduit que ni le gène TK, ni le gène HPRT de l'huitre ne s'exprimait ou bien que ni la protéine TK ni la protéine HPRT d'huitre n'était suffisamment fonctionnelle dans l'environnement cellulaire des cellules de hamster et donc qu'un autre système de sélection devait être recherché.

Deux solutions ont été envisagées :

- une transfection des cellules d'huitres préalable à leur fusion avec un gène de résistance de rongeur.
- une levée de la pression de sélection HAT.

1-Transfection avec un gène de résistance TK

Des cellules acclimatées ont été transfectées avec un plasmide susceptible d'exprimer un gène TK humain. L'efficacité de plusieurs transfectants a été testée, mais seule la Lipofectamine 2000 nous a permis d'obtenir des clones, hélas en petit nombre, 46. L'ADN de ces 46 clones a été extrait, amplifié par Genomiphi puis sur chacun d'eux nous avons recherché par PCR la présence de 32 marqueurs d'huître *C. gigas*. L'analyse sur gel d'agarose nous a montré que seuls 9 clones étaient porteurs de quelques marqueurs avec par ailleurs des profils d'interprétation difficile. Cette approche qui aurait nécessité un long travail pour la rendre opérationnelle a été abandonnée, du moins provisoirement.

2- Levée de la pression de sélection HAT

Puisque ni le gène TK, ni le gène HPRT n'était en mesure d'assurer la croissance des cellules hybrides dans du milieu sélectif HAT, nous avons parallèlement aux essais de transfection imaginé d'alterner les périodes sous sélection HAT pour permettre la disparition des cellules de hamster non fusionnées tout en préservant en vie les hybrides cellulaires, avec des périodes d'incubation dans du milieu normal DMEM pour permettre aux cellules de croître. Ainsi après 15 jours de sélection sous milieu HAT, on a levé la pression de sélection en remplaçant le milieu HAT par du DMEM dans lequel les cellules hybrides devaient se multiplier. On a pu ou cru alors observer une légère croissance cellulaire. L'ensemble des cellules de chaque boîte a été récolté, l'ADN extrait puis amplifié par Genomiphi. Sur chacun des ADN récoltés, une détermination par PCR de la présence/absence de 32 marqueurs microsatellites de *C. gigas* a été faite.

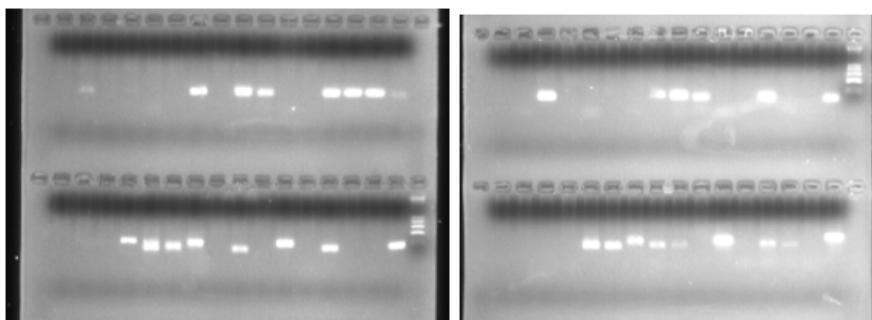


Figure 19 : Exemples de profils représentatifs obtenus avec deux clones.

Ces deux profils sont à la fois différents et montrent une bonne rétention de marqueurs indiquant une situation favorable pour la construction d'une carte RH.

Comme pour les autres panels nous avons alors sélectionné sur la base de leur rétention quantitative et qualitative des marqueurs d'huître, 116 ADN extraits à partir des clones obtenus. Ces extraits ont constitué le panel nécessaire à la cartographie du génome d'huître.

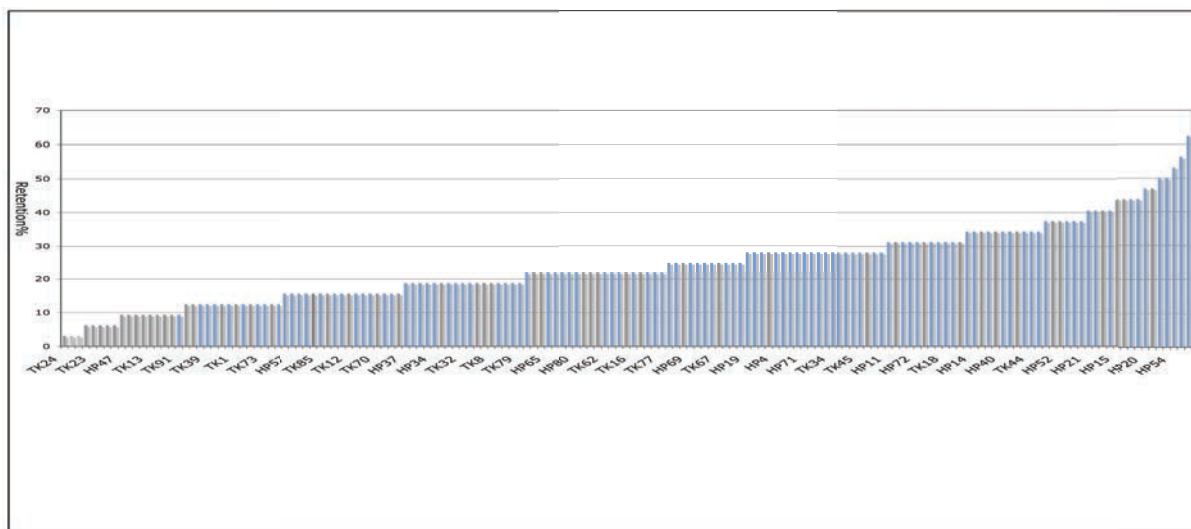
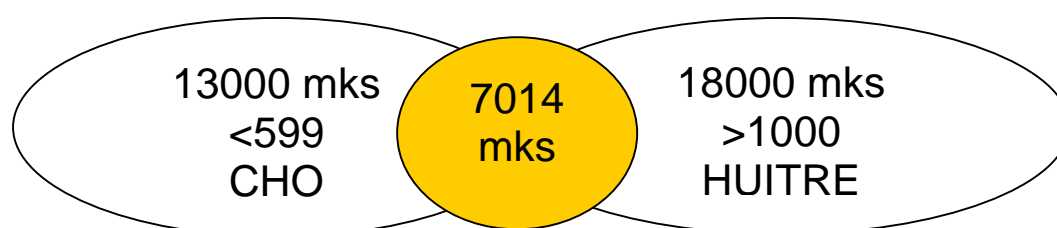


Figure 20 : Graphe indiquant le % de rétention des cellules hybrides huitre/hamster
 160 clones sur les 168 récoltés ont montré la présence de marqueurs et 116 ADN (en bleu sur le graphe) ont été sélectionnés sur des critères quantitatifs et qualitatifs pour être analysés plus avant par hybridation sur des puces à oligonucléotides. La rétention moyenne de ce panel est de 24 % avec les valeurs extrêmes de 3,1% et 62,5 %.

3- Génotypage et Cartographie

Afin de positionner un très grand nombre de marqueurs nous avons décidé de mettre à profit les capacités d'hybridation hautement parallèle des puces à oligonucléotides et nous nous sommes adressés à la Société Agilent pour l'obtention de puces sur lesquelles étaient synthétisés 38014 oligonucléotides de 60 nucléotides, dessinés à partir de séquences EST d'huitre. Dans une première série d'expériences nous avons hybridé sur ces puces de l'ADN de hamster d'une part et de l'ADN d'huitre d'autre part. Par le croisement de ces deux résultats nous avons sélectionné 7014 oligonucléotides donnant des valeurs d'hybridation faibles avec l'ADN de hamster et fortes avec l'ADN d'huitre. Nous avons alors demandé à la société AGILENT de nous fournir des puces de 60 mers sur les lesquelles ces 7014 oligonucléotides étaient déposés 8 fois chacun. Avec ces nouvelles puces nous avons réalisé des hybridations pour 100 lignées dont un certain nombre faites en duplica ou triplicata.



Les 7014 oligonucléotides, sur les 38014, susceptibles de donner un signal d'hybridation interprétable ont été déposés sur les puces.

3-1-Analyse des marqueurs et cartographie

Le logiciel « Agilent Scan control » a été utilisé pour scanner les puces après hybridation et le logiciel « Feature Extraction » pour extraire les données. A partir du « rmedian signal » (non processed signal) nous avons calculé les moyennes des 8 répliquas de chaque puce puis normalisé ces valeurs en les comparant à la médiane de chaque puce. A partir de ces valeurs normalisées nous avons défini les différents vecteurs caractéristiques de la rétention par chaque lignée des 7014 marqueurs sélectionnés. Pour chaque lignée, on a construit un graphe avec en ordonnée y la valeur du rapport hybridation de lignée/ huître et en abscisse x le nom des marqueurs. Les deux graphes suivants montrent la valeur de ce rapport pour deux lignées hybrides, permettant d'identifier pour chacune d'entre elles les différents vecteurs.

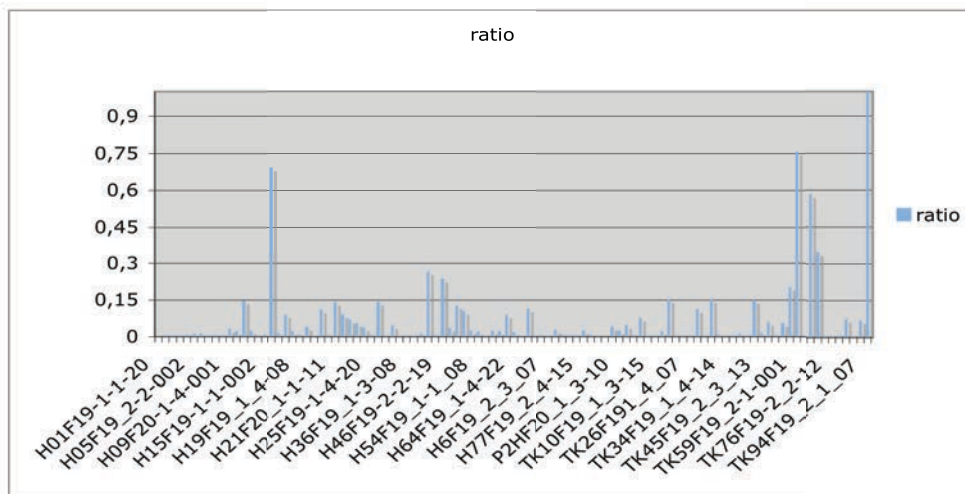


Figure 21 : Graphe du ratio d'hybridation des cellules hybrides pour le marqueur AM861117

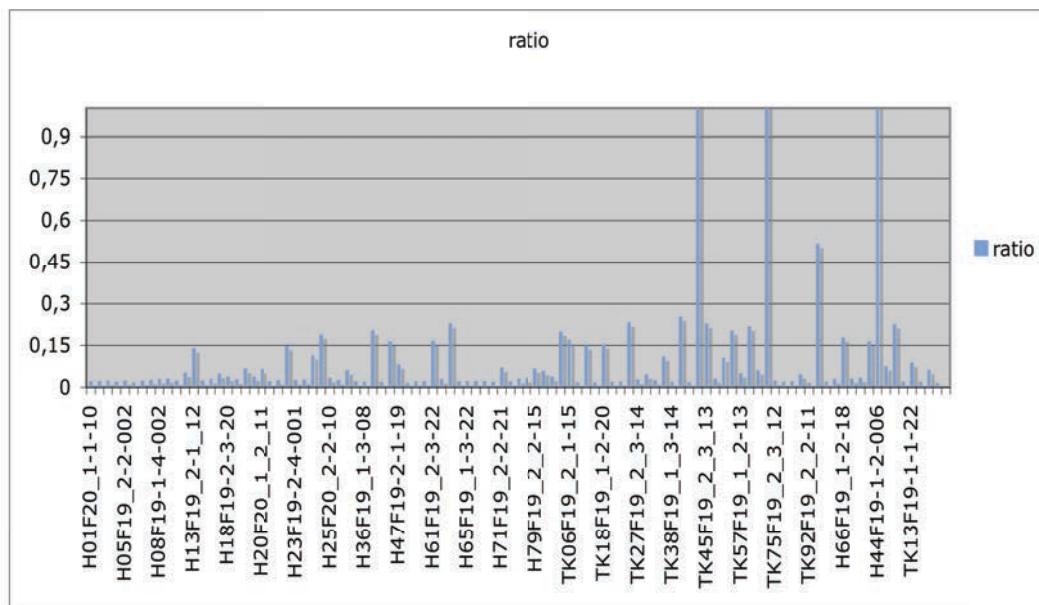


Figure 22 : *Grappe du ratio d'hybridation des cellules hybrides pour le marqueur AM85176*

A notre grande surprise, l'analyse en deux points de ces vecteurs avec le logiciel Multimap ne nous a pas permis d'obtenir des groupes de marqueurs liés et ce en dépit de l'utilisation de LOD relativement faible. Nous avons alors fait l'hypothèse que les fragments retenus dans les lignées étaient très petits, trop petits. Nous avons testé cette hypothèse en génotypant sur le panel RH des paires de marqueurs définis à partir de la séquence de plusieurs BACs et de mRNA de façon à connaître la distance réelle séparant les deux marqueurs de chaque paire. Les résultats de ces typages par PCR nous ont confirmé que les fragments retenus dans les cellules hybrides étaient très petits de l'ordre du millier à quelques milliers de nucléotides.

Nous avons tout d'abord rejeté l'hypothèse que le DNA d'huitre serait très peu résistant au rayonnement et aurait été pulvérisé par l'irradiation à 3500 rads et avons plutôt retenu l'hypothèse pour expliquer ce résultat que nous n'avons jamais observé au préalable que les conditions particulières de culture suivies pour l'obtention du panel en étaient la cause. Nous avons alors produit un autre panel, cultivé dans des conditions différentes susceptibles selon notre hypothèse de travail de ne pas affecter (autant) la taille des fragments.

Pour ce nouveau panel l'alternance des périodes sous HAT et DMEM a été modifiée selon le schéma : 48h milieu normal // 72 h HAT // 48h milieu normal // 72 h HAT // 48h milieu normal // 72 h HAT // et ce pendant deux semaines, ensuite on a récolté les cellules présentes dans chacun des puits. Une extraction ADN et une amplification Genomphi ont été faites suivies d'une analyse de marqueurs d'extrémités de BAC. Malheureusement l'analyse de la distribution dans ce nouveau panel des extrémités de BAC et plasmides a montré que le problème n'avait pas été résolu. Nous avons alors testé la capacité de lier de ce panel sur des marqueurs proches distants les

uns des autres de 500 Pb déterminés à partir d'une séquence de 20 Kb d'un BAC. Les résultats de cartographie de ces marqueurs ont montré que des marqueurs suffisamment proches se liaient et que ce panel s'il n'était pas utilisable pour la construction d'une carte de tout le génome pouvait néanmoins être utile pour l'analyse de la structure de régions d'intérêt.

C-2 Esturgeon (*Acipenser baerii*)

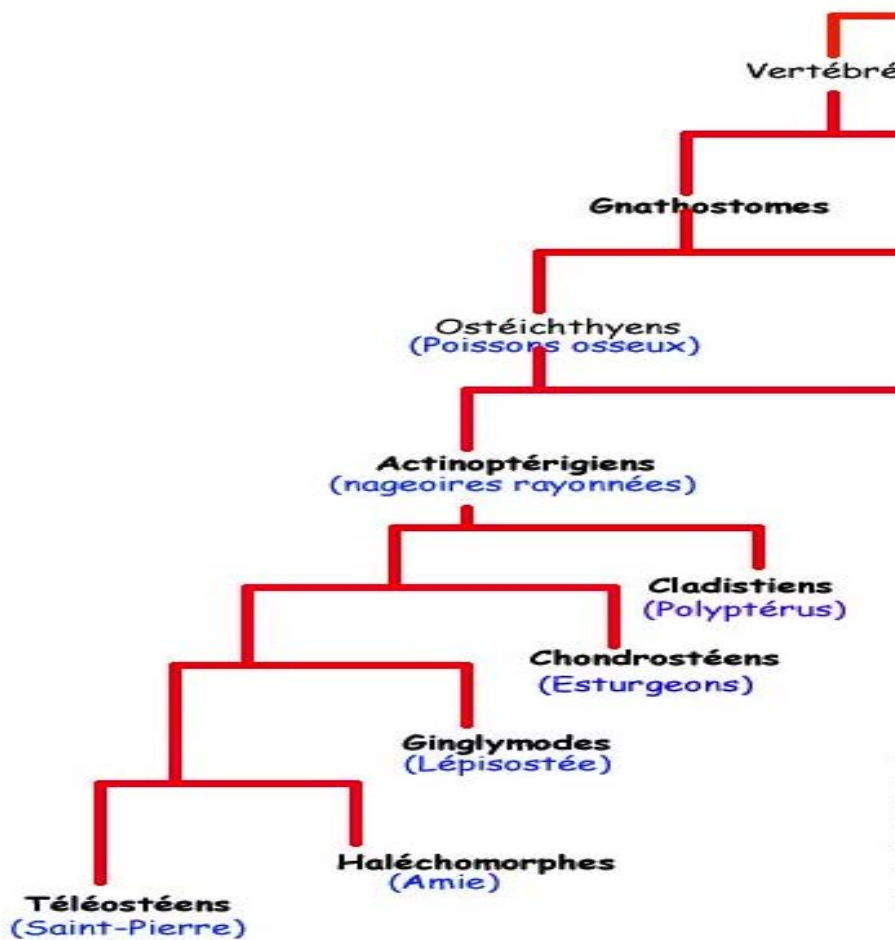


Figure 23 : Arbre phylogénique montrant la place des esturgeons (chondrostéens) et celle des téléostéens. (d'après G. Lecointre & H. Le Guyader, modifié)

Les *Acipenseridae* forment une famille de poissons peuplant exclusivement l'hémisphère nord et appelés communément esturgeons lesquels renvoient une image de grand luxe à cause du caviar qu'ils produisent, ce qui a conduit à leur surpêche. Parmi les 24 espèces qui composent cette famille, on peut distinguer l'*huso huso*, le plus gros esturgeon, appelé également grand esturgeon ou esturgeon européen, à partir duquel on commercialise le caviar beluga, les esturgeons *Acipenser gueldenstaedtii*, *Acipenser persicus* qui produisent le caviar osciètre) ou encore *Acipenser stellatus*, le plus petit des esturgeons qui produit le sévruga. On trouve ce dernier en mer Noire, mer d'Azov et mer Caspienne.

L'esturgeon de l'Atlantique *acipenser studio* est le plus grand poisson migrateur anadrome de France. On le trouve en France dans l'estuaire de la Garonne et dans quelques autres lieux. L'esturgeon européen était abondant le long des côtes européennes ainsi que dans la majorité des grands fleuves. Il a disparu de la plupart des eaux européennes à la fin du 19^e siècle. Aujourd'hui, seule une population demeure, en France, dans la Gironde, avec une présence dans les eaux côtières depuis le golfe de Gascogne jusqu'à la mer du Nord. Cependant, en dépit de son statut d'espèce protégée en France depuis 1982 et en Europe depuis 1998, et malgré des efforts significatifs entrepris en France depuis 26 ans, sa population continue à décliner. Elle a connu une forte régression par surexploitation, par surpêche et peut-être aussi à cause de la pollution qui l'a conduit au bord de l'extinction. En France, elle fait depuis 2007 l'objet d'un plan de restauration avec réintroduction d'alevins.

Ce poisson longiligne a un museau en forme de groin, doté de barbillons sensitifs, se terminant par une longue pointe. Les nageoires dorsales sont situées très en arrière du corps, qui comporte cinq rangées longitudinales de grosses plaques osseuses. Le ventre est jaunâtre et le dos gris cendré à gris brun. Les esturgeons se nourrissent d'animaux benthiques et principalement d'invertébrés. En eau douce, ils s'alimentent de vers, de larves d'insectes et de crustacés ; en eau saumâtre, de vers et de crustacés ; en eau de mer, essentiellement de vers. Les poissons représentent une part notable de proies pour les adultes. L'esturgeon est un amphihalien potamotoque, c'est-à-dire qu'il effectue sa reproduction en eau douce et une partie de sa croissance en lacs et rivières et qu'il fréquente les eaux littorales. La reproduction naturelle a lieu de mai à juillet dans le cours moyen des fleuves sur le fond des graviers. Les adultes qui vivent dans les eaux littorales commencent à remonter le fleuve en avril (comportement migratoire). La femelle, chargée d'œufs, rase le fond. Elle s'aide du flux montant et pond en plusieurs fois, sur un emplacement profond et graveleux, entre 300 mille et 2 millions d'œufs (ce qui représente 10 % de sa masse) ; Les mâles répandent ensuite leur semence à proximité. Vifs et alertes, ils sont capables de sauter par dessus les filets et de se défendre contre les prédateurs, tels les barbots et les anguilles. L'adulte mesure de un à deux mètres de long (au maximum trois mètres et demi) et pèse 300 jusqu'à 500 Kg. Il peut vivre plus

de 80 ans avec une maturité sexuelle tardive, s'exprimant après 10 ans pour les mâles et après 13 ans pour les femelles. Les esturgeons ne possédant pas d'écaillés et la lecture de leurs écussons épais et opaques étant difficile, on détermine couramment leur âge en observant à la loupe binoculaire une coupe mince du premier rayon des nageoires pectorales. Cependant, cette technique qui atrophie la nageoire pectorale peut déséquilibrer le poisson.

L'esturgeon est une espèce d'un grand intérêt économique en raison des qualités gustatives du caviar qui se récolte chez des femelles adultes. La maturité sexuelle tardive et une absence de dimorphisme sexuel chez ce groupe d'espèces compliquent sérieusement le suivi des populations et les tentatives de restauration de ces espèces dans leur milieu naturel (en particulier en Gironde), tout comme le développement de son aquaculture, sans parler de l'intérêt biologique que représenterait une meilleure connaissance du déterminisme et de la différenciation sexuels chez ce groupe d'espèces.

Malheureusement, toutes les tentatives pour trouver des marqueurs génétiques spécifiques du sexe des esturgeons par RAPD, AFLP, ISSR ont été des échecs (Wuertz et al. 2006; McCormick *et al.*, 2008). L'approche gène candidat à partir des gènes Sox (exemple Sox9) n'a pas non plus montré de spécificité de sexe, elle a conduit néanmoins à des observations intéressantes (Hett *et al.*, 2005; Hett & Ludwig, 2005). Par criblage avec de oligonucléotides dégénérés, les auteurs de ces études ont identifié 8 gènes Sox que des analyses phylogéniques ont montrés très proches de ceux de la souris. En revanche ils n'ont observé aucune différence de structure reliée au sexe. Dans le cas des gènes *Sox4*, *Sox17* et *Sox21*, des évidences de duplication génique ont été trouvées.

Récemment, l'analyse des mRNA par pyroséquençage de gonades d'esturgeons a été réalisée aux USA mais sans succès quant à l'identification de marqueurs sexuels (Hale *et al.*, 2009). Une étude comparative protéomique des gonades a également été effectuée, mais là encore sans mettre en évidence de différences entre les sexes (Keyvanshokoo *et al.*, 2009). Actuellement seules quelques méthodes fondées sur un examen individuel par ultrason ou l'usage d'un borescope des voies génitales permettent des l'âge de deux ans et avec une assez bonne fiabilité de déterminer le sexe des individus (Kynard and Kieffer, 2002; Wildhabert *et al.*, 2005). Les analyses de séquence d'ADN et de mRNA d'esturgeon ont montré une homologie de séquence avec d'autres vertébrés en particulier avec les téléostéens. Ceci suggère que des comparaisons avec des génomes téléostéens, tels que daurade, tilapia, médaka, zebrafish, tetraodon et salmonidés grâce à la production d'une carte RH de l'esturgeon seraient une bonne stratégie permettant de définir des régions de synténie dans lesquelles des gènes d'intérêt pourraient être recherchés.

C'est dans ce contexte qu'un projet collaboratif avait été proposé dans lequel une carte RH du génome de *A. baerii* aurait été produite pour aider à l'identification de marqueurs génétiques du sexe. Pour cela, nous avons reçu trois jeunes poissons d'environ 400g chacun de l'Écloserie de Guyenne auxquels nous avons prélevé la rate et du sang. Dans un premier temps nous avons testé la possibilité d'utiliser des splénocytes et des lymphocytes comme cellules donneuses et fait quatre fusions : deux avec des cellules de hamster (CHO) TK⁻ et deux avec des cellules de hamster (CHO) HPRT⁻. Après deux semaines de culture, nous n'avons eu aucun clone pour les fusions faites avec les cellules TK⁻, mais un petit nombre de clones avec celles faites avec le gène HPRT⁻. A la suite de ces résultats, trois autres fusions ont été réalisées entre des splénocytes irradiés à 3500 rads et des cellules CHO HPRT⁻. Au total nous avons obtenu 474 clones indépendants qui ont été repiqués et chacun mis isolément en culture dans des boîtes de Petri de 6 cm de diamètre. A confluence les ADN ont été extraits, titrés puis tester avec 48 marqueurs de gènes uniques et des paires de marqueurs identifiés à partir de 12 séquences de mRNA. A partir de ces résultats nous avons sélectionné sur la base de leurs rétentions quantitative et qualitative 94 clones correspondant à un panel ayant une rétention moyenne de 25% (13% -50%). Le typage sur ce panel de un à plusieurs milliers de marqueurs de gènes devrait permettre la construction d'une carte RH utile pour la recherche de marqueurs de différenciation sexuelle ou pour toute autre étude de génomique.

Malheureusement ce projet n'a pas été soutenu par l'ANR ni aucune autre agence à ce jour et nous avons du l'arrêter.

Publication N°3 :

Construction of Radiation Hybrid panels for two major aquaculture species : sturgeon and Oyster

Azzouzi N^{1,2}, Ennaji MM², Benchekroun MN³, Rakotomanga M¹, Baroilier JF⁴ and Galibert F¹

Journal of Aquaculture Research & Développement 2013, 4

Construction of Radiation Hybrid Panels for Two Major Aquaculture Species: Sturgeon and Oyster

Azzouzi N^{1,2*}, Ennaji MM², Benchekroun MN³, Rakotomanga M¹, Baroillier JF⁴ and Galibert F¹

¹Institut Génétique et Développement (UMR 6290) CNRS/Université de Rennes 1, Faculté de Médecine, 35000 Rennes, France

²Laboratoire de virologie et Hygiène & Microbiologie, Faculté des Sciences et Techniques Mohammedia (FSTM) Université Hassan II, 20650 Mohammedia-Casablanca, Morocco

³Laboratoire de Biotechnologie, de l'Environnement et de la Santé, Faculté des Sciences et Techniques Mohammedia (FSTM) Université Hassan II, 20650

Mohammedia-Casablanca, Morocco

⁴INTREPID (INTensification Raisonnée et Ecologique pour une Plsiculture Durable) (UMR110) Cirad/Ifremer, Campus International Baillarguet, 34398 Montpellier, France

Abstract

Aquaculture will probably have to increase the production of many marine fishes and other products in the coming years to contribute to fulfilling the needs of an increasing population against a background of the impoverishment of natural resources. Genomics tools for improving the performance of marker assisted selection (MAS) would be valuable assets for increasing production. Radiation hybrid mapping is one such tool; however, the construction of fish radiation hybrid (RH) panels has long been very difficult in practice. Here, we present the construction of a sturgeon RH panel and an oyster RH panel. By typing a number of markers corresponding to mRNA and bacterial artificial clone (BAC) ends, we showed that the fish panel performs well and would allow the construction of RH maps of several thousand markers. The DNA fragments retained by the hybrids of the oyster panel were very small. However, by typing 40 markers designed from a BAC sequence, we demonstrated this oyster RH panel could be useful for constructing maps of very high marker density for regions of particular biological interest.

Keywords: Sturgeon; Oyster; Radiation hybrid panel; Marker assisted selection

Abbreviations: RH radiation hybrid; MAS marker assisted selection; HPRT hypoxanthine-guanine phosphoribosyltransferase; TK thymidine kinase; CHO Chinese hamster ovary; HAT hypoxanthine-aminopterin-thymidine medium; WGA whole genome amplification; EST expressed sequence tag

Introduction

Both fish and molluscs are important sources of protein in the human diet [1-3]. Currently, more than half of the fish and molluscs consumed are produced by aquaculture and this percentage may well increase due to environmental constraints and the decline of the natural stock in ocean coastal area. Indeed, according to the Convention on Biological diversity, the extent of coastal fish depletion has nearly doubled every 10 years since 1960 (www.fao.org/docrep/016/i27277/i27277f01.pdf-2012). This alarming reality partially explains the development of international projects aimed at determining the genome sequences of fish species, including salmon, trout and tilapia, and molluscs, like the oyster [4]. The genome sequences may allow the identification of genomic markers useful for marker assisted selection (MAS) [5], genomic selection, and more generally the development of “omics” approaches for a better understanding of fish biology and evolution. Over recent years, we have constructed RH panels and corresponding RH maps for three important aquaculture species: seabream (*Sparus aurata*), seabass (*Dicentrarchus labrax*) and tilapia (*Oreochromis niloticus*) [6-8].

RH fish panels have long been considered to be very difficult if not impossible to produce, and only two zebrafish RH panels were produced; both were constructed starting with a permanent zebrafish cell line as donor cells. This source of cells is not optimal, because the genomes of most permanent cell lines are rearranged and therefore do not accurately reflect the genome of the species [9,10]. Here, we report the construction of two RH panels, one for a fish, the sturgeon (*Acipenser bareii*), and the other for a mollusc, the Pacific oyster (*Crassostrea*

gigas). Both are major aquaculture products with an annual production of 4,500 (http://www.fao.org/fishery/culturedspecies/Acipenser_bareii/fr) and 4.5 million tons, (http://www.fao.org/fishery/culturedspecies/Crassostrea_gigas/fr) respectively.

Materials and Methods

Panel construction

Juvenile sturgeon (*Acipenser bareii*) averaging 200 g were obtained from Sturgeon SA. Radiation hybrid panels were constructed as previously described [8]. Briefly for each fusion, one fish was killed with an overdose of 2-phenoxyethanol and rinsed briefly in 70% ethanol. Spleens were rapidly dissected and rinsed several times in washing medium (Leibovitz L-15; 1% Penicillin/Streptomycin/Fungizone) then cut into small pieces. The pieces were fragmented with a Potter device and the resulting cell suspensions were γ -irradiated at 3500 rad. The cells were fused with HPRT-derivative CHO cells as previously described [6-8]. Briefly HPRT-derivative CHO cells and splenocytes were mixed in a 1:5 ratio (CHO/splenocyte) in the presence of polyethylene glycol 1500 (Roche, Mannheim, Germany). Cells were seeded in 12-well microplates at a total concentration of 150,000 cells per well and cultivated with HAT medium for 2 to 3 weeks until hybrid

***Corresponding author:** Azzouzi N, 1Institut Génétique et Développement (UMR 6290) CNRS/Université de Rennes 1, Faculté de Médecine, 35000 Rennes, France; Tel: +(33) 223234782; Fax: + (33) 223234478; E-mail: galibert@univ-rennes1.fr

Received March 22, 2013; **Accepted** May 24, 2013; **Published** June 04, 2013

Citation: Azzouzi N, Ennaji MM, Benchekroun MN, Rakotomanga M, Baroillier JF, et al. (2013) Construction of Radiation Hybrid Panels for Two Major Aquaculture Species: Sturgeon and Oyster. J Aquac Res Development 4: 185 doi:[10.4172/2155-9546.1000185](https://doi.org/10.4172/2155-9546.1000185)

Copyright: © 2013 Azzouzi N, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

clone appearance. Clones were recovered individually and further cultivated under HAT selection in 60 mm diameter Petri dishes. The cells were trypsinised and DNA was extracted from individual clones using the NucleoSpin Tissue kit (Macherey-Nagel, Düren, Germany). DNA concentrations were quantified by fluorescence using the Quant-iT Picogreen assay kit and a Qubit spectrofluorometer (Invitrogen, Carlsbad NM, USA). DNA from each hybrid cell line was amplified by a Whole Genome Amplification (WGA) procedure as necessary; in these cases, two separate WGA were performed with 10 ng of DNA each using V2 GenomiPhi kits (GE healthcare, Fairfield CT, USA). WGA products were pooled providing ~10 ug of material for subsequent genotyping. We previously showed that WGA is an important step, which limits or over pass cell culture difficulties while providing a faithful representation of the input DNA [7,8].

Commercial oysters were bought from the market and bathed in seawater for a week. Thereafter, a volume of fresh water equivalent to 10% of the bath volume was added every week for a period of 8 weeks. During this period 1 to 3 oysters died every week. At the end of the adaptation period, living oysters were opened and muscles recovered. Muscles from 6 to 12 oysters were pooled, suspended in 50 ml of sterile diluted seawater and gently dissociated with a Potter device. The samples were centrifuged for 5 min at 1200 rpm and the pellets re-suspended and washed twice in sterile diluted seawater (25% seawater/75% water). Finally, the cell pellet was γ -irradiated at 3500 rads. The irradiated oyster cells were suspended in diluted seawater and counted. CHO derivative cells were added to a ratio of 8/1 (oyster/CHO cells) and allowed to fuse in the presence of polyethylene glycol 1500 (Roche, Mannheim, Germany). The cells were seeded in 12-well microplates and cultivated in HAT and DMEM media, alternately. Two weeks later, the contents of each well were recovered, and the DNA extracted and directly amplified and quantified [8].

Genotyping markers

Sturgeon and oyster expressed sequences (ESTs) and mRNA sequences, and oyster BAC sequences were downloaded from NCBI (<http://www.ncbi.nlm.nih.gov>). Oligonucleotide primers of 23 nt average length for polymerase amplification (PCR) were designed with Primer3 v0.4.0 software [11]. (<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/>). Markers, 60 to 150 bp long (including primers) were amplified and electrophoresed on 2% agarose slab gels. Markers were usually genotyped in duplicate with no visible significant difference.

Results

Panel construction

The first step in RH panel construction is the fusion of donor and acceptor cells. To avoid the use of permanent cell lines, we prepared sturgeon donor cells from the spleen organs. Spleen cells present numerous advantages. The spleen is a sort of bag full of splenocytes that can be liberated by very gentle mechanical disruption in L15 Leibovitz medium with a Dounce or a Potter device; thus several hundred millions of splenocytes can be readily obtained and can be directly γ irradiated without the need for cell culture before fusion. The construction of the oyster panel presented two specific difficulties not encountered with the fish panel. Oysters have no spleen organ and, being an open organism, oyster cells are bathed and live in a medium with a naturally high osmotic pressure (NaCl ~35 g/litre) incompatible with the osmotic pressure suitable for CHO receptor cells (NaCl ~9 g/litre). To overcome these two difficulties, we tested several oyster

organs for their suitability as donor cells. We found that the muscle was appropriate tissue, because, like the fish spleen, it can be gently dissociated in sterile sea water with a Potter device, and a pool of six to nine oyster muscles can provide more than 10×10^7 cells, sufficient for fusion protocols. Although oysters live mostly in conditions of high osmotic pressure, oysters in their natural environments tolerate a range of different osmotic pressures, for example in the oyster beds where they are immersed during the finishing period, or near estuaries where they like to settle. We therefore tested whether they could adapt to lower osmotic pressure: we immersed oysters in a basin filled with seawater and every week added a volume of tap water equal to one tenth of the total volume. During this adaptation period, some oysters died but most remained viable. After seven weeks of adaptation, the osmotic pressure of the water bath was close to NaCl ~9 g/litre. More than half (65%) of the muscle cells prepared from adapted oysters tolerated the temperature and the osmotic pressure of the medium in which they were mixed with CHO receptor cells before fusion. Conversely cells prepared from non-adapted oysters did not survive when subjected to

	Gene used for selection	# fusions	#a hybrid clones	Panel size
Sturgeon	HPRT	3	474	94
Oyster	HPRT	2	77	48
Oyster	TK	2	91	46

Table 1: Number of hybrid cell lines obtained with the different fusion and RH panel characteristics.

The number of fusions indicated in this table includes only the productive fusions. There were several attempted fusion before appropriate experimental conditions were found. As it was not known whether the oyster HPRT and TK proteins complement the hamster mutant, we used the two systems in parallel. The working oyster panel was then constituted of 48 HPRT and 46 TK hybrid cell lines. This strategy of using both TK- and HPRT-rescued clones may represent an advantage over using one gene only: markers located close to the TK or HPRT receptor gene cannot be mapped because they are present in all or nearly all the hybrid cells rescued with the corresponding gene.

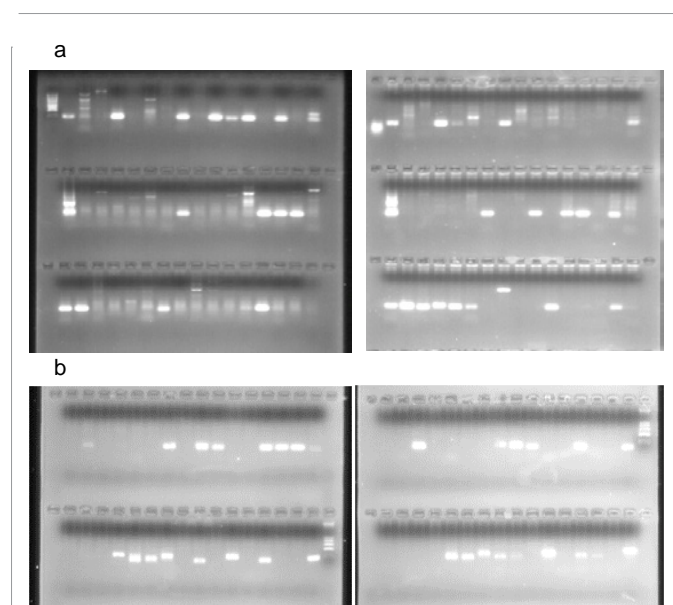
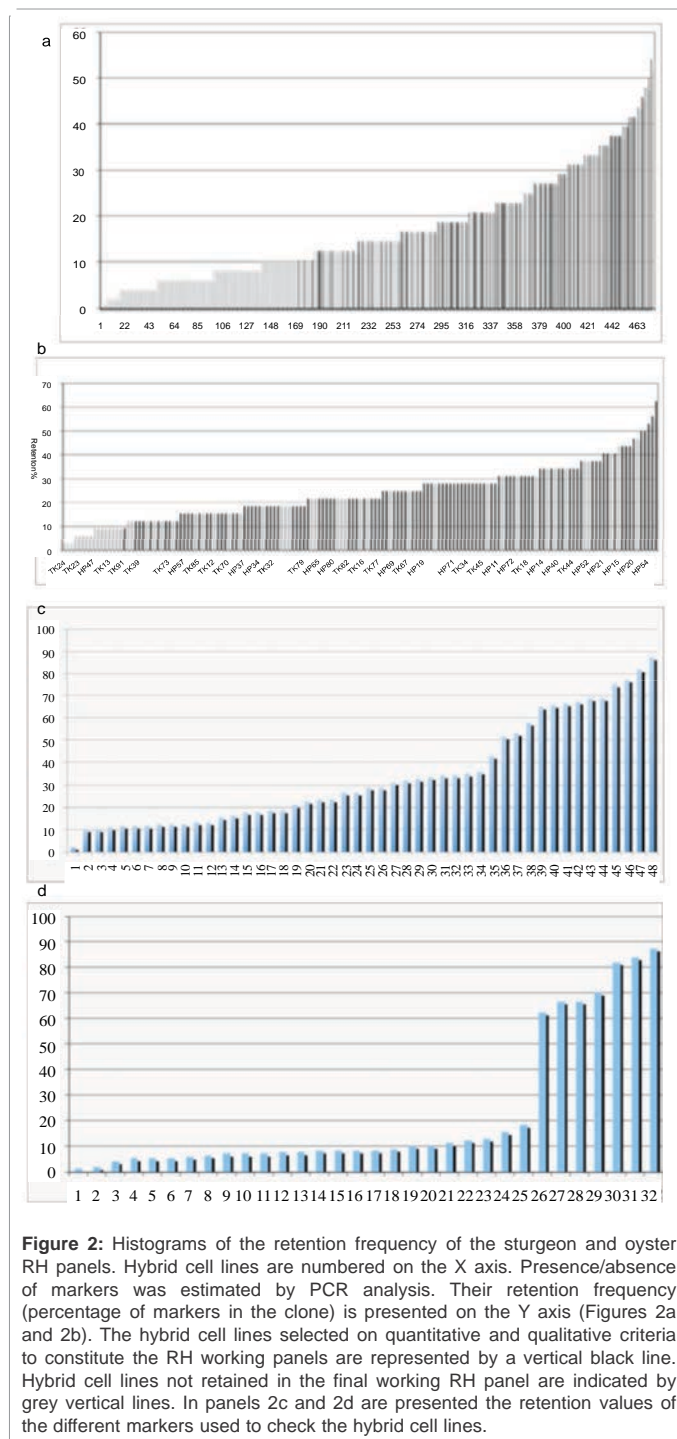


Figure 1: Agarose gel electrophoresis of PCR products obtained with various cell lines.

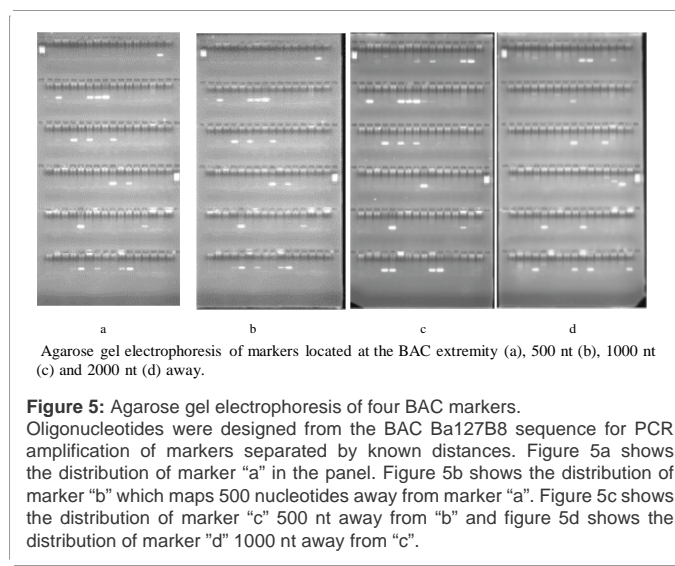
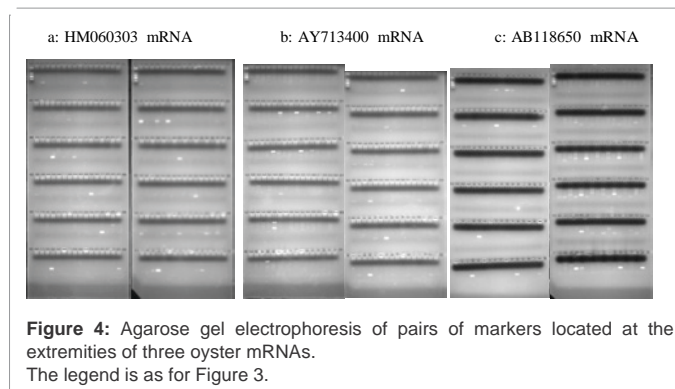
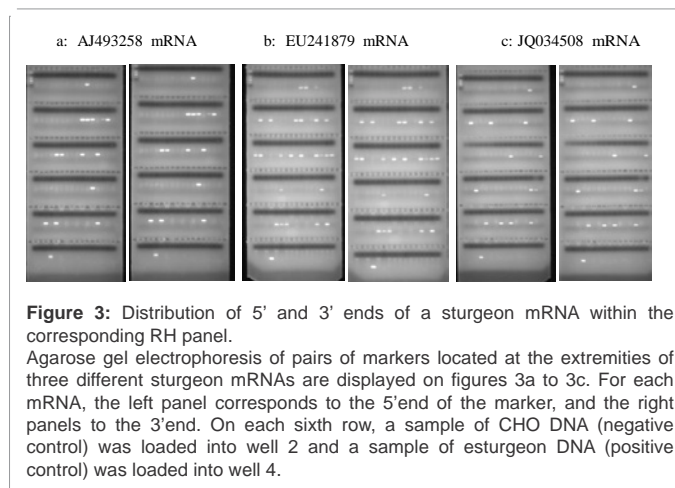
The left panel of Figure 1a corresponds to one sturgeon hybrid cell line and the right panel to another cell line. Similarly, Figure 1b corresponds to two oyster hybrid cell lines. Each slot corresponds to the products of PCR amplification of one marker.



the same conditions. Sturgeon and oyster cells prepared as described were irradiated with a dose of 3500 rads, which past experience suggests is suitable for constructing panels of manageable resolution [6-8,12,13]. Following fusions, the cells were incubated in HAT medium to allow the growth of the hybrid cells only. Little is known about the sturgeon and oyster TK and HPRT genes and proteins, so for both donor cell types, we performed two fusions in parallel, one with CHO TK-cells as the receptor cells and the other with CHO HPRT-cells. We then checked the two selection systems for their ability to support the growth of the hybrid cells in the HAT selective medium. Subsequent fusions were

	Mean Retention Values		Panel Retention Values	
	All clones	Panel	Min	Max
Sturgeon	14%	25%	13%	50%
Oyster	22.7%	24%	3.1%	62.5%

Table 2: Marker retention values. Retention values were calculated by genotyping 48 sturgeon and 32 oyster markers in each of the corresponding hybrid cell lines.



mRNA	Length (bp)	Oligonucleotide pairs at the 5' end		Oligonucleotide pairs at the 3' end	
		Forward	Reverse	Forward	Reverse
EU241879	992	GGCAACCTGAAGAAGGCTCT	gtgccttgtgtAACcGTCCT	GCCTACATGGACCAACAAGC	ATGTGAATCGGCTGTTTCCT
U65435	1091	gcagaagacctgtgcagtt	gtgaaattgcgcggaagag	CCGCAGATGAAAGTTCAAGAG	GCCTCCTGCCTATGTACCAG
HM348777	2343	ATGACCAgGGcAACAGGAC	TGTGTTGGTGGGATTCATtG	CAGACTGcTGAAGGAGGA	ATGGTaGGGCCGgAtgag
FJ428827	1684	ggatatctgacccagagca	acgtcaattgcaaggaaacc	GATGACCTGTGGGTGGAGTT	TCAGGTTCTCTGAGCCAAGC
AB111408	1634	accctacggaccctatttt	tccattagcagttggcgta	ACTTCCCTGTTTGACGAGGA	TGATGAAATCGACAGGCAAA
AJ493258	1490	aggcaagaagaagcaacg	caatgcaagaagagccatt	AGGAAGTGAAAGCCAAAGCA	TCCTCCCTGGTATTGACCAC
JQ288774	1205	ATTGCAGATCTCCAGCCTGT	ATGTCCGTCATGCTGCTGT	gaaggggtggactgtgaaga	agcattcaatgTGGGGAAAC
JQ034508	2325	CGAGAGGAacCGcTATGAGA	gAtCAGgGAGaccAccAtA	tgtccttgtgcttctgctc	cagcataccgcatgtgcat
Y13253	1950	tcGACAagttCCTGGAAAgG	cACgGCAGTGCTTCCAGTCT	tcatgagaccatccaaccac	acaggggttagcctgtcag
X90557	1027	tctcccagagacacagta	TCTTCAGcCTGGACTCCA	tcccgctgtgtaagtgc	aagggtcaggcacattgaaac
AB111407	3359	GGTGAGAcAGCGGAGATGC	CTGAAtCCGTTGCTGTGgAc	CAGAGCTGGAGTGGGATTGT	GCATAGCACTGAGCATCAACA
AJ493265	2145	CTgAATGACCGTTtGtgAA	cCTcTGgTTGGTgAggAtgT	ATCCCATGTCACGGCTGTAG	CACACACCTGTTTGCTGATG

Table 3: List of sturgeon mRNAs selected for panel analysis. Forward and reverse oligonucleotides were designed with software Primer 3 for the amplification of DNA fragments of roughly 100 nt.

mRNA	Length (bp)	Oligonucleotide pairs at the 5' end		Oligonucleotide pairs at the 3' end	
		Forward	Reverse	Forward	Reverse
AB118650	1228	CTTAGCGAGGGGGAGAGC	CCAAAGTGCCTCCCAATTTA	CCCCACTGTGTGTGACC	AAGAGAGAAATTCGGCCACAAG
JN251132	1855	CCTCCTCAGAACGCACA	AGCTGATGGGTGGCAATTT	AATCCAGTCTCAGGGCGTTT	TCTCCACGTTGGACAGACTC
HM060303	2424	AGATGAGTGCAGAGGTTCC	CATCGGTCCCCGTCAGTT	GGATGTACAACCTCAAGCTTCAAC	TGCTCCATGCGTTGTATTGT
HQ457439	2018	CGAAACCGGGTTTGTATT	TCTTCCATGTTTCTTTCCTTT	TGTTGTTTGGTACATTGTGATTAGA	TTGAAATGATGTGTTTGCATTTT
AY713400	1128	ACATCCAGGCCGTCAGT	TCACCATGGTATGTGCCAGT	TGATGTGCAGATTTGAACG	ACGTGGTGGGATGTACACAG
DQ336097	921	GTATAGCTGACCTGGCCCTTG	TTTTACAGACCCGGAGCTACA	GCTAATGACTAATGTGCAACTGG	TTGGGGCAAGTCTACCTC

Table 4: List of oyster mRNAs selected for panel analysis. Legend as for Table 3.

made with the CHO HPRT-mutant to construct the sturgeon panel: individual sturgeon hybrid clones were obtained and could be grown. By contrast, no oyster hybrid clones were obtained with either of the two CHO mutants. However, it seemed likely that a small number of hybrid cells could have survived after three weeks under HAT medium in a number of titre plate wells because in the absence of any fusion the bottoms of the wells should be totally clear with no remaining cells. However, the aspect of some of the wells was unusual, but suggestive of the presence of cells. We suspected that there had been cell fusion, but that the TK and the HPRT proteins were produced at only a low level or were poorly active in the CHO context, such that hybrid clones survived but did not grow properly. We therefore modified the protocol of selection: the cells were cultivated in HAT and DMEM media alternately. The aim was to eliminate the oyster cells and the CHO cells but not the hybrid cells during the HAT periods (48 hours) and then allow the hybrid cells to grow during the DMEM periods (48 hours). After two weeks of this selection procedure, the content of each well was independently recovered, and the DNA extracted and amplified with Genomiphi [7,8]. From three fusions for sturgeon we recovered 474 hybrid cell lines and from four fusions for oyster, we recover 168 lines (Table 1).

Panel characterization

The hybrid retention value (the percentage of markers per RH clone) and the marker retention value (the percentage of each individual marker in each hybrid cell) are two metrics that characterize the ability of an RH panel to produce good and useful RH maps [8]. To estimate these two values for our RH lines, we selected from GenBank 48 sturgeon microsatellites and 32 oyster markers composed of ESTs and microsatellites. The banding pattern illustrating the distribution of each of the 48 microsatellite markers in two sturgeon hybrid cell lines is shown in Figure 1a (Figure 1a). Similarly, the distribution of the 32 markers in two oyster hybrid cell lines is shown in Figure 1b (Figure

1b). From these experiments, we selected 94 cell lines to constitute each of the two panels. Briefly we first selected the hybrid clones with each containing between 10% and 70% of the markers that were typed. Then based on the marker retention values, we added to these first groups of cells various other clones such that the overall retention values for markers poorly represented in the panel were increased (Figures 2a-2d and Table 2). The capacity of the sturgeon and oyster panels to link markers, and make RH groups and RH maps was tested by typing a number of pairs of markers with known nucleotide inter-distances. We retrieved from GenBank the nucleotide sequences of 12 sturgeon mRNAs (Table 3), designed oligonucleotide pairs at their extremities and performed PCR amplification with each of the 94 cell lines of the sturgeon RH panel. The PCR products were resolved by electrophoresis in a 2% agarose slab gel. Visual inspection of the gel patterns indicated that markers at the two extremities of a mRNA gave the same pattern, indicating the capacity of this panel to link markers and make RH groups (Figures 3a-3c). Similarly, to test the quality of the oyster panel we selected six mRNAs from GenBank. However, in these cases the markers located at the two extremities of an oyster mRNAs gave different banding patterns (Figures 4a-4c and Table 4), suggesting that the DNA fragments in the oyster hybrid lines were short. To estimate the sizes of the fragments in the hybrid lines, we obtained oligonucleotides corresponding to sequence of BAC Ba127B8 located 500 nt apart. Markers 500 or 1000 nt apart had strictly or nearly identical distributions in the panel, as assessed by visual inspection, whereas markers 2000 nt apart had different distributions (Figures 5a-5d). These results indicate that the size of the DNA fragments retained by the hamster/oyster hybrid cells are shorter than 2000 nt and too short for the construction of an RH map of thousands of markers. However, we tested whether these RH lines could be used for the construction of maps of small (1 megabase or shorter) regions of interest in the genome. We obtained 40 pairs of oligonucleotides for amplification of markers, 500 nt apart within a region of 20 KB of BAC Bac188K9

Marker names	Left Primer	Right Primer
Ma1	CACGCAACATTAAGCTTTGG	GCGTCCCTGAACAGGTAATG
Ma2	TGCGTTCTGTACGTGAACCT	CGCCGCTCATAGGTAGAAAC
Ma3	CGAGGGCACGAATAGACTGT	GGGACTTGTGGTCTGCTGTT
Ma4	TCGGAATAACGTGACTCGAT	AGTAGGATTCCGCTGATTGC
Ma5	AGGCTCGATTGTCATCATT	AAACATGTCCCCTCGTGTTT
Ma6	GTTCCCATGAGGAGGGATT	TGTCGGCTAGTGTGGAACAG
Ma7	GCTTCTTACCTTCGCACTC	CGACTGGTGAGATGCTTCAA
Ma8	GTGGGTGCATCATTTACAG	CGGCTTATGGAAGGTGATGT
Ma9	TTGTACGTGTGGGGAAAGT	TGCTTTCGGATCTCCTCAGT
Ma10	AATCCAAGATGGCGAGTGT	GGTCAATTCGGCATGTTTTT
Ma11	CAGGAGACGATTTGGCTGTT	GCGACATTGCAGGAAAATTA
Ma12	CGGGTGCACCTTCACTGATT	ACATGGTTTTTTCAGCCTGGAC
Ma13	GATGTCCGTGAAAGTACTCC	AAAAATCAGGTTTAGTTCCGACAGG
Ma14	CTGTTGGCCTTACATGTTGTG	ACAGGGATTGGACACAAGTTC
Ma15	CCGGCGATATGACAGAGTAGA	CGCCTAGGGTGTGACATTTT
Ma16	ATTGCCCGCTGTTTCACT	GCGTTTTTCCAATTTTTCCA
Ma17	ATCAGAGACTGGCCCATGTT	GCCTGATGGTGGTGATCATA
Ma18	GAATCATGGACAGGGACTGG	CTCAAAACCATCCCCTACGA
Ma19	TCTTCTCCGTACCCATTCA	TTTAGTCAAGTGAGATGTTGTTTGC
Ma20	TTGCTAGCGCATCTCACC	CGGCTGCATACTTTCATTG
Ma21	GTGCGCCATGAAAAGGATAC	GATCCGCAAAAAGCTGAAAC
Ma22	TTGCATACAAGGTGGTGAGC	GAAAATCAACCCTGAACCAAGA
Ma23	CCCTACCAGACAGGACAATCA	CAAACCACTGCCAATCACTT
Ma24	TCACAGGCAAAGACACTGGT	TCCACTTTTGTGCTGTTAGG
Ma25	CAAGCCAGTCCACACTTTCT	ATTGCCTGGACATTGGTGAT
Ma26	TCCACCATGGCAGTCAAAG	TGCACTGTGCTTTTCCCTTCTG
Ma27	GGAATCCCATGAAACATCAGA	AGCCCTCCTACAGCTCCTTT
Ma28	TCGTGCCACACTGAGAGACT	TTCCAGGGAGACAGTGCTTT
Ma29	GTACAATACGCACCCCAAC	TCAGGATTACCGTTTGGTCA
Ma30	AAAAACGGACGGGTGACAGG	ATTTTATTATGTTCCAGGCCTCT
Ma31	TGTCAAAGCTCACCTCCCTTA	CCAGTACGGGTATTTGCAT
Ma32	CACCACCGGGTTAGAGAAGA	TCTGACAGCTTTCTCCGGTA
Ma33	AATGCACAGCCAAATGTAAC	CCAAGTTTTCTACTGTGCAATCA
Ma34	AGACTCCTGCAGCCCTGAC	TTCACTGTGCGACCTTTGAC
Ma35	AGGTCATCTCTGAGGGTCAAG	TTCTCAGGTGATGCTGACA
Ma36	CCTGGTAAAATTCTATGCCACA	ACCTCCTTTACAAGCATCCAG
Ma37	CCAGAGTTGTTTCCCTGTC	CAGTGTTACACCACCAAAAG
Ma38	TTGAGCAGTATTACGTGCTTATTG	CGGGGTATTAGTCCCATTA
Ma39	AGCCAAGCAATGCTCTTCTG	AGGAGAAAGTGAGGAAAATTTGAA

Table 5: List of oyster markers designed on BAC Ba188K9. Forty sequences regularly spaced over 20 Kb were selected. Forward and reverse oligonucleotides were designed with Primer 3 to amplify short DNA fragments as markers named : Ma1 to Ma 40.

(Table 5) and genotyped with them the 94 hybrid cell lines of the RH oyster panel. The vector suites were determined and computed as usual [14,15]. Results of the two-point analysis are summarized in Table 6 and those of the multipoint analysis determining the order of markers, within the three largest groups are given in Table 7. Both analyses indicated that a dense regional map can be constructed.

Discussion

With the advent of the high throughput sequencing technologies, collectively named NGS (next generation sequencing), and the substantial reduction in the associated costs, it could be believed that the construction of physical maps of genomes is no longer required. Indeed, genome sequencing strategies have progressed from sequencing BAC clones organized into a physical map [16], to a global shotgun strategy as was used for the human genome sequence [17,18] and then

#of Markers /linkage groups	7	6	6	2	2	7 unlinked Markers	10 non genotyped Markers
Marker names	Ma34 Ma33 Ma32 Ma27 Ma26 Ma25 Ma23	Ma8 Ma6 Ma3 Ma1 Ma2 Ma5	Ma15 Ma19 Ma20 Ma21 Ma17 Ma18	Ma31 Ma28	Ma29 Ma30	Ma4 Ma7 Ma10 Ma11 Ma16 Ma36 Ma37	Ma9 Ma13 Ma14 Ma15 Ma22 Ma24 Ma35 Ma38 Ma39 Ma40

Table 6: Results of the two-point analysis performed on BAC 188K9. Five RH linkage groups were obtained with Multimap software at a lod score of 3.0. (Matisse et al. 1994). As indicated in the legend to Table 3, markers were named according to their positions on the BAC sequence. Note that RH groups are constituted of markers located close together on the BAC sequence.

RH group 1:

Marker names	Relative position in the RH group	Distance between markers	Cumulative distance	Theta (%)	2pt LOD
Ma 23	1	0.0 cR			
Ma 25	2	69.4	69.4 cR	50.0 %	3.2
Ma 26	3	15.0	84.4 cR	13.9%	10.8
Ma 27	4	9.9	94.3 cR	9.4 %	11.8
Ma 32	5	28,9	123.2 cR	25.1 %	6.9
Ma 33	6	35,2	158.4 cR	29.6 %	6.3
Ma 34	7	21,2	179.6 cR	19.1 %	9.0

RH group 2 :

Marker names	Relative position in the RH group	Distance between markers	Cumulative distance	Theta (%)	2pt LOD
Ma 8	1	0.0 cR			
Ma 6	2	68.2 cR	68.2 cR	49.4 %	1.3
Ma 3	3	16.8 cR	85.1 cR	15.5 %	5.0
Ma 1	4	16.9 cR	102.0 cR	15.5 %	7.1
Ma 2	5	48.4 cR	150.4 cR	38.4 %	5.5
Ma 5	6	55.5 cR	205.8 cR	42.6 %	4.3

3rd group :

Marker names	Relative position in the RH group	Distance between markers	Cumulative distance	Theta (%)	2pt LOD
Ma15	1	0.0 cR			
Ma19	2	24.8 cR	24.8 cR	22.0 %	6.4
Ma 20	3	53.7 cR	78.5 cR	41.6 %	4.7
Ma21	4	61.4 cR	139.9 cR	45.9 %	3.0
Ma17	5	37.1 cR	177.0 cR	31.0 %	3.8
Ma18	6	53.3 cR	230.3 cR	41.3 %	3.8

Table 7: Order of markers within the three largest RH groups. Multipoint analyses were carried out with CarthaGene software. Distances are expressed in centirays (cR3500) (Schiex et al. 1997). The order of markers as defined by the sequence nucleotide (1st column) appear to be exactly the same as that defined by RH mapping (2nd column) for RH group 1. In RH group 2 and 3, there are a small number of inversions between the two methods but overall the two mapping strategies gave collinear results all along the 20 Kb of BAC188K9 analyzed.

the canine genome sequence [19]. However, despite the development of very sophisticated algorithms to assemble NGS reads, the recent releases from whole genome sequencing projects based on a NGS approach showed highly discontinuous sequences composed of hundreds of thousands contigs [20]. This is due to the reads being short (relative to the entire genome) and an uneven coverage of the whole genome. To overcome this problem of discontinuity, genome maps constructed with a large number of markers were, and still are, a necessity [21-23].

RH maps are by definition not physical maps, as distances between markers are measured in centirads, which result from a statistical treatment of the data. Nevertheless, they are valuable, partly because they are straightforward to construct due to the development of high throughput genotyping methods; also, diverse markers, both polymorphic and non polymorphic, with a large range of marker density, can be used as long as the level of radiation applied to the donor cell before fusion is appropriate [24].

There are, nevertheless, various difficulties associated with the construction of RH panels: some are associated with particular steps of the construction, and some differ between projects. Constructing panels for mammal species does not present any particular difficulty, because donor cells are easy to obtain and their fusion with rodent cells poses no problem.

By contrast, the construction of RH fish panels was long considered to be difficult, or even impossible, mainly because of the absence of a suitable source of donor cells. The strategy we developed using the spleen as the starting material proved to be very effective [7,8]. Splenocytes can be obtained in very large numbers, ready to use after cutting the whole organ into pieces with scissors and gentle dissociation in a Potter device. Fusion between fish splenocytes and hamster cells was straightforward: we have been able to produce hybrid cells with all the species we have tested so far, including seabream [6], seabass [7], tilapia [8], sturgeon (this work) and trout (unpublished data). For oyster RH lines, once we had overcome the technical issue of the difference in the osmotic pressures appropriate for oyster cells and the hamster cells, we had no major difficulty in obtaining enough donor cells from oyster muscle.

The main problem in any RH panel construction is the uncertainty about whether the TK or HPRT proteins of the donor cells are expressed and active in the hamster environment, so as to allow the growth of the hybrid cell lines and prevent the growth of the receptor cells. Unfortunately, this is highly unpredictable. For all the fish panels we have constructed, the donor HPRT gene was used to select the hybrid cells, and for some of them the TK gene could also have been used. In an attempt to make a RH panel for crocodile (*Crocodylus porosus*), we prepared fresh white blood cells from 50 ml of blood. These cells were then irradiated and fused with acceptor hamster cells. The number of hybrid clones was the same whether the HPRT or TK gene was used for selection (unpublished data). Conversely, neither the oyster TK nor the HPRT proteins were able to allow strong growth of the hybrid cells, or their selection from the non-fused hamster acceptor cells. Nevertheless, oyster markers were detected in the DNA extracts demonstrating that there had indeed been cell fusion, and the oyster TK and HPRT proteins allowed the hybrid cells to survive under HAT selection (Figure 1b). Selection of the hybrid cells is an essential step in the construction of any panel, but currently, only the TK and HPRT genes can be used for this purpose. To overcome this limitation, we attempted to transform the oyster muscle cells, prior to fusion, with a plasmid (Addgene plasmid 20652) carrying the human TK gene, which, intrinsically, is able to complement the CHO deficiency. Although transformants were obtained, their numbers were too low to produce a RH panel.

As demonstrated by the typing of 40 markers of BAC Ba188K9 and the corresponding mini RH map, the oyster RH panel we produced could nevertheless be useful to make very dense RH maps of small regions of particular biological interest. Whole genome sequencing with NGS technologies provides very short contigs, many in the range

of one to two kb. These contigs are usually assembled in scaffolds made of several contigs. But checking the exactness of some scaffolds of particular biological interest could be important and would be feasible with such oyster panel

In this study we demonstrate that the construction of RH panel for a wide variety of species is entirely feasible starting from various tissues as spleen, blood or muscle as sources for the donor cells. We also show that the issue of the ability of donor TK or HPRT proteins to rescue the hybrid cells remains problematic. Consequently, the development of alternative systems for selection would be very useful, given the importance of RH maps as resources for aquaculture, and more generally their utility in whole genome sequence assembly

Acknowledgments

This work was supported in part by ANR (Agence Nationale de la Recherche, France-Gametogenes project (08-GENM-041-03).

Conflicts of Interest

Authors declare no conflict of interest.

References

1. Calderon-Garcia JF, Moran JM, Roncero-Martin R, Rey-Sanchez P, Rodriguez-Velasco FJ, et al. (2012) Dietary habits, nutrients and bone mass in Spanish premenopausal women: the contribution of fish to better bone health. *Nutrients* 5:10-22.
2. GISSI (1999) Dietary supplementation with n-3 polyunsaturated fatty acids and vitamin E after myocardial infarction: results of the GISSI-Prevenzione trial. *The Lancet* 354: 447-455.
3. Bucher HC, Hengstler P, Schindler C, Meier G (2002) N-3 polyunsaturated fatty acids in coronary heart disease: a meta-analysis of randomized controlled trials. *Am J Med Mar* 112: 298-304.
4. Zhang G, Fang X, Guo X, Li L, Luo R, et al. (2012) The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* 490: 49-54.
5. Guimarães EP, Ruane J, Scherf BD, Sonnino A, Dargie JD (2007) Marker-Assisted Selection: Current status and future perspectives in crops, livestock, forestry and fish. Food and Agriculture Organization of the United Nations, Rome.
6. Senger F, Priat C, Hitte C, Sarropoulou E, Franch R et al. (2006) The first radiation hybrid map of a perch-like fish: the gilthead seabream (*Sparus aurata* L). *Genomics* 87: 793-800.
7. Guyon R, Senger F, Rakotomanga M, Sadequi N, Volckaert FA, et al. (2010) A radiation hybrid map of the European sea bass (*Dicentrarchus labrax*) based on 1581 markers: Synteny analysis with model fish genomes. *Genomics* 96: 228-238.
8. Guyon R, Rakotomanga M, Azzouzi N, Coutanceau JP, Bonillo C, et al. (2012) A high-resolution map of the Nile tilapia genome: a resource for studying cichlids and other percomorphs. *BMC Genomics* 13: 222.
9. Hukriede NA, Joly L, Tsang M, Miles J, Tellis P, et al. (1999) Radiation hybrid mapping of the zebrafish genome. *Proc Natl Acad Sci USA* 96: 9745-9750.
10. Kwok C, Critcher R, Schmitt K (1999) Construction and characterization of zebrafish whole genome radiation hybrids. *Methods cell boil* 60: 287-302.
11. Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. *Meth Mol Biol* 132: 365-386.
12. Priat C, Hitte C, Vignaux F, Renier C, Jiang Z, et al. (1998) A whole-genome radiation hybrid map of the dog genome. *Genomics* 54: 361-378.
13. Vignaux F, Hitte C, Priat C, Chuat JC, Andre C, et al. (1999) Construction and optimization of a dog whole-genome radiation hybrid panel. *Mamm Genome* 10: 888-894.
14. Matise, Perlin M, Chakravarti A (1994) Automated construction of genetic linkage maps using an expert system (MultiMap): a human genome linkage map. *Nat Genet* 6: 384-390.
15. Schiex T, Gaspin C (1997) CARTHAGENE: constructing and joining maximum likelihood genetic maps. *Proc Int Conf Intell Syst Mol Biol* 5: 258-267.

16. Krzywinski M, Bosdet I, Smailus D, Chiu R, Mathewson C, et al. (2004) A set of BAC clones spanning the human genome. *Nucleic Acids Res* 32: 3651-3660.
17. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, et al. (2001) The sequence of the human genome. *Science* 292: 1304-1351.
18. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. (2001) Human Genome. *Nature* 409 : 860-921.
19. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, et al. (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438: 803-819.
20. Gnerre S, MacCallum I, Przybylski D, Ribeiro F, Burton J, et al. (2010) High- quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proceedings of the National Academy of Sciences* 4: 1513-1518.
21. Hitte C, Madeoy J, Kirkness EF, Priat C, Lorentzen TD, et al. (2005) Facilitating genome navigation: survey sequencing and dense radiation-hybrid gene mapping. *Nat Rev Genet* 6: 643-648.
22. Lewin HA, Larkin DM, Pontius J, O'Brien SJ (2009) Every genome sequence needs a good map. *Genome Res* 19: 1925-1928.
23. Servin B, Faraut T, Iannuccelli N, Zelenika D, Milan D (2012) High-resolution autosomal radiation hybrid maps of the pig genome and their contribution to the genome sequence assembly. *BMC Genomics* 13: 585.
24. Cox DR, Burmeister M, Price ER, Kim S, Myers RM (1990) Radiation Hybrid Mapping: A Somatic Cell Genetic Method for Constructing High-Resolution Maps of Mammalian Chromosomes. *Science* 250: 245-250.

II- Le Génome des cichlidés

Oreochromis niloticus a été choisi comme espèce modèle pour le séquençage complet de son génome. Celui-ci a été réalisé par le BROAD Institute (Boston) qui a utilisé la méthode développée par Illumina. En bref il s'agit d'une approche en shotgun du génome complet et d'un séquençage hautement parallélisé connu sous le vocable de NGS. L'avantage de cette approche est qu'elle est rapide et d'un coût très inférieur. L'inconvénient est que la longueur des lectures est courte (surtout à cette époque de l'ordre de 75 nt seulement). Il s'ensuit que l'assemblage est délicat, nécessite des algorithmes particuliers et malgré cela délivre une séquence génomique très fragmentée en plusieurs dizaines de milliers de contigs. Nos résultats de cartographie et ceux du premier assemblage du Broad ont été comparés. Ceci a permis de relever plusieurs incohérences et de corriger un certain nombre d'erreurs dans ce premier assemblage. Par ailleurs la carte RH étant intégrée au caryotype, cela a permis d'attribuer un grand nombre de contigs et de supercontigs aux divers chromosomes.

Les lacs africains sont des biotopes naturels exceptionnels contenant un grand nombre d'espèces de la famille des cichlides (1000 espèces dans le Lake Malawi seul). Ces espèces présentent des rythmes évolutifs très rapides. Elles constituent ainsi un excellent modèle pour étudier les mécanismes d'évolution chez les vertébrés. C'est pourquoi, en plus du Tilapia du Nil, quatre autres cichlidés ont fait l'objet de séquençage par le Broad. Il s'agit de : *Pundamilia neyereri*, *Haplochromis burtoni*, *Metriaclima zebra* et *Neolaprologus brichardi*.

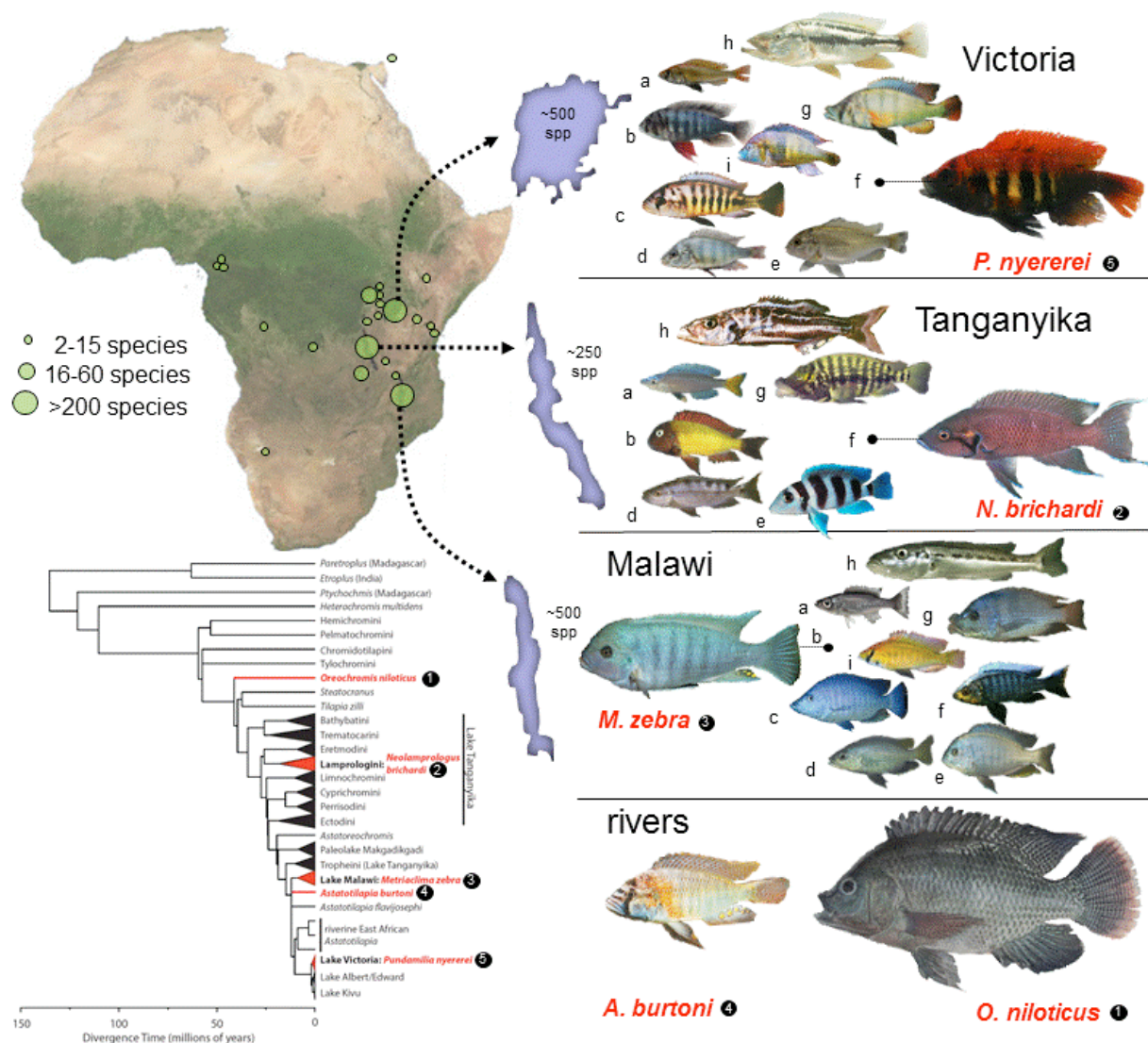


Figure 24: Les grands lacs de l'Est africain regroupent beaucoup d'espèces de cichlidés (Wagner CE et al. 2012)

• *Pundamilia nyererei*



Pundamilia nyererei est une espèce endémique de l'Afrique et du lac Victoria. Il se nourrit principalement de larves d'insectes et de bryozoaires.

• *Haplochromis Burtoni*



Le Burtoni est un cichlidé du Lac Tanganyika. On retrouve la ligne typique à cette espèce : un corps allongé et aplati, dans la nature il se nourrit principalement de proies vivantes.

• *Metriaclima zebra*



Cette espèce située au Lac Malawi a pour principale source de nourriture des algues filamenteuses qui poussent sur des surfaces rocheuses exposées au soleil.

• *Neolaprologus brichardi*



Brichardi est une espèce endémique du Lac Tanganyika. Relativement pacifique, vivant en bancs, il se nourrit de matières végétales, comme la spiruline ou d'épinards blanchis.

Pour ces quatre cichlidés nous n'avons pas construit de carte RH. La séquence de leurs génomes a été réalisée par le BROAD (publication N°6) et les cinq séquences comparées entre elles. Par contre, nous avons participé à l'annotation des 5 séquences des génomes de ces cichlidés en focalisant notre analyse sur l'identification des répertoires des gènes codant les récepteurs olfactifs (OR) et les récepteurs d'amines « Trace Amine Associated Receptors » ou TAAR. Les résultats de ces analyses sont décrits avec les publications 4 et 5.

1- Annotation des génomes

L'annotation d'un génome consiste à traiter l'information brute contenue dans la séquence. A cet égard on peut distinguer :

- (1) L'annotation structurale ou syntaxique dont le but est de prédire le contenu en gènes, la position de ceux-ci, ainsi que leur organisation.
- (2) L'annotation fonctionnelle dont le but est de prédire la fonction potentielle des gènes.
- (3) Et l'annotation relationnelle qui établit la relation entre les entités, soit au sein du même génome (voies métaboliques, familles de paralogues, etc), soit au sein de plusieurs génomes (groupes de synténie, familles d'orthologues, etc).

L'annotation peut être automatique, elle s'appuie alors uniquement sur des algorithmes recherchant des similarités de séquence, de structure, de motifs protéiques. Elle aboutit au transfert « automatique » de l'information figurant dans l'étiquette d'un gène « similaire » d'un génome déjà annoté au génome en cours d'annotation. L'annotation automatique est une nécessité compte tenu de l'afflux massif de données issues des nouvelles techniques de séquençage. Toutefois l'annotation automatique doit être complétée par une annotation manuelle par des experts qui valident ou invalident la prédiction en fonction de leurs connaissances ou de résultats expérimentaux. Par ailleurs l'annotation peut être d'emblée manuelle, notamment si l'on s'intéresse à une fonction ou une famille de gènes en particulier.

Identification des gènes OR et TAAR

Dans le cadre du consortium international réunissant une quinzaine de laboratoires autour du projet cichlidé, nous avons participé à l'annotation des données de séquences des génomes de cinq poissons (*O. niloticus*, *P. nyererei*, *H. burtoni*, *N. brichardi* et *M. zebra*) en concentrant notre analyse sur l'identification des gènes codant les récepteurs olfactifs (OR) et les « Traces Amine Associated » ou TAAR. Le tableau 1 donne la répartition des séquences fournies par le BROAD.

Tableau 1 : Répartition des séquences fournies par le BROAD

	<i>O. niloticus</i>	<i>H. Burtoni</i>	<i>P. Nyererei</i>	<i>M. zebra</i>	<i>N. Brichardi</i>
#Total CONTIGS	77754	69074	68053	71043	118196
#Taille CONTIGS	819Mb	699Mb	699Mb	713Mb	686Mb

Pour la réalisation de cette étude, les principales étapes d'annotation ont été automatisées par la réalisation de différents scripts permettant d'enchaîner facilement les traitements à chaque étape.

L'identification des gènes dans une séquence nouvelle pose de nombreux problèmes résolus par l'application de plusieurs règles. Pour les gènes codant les OR et les TAAR, nous avons suivis deux stratégies. La première par recherche de motifs d'acides aminés caractéristiques de ces gènes, la deuxième par BLAST.

1- 1- Recherche des gènes OR et TAAR par motifs d'acides aminés

Pour cette analyse, nous sommes partis du fait, apparemment bien établi selon lequel les gènes OR et TAAR des mammifères et des poissons ne sont constitués que d'un seul exon codant d'une taille comprise entre 850 et environ 1000 nucléotides (Buck L.B. & Axel R. 1991 ; Liberles S.D. & Buck L.B. 2006; Malnic B *et al.*, 2010) ; Quignon *et al.*, 2005). Nous avons d'abord sélectionné toutes les phases ouvertes (ORF) répondant à ce critère de taille à l'aide d'un programme créé au laboratoire et appelé ORF.py qui traduit les contigs dans les six phases de lecture. Nous n'avons alors conservé que les ORF ayant une capacité de codage d'au moins 280 acides aminés. Puis, nous avons identifié avec le programme « MEME suite » les 5 meilleurs motifs spécifiques des OR et des TAAR des poissons modèles (poisson zèbre, stickleback et médaka) trouvés dans les bases de données (NCBI et Ensembl Genome Browser). Chacun de ces motifs a été alors recherché indépendamment avec le programme « MAST » dans les protéines traduites des ORF sélectionnées. Toutes les ORF possédant au moins deux des cinq motifs ont été sélectionnées pour la construction d'un arbre phylogénique avec le programme ClustalX. Plusieurs voire toutes les séquences de chaque branche de l'arbre ont alors été analysées par un Psi Blast pour ne retenir que celles donnant une réponse significative avec un gène OR ou TAAR présents dans les bases de données.

A ce stade les résultats obtenus ont été les suivants :

Pour les OR :

Tableau 2 : Résultats des OR pour les 5 cichlidès selon les motifs recherchés

	<i>O. niloticus</i>	<i>H. burtoni</i>	<i>P. nyererei</i>	<i>N. brichardi</i>	<i>M. zebra</i>
Motif 1	270	167	149	148	173
Motif 2	171	100	96	78	107
Motif 3	129	87	86	69	94
Motif 4	144	79	85	62	90
Motif 5	113	61	65	38	67

≥3 Motifs	149	93	91	74	98
2 Motifs	33	10	9	9	13
Total OR	182	103	100	83	111

Pour les TAAR :

Tableau 3 : Résultats des TAAR pour les 5 cichlidès selon les motifs recherchés

	<i>O. niloticus</i>	<i>H. burtoni</i>	<i>P. nyererei</i>	<i>N. brichardi</i>	<i>M. zebra</i>
Motif 1	57	44	43	69	33
Motif 2	78	57	61	44	52
Motif 3	87	60	70	38	46
Motif 4	61	68	64	47	52
Motif 5	53	57	48	28	37

≥3 Motifs	63	52	55	39	51
2 Motifs	13	7	18	7	0
Total TAAR	76	59	73	46	51

Dans la deuxième partie des tableaux 2 et 3 ne sont comptabilisés que les gènes ayant au moins 2 motifs dont le premier.

1-2- Recherche des gènes OR et TAAR par BLAST

Dans cette deuxième approche nous ne sommes partis d'aucun *a priori* sur la structure des gènes OR et TAAR des cichlidés. Nous avons d'abord récupéré à partir de la base de données (NCBI) 891 gènes OR de différents poissons, 395 gènes TAAR également de nombreux poissons et 407 gènes GPCR de poissons. Avec ces séquences nous avons constitué deux sets positifs, un pour les OR et un pour les TAAR, et un set négatif constitué de GPCR non OR ni TAAR.

Chacun des deux sets positifs a été utilisé pour identifier par TBLASTN à partir de l'ensemble des séquences du BROAD les gènes candidats. Chacune des deux listes de gènes candidats a alors été comparée par BLAST avec le set négatif pour éliminer les gènes candidats faux positifs. Toutes ces comparaisons ont été faites avec une limite de « e-value » de 10^{-50} . Les OR et TAAR ainsi identifiés sont classés en : gènes complets, pseudogènes, « edge » correspondant à des gènes incomplets car situés en limite de contig et à notre grande surprise d'un certain nombre de gènes constitués de plus de 1 exon codant. De plus par cette approche nous avons identifié des gènes incomplets qui correspondent à des fragments de gènes inclus à l'intérieur de contigs. Ces derniers peuvent être dus soit à des problèmes techniques de séquençage ou d'assemblage soit correspondre à des réalités biologiques dont la signification est inconnue.

Les résultats des deux approches méthodologiques ont été pratiquement identiques pour l'identification des gènes complets, avec seulement quelques gènes trouvés par une méthode et pas par l'autre et inversement. En revanche la seconde approche basée sur l'analyse TBLASTN permet l'identification d'autres catégories et en particulier des gènes formés de plusieurs exons codants. Dans le détail des résultats nous ne présenterons que ceux obtenus par cette seconde approche. Les analyses de Blast ont été faites avec le logiciel KoriBlast et la caractérisation des candidats en gènes complets, pseudogènes, « edge », gènes avec plus de deux exons et des gènes incomplets avec le programme GENEIOUS 6.1. Afin de définir précisément la structure des gènes, une analyse manuelle par BLAST a ensuite été faite.

A- Le répertoire des gènes OR (Publication n°4)

Les résultats sont présentés sous forme de tableaux et de graphes avec pour chaque poisson les informations suivantes :

- La répartition des différentes catégories de gènes OR pour chaque poisson :
- les noms des gènes, leurs positions sur le contig, leurs tailles et leurs localisations sur les scaffolds (tableau excel, Article 4).
- Les séquences complètes en nucléotides et en acides aminés
- Un graphe de cartographie des gènes sur les contigs

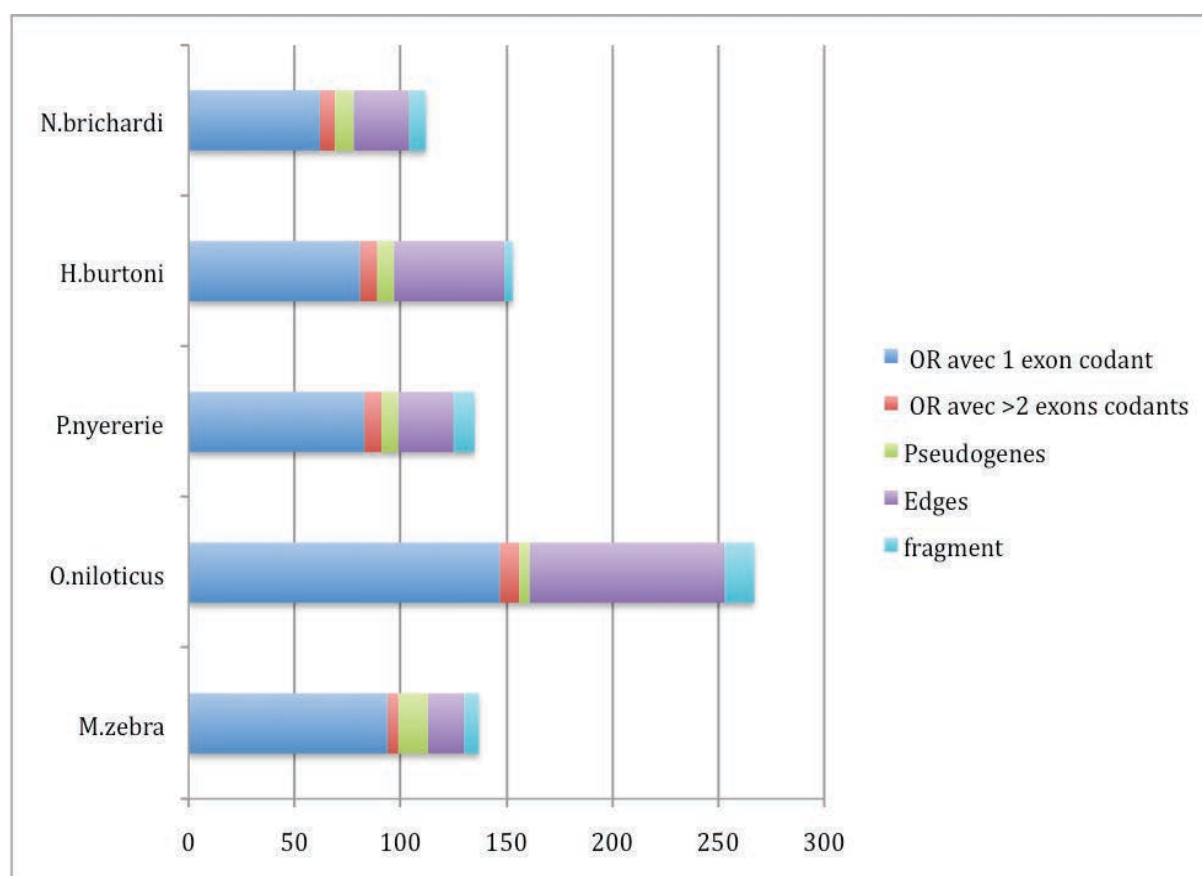


Figure 25 : Répartition de l'ensemble des catégories d'OR chez les cinq cichlidés.

A noter que les gènes identifiés comme « edge » sont destinés à devenir des membres de l'une ou l'autre des catégories au fur et à mesure de l'amélioration de la séquence.

Les données de la littérature qui se limitent aux gènes entiers et pseudogènes concernant les poissons modèles sont rapportées dans le tableau suivant.

Tableau 4 : Décompte des OR et des pseudogènes chez les 5 cichlidés comparé à ceux des poissons modèles (tableau 1 Article 4).

	<i>M.zebra</i>	<i>O.niloticus</i>	<i>P.nynerie</i>	<i>H.burtoni</i>	<i>N.brichardi</i>	<i>D.rerio</i>	<i>G.aculeatus</i>	<i>O. Latipes</i>	<i>T.rubripes</i>
# OR complets	102	158	88	90	69	136 ⁽¹⁾	97 ⁽²⁾	62 ⁽²⁾	40 ⁽¹⁾
#Pseudogenes	11	6	9	6	12	10 ⁽¹⁾	46 ⁽²⁾	28 ⁽²⁾	54 ⁽¹⁾
#Edges	30	102	37	55	37				

(1) Niimura & Nei 2005 (2) Chen Ming *et al.*, 2010

La comparaison de toutes ces données chiffrées est difficile et relativement peu pertinente dans la mesure où les séquences génomiques ne sont pas totalement finalisées. Toutefois, il semble assez évident que les nombres de pseudogènes trouvés chez les cichlidés sont bien inférieurs à ceux des poissons modèles.

1- Etudes phylogénétiques des OR

Les dendrogrammes phylogénétiques sont habituellement construits selon trois méthodes différentes : le Neighbour Joining (NJ), le maximum de vraisemblance (ML), et le maximum de parcimonie (MP). Pour notre analyse nous avons opté pour l'algorithme Neighbour Joining (NJ) qui est une matrice de distances calculées avec la correction de Kimura à deux paramètres. Pour cela nous avons aligné les séquences des gènes avec le programme MAFFT version 7 et construit l'arbre avec PHYML. Les valeurs de bootstraps ont été obtenues en réalisant 1000 répétitions. Les dendrogrammes phylogénétiques ont été mis en forme en utilisant Fig Tree (version 1.3.1, 2009).

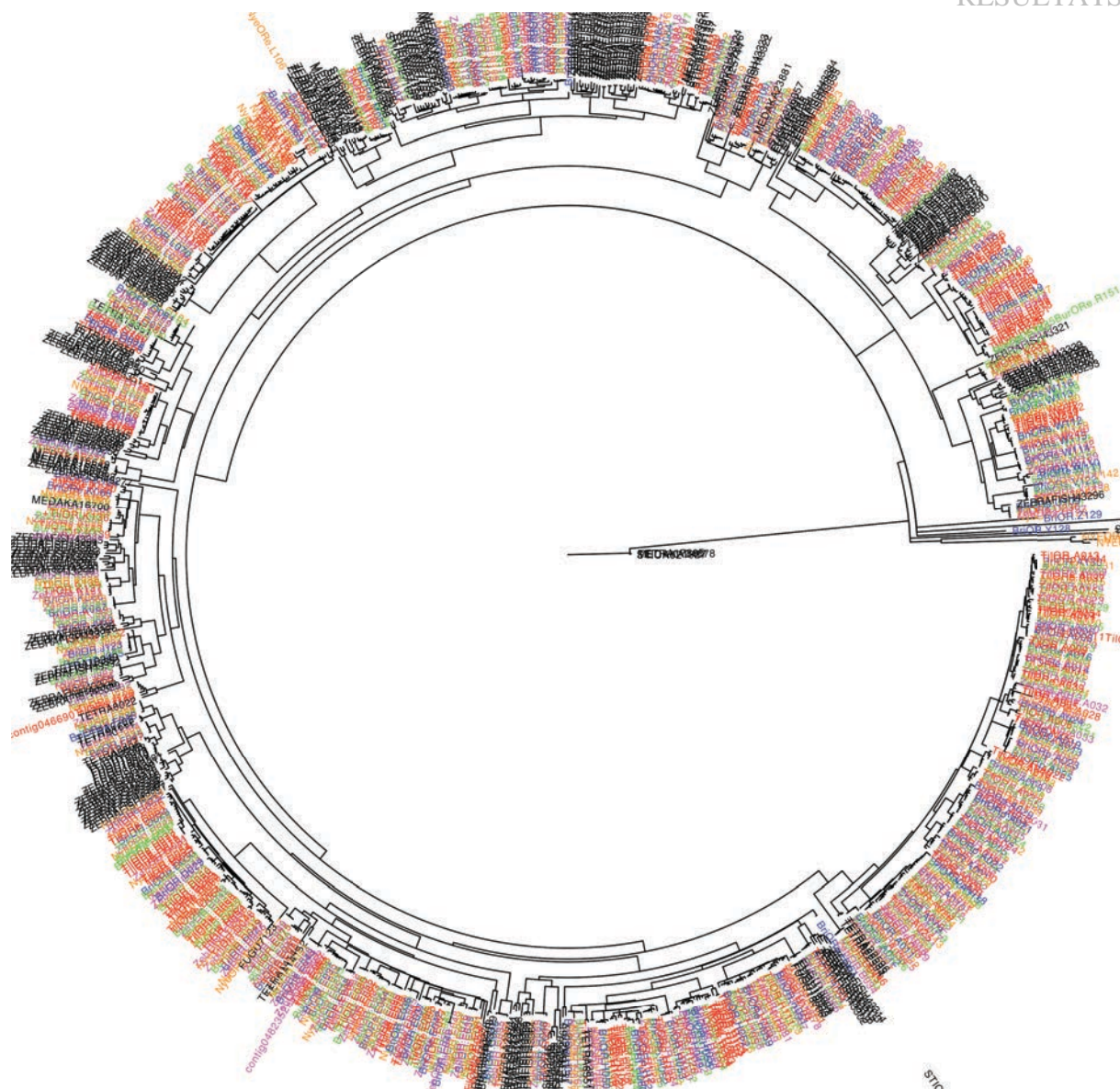


Figure 26 : Arbre phylogénique des OR des 5 cichlidés et des poissons modèles (poisson zèbre, fugu, tetraodon, medaka et stickelback dessiné avec le code couleur suivant : bleu *N. brichardi* ; rose *M. zebra* ; vert *H. burtoni* ; orange *P. nyererie* ; rouge *O. niloticus* et noir poissons modèles (figure 1 Article 4).

Afin d'identifier les familles et sous-familles, nous avons calculé le pourcentage d'identité protéique entre les séquences de la même branche avec le site EMBL-EBI (EMBOSS Needle - Pairwise Sequence Alignment) et un alignement multiple obtenu avec le programme CLUSTALW (Multiple Sequence Alignment by CLUSTALW) (tableau 5)

Tableau 5 : Répartition des familles et sous-familles OR pour chaque poisson (tableau 3 Article 4)

	<i>N.brichardi</i>	<i>N.burtoni</i>	<i>P.nyerelei</i>	<i>O.niloticus</i>	<i>M.zebra</i>	Poissons Modèles		
						Même sous-famille	Même famille	familles différentes
A1	3(e2,p1)	6(e2,p2)	6(e2,p1)	14(e8,p1)	8(e1,p1,f1)		3	
A2	(e2,p2)	2	2	2	2			
A3	1	2(e1)	3	2(e1)	2(e1)			
A4	4(p2)	4(e1)	4(p1)	4(e3)	5(p1)			
A5	3(e1,1)	3(e2,p1)	5(e2)	4(e2)	6(e2)			
A6	(e1)	1(p1)	1		1			
B1		1	1	1(e1)	1			
C1	1	1	1	(e1)	1		1	
D1	3(e2)	4(e6)	7(e3,p3)	11(e14)	7(p2)		1	
E1	7(e3)	12(e1,p1)	8(e2,p1)	13(e5,p2)	7(e7,p1)		5	
E2	(e1)	2	2	3	2			
F1	1	1	(p1)	1	1	8	12	
F2	(p1)	1	1	1	1	3		
G1	1	1	1	1	1		13	
H1	1	1	1	1	1		15	
H2	2(e2)	5	6	6(e3)	3(e1,p1)			
H3	2(e1)	(e3)		4(e4)	4			
H4	2(e2,p1,f1)	3(e2,p1)	4(e1,p1)	10(e1,p2)	4(e1)			
I1		1	1	3	1	1		
J1	1	(e1)	1	1	1		6	
J2	1(e1)	(e2)	1(e1)	(e1)	1	1		
J3	1	1	1	1	1	2		
K1	1	1	1	1	1		12	
K2	1	1(e1)	2	2	1(e1)			
K3	(e1)	1(e3)	2(e1,p1)	5(e2)	2(p1)	1		
K4	(e1)							
L1	5(e2,f1)	4(e7,p1)	5(e4,p1)	16(e15,p1)	6(e2)		12	
L2			(e1)					
M1	(e1)	(e3)	2	(e2)	1	9	15	
N1				(e2)			15	
N2	3	2(p1)	2(e1)	7	1(e3)	5		
N3	2(e2)	2(e2)	2(e1)	1(e3)	2(p1)	17		
N4	1(e2,p2)	1(e4,p1)	2(p1)	3(e4)	6(p1)			
N5	(e1)							
O1	2	1(e1)	2(e1)	2(e5)	1(e4)	1	13	
O2	1(e1)	(e1)		(e4)	1	3		
O3	1(e1)	(e3)	1(e1)	1(e1)	1	1		
P1	1	1	(e1)	1	1	1	4	
P2	1	1	(e1)	3(e1)	1			
P3	1	1	1	1	1	7		
Q1	1	1	1	(p1)	1		9	
R1		1(e1)	1		1			
R2	1			(e3)				
R3	(p1)	1		2	(p1)			
R4	(e2)	3(e1)	(e2)	7(e8)	3			
S1	2	2(e1)	1(e1)	4(e2)	1(e2)			
S2	1(e3,p1)	3(e1)	1(e2)	11(e1)	3(e2,p1)			
T1		1	(p1)				8	

U1	1	1	1	1	1	1		
V1	1	1	(e2)	1	1			
W1	1	1	1		1			
W2	1(e1)	1		(e1)	1			
W3	2	2	2	4(e2)	1			
W4	1	1	1	1	1			
W5	1	2						
X1		1(e1)	(e1)	(e2)	(e1)		1	
Y1	1							
Z1	1							
AB1	(f1)	1		1	(p1)		5	
AC1			(e3)					
total	121	151	134	267	143	61	150	16
details	69(e36,p13,f3)	90(e52,p9,f0)	88(e34,p12,f0)	158(e102,p7,f0)	102(e28,p12,f1)			
S/F	52	52	48	49	50			
F	23	24	25	23	24			

De ces analyses il ressort que les répertoires des 5 cichlidés sont très proches. En effet les récepteurs olfactifs des différents cichlidés se regroupent dans les mêmes familles et sous-familles. Par ailleurs la répartition des récepteurs du poisson zèbre montre que 13 familles du poisson zèbre sur les 25 qui constituent le répertoire sont communes à celles des cichlidés et que seulement deux sous-familles des cichlidés comptent des membres du répertoire du poisson zèbre.

2- Calcul du rapport dN/dS et analyse des résultats du répertoire OR

Pour déterminer le type de sélection qui s'exerce sur les séquences codant les gènes OR nous avons calculé le rapport dN/dS pour des couples de protéines appartenant aux mêmes sous-familles. Dans ce rapport, le nombre de mutations non silencieuses rapporté au nombre de sites non synonymes est appelé dN et le nombre de mutations silencieuses rapporté au nombre de sites silencieux est appelé dS. Pour le calcul du ratio dN/dS nous avons utilisé la méthode Nei-Gojobori modifiée par Zhang (Zhang *et al* 1998 PNAS).

Les différentes étapes du calcul ont été commandées par un script et l'ensemble des scripts a été regroupé au sein d'un pipeline bioinformatique.

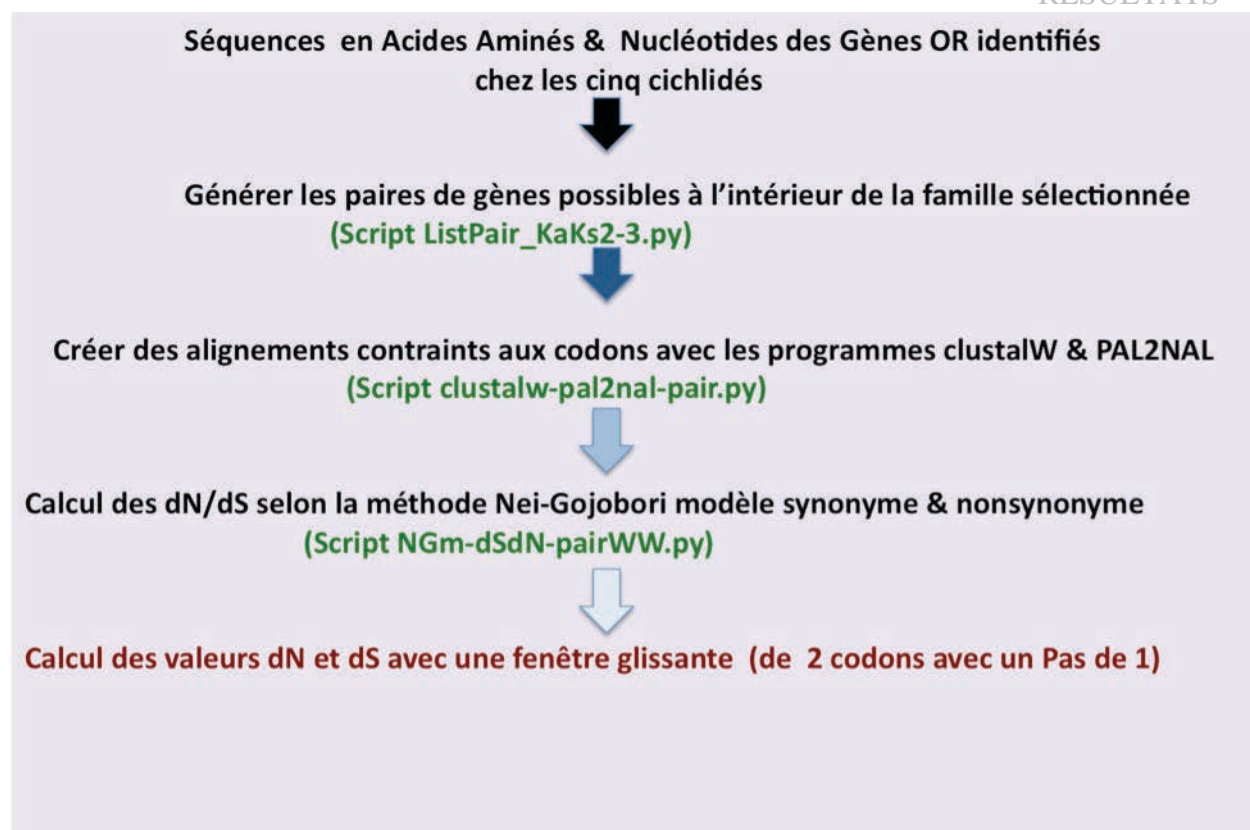


Figure 27 : Diagramme du pipeline de calcul des valeurs dN et dS.

Les résultats des calculs obtenus pour chaque paire de gènes des différentes familles ont été moyennés et sont rassemblés dans le tableau 6.

Tableau 6 : Les résultats des moyennes de calcul dN/dS obtenus pour chaque paire de gènes des différentes familles (tableau 6 Article 4)

Famille	Nombre de sous-familles	Nombre de gènes	Moyenne	Min.	Max.
Fam A	6	100	0,40 ± 0,09	0,00	> 10
Fam D	1	32	0,44 ± 0,10	0,15	1,30
Fam E	2	56	0,40 ± 0,11	0,10	2,27
Fam G	1	5	0,28 ± 0,10	0,18	0,41
Fam H	4	50	0,41 ± 0,14	0,00	1,14
Fam I	1	6	0,43 ± 0,29	0,00	> 10
Fam K	4	22	0,29 ± 0,18	0,12	> 10
Fam L	2	36	0,50 ± 0,12	0,04	1,20
Fam N	5	37	0,39 ± 0,14	0,18	1,79
Fam O	3	14	0,44 ± 0,10	0,12	0,83
Fam P	3	15	0,37 ± 0,10	0,07	0,86
Fam R	4	20	0,43 ± 0,08	0,19	0,88
Fam S	2	26	0,39 ± 0,09	0,14	1,19
Fam W	5	24	0,32 ± 0,13	0,00	1,48

La moyenne de chaque famille est comprise entre 0,28 et 0,50 avec une moyenne générale de la valeur du rapport dN/dS égale à 0.40. Cette valeur est notablement plus élevée que celle calculée par Makalowski et Bogulski (Proc Natl Acad Sci USA 1998) et égale à 0.11 pour 1880 paires de gènes orthologues homme et rongeur. Ainsi, une sélection positive ou au moins neutre doit être envisagée pour l'évolution de la famille des gènes OR à l'image de ce qui a été postulée pour les récepteurs de mammifères (Gilad *et al.*, 2000 ; Zhang *et al.*, 2004 ; Robin *et al.*, 2009).

Afin d'apprécier dans quelle mesure les différentes parties des récepteurs OR étaient la cible de mutations silencieuses ou non silencieuses, nous avons dans un deuxième temps procédé aux mêmes calculs mais en imposant une fenêtre glissante d'analyse d'un codon. Dans les panneaux suivants sont indiquées les positions des mutations non silencieuses (en rouge) et silencieuses (en bleu) ainsi que la position de celles-ci par rapport à la structure 2D de ce même couple de récepteurs

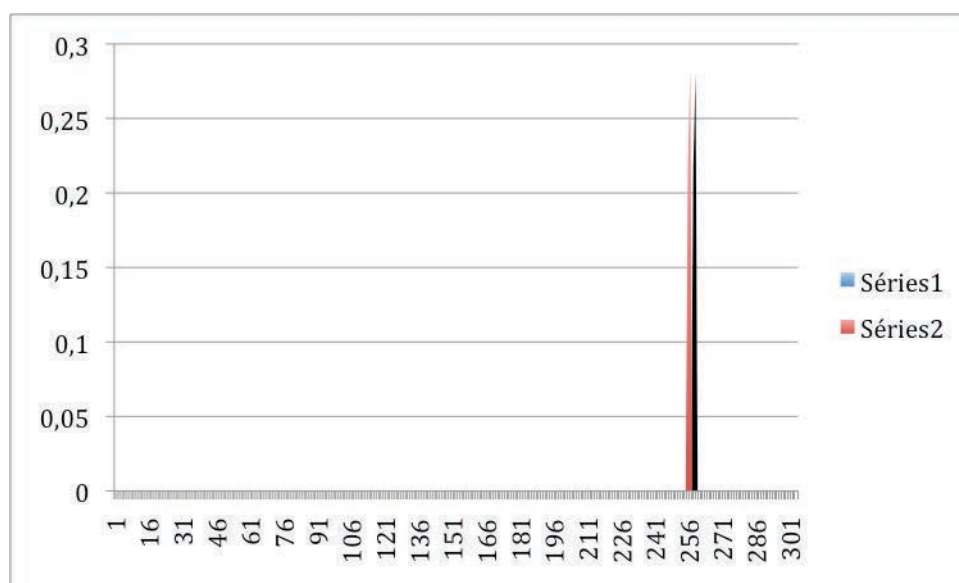


Figure 28 : ce graphe représente les positions des mutations non silencieuses (en rouge) du couple NYEsplice03/ZEBsplice05.

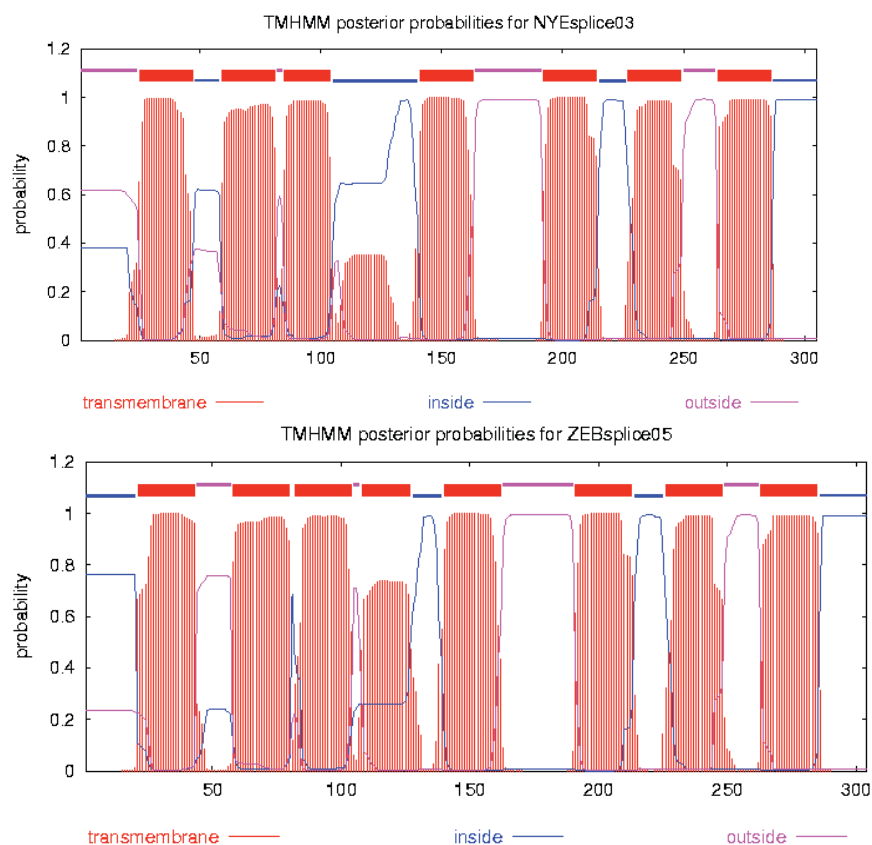


Figure 29 : schéma présentant la structure 2D des récepteurs du couple NYEsplice03/ZEBsplice05.

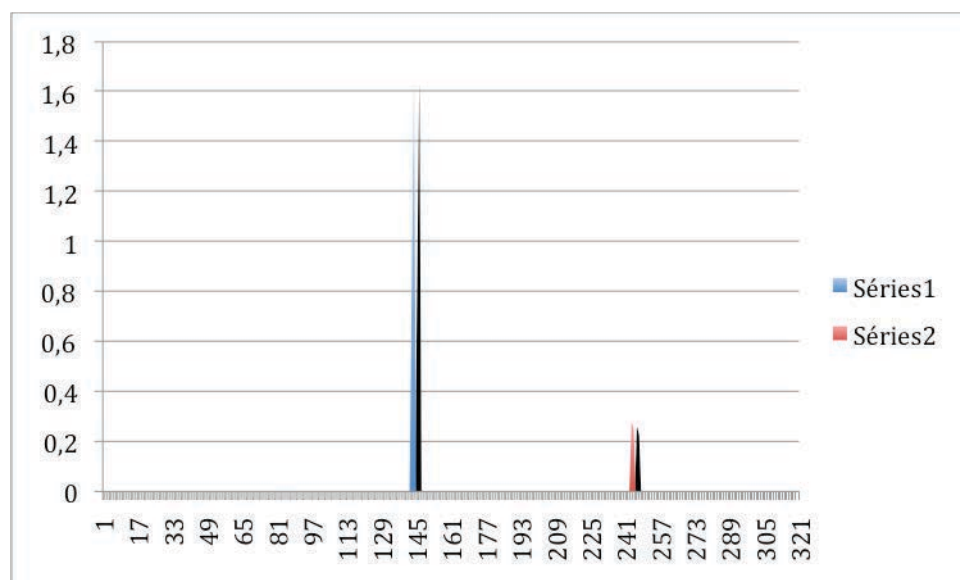


Figure 30 : ce graphe présente les positions des mutations non silencieuses (en rouge) et silencieuses (en bleu) du couple NYE066/BUR78.

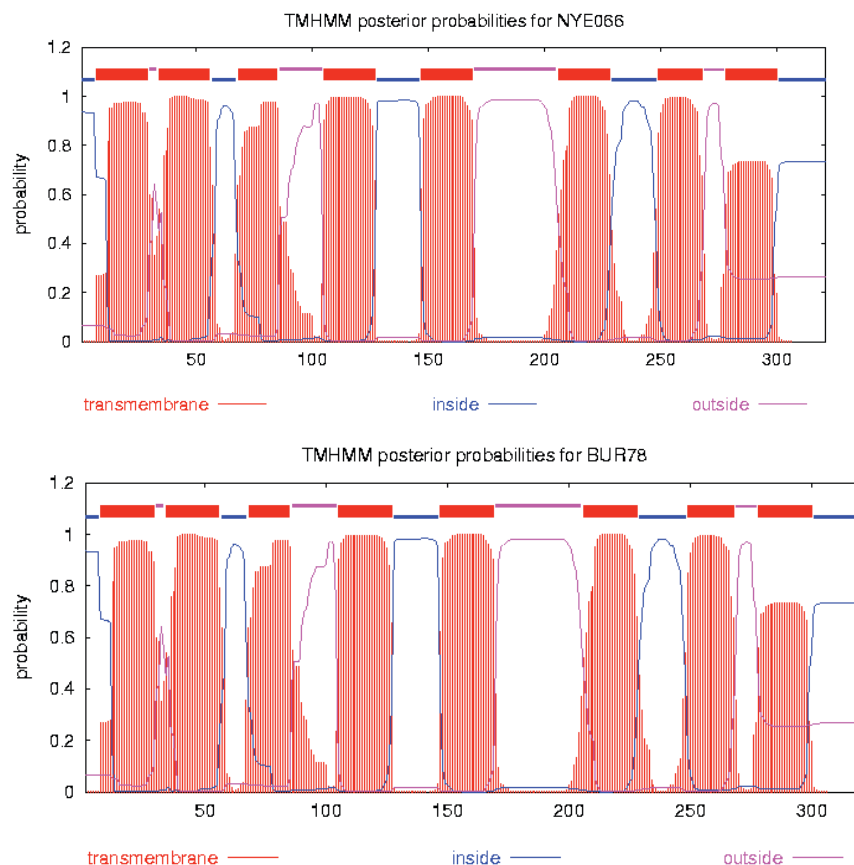


Figure 31: ce schéma présente la structure 2D des récepteurs NYE066/BUR78.

Sur 14 paires de gènes OR analysés on a trouvé des valeurs dS et dN réparties sur les différents domaines de la structure 2D (tableau 7).

Tableau 7 : Analyse des valeurs dN et dS des différents domaines protéiques de 14 paires de récepteurs.

structure 2D	dS	dN
OUTSIDE	4	0
TM1	4	4
IL1	0	0
TM2	2	2
EL1	2	1
TM3	0	3
IL2	5	1
TM4	1	0
EL2	4	3
TM5	0	0
IL3	3	1
TM6	1	1
EL3	2	1
TM7	2	4
INSIDE	0	0
Moyenne	2,00+/-1,69	1,40+/-1,45

2-3- Les motifs conservés des séquences protéiques

Les OR qui appartiennent à la famille A, celle de la rhodopsine, se caractérisent et se distinguent des autres membres de cette famille par un certain nombre de motifs protéiques que l'on retrouve chez tous les OR à des positions bien caractérisées. Le plus emblématique de ces motifs est le MAYDRY. C'est en s'appuyant sur l'existence partagée de ces motifs que les répertoires des gènes OR de chien et de rat ont été identifiés (Quignon *et al.*, 2005). En utilisant le logiciel MEME nous avons recherché les cinq meilleurs motifs caractéristiques des cichlidés et comparé ceux-ci à ceux du poisson Zèbre (*Danio rerio*) :

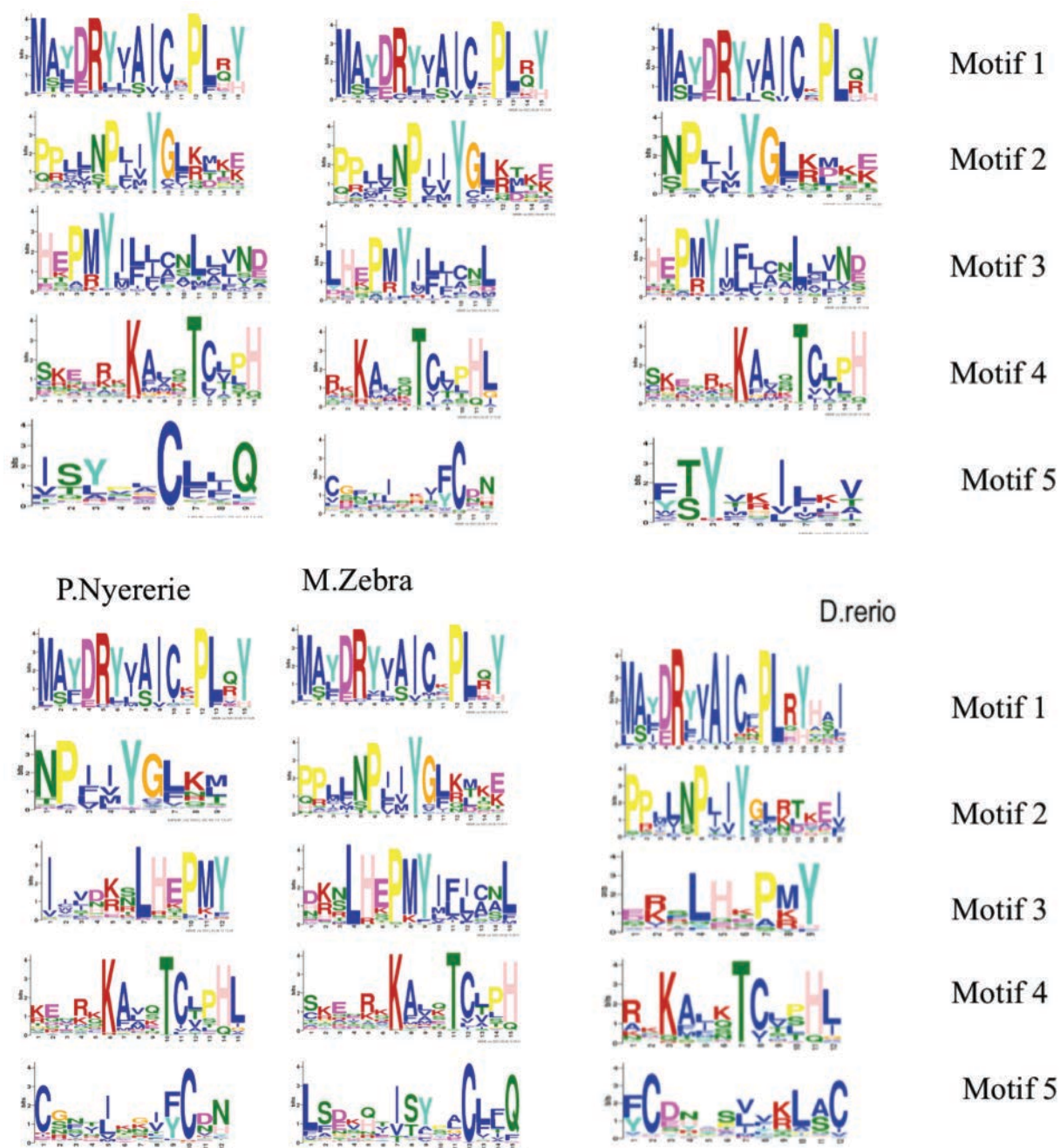


Figure 32 : Les cinq meilleurs motifs caractéristiques des 5 cichlidés et de *Danio rerio* (figure 2 Article 4).

Les logos des motifs dessinés avec le site MEME Suite, montrent que seuls les quatre premiers motifs sont conservés entre les cichlidés et le poisson zèbre.

Les quatre premiers motifs occupent des positions comparables pour les cinq cichlidés par rapport à la position des régions transmembranaires avec le motif 1 sur TM3_IL2, le motif 2 sur TM7_Inside, le motif 3 sur IL2_TM4 et pour le motif 4 sur IL3_TM6. En revanche le motif 5 n'est pas commun aux différents poissons et il a un positionnement variable. Il n'est donc pas pertinent. Ces observations sont semblables avec celle de la littérature (Liu *et al.*, 2003 ; Mombaerts 2004).

4- Les gènes épissés

Depuis leur découverte on a considéré que les gènes OR des vertébrés n'étaient constitués que de deux exons dont seul le deuxième était codant. Pourtant au cours de notre étude nous avons identifié plusieurs gènes formés de un à quatre exons codants comme indiqué sur le Tableau 8.

Tableau 8 : Décompte des gènes formés d'un et plusieurs exons codants.

	<i>M. zebra</i>	<i>O. niloticus</i>	<i>P. nyererie</i>	<i>H. burtoni</i>	<i>N. brichardi</i>
# OR avec 1 exon codant	94	146	81	78	62
# OR avec ≥ 2 exons codants	8	12	7	12	7

L'analyse des sites de reconnaissance des limites exon/intron, ainsi que la position de ces sites dans la séquence des gènes sont des arguments forts sur la réalité de ces gènes, dont toutefois la fonctionnalité ne pourra définitivement être démontrée que par l'analyse du transcriptome de l'épithélium olfactif, comme d'ailleurs pour tous les gènes.

4-1- Alignement de séquences de gènes multi-exons sur des séquences génomiques

- Exemple d'un gène OR avec deux exons codants :

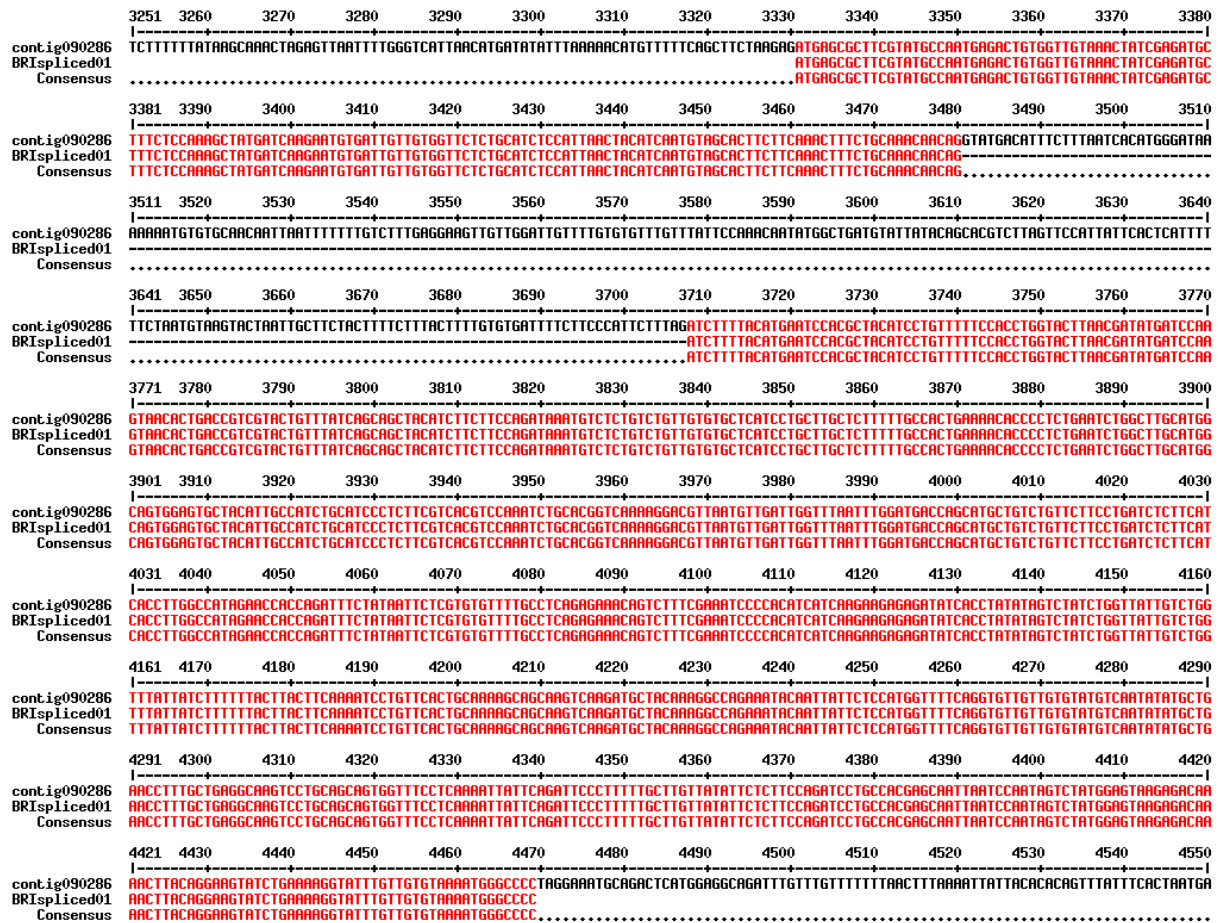


Figure 33 : Alignement de la séquence du gène BRISpiced01 sur son contig.

Dans cette figure la partie codante du gène BRISpiced01 est alignée sur la séquence du contig qui contient ce gène. Les positions 3480 et 3708 signalent les bornes de l'intron interrompant la phase de lecture.

L'alignement de la partie codante du gène *Tilsplice01* sur le contig 046690 délimite les bornes des deux introns interrompant la partie codante du gène.

- Exemple d'un gène OR constitué de 4 exons codants :

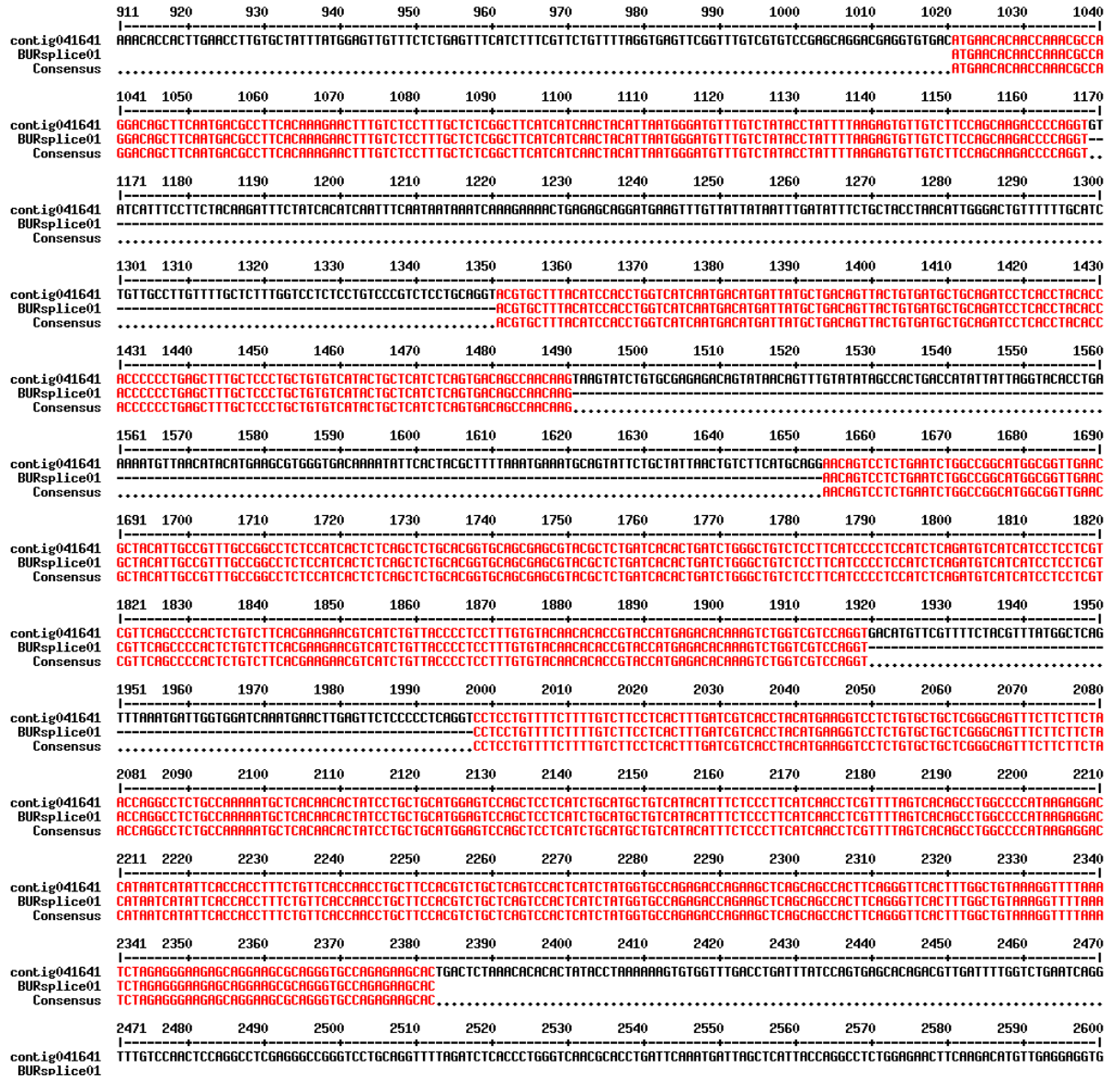
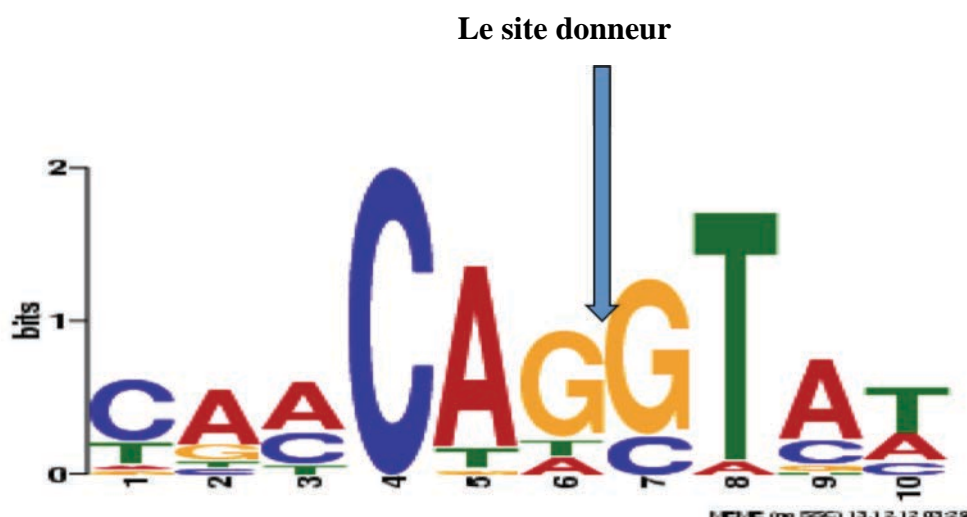


Figure 35 : Alignement de la séquence du gène *BURsplice01* sur son contig. Dans cette figure l'alignement de la partie codante du gène *BURsplice01* sur le contig délimite les bornes des trois introns interrompant la partie codante du gène.

4-2- L'analyse des sites d'épissages des gènes OR

Les séquences à cheval sur les sites donneurs d'une part et sur les sites accepteurs d'autre part ont été comparées avec le logiciel Mème Suite.



La limite exon/intron se situe entre les deux G de la séquence CAGGT.

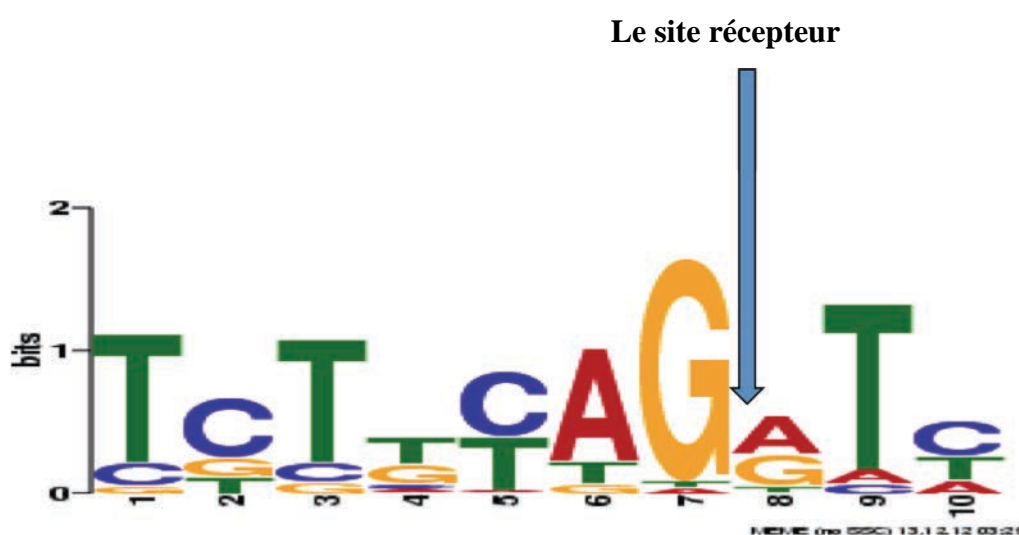


Figure 36 : Motifs de sites donneurs et récepteurs des gènes OR épissés des 5 cichlidès (figure 4 Article 4)

La limite intron /exon se situe entre le G et le A de la séquence pyAGpuT.

Le tableau 9 montre les positions des sites donneurs/accepteurs par rapport à la structure 2D de l'OR, ainsi que le dernier codon et sa position par rapport à l'ATG

Tableau 9 : Positions des sites donneurs/accepteurs par rapport à la structure 2D des OR (tableau 7 Article 4).

NOMBRE CODANTS	D'EXONS	nom des OR	dernier codon	N° du codon	Position sur la structure transmembranaire	
2		contig034988-NyeORs.A033	GTC.AG	159	TM4	
		contig050024-NyeORs.W129	CAA.CAG	50	IN 1	
		contig050025-NyeORs.W131	AAC.AAG	50	IN 1	
		contig050025-NyeORs.W130	CAC.CAG	50	IN 1/TM2	
		contig050026-NyeORs.W132	CAC.CAG	52	IN 1	
		contig090286-BriORs.W112	CAA.CAG	50	IN 1	
		contig090288-BriORs.W113	CAA.CAG	50	IN 1	
		contig090291-BriORs.W114	AAC.AA	41	TM1	
		contig090292-BriORs.W115	CAG.CAG	52	IN 1	
		contig090301-BriORs.W116	TAT.CAG	49	IN 1	
		contig046002-ZebORs.K090	AAG.TAT	24	N ter	
		contig067811-ZebORs.W140	CAA.CAG	50	IN 1	
		contig062664-ZebORs.W141	AAA.CA	43	IN 1	
		contig025842-ZebORs.W142	AAA.CAC	51	IN 1	
		contig025841-ZebORs.W139	AGT.ATC	52	IN 1	
		contig045454-BurORs.W131	CAA.CAG	50	IN 1	
		contig066785-BurORs.W148	TAT.CAG	49	IN 1	
		contig045453-BurORs.W132	CAC.CAG	50	IN 1	
		contig045452-BurORs.W133	CAC.CAG	52	IN 1	
		contig045453-BurORs.W134	AAC.AAG	50	IN 1	
		contig041638-BurORs.W135	AAA.CA	43	IN 1	
		contig041640-BurORs.V144	CGA.CAC	59	IN 1	
		contig049605-BurORs.AB153	AAC.AGT	77	IN 1	
		contig046708-TilORs.K143	AAG.TAT	24	N ter	
		contig027203-TilORs.W238	CAC.CAG	50	IN 1	
		contig027204-TilORs.W239	AAC.CGG	50	IN 1	
		contig027206-TilORs.W240	CAC.CAG	50	IN 1	
		contig027209-TilORs.W241	CAC.CAG	50	IN 1	
		contig027202-TilORs.W243	AAA.CAC	51	IN 1	
		contig046717-TilORs.AB275	TAT.GTG	72	TM1	
		contig027194-TilORs.V262	CGA.CAC	59	IN 1	
	3		contig046495-NyeORs.I079	GAG.AGG	121	IN 2
				ACA.ATC	232	OUT3
			contig025847-ZebORs.U137	TAT.CAG	16	N ter
			CAC.CAG	54	IN 1	
		contig026932-ZebORs.I082	GAC.AG	125	IN 2	
			GAC.ATC	200	OUT2-TM5	
		contig048321-BurORs.I076	GAC.AG	125	IN 2	
			ATC.TAT	201	OUT2	
		contig046690-TilORs.I128	GAC.AG	120	IN 2	
			AAC.AT	194	OUT2-TM5	
		contig046694-TilORs.I129	GAC.AG	125	IN 2	
			ATC.TAT	201	OUT2-TM5	
		contig046695-TilORs.I130	GAC.AG	120	IN 2	
		ATC.TAT	196	OUT2-TM5		
	contig090301-BriORs.U109	TAT.CAG	16	N ter		
		CAG.GAT	56	OUT1		

	contig027194-TilORs.U236	TAT.CA	15	N ter
		CAC.CAG	54	IN 1
	contig041640-BurORs.U130	TAT.CA	15	N ter
		CAC.CAG	54	IN 1
	contig051999-NyeORs.U128	TAT.CA	15	N ter
		CAC.CAG	54	IN 1
4	contig090302-BriORs.V122	CAC.AG	80	IN 1
		CTT.CTG	127	OUT1
		GTG.CAG	269	TM6
	contig025847-ZebORs.V149	CAC.AG	62	IN 1
		CTT.CTG	109	OUT1
		GTG.CAG	251	TM6
	contig041641-BurORs.T129	CCC.AG	48	IN 1
		AAC.AAG	96	OUT1
		GTC.CAG	184	TM5

Ce tableau met clairement en évidence l'homogénéité de structure des différents gènes OR épissés, ce qui aurait peu de chance d'être le cas si ces gènes étaient le résultat d'artefacts de séquence ou d'analyse.

Article N° 4 à soumettre à BMC Genomics

The Cichlid Olfactory Receptor Genes
Naoual Azzouzi, Frederique Barloy-Hubler, Francis Galibert

The Cichlid Olfactory Receptor Genes

Naoual Azzouzi, Frederique Barloy-Hubler, Francis Galibert

Introduction

With more than 2000 species, the cichlid family is by far the largest fish family. Members of this family occupy all sorts of ecological niches everywhere in the world with a remarkable concentration of species in the great African lakes (1, 2). As such they represent a good model for evolution and adaptation studies in addition to the economical importance of tilapia, *Oreochromis niloticus*, which is second in aquaculture (3). This prompts the scientific community to undergo the complete sequencing of 5 cichlid genomes namely those of *O. niloticus*, *P. nyererei*, *H. burtoni*, *N. brichardi* and *M. zebra* (genome paper) to be published)

All animal species whatever their ecological niches have developed sophisticated systems to sense the outside world in order to prevent attack from predators, find food and select appropriate partners to mate and reproduce themselves. Volatile and soluble odorant molecules are targets of several of these biological systems among which the olfactory receptors (OR) were the first to be discovered (4). OR are G protein coupled receptors (GPCR) (5, 6). They are expressed at the cilia membrane of olfactory neurons (OSN) (7- 9) themselves embedded in the olfactory epithelium. OR constitute the largest gene family with some hundred members identified in the genome sequences of insects and up to around 1000 in mammals (10-12). According to their amino-acid sequence analogy, OR are classified into classes, families and sub-families (13). Mammal ORs are organized in two classes named I and II whereas fish ORs were shown to belong to class I only, also called fish –like class. (14, 15). This class I has not to be confounded with the GPCR A family or rhodopsine-like also named class I, to which belong the entire OR family. Given the importance in behavior played by the olfactory system, it is thought they have played an important role in shaping species evolution.

As such we were interested to identify the complete OR gene repertoires present in the genomes of five members of the cichlid family *O. niloticus*, *P. nyererei*, *H. burtoni*, *N. brichardi* and *M. zebra* that are themselves viewed as a good model for evolution studies and were recently sequenced (paper to be published)

Results.

Cichlid OR repertoires

A comprehensive search of the genome sequences of five fishes belonging to the *cichlid* family (*Oreochromis niloticus*, *Pundamilia nyererei*, *Haplochromis (Astatotilapia) burtoni*, *Neolaprogus brichardi*, *Mitriaclima zebra*) has been undertaken in order to identify their OR gene repertoires. First we retrieved from Genbank and Ensembl databases 891 non-redundant fish ORs to construct a query set that was used for a TBLASTN search on each of the cichlid genome sequence

determined by the BROAD Institute. This search performed with a cut-off of $1e-50$ identified 820 candidates OR genes distributed over 733 contigs. These candidate genes were checked by a TBLASTN against a set of 248 (Additional file 1) non OR GPCRs in order to eliminate false positive hits. Finally candidate genes were checked by TBLASTX against the protein fish database (NCBI, taxid : 7898)

In table 1 are displayed the number of genes identified in each of the five cichlid genomes as well as those of four fish models retrieved from the literature and GenBank. Their nucleotide and amino-acid (AA) sequences as well as their genome localization are given in supplementary materials (Additional files 2-4) In addition to the so-called complete and potentially functional genes, we identified a number of pseudogenes, edge genes and gene fragments. Pseudogenes are common to any olfactory repertoires (11-14). OR pseudogenes are not retrogenes they arose from gene duplication and their elevated proportion in vertebrate genomes is thought to result from a balance between gene duplication and nucleotide misincorporation during the course of replication and responsible of a nonsense or frameshift mutations altering the coding frame. As compared to other known fish olfactory repertoires they appear less numerous in the cichlid family (16, 17). Most of these 52 pseudogenes have derived from frameshift mutations, whereas 20 pseudogenes are due to nucleotide misincorporation during replication changing a sense codon into a stop codon (table 2).

	Tilapia	Zebra	Nyererei	Brichardi	Burtoni
Frameshift	4	8	9	5	6
Stop in frame	3	4	3	7	3

Table 2

The distribution of pseudogenes appears largely random affecting many different sub-families (21 out of 60 see table 3) whatever their size in the different cichlids although some rather large sub-families like R4, O1 and H3 with 26, 19 and 18 genes respectively are devoid of pseudogene or N2 with only one pseudogene out of 20 genes.

Fragment genes are sequences with high percentage of analogy to a restricted part of a functional gene. They have been identified in many complete genome sequences (18,19). Their significance if any is unknown, they may correspond to "dead" genes or more prosaically result from sequencing problems. Although it is difficult to depart between these two possibilities, it appeared from the phylogenetic analysis (Fig 1) that some of these gene fragments may be due to sequencing errors although most of them probably correspond to remains of ancient genes heavily pseudogenized. During this search we have also identified another type of gene fragments. These fragments are located at the edges of contigs and correspond to either the 5' or 3' end of an OR. As such they potentially corresponded to actual OR genes whose sequences were interrupted by the genome fragmentation into many and short contigs. Would these genome sequences be completed these edge genes would become complete genes, pseudogenes or gene fragments.

We used MAFFT (20) and PHYML (21) to align the OR AA sequences and construct a phylogenetic tree with the OR repertoires of the five cichlids and 227 OR amino-acid sequences identified in the genome of zebra fish (*Danio rerio*), medaka (*Oryzias latipes*), stickleback (*Gasterosteus aculeatus*), takifugu (*Takifugu rubripes*) and tetraodon (*Tetraodon nigroviridis*) (Additional file 5). OR repertoires are usually clustered in classes, families and subfamilies according to the percentage of AA

identities shared by the different OR (13). In this study, we used the same 40 and 60 % AA identity thresholds and identified from 23 to 25 families and between 48 and 52 sub-families per cichlid (Fig 1, Table 3, Additional file 4.) From this data, it appears that all the large sub-families are shared by the five cichlid fishes. It is also worth noting 7 species-specific sub-families, each with one member only (K4, N5, Y1 for *N. brichardi*; L2, AC1 for *P. nyererei* and N1 for *O. niloticus*) and 2 sub-families shared by two cichlids only out of a total of 59 sub-families. As indicated in Fig 1 and Table 3, out of the 820 cichlid and 227 model fish AA sequences that make the tree, only 147 cichlid and 61 model fish sequences share 15 sub-families. This low level of sub-family overlapping between the cichlid and the fish model is suggesting of a rather high level of divergence between the cichlid and the other teleost repertoires.

Of interest regarding the evolution of Nile tilapia and lake cichlids is the existence of a number of pairs or triplets and even quadruplets of orthologous genes sharing >99 % amino-acid sequence identity (Table 4 and Additional file 6. As compared to this we found less paralogous OR pairs except for tilapia for which we identified 7 pairs, 1 triplet and 2 quadruplets of paralogous genes with an AA identity above 99%. This last finding is in light with the larger number of OR present in this species. This observation recalls that recently made by Nikaido et al (22) who analyzing the expansion of vomeronasal type 2 receptor-like (OlfC) genes in cichlids noted that recently duplicated paralogs are more variable than orthologs. Accordingly but not necessarily we found that these paralogous genes share more than 99% of identity at the nucleotide level.

Evolution dN/dS ratio

The dN/dS ratio also named Ka/Ks is a common metrics to measure the selective pressure exerted on genes during evolution. We used the method of Nei-Gojobori modified by Zhang (23) to calculate this ratio for each pair of OR genes from the 14 cichlid OR families with four or more genes. The dN/dS mean values for these families extend from 0.28 for family G made of 1 sub-family to 0.50 for family L made of two sub-family (Table 5 panel a). These values are clearly above the 0.11 mean value calculated for 1880 human rodent orthologous gene pairs (24) and similar to that obtained for medaka and stickleback OR (17). Although below 1, the theoretical limit between negative and positive evolution trends, they indicates a clear tendency for a positive selection favoring the OR repertoire diversification as previously noted for other fishes (16, 25) and mammals (26, 27). However it is important to note that the different OR pairs behaved very differently. As shown in Table 5 panel a and additional file 7 we identified a number of OR gene pairs with only synonymous mutations as, in families A, H and W and OR pairs with only non-synonymous mutations as in families A, I and K. Table 5 panel b are given the number of OR pairs with dN/dS ratio above 1. Interestingly, intra-species dN/dS ratios (paralogous comparison) have similar values as inter-species values (orthologous comparisons) as indicated by a ratio close to 1 suggesting a similar evolution of the five cichlid OR repertoires (Table 5, panel c).

We also calculated the dN/dS ratio of the different OR protein domains (TM regions, internal and external loops) for five families (D, E, H, L and N) selected for their high number of genes. As shown in Table 6 and additional file 8 a to e the dN/dS values for the TM regions have a tendency to be higher than that of the entire molecules, although this is not always the case (see TM 6 and TM7 of family E) and TM4 to have the highest values except in family E again. On another hand no clear tendency can be drawn out for the dN/dS ratio of the internal and external loops, although one should note that the standard deviations are in all cases very high, indicating that the different OR pairs behaved differently (17)

Conserved Amino-acid motifs

OR are GPCR belonging to the rodhopsin sub-family. They are characterized by a number of amino-acid patterns of which the MAYDRY located in the internal loop 1 is the most characteristic. Often these patterns have been used as criteria when mining whole genome sequences for OR identification (11). We used the program MEME (28) to search the five best motifs for each of the five cichlids and *D. rerio*. Apart from very minute variations the four first motifs are identical in all fishes with a similar location within the 2D structure (Fig 2). In spite of the long time that elapsed since fish divergence, the motifs identified here are not very different from those identify in OR mammals (12 ,29 -31). However, no clear fifth pattern was identified in the fish OR. OR are trans-membrane proteins. Although no signal peptide has been identified in their sequences using Polyphobius (32), an N-glycosylation site, Asn N Ser/Thr has been detected in all OR identified to date (33). OR protein sequences identified in the five cichlid genomes were inspected for the presence of such putative glycosylation sites using NetNGlycserver (34). For each of them, one to several putative sites were proposed (additional file 9). But interestingly in all cases a site with a very high score was present at close proximity to the extracellular N-terminal part in agreement with the importance of the glycosylation site of this region for the correct expression of the OR and membrane trafficking (35). OR proteins are also characterized by the presence of cysteine residues located at fixe positions in particular in the extracellular loop 1 and 2 (EC1 and EC2) regions as observed for all *D. rerio* (33) and mouse OR genes (29, 36). We observed a similar situation for all complete cichlid receptors identified in this study. But, we noted the existence of two subgroups of OR. One sub-group of OR with one cysteine residue only in EC2 and one sub-group with three cysteine residues. Interestingly, these two groups have slightly different MAYDRY motifs with an E replacing D in the sub-group with one cysteine (Fig. 3 and additional file 10a and b). These two groups differ also by motif 2 located in the cytoplasmic C terminal extremity. Whether these differences affect the recognition and binding of the G alpha subunit and the transduction signal is a matter of interest (37).

In addition to these features inspection of the AA alignments (additional file 10a and b) allow to identify a number of positions occupied by the same or nearly the same AA pointing to positions submitted to strong purifying constraint during evolution as for example another cysteine residue in the N terminal extension also noted previously (38).

Spliced OR genes.

During the mining of the fish genome sequences, among the complete OR genes we unexpectedly identified a number of spliced OR. Out of the 507 complete OR identified in the five cichlid repertoires 45 OR (9%) would be made of two to four coding exons (Table 1). Although the definitive proof that cichlid OR spliced genes are functional will be given by transcriptional analysis of olfactory epithelium and by the identification of their binding odorant properties, four arguments already strongly suggest they are actual OR genes : the splice site junctions, the intron position, the BLASTX analysis results and their position within the phylogenetic tree.

(a) The first argument comes from the analysis of the exon-intron boundaries. Amino-acid sequences overlapping the exon-intron boundaries identified by the alignment of the spliced OR sequences onto their cognate contig sequences were analyzed with the MEME suite to identify the presence of any nucleotide motifs (Fig 4). As shown in Fig 4, two nucleotide motifs, one at the donor site and one at the acceptor sites could

be identified. The strength of these two motifs is attested by the respective size of the letter. Importantly by delimiting the positions of the exon boundaries as indicated by the two arrows one maintains the reading frames open through the junction of the adjacent exons. Of note these two motifs appeared very much similar to those found at mammal gene exon/intron boundaries (39). Additionally, most of these donors/acceptors sites were predicted using the FSPLICE program (40) and the FISH model weight matrix (data not shown).

(b) By comparing the gene nucleotide sequences and the amino acid sequences, we inferred the intron positions relatively to the 2D OR structure and noticed a nearly fixed position (Table 7). Out of the 31 OR genes with one intron interrupting the coding frame, for 27 of them, the intron does not interrupt the last codon of the first coding exon and for 26 the intron is located within the sequence coding the first internal loop within the MAYDRY motif. Similarly out the 11 OR genes with two introns, 5 have their first intron within the sequence encoding the extracellular part and their second intron splicing out the external loop 2, whereas 5 genes have their first intron splicing out the internal loop 2 and their second intron the external loop 2. These observations do not prove these genes are actual OR genes, but at minimum they prove these genes are actual genes and not sequencing or mining artifacts.

(c) The third argument comes from the BLASTX analysis results. As indicated above, all proteins identified by TBLASTN search were submitted to a BLASTX search against the non-redundant NCBI protein database and only proteins giving a strong hit with OR proteins and no hit or meaningless hit with other GPCR were kept as actual OR.

(d) A fourth argument arose from a phylogenetic analysis. We aligned with MAFFT the amino-acid sequences of 507 cichlid OR and of 279 non OR class A (Additional file 1) GPCR identified in GenBank and constructed a tree with PHYML. As indicated in Figure 5, this phylogenetic tree drawn with FigTree (41) shows that all OR whatever their number of coding exons formed a branch clearly separated from the non OR class A GPCR, enhancing the fact they are actual OR genes and cannot be confounded with another type of GPCR. Additionally a more closely look at the tree showed that most, 39 out of the 45 cichlid spliced OR, are grouped into families or sub-families not shared with one coding exon OR and that spliced OR genes are grouped according to the number of their coding exons. For example family W is made of 24 OR with two coding exons whereas families I and U contain six and five OR respectively with three coding exons (table 8). The grouping of the OR with 2 or more coding exons into specific families indicates that splice OR genes arose before cichlid species separation and have evolved independently from the one coding exon OR although horizontal transfer cannot be formerly discarded.

Conclusion

The olfactory system, which allows animals to sense the outside world is of prime importance for the search of food and sexual partners. It is also a crucial allied in the prevention of attack by enemies and other species themselves in search of food. As such olfaction is a vital function. Olfactory receptors (OR) are transmembrane proteins expressed by olfactory neurons, where they are the first component of the olfactory transduction cascade. They bind odorant molecules inducing a cascade of protein interactions that transformed a primary chemical signal into an electrical signal that is conveyed to the brain, which decoded, stored it and ultimately sends an appropriate response (42-44). In this study, we identified the near complete repertoires of OR genes encoded in the genomes of five cichlids, which were recently

sequenced by a consortium lead by the BROAD institute because of their importance for aquaculture in the case of tilapia (*O. niloticus*) or for evolutionary studies (Ref genome paper).

In order to identify as completely as possible these five olfactory repertoires, we followed the same strategy as that followed by Chen et al (17) for the identification of the medaka and stickleback repertoires or Alioto and Ngai (33) in the mining of the OR zebrafish repertoire. We performed an exhaustive TBLASTN search with a query made of a very large set (> 800) of fish olfactory receptors retrieved from GenBank. Although without proven exception the vast majority of OR from vertebrates are encoded by an uninterrupted open reading frame (ORF) (10-13) we did not limit our search to positive hits longer than 700 nucleotides as Alioto and Ngai did (33) but retrieved all hits with an e-value cut-off of $1e^{-50}$. Then we checked each of these candidate genes against a set of non-OR non-TAAR class A GPCR. All genes retained as actual ORs shared a number of properties characterizing this class of molecules (4). They all have a 2D structure defined by an extracellular N terminal extremity, 7 hydrophobic transmembrane paths of 21 to 26 amino-acid each and an intracellular C extremity (Additional file 11). The N terminal ends are decorated by an N-glycosylation site, Asn N Ser/Thr, the implication of which for addressing these proteins to the cellular membrane has been demonstrated (35). Characteristic motifs of amino-acids have been identified in all OR. These patterns of AA as well as their localization inside the molecules have often been used as a mean for retrieving the corresponding genes from newly determined genome sequences (12). Here, starting with a different perspective that consisted in the characterization of these proteins as actual OR and not in their mining, we identified with MEME, four AA motifs, shared by the five cichlids as well as *D. rerio*. Although, minute variations can be observed when comparing the different motives between these fishes, they looked very much the same. Interestingly, some striking similarities can be observed with the AA motives characterizing the rat and dog OR (12). Finally, we noted the presence of cysteine residues at positions shared by all OR (Additional file 10a and 10b). These residues are of prime importance for the correct folding of proteins and mutations changing these cysteine into another AA have been shown to often impair their function (45, 46). Additionally these alignments allow the identification of regions or amino-acid positions strongly conserved and other highly variable as in the ligand binding pocket. (47-49). The birth and death hypothesis as previously described (50, 51) with a relaxed purifying selection favoring multiple amino-acid changes explain the large repertoire of OR found in numerous species (10 -13) as well as their partition in many families and subfamilies. Their multiplicity, which allows the detection of a large number of odorant molecules favors the search of food or sexual partners and as such is a good witness of evolution. Considering the size of the OR repertoires, even in the absence of formal and definitive numbers due to the non-completeness of the genome sequences, it seems that the numbers of potentially active OR genes found in these five cichlids would be higher to that identified in other teleosts (16, 17, 33), with in addition less pseudogenes identified in the cichlid genomes. But perhaps more importantly for an evolutionary point of view, the OR cichlid repertoires are dispersed in many more families, 23 to 25 depending upon the species than the other teleost repertoires for which no more than 6 to 8 families have been registered (33). In the light of this higher variability in OR sequences, it is important to oppose the surprising high number of pairs of ortholog genes with nucleotide and AA sequences over 99% of identity (Table 4) suggesting that the evolution of the cichlid OR repertoires predated species radiation. This rather high number of families and sub-families on another hand are in agreement with the elevated dN/dS mean ratio.

The discovery in the cichlid genomes of a subset of OR genes made of more than one coding exon came as a surprise, as so far it was assumed based on their original discovery that vertebrate OR genes were made of two exons, a 5' non coding exon and a second exon compassing an ORF coding for a protein of 300-330 AA (4) beside the fact they belong to the class A of the GPCR superfamily within which numerous genes are made of several coding exons. It was due to this belief that several authors mining newly determined genome sequences restricted their search to long ORF or eliminated afterwards short ORF (12, 17, 33). Several arguments developed in the result section indicate these multi coding exon genes code actual OR. We wondered then whether they were specific of the cichlids or whether they were overlooked during the mining in the genome of other fishes. To answer this question, we performed a TBLASTN search of the OR fish repertoires present in NCBI and ENSEMBL databases with the set of cichlid OR with 2 or more coding exons as query. We also inspected one by one the AA and gene sequences of the medaka, stickleback and zebrafish OR genes found in ENSEMBL database. These analyses allowed the detection of a number of OR belonging to different fish species and encoded by two or more coding exons (Additional file 12). These analyses would need to be repeated in a more systematic way and supported by functional studies, but at this point they strongly suggest that fish OR at large can be encoded by multi coding exon genes and that this structural feature is not restricted to OR cichlid genes. Then given the fact that invertebrate (52, 53) and fish OR have or could have more than one coding exon, a more general question would be, why mammal OR would have only one coding exon and be the only sub-group of GPCR having this characteristic? Would the peptides, corresponding to one or a subset of exons that made multi coding exon OR genes, have an additional function lost during mammalian evolution and then leading to the lost of these OR genes is a matter of speculation?

Materials and Methods

The sequences of the five genomes were determined and obtained by the BROAD Institute (<http://www.broadinstitute.org/>). A positive dataset composed of 891 fish OR was used as bait for an exhaustive TBLASTN search (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Candidate regions were then compared to a negative data set made of 248 non OR and non TAAR animal GPCR. Both positive and negative datasets were designed using genome sequences accessible through the NCBI and ENSEMBL databases (Additional file 1).

TBLASTN results were filtered with a homemade python script so that OR candidate sequences retained the following rules: (1) one or more matches with the positive dataset and (2) no match with the negative dataset using an e-value cut-off of $1e^{-50}$. Selected candidates were re-checked using both BLASTX and BLASTP against the protein fish database (NCBI, taxid : 7898) using default parameters with the expect value cut-off of $1.e^{-100}$.

All genes were manually collected, biocurated and translated into protein sequences using Geneious software 6.1(54). Incomplete OR genes found at the boundaries of contigs were annotated as “edges” whereas incomplete OR genes

found inside contigs were considered as “fragments”. Genes with disruptive frameshifts or stop codons were annotated as pseudogenes. For spliced OR genes, predicted sequences and splice sites were manually corrected using both multiple alignment using MAFFT 7 (20) and FSPLICE (40). The list and sequences of complete cichlid OR genes (spliced or not spliced), pseudogenes, edges and fragments are available as supplementary information (Additional files 2 and 3)

Positions of transmembrane domains in selected OR predicted proteins were analyzed using both TMHMM (55) and Phobius (32) to check for the expected typical 7-transmembrane domain topology.

The deduced amino acid sequences of 820 cichlids ORs genes and 141 zebrafish, 27 stickelbach, 25 tetraodon, 18 fugu and 18 medaka OR were aligned using MAFFT 7 using the E-INS_I version (optimal for sequences with conserved motifs and carrying multiple domains) with default parameters (20). A classification was proposed based of relatedness estimated using a bootstrapped maximum-likelihood unrooted tree generated by PHYML (1000 rounds of bootstrapping) and drawn using FigTree 1.3.1. Thresholds of 40% and 60% amino acid similarity were used to distinguish between families and subfamilies, respectively, as described by Glusman et al (13). The names of the Cichlid OR sequences were created regarding their phylogenetic associations and were composed as followed: Fish_Symbol (Bri, Bur, Nye, Til or Zeb for *N. brichardi*, *H. burtoni*, *P. nyererei*, *O. niloticus* and *M. zebra* respectively) then “OR”, then p for pseudogene, e for edge or f for fragment. Finally a letter and a 3 digit number to designate the family and the gene itself. As an example, BRIORe.E041 indicates the 041 *H. burtoni* edge OR of family E.

Ratios of non-synonymous and synonymous nucleotide substitutions ($\omega=dN/dS$) were calculated with the method of Nei-Gojobori modified by Zhang et al 1998 (23) using perl and python scripts to automate the whole process. These ratios were calculated for both the entire proteins and sub-regions (i.e. individual transmembrane domains or loop regions).

Conserved motifs were identified in predicted OR protein sequences with the Multiple Expectation Maximization for Motif Elicitation (MEME) online program v.4.9.0. Potential N-glycosylation were predicted using with NetNGlycserver. Only N-glycosylation sites with a “potential” score > 0.5 and a jury agreement of “++” or higher) were considered as positive in our analyses.

Les tableaux 1, 3, 6 et 7 ainsi que les figures 1, 2 et 4 sont dans le résumé qui précède l'article à soumettre, les autres tableaux et figures sont donnés ci-dessous

Table 4 : Distribution of orthologous OR pairs, triplets and quadruplets compared with their paralogous counterparts. Olfactory receptors sharing 99% of AA identity were identified from the phylogenetic tree. In *O. niloticus* 7 pairs, 1 triplet and 2 quadruplets of paralogous genes were identified in support of the higher level of duplication that underwent this repertoire. The list of genes is given in S 6.

pairs		Bur	Zeb	Bri	Nye	Til
	Bur	2	9	0	11	0
	Zeb		0	0	10	0
	Bri			0	1	0
	Nye				0	0
	Til					7
triplets						
	Bur	Zeb	Nye	37		
	Bur	Bur	Nye	1		
	Til	Til	Til	1		
quadruplets	Bur	Zeb	Nye	Bri	2	
	Bur	Zeb	Nye	Til	1	
	Til	Til	Til	Til	2	

Table 5: dN/dS ratios of the 7 TM regions, the 3 external and 3 internal loops for the 4 largest families were calculated. TM and loops were identified with PolyPhobius.

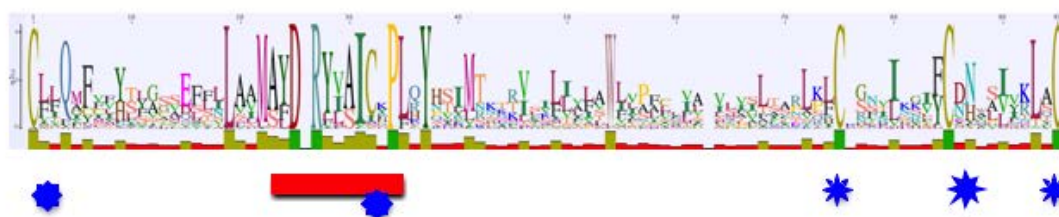
Family L	entire molecule				
40 genes	0.50 ± 0.12	TM 1	0.56 ± 0.42	IN 1	0.23 ± 0.11
		TM 2	0.59 ± 0.58	IN 2	0.41 ± 0.38
		TM 3	0.76 ± 0.43	IN 3	0.59 ± 0.33
		TM 4	1.48 ± 1.54	OUT 2	0.29 ± 0.34
		TM 5	0.63 ± 0.36	OUT 3	0.35 ± 0.22
		TM 6	0.72 ± 0.42	OUT 4	0.40 ± 0.34
		TM 7	0.56 ± 0.58		
Family N	entire molecule				
40 genes	0.39 ± 0.14	TM 1	0.45 ± 0.35	IN 1	0.57 ± 0.35
		TM 2	0.79 ± 0.37	IN 2	0.60 ± 0.55
		TM 3	0.68 ± 0.57	IN 3	0.41 ± 0.33
		TM 4	0.95 ± 0.54	OUT 2	0.35 ± 0.27
		TM 5	0.56 ± 0.41	OUT 3	0.24 ± 0.11
		TM 6	0.60 ± 0.41	OUT 4	0.40 ± 0.13
		TM 7	1.20 ± 1.07		
Family E	entire molecule				
56 genes	0.40 ± 0.11	TM 1	0.57 ± 0.27	IN 1	0.24 ± 0.01
		TM 2	0.52 ± 0.48	IN 2	0.42 ± 0.37
		TM 3	0.73 ± 0.79	IN 3	0.32 ± 0.37
		TM 4	0.69 ± 0.39	OUT 2	0.43 ± 0.24
		TM 5	0.70 ± 0.44	OUT 3	0.30 ± 0.23
		TM 6	0.23 ± 0.21	OUT 4	0.41 ± 0.38
		TM 7	0.21 ± 0.12		
Family D	entire molecule				
32 genes	0.44 ± 0.10	TM 1	0.28 ± 0.18	IN 1	0.36 ± 0.33
		TM 2	0.24 ± 0.23	IN 2	0.69 ± 0.72
		TM 3	0.64 ± 0.46	IN 3	0.26 ± 0.26
		TM 4	0.83 ± 1.01	OUT 2	0.22 ± 0.22
		TM 5	0.58 ± 0.39	OUT 3	1.14 ± 1.06
		TM 6	0.42 ± 0.41	OUT 4	0.80 ± 0.56
		TM 7	0.36 ± 0.35		

Table 8 : Family distribution of OR genes with more than 1 coding exon.

FAMILIES	sub-fam	# genes	# genes with >1 coding exon	# exons
fam V	1	4	4	2gene= 2exons 2gene= 4exons
fam U	1	5	5	3
fam W	5	24	24	2
fam I	1	6	6	3
fam T	1	2	1	4
fam AB	1	4	2	2
fam A	6	100	1	2
fam K	4	22	2	2

Fig 3: WebLogo representation of the AA conservation around the MAYDRY motif. On a multi-alignment performed with MAFFT followed by PHYML clustering, two subgroups of cichlid OR were identified. One with a classical MAYDRY motif followed by 3 cysteine residues indicated by a blue star and a second with an altered MAYDRY motif in which the aspartic residue (D) is replaced by a glutamic residue (E). The genes of each of these two groups are given in S10 along with a complete alignment of their sequences

Sub-group of cichlid OR with a motif of 3 cysteins



Sub-group of cichlid OR with 1 cystein only

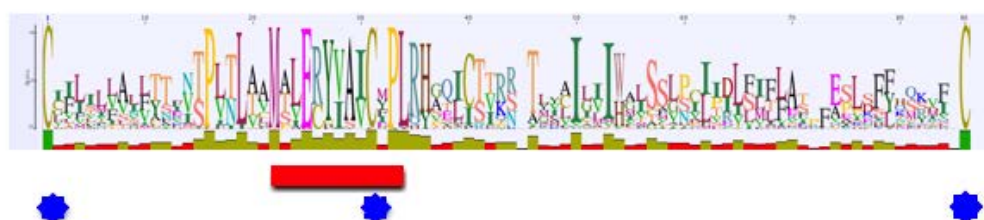
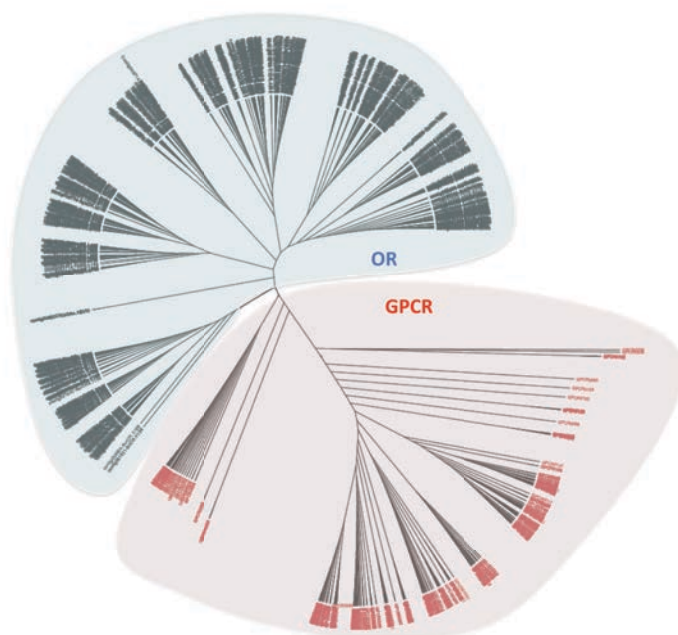


Fig 5: Phylogenetic tree constructed with the cichlid OR repertoires (in blue) and 279 non OR class A GPCR (S1) (in red). This tree clearly shows that the cichlid OR cannot be confounded with the non OR class A GPCR.



References

1. Turner GF, Seehausen O, Knight ME, Allender CJ, Robinson RL. How many species of cichlid fishes are there in African lakes? *Mol Ecol.* 2001; 10:793-806.
2. Kocher TD. Adaptive evolution and explosive speciation: the cichlid fish model. *Nat Rev Genet.* 2004; 5:288-298.
3. FAO: The State of World Fisheries and Aquaculture. <http://www.fao.org/docrep/013/i1820e/i1820e00.htm> 2010.
4. Buck L and Axel R. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell.* 1991; 65 (1): 175–87.
5. Raming K, Krieger J, Strotmann J, Boekhoff I, Kubick S, Baumstark C, Breer H. Cloning and expression of odorant receptors. *Nature.* 1993; 361(6410), 353-356.
6. Pace U, Hanski E, Salomon Y, Lancet D. 1985. Odorant-sensitive adenylate cyclase may mediate olfactory reception. *Nature.* 316(6025), 255-258.
7. Bakalyar HA and Reed RR. Identification of a specialized adenylyl cyclase that may mediate odorant detection. *Science.* 1990; Dec 7; 250(4986):1403-6
8. Asanuma N and Nomura H. Cytochemical localization of adenylate cyclase activity in rat olfactory cells. *Histochem J.* 1991; Feb; 23(2):83-90.
9. Laberge and Hara. Neurobiology of fish olfaction: a review. *Brain Research Reviews.* 2001; 36 46–59
10. Godfrey PA, Malnic B, Buck LB. The mouse olfactory receptor gene family. *Proc. Natl. Acad. Sci. USA.* 2004; 101:2156–61
11. Malnic B, Godfrey PA, Buck LB. The human olfactory receptor gene family. *Proc. Natl. Acad. Sci. USA.* 2004; 101(8), 2584-2589.
12. Quignon P, Giraud M, Rimbault M, Lavigne P, Tacher S, Morin E, Retout E, Valin A.S, Lindblad-Toh K, Nicolas J, Galibert F. The dog and rat olfactory receptor repertoires. *Genome Biology.*

- 2005; 6(10), R83.
13. Glusman G, Bahar A, Sharon D, Pilpel Y, White J, Lancet D. The olfactory receptor gene superfamily: data mining, classification, and nomenclature. *Mamm Genome*. 2000; Nov;11(11):1016-23.
 14. Freitag J, Ludwig G, Andreini I, Rössler P, Breer H. Olfactory receptors in aquatic and terrestrial vertebrates. 1998; *J. Comp. Physiol. A* 183, 635–50.
 15. Ben-Arie N, Lancet D, Taylor C, Khen M, Walker N, Ledbetter D.H, Carrozzo R, Patel K, Sheer D, Lehrach H, North MA. Olfactory receptor gene cluster on human chromosome 17: possible duplication of an ancestral receptor repertoire. *Human Molecular Genetics*. 1994; 3(2), 229-235.
 16. Niimura Y and Nei M. Evolutionary dynamics of olfactory receptor genes in fishes and tetrapods. *Proc. Natl. Acad. Sci. USA*. 2005; 102: 6039–6044 (2005)
 17. Chen M, Peng Z, He S. Olfactory receptor gene family evolution in stickleback and medaka fishes. *Sci China Life Sci*. 2010; Feb;53(2):257-66. doi: 10.1007/s11427-010-0025-4.
 18. Gibbs RA, et al. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature*. 2004; Apr 1;428(6982):493-521.
 19. Zhang X, Rodriguez I, Mombaerts P, Firestein S. Odorant and vomeronasal receptor genes in two mouse genome assemblies. *Genomics*. 2004; May; 83(5): 802-11.
 20. Katoh, Standley 2013 (*Molecular Biology and Evolution* 30:772-780) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. (<http://mafft.cbrc.jp/alignment/server/>)
 21. Boc, A., Diallo, Alpha B. and Makarenkov, V. (2012), T-REX: a web server for inferring, validating and visualizing phylogenetic trees and networks, *Nucleic Acids Research*, 40(W1), W573- W579. (PHYML <http://www.trex.uqam.ca/>)
 22. Nikaido M, Suzuki H, Toyoda A, Fujiyama A, Hagino-Yamagishi K, Kocher TD, Carleton K, Okada N. Lineage-specific expansion of vomeronasal type 2 receptor-like (OlfC) genes in cichlids may contribute to diversification of amino acid detection systems. *Genome Biol Evol*. 2013; 5(4):711-22. doi: 10.1093/gbe/evt041.
 23. Zhang Z, Li J, Zhao XQ, Wang J, Wong GK, Yu J. KaKs Calculator: calculating Ka and Ks through model selection and model averaging. *Genomics Proteomics Bioinformatics*. 2006; 4: 259-263.
 24. Makalowski W, Bogulski MS. Evolutionary parameters of the transcribed mammalian genome: an analysis of 2,820 orthologous rodent and human sequences. *Proc. Natl. Acad. Sci. USA*. 1998; Aug 4; 95(16):9407-12.
 25. Kondo R, Kaneko S, Sun H, Sakaizumi M, Chigusa SI. Diversification of olfactory receptor genes in the Japanese medaka fish, *Oryzias latipes*. *Gene*. 2002; 282, 113-120.
 26. Gilad Y, Bustamante CD, Lancet D, Paabo S: Natural selection on the olfactory receptor gene family in humans and chimpanzees. *Am J Hum Genet* 2003; 73(3):489-501.
 27. Robin S, Tacher S, Rimbault M, Vaysse A, Dréano S, André C, Hitte C, Galibert F. Genetic diversity of canine olfactory receptors. *BMC Genomics*. 2009; Jan 14;10:21. doi: 10.1186/1471-2164-10-21.

28. MEME <http://meme.nbcr.net/meme/cgi-bin/meme.cgi>
29. Liu AH, Zhang X, Stolovitzky GA, Califano A, Firestein SJ. Motif-based construction of a functional map for mammalian olfactory receptors. *Genomics* 2003; 81, 443–456.
30. Mombaerts P. *Nat Rev Neurosci*. Genes and ligands for odorant, vomeronasal and taste receptors. 2004; 5(4):263-78. 29.
31. Zozulya S, Echeverri F, Nguyen T. The human olfactory receptor repertoire *Genome Biol*. 2001; 2(6): RESEARCH0018. Epub 2001 Jun 1.
32. Phobius : <http://phobius.sbc.su.se/>
33. Alioto TS and Ngai J. The odorant receptor repertoire of teleost fish. *BMC Genomics*. 2005; 6: 173.
34. NetNGlycserver <http://www.cbs.dtu.dk/services/NetNGlyc/>
35. Katada S, Tanaka M, Touhara K. Structural determinants for membrane trafficking and G protein selectivity of a mouse olfactory receptor. *J Neurochem*. 2004; Sep; 90(6):1453-63.
36. Young JM, Friedman C, Williams EM, Ross JA, Tonnes-Priddy L, Trask BJ. Different evolutionary processes shaped the mouse and human olfactory receptor gene families. *Hum Mol Genet*. 2002; Mar 1; 11(5):535-46.
37. Masters SB, Stroud RM, Bourne HR. Family of G protein alpha chains: amphipathic analysis and predicted structure of functional domains. 1986; *Protein Eng*. 1986 Oct-Nov;1(1):47-54.
38. Zhao H, Firestein S. Vertebrate odorant receptors. *Cell Mol Life Sci*. 1999; Nov 15; 56(7-8): 647-659.
39. Wahl MC, Will CL, Lührmann R. The spliceosome: design principles of a dynamic RNP machine. *Cell*; 2009; Feb 20; 136(4): 701-18. doi: 10.1016/j. cell. 2009.02.009. review.
40. FSPLICE: <http://linux1.softberry.com/berry.phtml?topic=fsplce&group=programs&subgroup=gfind>)
41. FigTree 1.3.1 <http://tree.bio.ed.ac.uk/software/figtree/>
42. Barber RD and Ronnet GV. Reconstructing smell. *Mol. Neurobiology*. 2000; 21 161-173.
43. Buck L. The molecular architecture of odor and pheromone sensing in mammals. *Cell*. 2000; 100, 611-618
44. Mombaerts P. Genes and ligands for odorant, vomeronasal and taste receptors. *Nat. Rev. Neuroscience*. 2004; 5 263-278
45. Tiedge M, Richter T Richter T. Importance of Cysteine Residues for the Stability and Catalytic Activity of Human Pancreatic Beta Cell Glucokinase. *Arch Biochem Biophys*. 2000 Mar 15;375(2):251- 260.
46. Chen ZW, Jiang CY, She Q, Liu SJ, Zhou PJ. Key role of cysteine residues in catalysis and subcellular localization of sulfur oxygenase-reductase of *Acidianus tengchongensis*. *Appl Environ Microbiol*. 2005 Feb;71(2):621-8.
47. Abaffy T, Malhotra A, Luetje ChW. The molecular basis for ligand specificity in a mouse

- olfactory receptor. *J. Biol. Chem.* 2007; 282,1216–1224.
48. Doszczak L, Kraft P, Weber HP, Bertermann R, Triller A, Hatt H, Tacke R. Prediction of perception: probing the hOR17-4 olfactory receptor model with silicon analogues of bourgeonal and linal. *Angew. Chem. Int.* 2007 ; Ed.46, 3367–3371.
 49. Singer MS. Analysis of the molecular basis for octanal interactions in the expressed rat I7 olfactory receptor. *Chem. Senses.* 2000 ; 25, 155–165.
 50. Niimura Y, Nei M. Extensive gains and losses of olfactory receptor genes in mammalian evolution. *PLoS ONE* 2007; 2:e708.
 51. Dong D, He G, Zhang S, Zhang Z. Evolution of olfactory receptor genes in primates dominated by birth-and-death process. *Genome Biol. Evol.* 2009; Aug 4;1: 258-64. doi: 10.1093/gbe/evp026.
 52. Vosshall LB, Amrein H, Morozov PS, Rzhetsky A, Axel R. A spatial map of olfactory receptor expression in the *Drosophila* antenna. *Cell.* 1999; 96(5) 725–736.
 53. Robertson HM. The large srh family of chemoreceptor genes in *Caenorhabditis* nematodes reveals processes of genome evolution involving large duplications and deletions and intron gains and losses. *Genome Research.* 2000; 10 (2), 192-203.
 54. GENEIOUS :<http://www.geneious.com> Geneious
 55. TMHMM : <http://www.cbs.dtu.dk/services/TMHMM/> v. 2.0

B- Le répertoire des gènes TAAR (Publication n°5)

Les résultats sont présentés sous forme de tableaux et de graphes construits comme pour les OR.

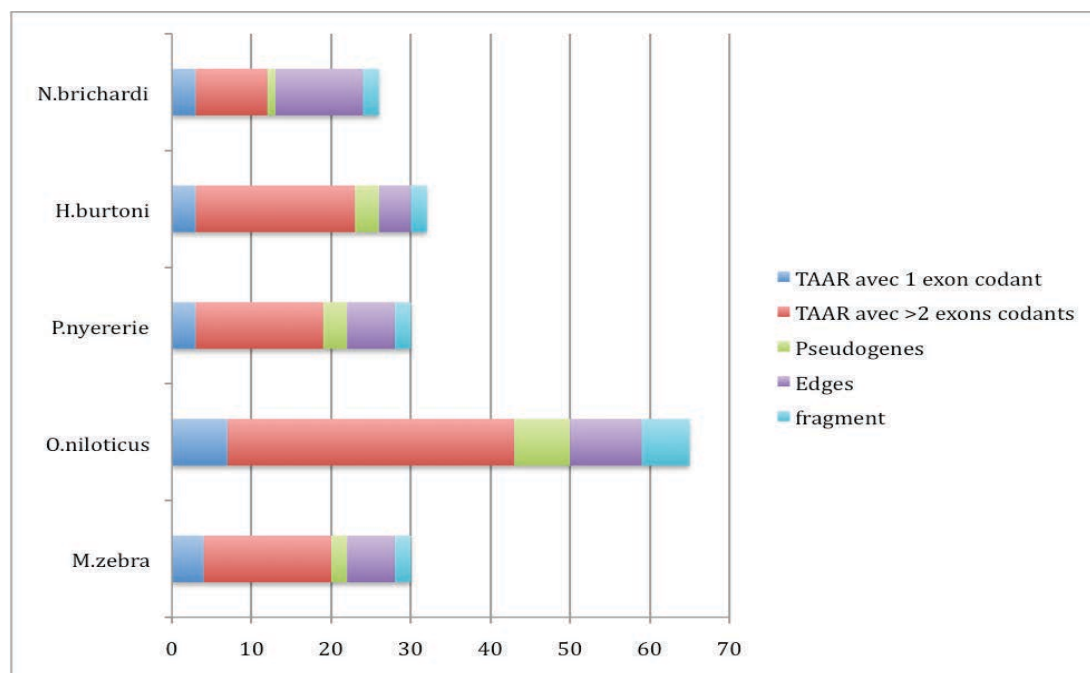


Figure 37 : Répartition de l'ensemble des catégories des TAAR chez les cinq cichlidés.

Par ailleurs le tableau 9 présente la comparaison du répertoire des gènes TAAR des cinq cichlidés, du poisson zèbre (*Danio rerio*) et d'autres poissons modèles (Hashiguchi & Mutsumi, 2007). Toutefois, comme pour les OR, il est difficile de commenter ces chiffres dans la mesure où les séquences génomiques ne sont pas finalisées bien qu'il semble assez évident que le *D. rerio* a un nombre de gènes beaucoup plus élevé que n'importe lequel des autres poissons et que les cichlidés dans leur ensemble auraient moins de pseudogènes (tableau 10).

Tableau 10 : Décompte du nombre de gènes TAAR et de pseudogènes chez les 5 cichlidés comparé à celui des poissons modèles.

	<i>M.zebra</i>	<i>O.niloticus</i>	<i>P.nynerie</i>	<i>H.burtoni</i>	<i>N.brighardi</i>	<i>D.rerio</i>	<i>G.aculeatus</i>	<i>O. Latipes</i>	<i>T.rubripes</i>
#TAAR complets	20	45	19	23	12	109 ⁽¹⁾	49 ⁽¹⁾	25 ⁽¹⁾	13 ⁽¹⁾
#Pseudogenes	4	7	2	3	1	10 ⁽¹⁾	15 ⁽¹⁾	7 ⁽¹⁾	6 ⁽¹⁾
#Edges	5	13	7	4	12				

⁽¹⁾ Hashiguchi Y. & Mutsumi, 2007

1-Etudes phylogéniques

Ces analyses ont été réalisées en suivant exactement la même démarche et le même protocole que pour les OR.

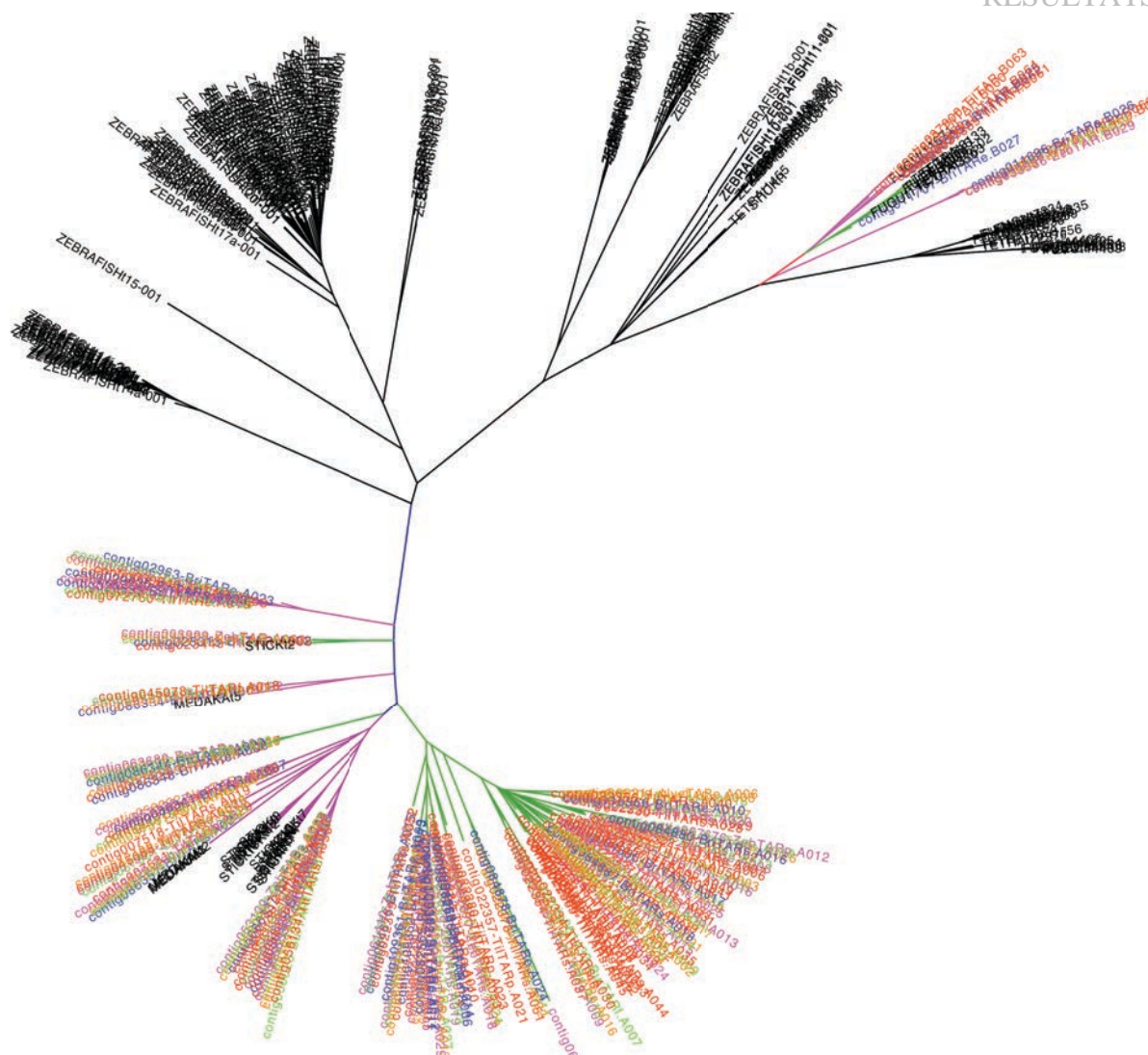


Figure 38 : L'arbre phylogénique des TAAR des 5 cichlidés et des poissons modèles (Poisson Zèbre, Fugu, Tetraodon, Medaka et Stickelback) est dessiné avec le code couleur suivant : Bleu *N. brichardi* ; rose *M. zebra* ; vert *H. burtoni* ; orange *P. nyererie* ; rouge *O. niloticus* et noir poissons modèles.

Comme pour les OR, on a calculé le pourcentage d'identité protéique entre les séquences de la même branche afin d'identifier les familles et sous-familles des gènes TAAR.

Le tableau 11 présente la répartition des familles et sous-familles des TAAR pour chaque poisson. Ce tableau montre par ailleurs que les TAAR sont organisés en familles et sous-familles partagées par les cinq cichlidés. En revanche les TAAR des poissons modèles sont bien isolés de ceux des cichlidés, ils ne partagent aucune famille ou sous-familles avec ces derniers ainsi qu'on peut le voir sur la figure 36 élargie.

Tableau 11 : La répartition des famille et sous-familles TAAR pour chaque poisson

	<i>N. Brichardi</i>	<i>H. Burtoni</i>	<i>P. Nyererei</i>	<i>Tilapia</i>	<i>M. zebra</i>
A1	3	3	6	6	6
A2	3	9	6	24	5
A3	3	7	4	5	4
A4	1	2	1	1	2
A5	1	1	1	4	2
B1	0	1	1	0	1
B2	1	0	0	2	0
# familles	2	2	2	2	2
# s-familles	6	6	6	6	6

2- Calcul du rapport dN/dS et analyse des résultats du répertoire TAAR

Comme pour les OR, le rapport dN/dS a été calculé selon la méthode Nei-Gojobori modifiée par Zhang (Zhang *et al.*, 1998 PNAS) en utilisant le même pipeline

Les résultats des calculs dN/dS des TAAR sont présentés comme pour ceux des OR

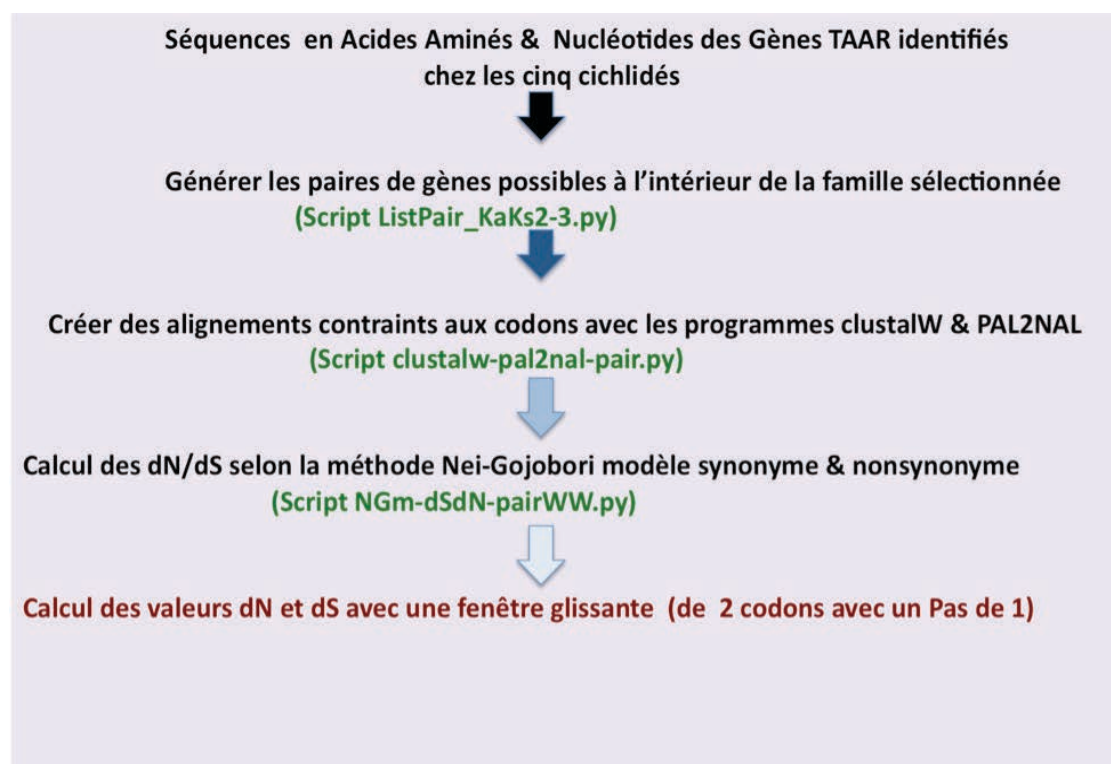


Figure 39 : Diagramme du pipeline de calcul des valeurs dN et dS pour les gènes TAAR

Les résultats des calculs sont présentés sous forme d'un tableau (12) indiquant la moyenne des valeurs de dN/dS pour chacune des 2 familles :

Tableau 12 : Résultats des moyennes de calcul dN/dS obtenus pour les deux familles des gènes TAAR

famille A	moyenne dN/dS	min	max
	0,44	0,16	1,22
famille B	moyenne dN/dS	min	max
	0,51	0,35	>10

Avec une moyenne générale égale à 0,47 une sélection positive doit être envisagée pour l'évolution de la famille des gènes TAAR comme pour les OR. Une analyse identique à celle des OR a été faite avec une fenêtre glissante d'un codon. Dans les panneaux suivants sont indiquées les positions des mutations non silencieuses (en rouge) et silencieuses (en bleu) ainsi que la position de celles-ci par rapport à la structure 2D de ce même couple de récepteurs.

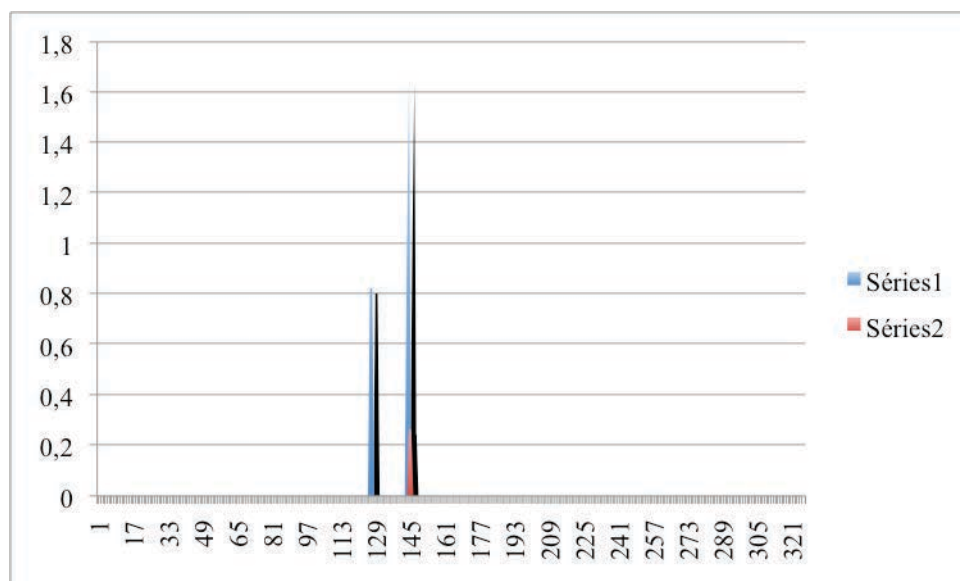


Figure 40 : Ce graphe présente les positions des mutations non silencieuses (en rouge) et silencieuses (en bleu) du couple BURspliced13/NYEsplined17

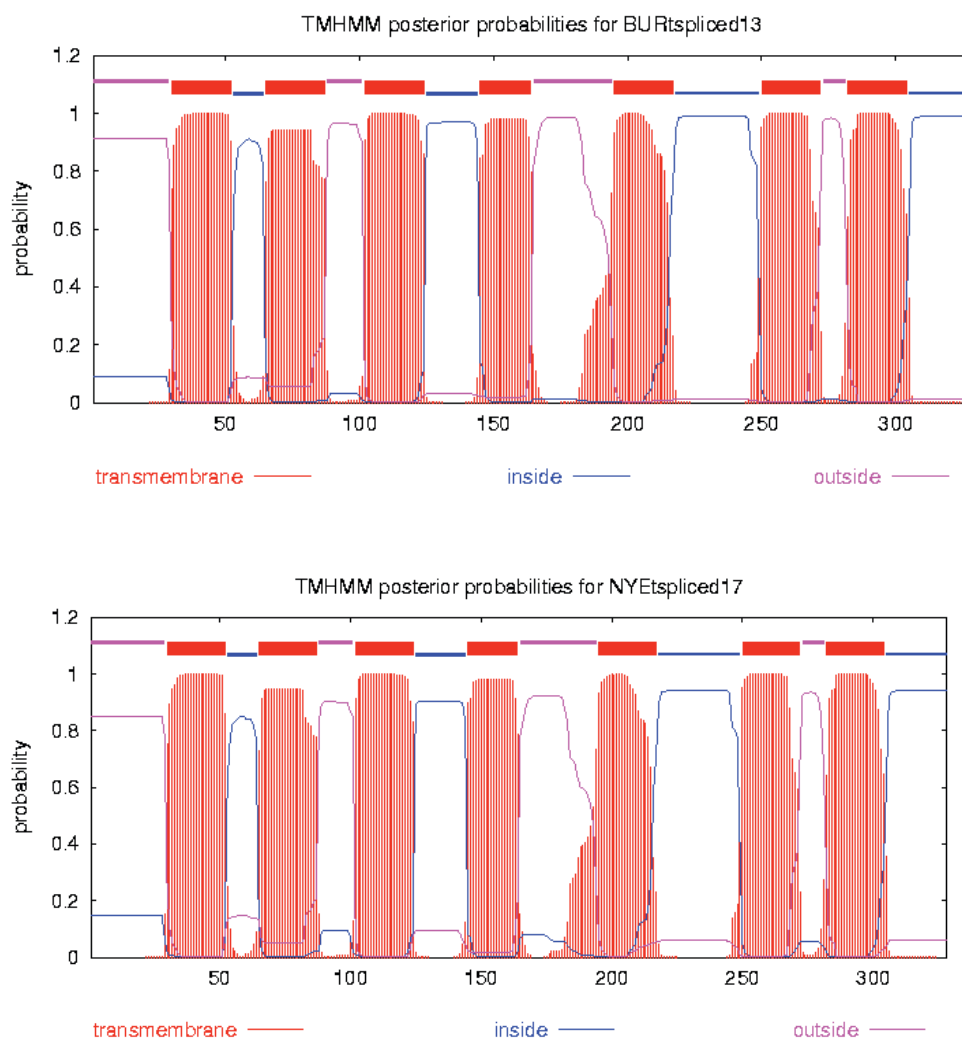


Figure 41 : Structure 2D du couple de récepteurs BURspliced13/NYEspliced17.

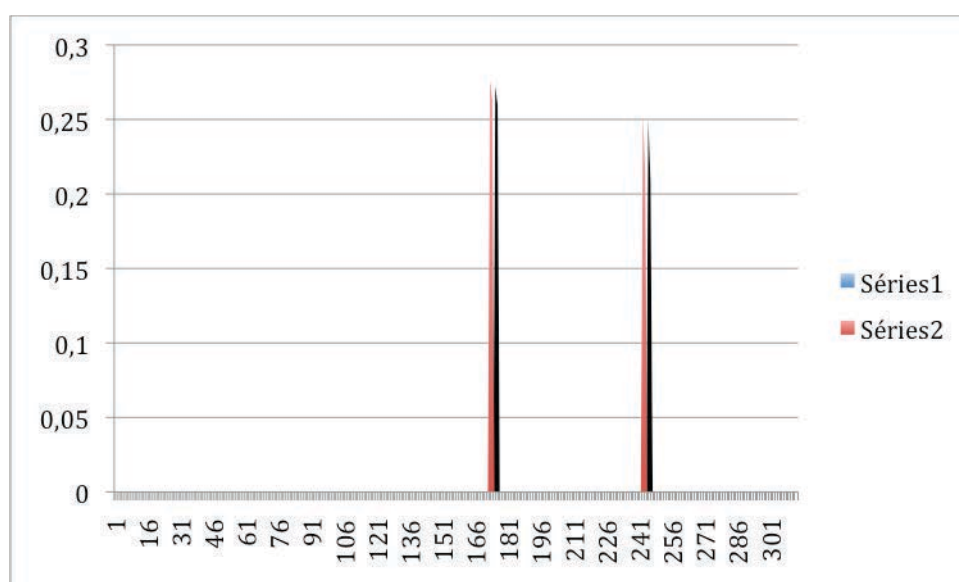


Figure 42 : Ce graphe présente les positions des mutations non silencieuses (rouge) et silencieuses (bleu) du couple BURt03/ZEBt03.

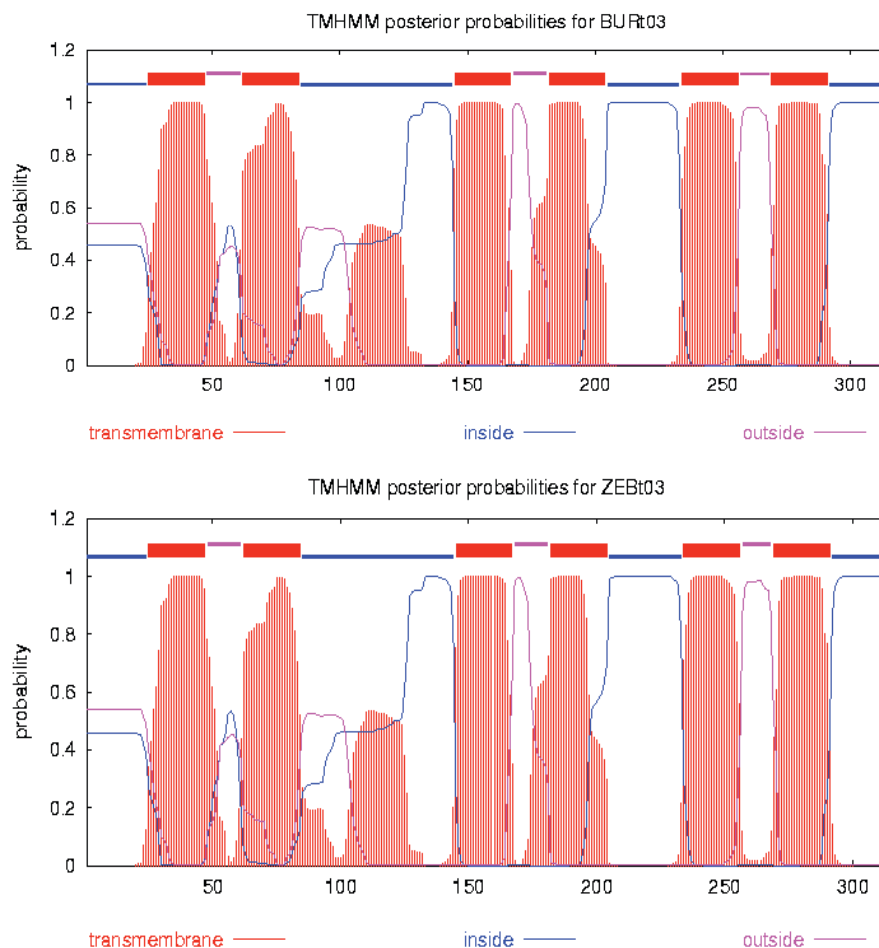


Figure 43 : Structure 2D du couple de récepteurs BURt03/ZEBt03.

Sur 14 paires de gènes TAAR analysées on a trouvé des valeurs dS et dN réparties sur les différents domaines protéiques selon le tableau 13 :

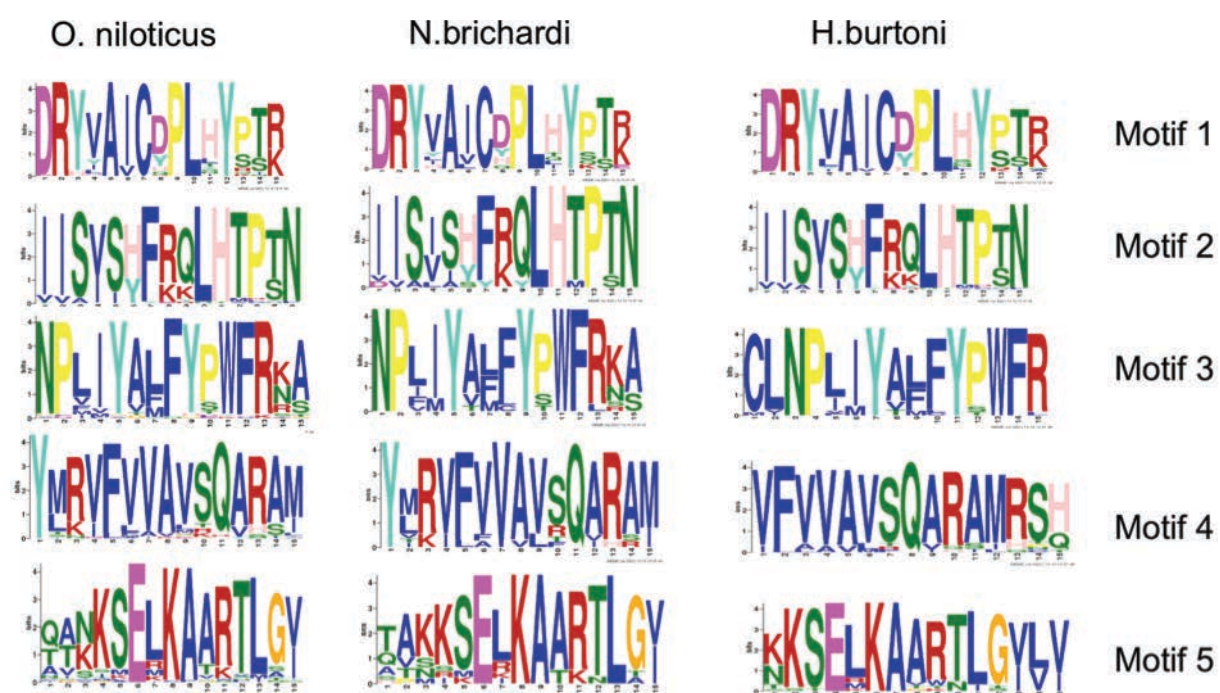
Tableau 13 : Analyse des valeurs dN et dS sur les différentes parties de la structure 2D pour 14 paires de gènes TAAR

Structure 2D	dS	dN
OUTSIDE	3	4
TM1	2	3
IL1	1	0
TM2	6	2
EL1	0	0
TM3	0	2
IL2	8	1
TM4	4	7
EL2	1	7
TM5	1	2
IL3	1	3
TM6	2	4
EL3	0	2
TM7	0	3
INSIDE	3	3
Moyenne	2,13 +/-2,3	2,87 +/-2,06

Dans ce tableau sont surlignées les valeurs largement au delà de la valeur de l'écart-type, suggérant l'existence de domaines protéiques propices aux mutations silencieuses et non silencieuses. A cet égard, il est intéressant de noter que Chen Ming *et al.*, ont identifié des dN élevés dans les régions transmembranaires TM4 et EL2 chez medaka et stickleback (Chen Ming *et al.*, 2010).

3- Les motifs conservés des séquences protéiques

Comme les OR, les TAAR se caractérisent par un ensemble de motifs protéiques que nous avons recherchés en suivant la même démarche.



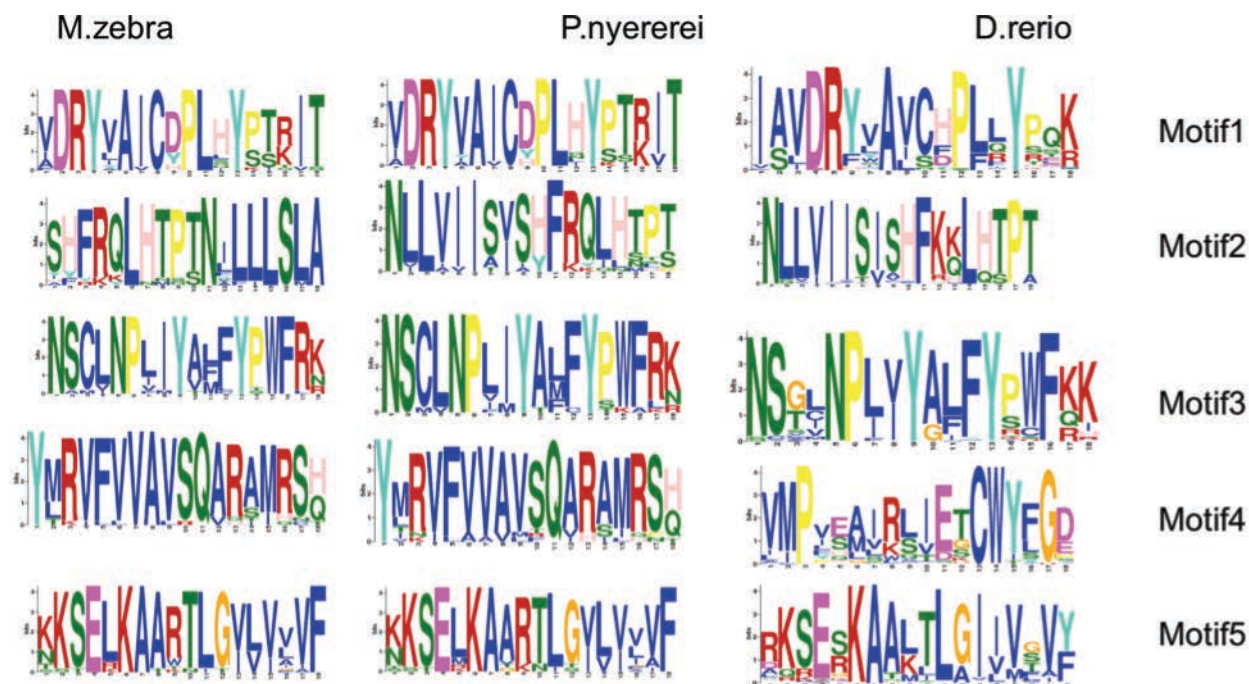


Figure 44 : Les 5 Motifs protéiques les plus pertinents des TAAR de cichlidés et du poisson zèbre.

Les logos des motifs dessinés par le site MEME Suite, montrent que les cinq motifs sont très conservés chez les cichlidés et que le poisson zèbre partage avec ceux-ci les trois premiers seulement. Ces cinq motifs occupent des positions comparables pour les cinq cichlidés par rapport à la position des régions transmembranaires avec le motif 1 sur IL2_TM4, le motif 2 sur IL1_TM2, le motif 3 sur TM7_INSIDE, et pour le motif 4 sur TM5_IL3, le motif 5 sur IL3_TM6.

4- Les gènes épissés

Comme pour les gènes OR, on a considéré dès leur mise en évidence chez les vertébrés que seul le deuxième exon était codant. Dans notre étude toutefois nous montrons qu'il n'en est rien et que même la grande majorité des gènes TAAR comportent deux exons codants comme indiqué sur le tableau 14. Par rapport aux gènes OR, on peut noter deux différences :

- (a) les gènes épissés sont majoritaires
- (b) ils ont tous deux exons codants.

Tableau14 : Décompte des gènes TAAR formés d'un et deux exons codants

	<i>M.zebra</i>	<i>O.niloticus</i>	<i>P.nynerie</i>	<i>H.burtoni</i>	<i>N.brighardi</i>
#TAAR avec 1 exon codant	4	9	3	3	3
#TAAR avec 2 exons codants	16	36	16	20	9

Comme pour les OR, les résultats des analyses des sites de reconnaissance des limites exon/intron et de ligature des exons, ainsi que de la position de ces sites dans la structure des gènes TAAR sont des arguments forts en faveur de la réalité de la nature de ces gènes :

4-1- Alignement des régions codantes d'un gène TAAR sur l'ADN génomique

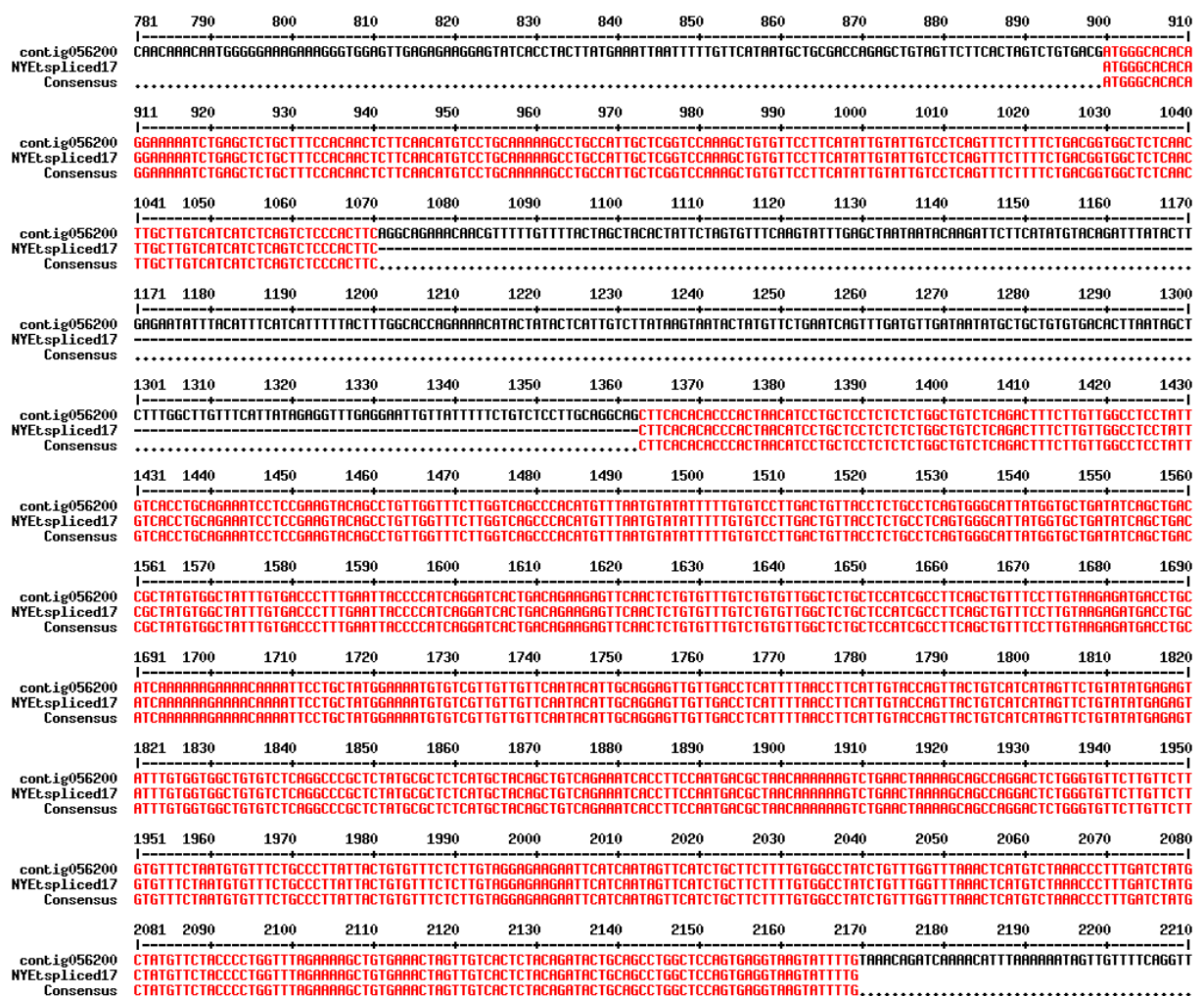


Figure 45 : Alignement de la séquence du gène *NYEspliced17* sur son contig056200. Dans cette figure la partie codante du gène *NYEspliced17* est alignée sur la

séquence du contig qui contient ce gène. Les positions 1070 et 1363 signalent les bornes de l'intron interrompant la phase de lecture.

Les limites des parties communes entre le contig et la partie traduite du gène indiquent d'une part la nature des séquences bornant le site d'épissage et la position de celui-ci par rapport à la structure 2D de la protéine. Tous les gènes TAAR étant construits sur le même modèle, un seul exemple est montré.

4-2- L'analyse des sites d'épissages des gènes TAAR

Nous avons procédé avec les gènes TAAR épissés de la même façon qu'avec les gènes OR pour définir les jonctions exon/intron. Les deux motifs très conservés permettent de prédire une position qui maintient l'intégrité de la phase de lecture.

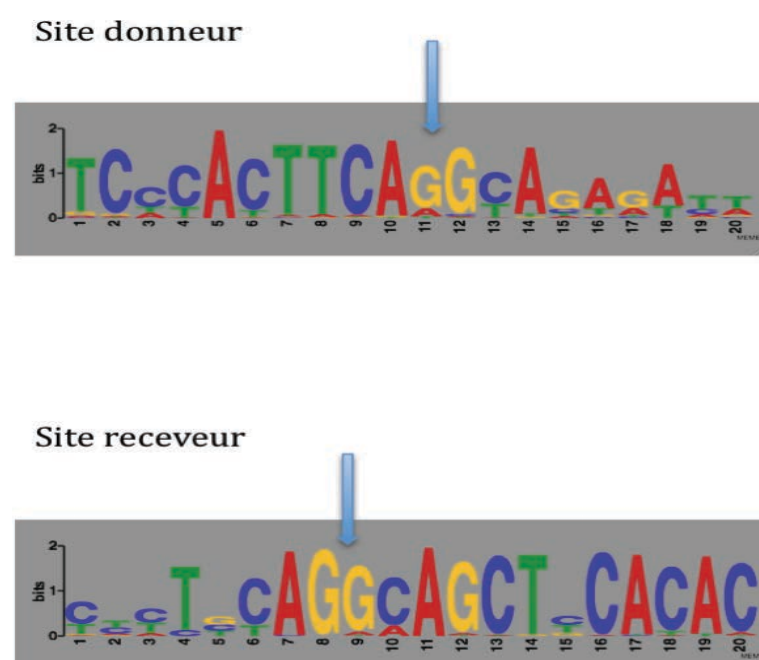


Figure 46 : Motifs des sites donneurs et récepteurs des gènes TAAR épissés des 5 cichlidés.

Pour les 5 poissons le site donneur est entre les deux G de la séquence CAGGcA, de même pour le site accepteur. Le tableau 15 montre les positions des sites d'épissage par rapport à la structure 2D de la protéine TAAR ainsi que la position du dernier codon du premier exon codant et son numéro.

Tableau 15 : Positions des sites d'épissage des TAAR

nom des TAAR	dernier codon	N° du codon	Position sur la structure transmembranaire
TiITARs.A019	TTC=F	56	INTLOOP1
TiITARs.A024	TTC=F	55	INTLOOP1
TiITARs.A025	TTC=F	57	INTLOOP1
TiITARs.A026	TTC=F	57	INTLOOP1
TiITARs.A027	TTC=F	59	INTLOOP1
TiITARs.A028	TTC=F	57	INTLOOP1
TiITARs.A029	TTC=F	57	INTLOOP1
TiITARs.A030	TTC=F	58	INTLOOP1
TiITARs.A031	TTC=F	58	INTLOOP1
TiITARs.A032	TTC=F	57	INTLOOP1
TiITARs.A033	TTC=F	57	INTLOOP1
TiITARs.A034	TTC=F	57	INTLOOP1
TiITARs.A035	TTC=F	57	INTLOOP1
TiITARs.A036	TTC=F	57	INTLOOP1
TiITARs.A037	TTC=F	57	INTLOOP1
TiITARs.A038	TTC=F	57	INTLOOP1
TiITARs.A039	TTC=F	57	INTLOOP1
TiITARs.A040	TTC=F	57	INTLOOP1
TiITARs.A041	TTC=F	57	INTLOOP1
TiITARs.A042	TTC=F	57	INTLOOP1
TiITARs.A043	TTC=F	57	INTLOOP1
TiITARs.A044	TTC=F	56	INTLOOP1
TiITARs.A045	TTC=F	54	INTLOOP1
TiITARs.A046	TTC=F	57	INTLOOP1
TiITARs.A047	TTC=F	55	INTLOOP1
TiITARs.A048	TTC=F	57	INTLOOP1
TiITARs.A050	TTC=F	55	INTLOOP1
TiITARs.A051	TTC=F	56	INTLOOP1
TiITARs.A052	TTC=F	57	INTLOOP1
TiITARs.A053	TTC=F	57	INTLOOP1
TiITARs.A054	TTC=F	57	INTLOOP1
TiITARs.A055	TTC=F	57	INTLOOP1
TiITARs.A056	TTC=F	57	INTLOOP1
TiITARs.A057	TTC=F	54	INTLOOP1
TiITARs.A058	TTC=F	54	INTLOOP1
TiITARs.A059	TTC=F	57	INTLOOP1
ZebTARs.A014	TTC=F	57	INTLOOP1
ZebTARs.A015	TTC=F	57	INTLOOP1
ZebTARs.A016	TTC=F	55	INTLOOP1
ZebTARs.A017	TTC=F	54	INTLOOP1
ZebTARs.A018	TTC=F	57	INTLOOP1
ZebTARs.A019	TTC=F	57	INTLOOP1
ZebTARs.A020	TTC=F	57	INTLOOP1
ZebTARs.A021	TAC=TYR	55	INTLOOP1
ZebTARs.A022	TTC=F	57	INTLOOP1
ZebTARs.A023	TTC=F	54	INTLOOP1
ZebTARs.A024	TTC=F	57	INTLOOP1
ZebTARs.A025	TTC=F	55	INTLOOP1
ZebTARs.A026	TTC=F	57	INTLOOP1

ZebTARs.A027	TTC=F	54	INTLOOP1
ZebTARs.A028	TTC=F	54	INTLOOP1
ZebTARs.A029	TTC=F	57	INTLOOP1

BurTARs.A012	TAC=TYR	54	INTLOOP1
BurTARs.A013	TTC=F	59	INTLOOP1
BurTARs.A014	TTC=F	57	INTLOOP1
BurTARs.A015	TTC=F	54	INTLOOP1
BurTARs.A016	CTC=LEU	42	OUT-SIDE
BurTARs.A017	TTC=F	54	INTLOOP1
BurTARs.A018	TTC=F	57	INTLOOP1
BurTARs.A019	TTC=F	59	INTLOOP1
BurTARs.A020	TTC=F	57	INTLOOP1
BurTARs.A021	TTC=F	57	INTLOOP1
BurTARs.A022	TTC=F	55	INTLOOP1
BurTARs.A023	TTC=F	57	INTLOOP1
BurTARs.A024	TTC=F	57	INTLOOP1
BurTARs.A025	TTC=F	55	INTLOOP1
BurTARs.A026	TTC=F	57	INTLOOP1
BurTARs.A027	TTC=F	57	INTLOOP1
BurTARs.A028	TTC=F	55	INTLOOP1
BurTARs.A029	TTC=F	54	INTLOOP1
BurTARs.A030	TTC=F	54	INTLOOP1
BurTARs.A031	TTC=F	53	INTLOOP1

BriTARs.A014	AAC=N	47	TM1
BriTARs.A015	TTC=F	57	INTLOOP1
BriTARs.A016	TTC=F	59	INTLOOP1
BriTARs.A017	TTC=F	55	INTLOOP1
BriTARs.A018	TTC=F	57	INTLOOP1
BriTARs.A019	TTC=F	54	INTLOOP1
BriTARs.A020	TTC=F	54	INTLOOP1
BriTARs.A021	TTC=F	53	INTLOOP1
BriTARs.A022	TTC=F	57	INTLOOP1

NyeTARs.A013	TTC=F	57	INTLOOP1
NyeTARs.A014	TTC=F	57	INTLOOP1
NyeTARs.A015	TTC=F	57	INTLOOP1
NyeTARs.A016	TTC=F	56	INTLOOP1
NyeTARs.A017	TTC=F	55	INTLOOP1
NyeTARs.A018	TTC=F	54	INTLOOP1
NyeTARs.A019	TTC=F	55	INTLOOP1
NyeTARs.A021	TTC=F	57	INTLOOP1
NyeTARs.A022	TTC=F	56	INTLOOP1
NyeTARs.A023	TTC=F	57	INTLOOP1
NyeTARs.A024	TTC=F	57	INTLOOP1
NyeTARs.A025	TTC=F	57	INTLOOP1
NyeTARs.A026	TTC=F	59	INTLOOP1
NyeTARs.A027	TTC=F	53	INTLOOP1
NyeTARs.A028	TTC=F	54	INTLOOP1
NyeTARs.A029	TTC=F	57	INTLOOP1

Tous les gènes TAAR épissés (sauf les deux gènes surlignés)

sont construits sur le même modèle avec deux exons codants et une position unique pour ce site d'épissage.

L'article N°5

The Cichlid TAAR gene repertoires

Naoual Azzouzi, Frédérique Barloy-Hubler and Francis Galibert

Cet article est en cours de rédaction

C- Le génome des cichlidés (publication N° 6 à soumettre) :

Combinaison des données d'assemblage de séquences et de carte RH.

Une publication par l'ensemble du consortium international sur l'analyse de la séquence des génomes des cinq cichlidés est en cours de rédaction. Dans ce document très général notre contribution se rapporte à l'intégration des données de cartographie et d'assemblage du génome de tilapia. Par ailleurs suite à l'annotation des répertoires des gènes OR et TAAR des cinq poissons, brièvement rapportée dans ce document, nous avons procédé à une analyse portant sur la localisation génomique de ces répertoires, comme résumé ci-dessous.

Il est admis que les répertoires de gènes OR ont évolué par des duplications principalement locales. Chez l'homme, le rat, la souris et le chien, ces gènes sont organisés en groupes importants répartis sur un grand nombre de chromosomes avec seulement quelques gènes isolés (Malnic B *et al.*, 2004 ; Gloriam D.E. *et al.*, 2005). Les gènes de classe 1 appelé aussi « fish like » sont en revanche regroupés dans un seul chromosome HSA11, MMU7, RNO1 et CFA21, respectivement. De même, chez l'homme, la souris, le rat les gènes TAAR sont organisés en un seul cluster sur les chromosomes HSA6, MMU10 et RN01 (Gloriam D.E. *et al.*, 2005). En revanche, les 109 gènes OR identifiés dans le génome du poisson zèbre sont répartis en cinq groupes principaux contenant entre 14 et 31 gènes chacun, avec deux groupes sur le chromosome 15, deux sur le chromosome 21 et un sur le chromosome 10. A cela s'ajoutent plusieurs petits clusters sur les chromosomes 8, 14 et 17 et quelques gènes isolés (Alioto T.S et Ngai J, 2005). De même les 47 gènes TAAR du poisson zèbre cartographiés sur les 57 gènes identifiés sont répartis sur 9 locus distribués sur six chromosomes (Gloriam D.E. *et al.*, 2005).

Les données de séquençage ont permis d'attribuer une partie des gènes, contigs et scaffolds à des chromosomes alors que d'autres, désignés sous le vocable de UNK pour « unlinked » sont restés non liés. Nous avons sélectionnés 96 gènes (65 OR et 31 TAAR) de cette dernière catégorie et les avons cartographiés sur le panel RH avec quelques autres gènes déjà cartographiés. Par des analyses en 2 points puis Multipoint (avec un lod 4) nous avons obtenu 12 groupes RH de composition suivante RH1= 16 gènes; RH2 = 14 gènes; RH3 = 14 gènes; RH4 = 7 gènes; RH5 = gènes; RH6 = 6 gènes; RH7 = 5 gènes; RH8 = 4 gènes; RH9 = 3 gènes; RH10, 11, 12 = 2 gènes et 13 gènes non liés.

Comme le montre le tableau 16, dans lequel est donnée la composition du groupe RH3 et du groupe RH7 tous les deux localisés sur le chromosome LG14, l'intégration de ces nouvelles données de cartographie améliore nettement la connaissance de la structure des répertoires des

gènes OR et TAAR. En effet le groupe RH3 est maintenant constitué de 51 gènes dont 14 UNK67, 15 UNK10 qui s'ajoutent aux 22 gènes précédemment attribués à LG14.

De même les scaffolds 953 (4 gènes OR) et 373 (3 gènes TAAR) peuvent être associés à LG16-21. De plus ces dernières analyses de cartographie RH permettent d'associer 21 gènes OR au groupe RH8.

Ainsi, chez tilapia la distributions des gènes OR rappelle celle observée chez le poisson zèbre avec deux principaux groupes de 50 et 56 gènes situés sur les chromosomes LG14 et LG16-21, plus un certain nombre de petits groupes répartis ailleurs. L'organisation des gènes TAAR de tilapia, avec 32 gènes dans un cluster situé sur le chromosome LG16-21 et 11 gènes répartis sur cinq locus est également similaire à celle du poisson zèbre.

Tableau16 : Composition du groupe RH3 et RH7 localisé sur le chromosome LG14.

Name	groupe RH	Contig	Scaffold	LG	Name	groupe RH	Contig	Scaffold	LG
TIL121*	RH3	contig046699	scaffold_256	UNK67	TILe57bis	RH3	contig039437	scaffold_192	LG14
TIL123	RH3	contig046706	scaffold_256	UNK67	TIL092	RH3	contig039450	scaffold_192	LG14
TIL124	RH3	contig046708	scaffold_256	UNK67	TIL091*	RH3	contig039450	scaffold_192	LG14
TIL125	RH3	Contig046714	scaffold_256	UNK67	TIL093	RH3	contig039451	scaffold_192	LG14
TIL126	RH3	contig046718	scaffold_256	UNK67	TILe35	RH3	contig039453	scaffold_192	LG14
TIL128	RH3	contig046723	scaffold_256	UNK67	TIL095	RH3	contig039460	scaffold_192	LG14
TIL129*	RH3	contig046724	scaffold_256	UNK67	TIL094	RH3	contig039460	scaffold_192	LG14
TILe67	RH3	contig046701	scaffold_256	UNK67	TIL097	RH3	contig039461	scaffold_192	LG14
TILe68	RH3	contig046700	scaffold_256	UNK67	TIL096	RH3	contig039461	scaffold_192	LG14
TILf09	RH3	contig046696	scaffold_256	UNK67	TIL098	RH3	contig039462	scaffold_192	LG14
TILf12	RH3	contig046717	scaffold_256	UNK67	TIL099	RH3	contig039465	scaffold_192	LG14
TILs01	RH3	contig046690	scaffold_256	UNK67	TILe36	RH3	contig039467	scaffold_192	LG14
TILs02	RH3	contig046694	scaffold_256	UNK67	TIL101	RH3	contig039469	scaffold_192	LG14
TILs03	RH3	contig046695	scaffold_256	UNK67	TIL100	RH3	contig039469	scaffold_192	LG14
TIL071*	RH3	contig028607	scaffold_118	UNK10	TILe82	RH3	contig039472	scaffold_192	LG14
TIL072	RH3	contig028611	scaffold_118	UNK10	TILe83	RH3	contig039474	scaffold_192	LG14
TIL073	RH3	contig028614	scaffold_118	UNK10	TIL103	RH3	contig039481	scaffold_192	LG14
TIL074	RH3	contig028617	scaffold_118	UNK10	TIL104	RH3	contig039484	scaffold_192	LG14
TIL075	RH3	contig028619	scaffold_118	UNK10	TILpsd03	RH3	contig039487	scaffold_192	LG14
TIL076	RH3	contig028637	scaffold_118	UNK10	TILe60	RH3	contig023711	scaffold_85	LG14
TIL077*	RH3	contig028639	scaffold_118	UNK10	TILe61	RH3	contig023712	scaffold_85	LG14
TIL078	RH3	contig028644	scaffold_118	UNK10	TIL069	RH3	contig023724	scaffold_85	LG14
TIL137	RH3	contig028644	scaffold_118	UNK10	TILe22	RH3	contig023726	scaffold_85	LG14
TILe29	RH3	contig028593-comp	scaffold_118	UNK10	TILe23	RH3	contig023727	scaffold_85	LG14
TILe30	RH3	contig028604-comp	scaffold_118	UNK10	TILe24	RH3	contig023727	scaffold_85	LG14
TILe31	RH3	contig028636	scaffold_118	UNK10	TILe25	RH3	contig023728	scaffold_85	LG14
TILe32	RH3	contig028636	scaffold_118	UNK10	TILe26	RH3	contig023730	scaffold_85	LG14
TILe72	RH3	contig028636	scaffold_118	UNK10	TIL070*	RH3	contig023731	scaffold_85	LG14
TILe73	RH3	contig028630	scaffold_118	UNK10	TIL130*	RH7	contig047714	scaffold_265	LG14
TIL167*	RH3	contig004997	scaffold_8	LG14	TIL131	RH7	contig047725	scaffold_265	LG14
TIL168	RH3	contig004998	scaffold_8	LG14	TIL132	RH7	contig047726	scaffold_265	LG14
TIL002	RH3	contig004999	scaffold_8	LG14	TIL133	RH7	contig047729	scaffold_265	LG14
TIL004*	RH3	contig005000	scaffold_8	LG14	TIL134	RH7	contig047734	scaffold_265	LG14
TIL005	RH3	contig005005	scaffold_8	LG14	TILf08	RH7	contig047735	scaffold_265	LG14

TIL006	RH3	contig005007	scaffold_8	LG14	TILe45	RH7	contig047736	scaffold_265	LG14
TIL080	RH3	contig039415	scaffold_192	LG14	TILpsd06	RH7	contig047738	scaffold_265	LG14
TIL082	RH3	contig039416	scaffold_192	LG14	TILe46	RH7	contig047817	scaffold_265	LG14
TIL081*	RH3	contig039416	scaffold_192	LG14	TIL135	RH7	contig047820	scaffold_265	LG14
TIL083	RH3	contig039419	scaffold_192	LG14	TIL136	RH7	contig047825	scaffold_265	LG14
TIL084	RH3	contig039425	scaffold_192	LG14	TIL138	RH7	contig047826	scaffold_265	LG14
TILf04	RH3	contig039426	scaffold_192	LG14	TIL139	RH7	contig047829	scaffold_265	LG14
TIL085	RH3	contig039426	scaffold_192	LG14	TIL140	RH7	contig047832	scaffold_265	LG14
TIL086	RH3	contig039428	scaffold_192	LG14	TIL146	RH7	contig065247	scaffold_585	LG14
TILe57	RH3	contig039429	scaffold_192	LG14	TIL147*	RH7	contig065253	scaffold_585	LG14
TIL087	RH3	contig039435	scaffold_192	LG14	TIL148	RH7	contig065254	scaffold_585	LG14
TIL088	RH3	contig039435	scaffold_192	LG14	TIL141	RH7	contig047833	scaffold_265	LG14
TILf15	RH3	contig039435	scaffold_192	LG14	TIL142*	RH7	contig047834	scaffold_265	LG14
TIL089	RH3	contig039436	scaffold_192	LG14	TIL001	RH7	contig065241	scaffold_585	LG14
TIL090	RH3	contig039437	scaffold_192	LG14	TILe49	RH7	contig065241	scaffold_585	LG14
TIL066*	RH3	contig023716	scaffold_85	LG14					

Les séquences des quatre autres génomes de cichlidés ne sont pas ancrées sur leurs caryotypes respectifs et donc le profil d'organisation définitive des répertoires des gènes OR et TAAR ne peut être connu. Toutefois en comparant le positionnement des gènes de tilapia, orthologues de ceux des quatre autres cichlidés on peut noter une forte tendance au regroupement des gènes orthologues au sein de scaffolds ainsi qu'on peut le voir dans le tableau 17 où sont positionnés sur la même ligne les gènes OR orthologues identifiés chez les différents cichlidés.

Tableau 17 : Les gènes OR orthologues identifiés chez les différents cichlidés.

TIL086	Sca._192							ZEB24	Sca.00046
TIL087	Sca._192								
TIL088	Sca._192			BRI15	Sca._45	BUR14	Sca._227		
TIL089	Sca._192			" "		" "			
TIL090	Sca._192								
TIL091	Sca._192	NYE037	Sca._213	BRI18	Sca._45	BUR11	Sca._227	ZEB026	Sca.00046
TIL092	Sca._192	NYE038	Sca._213	BRI17	Sca._45	BUR12	Sca._227	ZEB025	Sca.00046
TIL093	Sca._192	NYE082	Sca._2620	BRI19	Sca._45			ZEB028	Sca.00046
TIL094	Sca._192	NYE035	Sca._213	BRI21	Sca._45	BUR76	Sca._1733	ZEB029	Sca.00046
TIL095	Sca._192								
TIL096	Sca._192								
TIL097	Sca._192								
TIL098	Sca._192	NYE034	Sca._213	BRI22	Sca._45	BUR85	Sca._4421	ZEB030	Sca.00046
TIL099	Sca._192								
TIL100	Sca._192								
TIL101	Sca._192								
TIL103	Sca._192			BRI24	Sca._45				
TIL104	Sca._192	NYE86	Sca._781	BRI25	Sca._45			ZEB038	Sca.00058

Les cases en vert soulignent l'appartenance à un même contig.

Dans ce tableau on trouve, tous les gènes *d'O.niloticus* appartenant au scaffold 192 positionné en LG14. Les orthologues positionnés dans chacun des autres poissons dans un même scaffold sont surlignés en vert. Les orthologues non surlignés appartiennent à d'autres scaffolds. La coïncidence de positionnement des orthologues sur des scaffolds orthologues suggère par ailleurs une conservation plus étendue des relations de syntenie permettant de proposer un placement des scaffolds orthologues comme cela apparaît dans les diagrammes obtenus avec le programme AutoGRAPH.

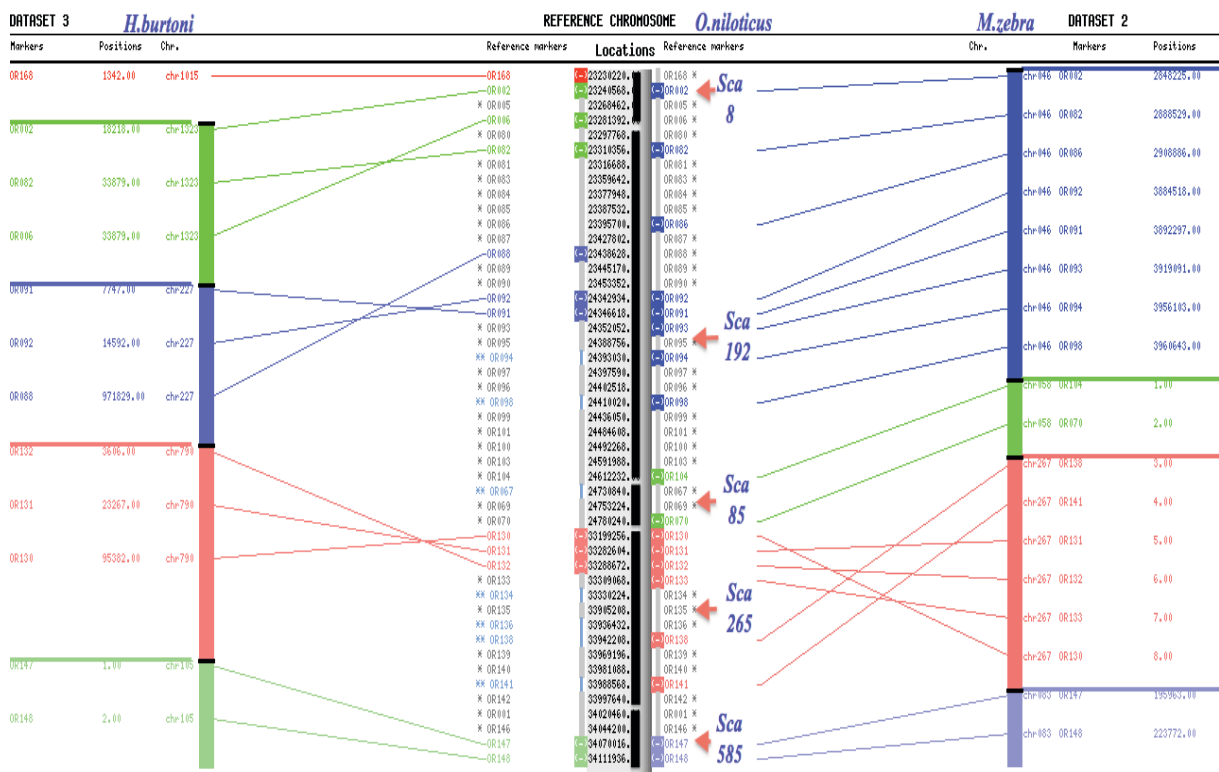
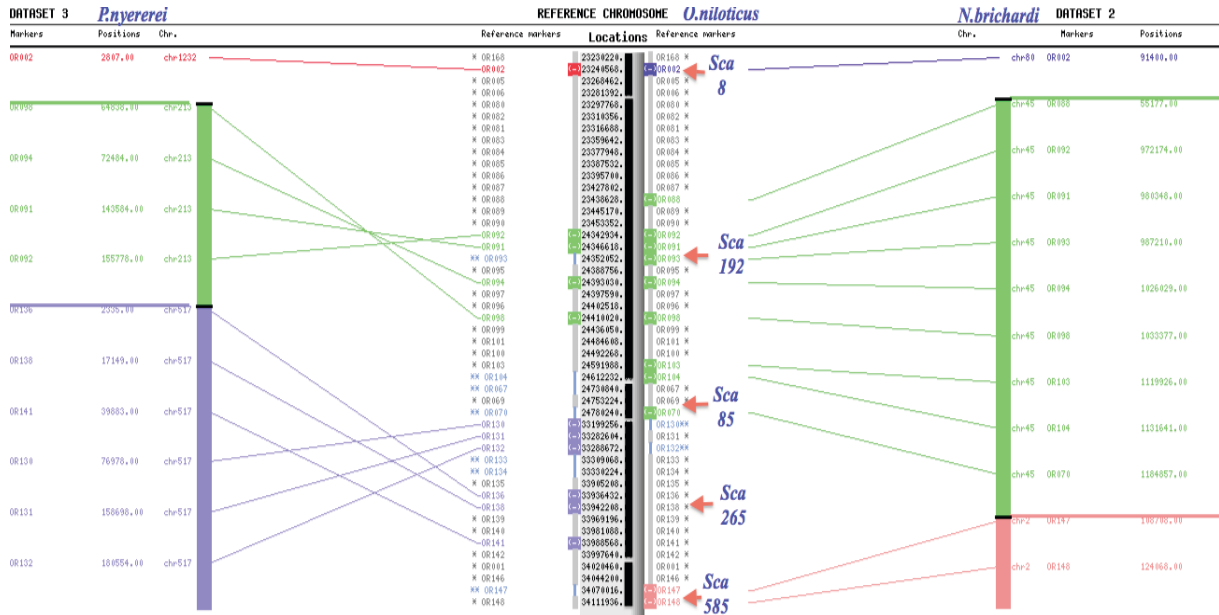
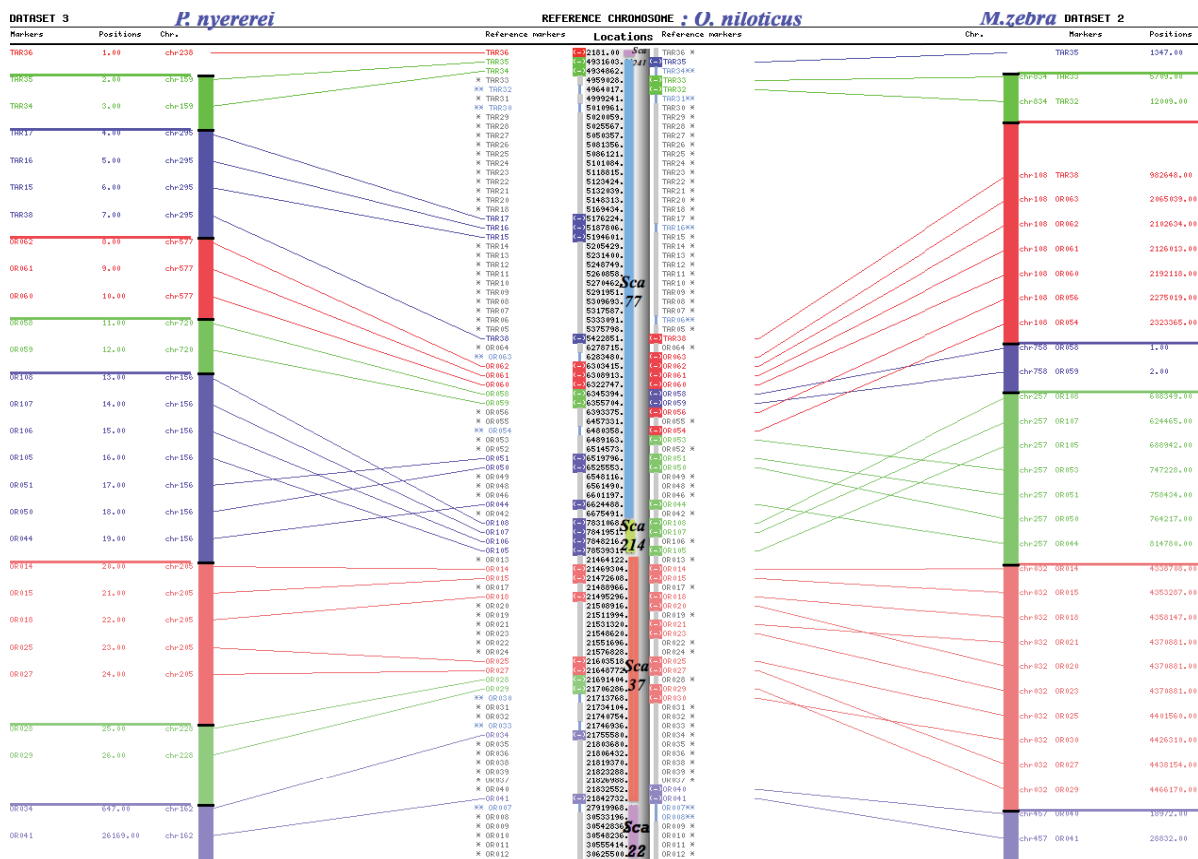


Figure 47 : Présentation avec AutoGRAPH des gènes OR identifiés sur le chromosome LG14 de *O. niloticus*. Ces gènes (ligne verticale du centre) sont alignés en fonction de leur position dans la séquence. Accolés à cette liste se trouvent les noms des scaffolds correspondant. A gauche sont listés les gènes identifiés chez *P. nyererei* et *H. burtoni*, à droite ceux identifiés chez *M. zebra* et *N. brichardi*. Les barres verticales indiquent l'étendue des différents scaffolds. Les lignes horizontales ou obliques indique la correspondance entre gènes orthologues.

La même analyse effectuée sur les gènes du chromosome LG16-21 où se trouvent un cluster d'OR et un cluster de TAAR donne des résultats similaires.



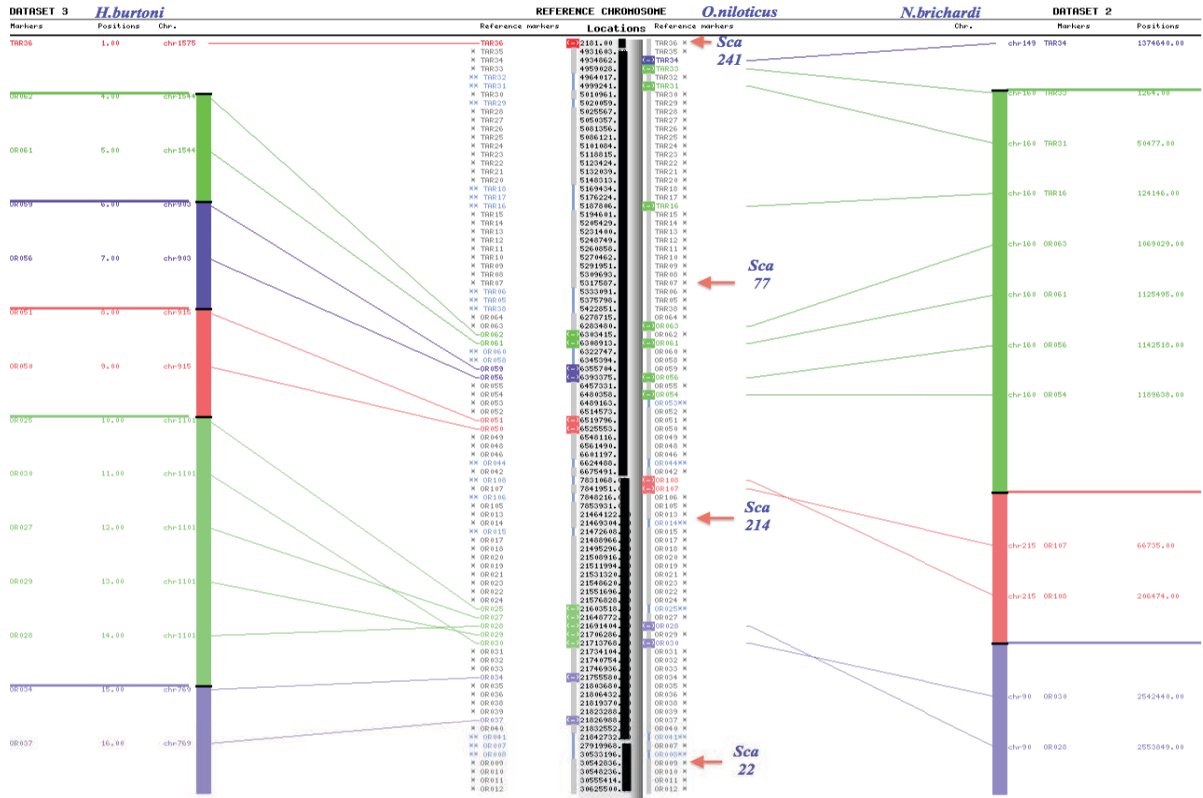


Figure 48 : relation de synténie des gènes OR et TAAR localisés sur le chromosome LG16-21 de *Tilapia* et de leurs orthologues chez *P. nyererei* et *M. zebra* d'une part et *H. burtoni* et *N. brichardi* d'autre part.

Un article (N°6) est en cours de rédaction par l'ensemble du consortium

Federica di Palma *et al.*, 2013

Discussion & Conclusion

I- Panel et carte d'Hybride d'Irradiation (RH) chez les espèces aquacoles

Dans un monde où la demande alimentaire croît, alors que les ressources terrestres sont limitées et que les impacts globaux des activités sur les écosystèmes doivent être pris en compte, une amélioration de l'efficacité globale de la production piscicole devient indispensable. Les objectifs et les méthodes de sélection doivent donc évoluer. Au cours de ces dernières années, les programmes d'élevages des espèces aquacoles se sont intéressés à l'amélioration génétique des poissons, comme pour les autres productions animales, élément clé du développement d'élevages performants.

Pour accompagner un développement durable de l'aquaculture, les critères de sélection doivent évoluer vers une amélioration de l'efficacité de la production (efficacité alimentaire, résistance aux maladies) et de l'adaptation au milieu extérieur. Les progrès de la sélection aquacole, peuvent s'appuyer sur les technologies génomiques pour à court terme, utiliser des pédigrés obtenus par génotypage de marqueurs et, à plus long terme, tirer parti d'une meilleure connaissance des bases génétiques des caractères d'intérêt et exploiter ces informations génétiques directement dans le processus de sélection (sélection génomique), sans négliger le potentiel incontestable des méthodes « classiques ». Une des approches génomiques qui permet de positionner des marqueurs d'intérêt pour l'aquaculture en particulier et pour l'élevage en général est la cartographie RH.

1- L'intérêt de la cartographie pour la sélection animale

La cartographie par hybrides d'irradiation permet de positionner des marqueurs génétiques extrêmement divers qu'ils soient polymorphes ou non polymorphes. Cette flexibilité permet aux cartes RH de faire la jonction entre les cartes génétiques et les données de séquençage. Par ailleurs le nombre de marqueurs à positionner sur une carte RH est lui-même extrêmement souple. On peut en effet construire des cartes de quelques centaines de marqueurs ou de plusieurs dizaines de milliers, dès lors qu'on adapte le niveau d'irradiation au but recherché ainsi que la méthode de typage qui peut être traditionnelle et faire appel à la PCR et à l'électrophorèse en gels d'agarose ou bien plus élaborée et faire appel à l'hybridation sur des puces comportant des milliers de marqueurs (oligonucléotides) ou bien encore au séquençage massif. La carte RH permet ainsi d'identifier des marqueurs utiles à la sélection de caractères phénotypiques d'intérêt agronomique au travers de la SAM ou sélection assistée par marqueurs ou bien aider à la recherche de gènes responsables de pathologie au travers d'une approche dite de gènes candidats.

1-1-Sélection assistée par marqueurs (SAM) et cartes RH

Pour les besoins de l'Aquaculture qui produit le tiers de la consommation mondiale de poissons, la carte RH peut contribuer à une sélection assistée par marqueurs. Celle-ci consiste à choisir précocement dans une descendance les sujets qui possèdent le ou les traits phénotypiques recherchés en s'appuyant sur l'identification de marqueurs situés au plus près des gènes conférant le ou les caractères d'intérêt comme la résistance au stress, à certaines infections ou liés à la croissance. Dans d'autres cas, comme l'esturgeon pour lequel la maturité sexuelle est très tardive (au delà de 10 ans) et pour lequel la valeur économique du mâle est faible par rapport à celle de la femelle eut égard à la production des œufs de caviar, il y aurait grand intérêt à déterminer le plus précocement possible le statut sexuel des animaux. Le problème consiste donc dans un premier temps à identifier les marqueurs pertinents puis à les utiliser au mieux au cours des étapes de production. Pour cela l'établissement de carte RH dense, riche de marqueurs polymorphes de type SNP pour ancrer la carte RH sur des données génétiques et de gènes dont les orthologues sont connus dans d'autres espèces déjà bien étudiées est une aide très précieuse. Pour revenir sur le cas de l'esturgeon pour lequel nous avons produit un panel de 94 clones hybrides à partir de 474 clones indépendants de cellules irradiées, la non acceptation par un comité ANR de subventionner la construction du panel et celle d'une carte de bonne densité nous a obligé à interrompre ce programme dont on peut pourtant prédire qu'il aurait eu des retombées très importantes pour la filière.

1-2-Identification de gènes impliqués dans l'expression d'un caractère phénotypique

Il existe deux stratégies majeures d'identification de gènes impliqués dans l'expression d'un caractère phénotypique d'intérêt. Elles sont bien développées en génétique humaine et sont également utilisées dans l'identification de gènes chez les animaux (Eggen, 2000). Il s'agit de l'approche appelée « gène candidat » et de l'approche par clonage positionnel.

1-2-1-Approche « gène candidat »

Cette approche s'appuie sur des arguments physiologiques. Elle consiste dans un premier temps à rechercher si parmi les gènes dont le rôle physiologique est connu, ceux d'entre eux qui seraient susceptibles d'expliquer le phénotype observé ou recherché puis de tester la validité des hypothèses.

1-2-2- Stratégie de clonage positionnel

Le clonage positionnel est l'approche de choix lorsque l'on ignore tout produit du gène responsable d'une maladie génétique ou d'un trait phénotypique d'importance et qu'aucun

gène n'apparaît comme un bon candidat. Dans ce cas on recherche par des analyses génétiques la région chromosomique susceptible d'abriter le ou les gènes candidats. Dans ces deux approches ainsi définies l'utilisation de cartes denses et informatives est indispensable

2-Intérêt des cartes RH pour le séquençage

Au cours de ces dernières années, plusieurs programmes de cartographie et de séquençage de génomes de poissons ont été menés. L'un des derniers concerne les génomes de cinq cichlidés : *O. niloticus*, *P. nyererei*, *H. Burtoni*, *N. brichardi* et *M. zebra*.

Comme cité ci-dessus, le séquençage par NGS (next generation sequencing or new genome sequencing) permet d'avoir une très bonne idée du niveau d'analogie entre espèces, mais il ne fournit que des informations partielles sur la structure des génomes, car l'assemblage de ces données de séquençage n'est que partielle en raison de la faible longueur des lectures de séquence mais aussi de la non ou faible représentativité de certaines régions des génomes séquencés.

La cartographie RH qui se fonde sur la fréquence de la répartition de marqueurs dans des lignées cellulaires hybrides est un complément très important sinon indispensable à l'assemblage de ces données et à l'analyse de la structure des génomes car elle permet d'une part de positionner des « contigs ou des scaffolds » les uns par rapport aux autres et d'autre part de confirmer certaines prédictions d'assemblage. Ainsi l'amélioration spectaculaire des capacités de séquençage des génomes loin de rendre obsolète la cartographie RH, demande au contraire une amélioration parallèle de cette dernière (Hitte C *et al.*, 2009).

La construction de cartes RH de mammifères terrestres est bien connue mais pas celle des espèces piscicoles qui pendant longtemps a eu la réputation d'être très difficile sinon impossible. Notre laboratoire a publié en 2006 une première carte RH de la daurade (Senger *et al.*, 2006). Par la suite nous avons produit une carte RH du bar puis du tilapia (Guyon *et al.*, 2010, 2012). Par ailleurs dans le cadre d'autres partenariats nous avons produit un panel RH d'esturgeon, d'huitre (Azzouzi *et al.*, 2013) et enfin de truite et de crocodile (résultats non publiés). Au cours de la construction de ces divers panels RH et cartes nous avons rencontré toutes sortes de difficultés dont il ressort que le premier problème est celui du choix de la cellule donneuse. D'une part on ne dispose pas toujours de lignées cellulaires à partir desquelles on pourrait procéder à l'irradiation mais par ailleurs les lignées cellulaires sont d'une façon générale une mauvaise solution, même si elles ont été utilisées par certains auteurs (Hukriede N *et al.*, 2001). En effet elles sont l'objet de remaniements chromosomiques et dès lors la carte RH qui pourrait en être obtenue ne serait pas un reflet

fidèle du génome de l'espèce d'origine mais de la lignée cellulaire. Il convient donc de partir d'un prélèvement frais à partir duquel on dérive plusieurs centaines de millions de cellules. Après avoir tester plusieurs tissus, il ressort que la rate qui peut par broyage très doux avec un broyeur de type « potter » fournir une très grande quantité de splénocytes ou le sang circulant duquel l'obtention de cellules nucléées est relativement aisée sont deux organes à privilégier. Dans le cas de l'huitre le broyage du muscle du pied de l'huitre s'est également révélé adéquat pour fournir en quantité suffisante des cellules isolées se prêtant directement à l'irradiation.

Le second problème majeur est celui du gène de résistance apporté par la cellule donneuse qui doit compléter le gène déficient de la cellule receveuse afin de permettre la sélection des cellules hybrides dans un milieu de sélection approprié. En réalité à ce jour on ne dispose que de deux gènes de sélection celui de la thymidine kinase et celui de l'hypoxanthine-guanine phosphoribosyltransferase (HPRT). De sorte que si aucun de ces deux gènes de la cellule donneuse n'est fonctionnel dans l'environnement cellulaire apporté par la cellule CHO receveuse, ou bien si la protéine produite n'est pas elle-même fonctionnelle on ne dispose actuellement d'aucune solution viable. Dans les cas de la daurade, du bar, du tilapia, de l'esturgeon ou du crocodile l'un ou l'autre voire ces deux gènes étaient fonctionnels. Dans le cas de la truite, nous avons obtenu à partir des splénocytes et une sélection avec le gène HPRT des clones cellulaires en nombre satisfaisant, toutefois ceux-ci ont eu beaucoup de difficulté à croître pour réaliser les deux passages nécessaires à l'obtention de cellules hybrides en nombre suffisant pour produire après extraction assez d'ADN. Cette difficulté a été contournée en procédant avant tout repiquage, directement à partir des clones isolés de quelques milliers de cellules à une extraction de l'ADN, suivie d'une amplification de celui-ci par Genomiphi. Le cas de l'huitre a été encore plus délicat à traiter. Aucun des deux gènes n'a permis d'obtenir des clones de belle taille comme dans le cas de la truite. La encore une amplification précoce par Genomiphi a été réalisée. Mais dans ce cas, sans qu'on en connaisse la raison précise, la taille des fragments retenus par les clones hybrides s'est révélée très petite. Sans empêcher en principe la construction d'une carte RH comme nous avons pu le montrer avec des marqueurs préparés à partir d'une séquence de BAC, la construction d'une carte RH de tout le génome aurait nécessité de positionner plusieurs centaines de milliers de marqueurs, tâche difficilement réalisable. Seule la construction d'une carte dense d'une région d'intérêt particulier aurait pu s'envisager avec ce panel. De ces expériences il ressort que le problème de la sélection des hybrides est le plus préoccupant d'autant que la capacité des gènes TK ou HPRT d'être fonctionnels est totalement imprévisible. Reste donc qu'il

faudrait développer dans ces cas là une approche totalement différente comme la transfection avant irradiation des cellules donneuses avec un clone exprimant un gène TK ou HPRT humain ou de souris et que l'on sait être fonctionnel dans un environnement CHO, ce que nous avons tenté de faire. Mais les rendements des transfections réalisées dans des conditions standard n'ont pas permis d'obtenir des clones en quantité suffisante pour être une solution viable, reste néanmoins que cette approche mériterait d'être reprise. Enfin un dernier aspect technique mériterait d'être amélioré. Il s'agit du typage sur chacune des lignées du panel des marqueurs préalablement choisis. La PCR a été pendant longtemps la méthode de choix. Toutefois celle-ci est longue, fastidieuse et finalement coûteuse si on souhaite positionnée plusieurs milliers de marqueurs. En effet cela revient à réaliser quelques 100.000 PCR suivies de leur analyse pour construire une carte de 1000 marqueurs. Pour la construction de la carte du tilapia nous avons développé une autre approche dérivée d'une méthode utilisée pour la détermination allélique des SNP, mais cette approche pour être plus efficace que la PCR, n'est pas encore parfaite. Une autre approche pourrait mettre à profit un système de microfluidique permettant une analyse PCR massivement parallèle (Colin B *et al.*, 2008). Pour être encore plus efficace, on pourrait aussi songer à une approche dérivée de l'utilisation des NGS. Celle-ci consisterait à séquencer à une faible profondeur le contenu en ADN de chaque lignée et à trier les marqueurs pertinents des séquences de hamster. Outre la possibilité de cartographier des dizaines de milliers de marqueurs en un temps raisonnable, l'approche NGS présenterait un autre avantage technique. En effet l'identification de la présence des marqueurs dans chaque lignée basée sur la détermination des séquences permettrait de distinguer les uns des autres des marqueurs paralogues de séquence proche, ce que ne peut faire que très difficilement la PCR. Cette capacité supplémentaire offerte par la séquence devrait ainsi permettre de résoudre avantageusement les problèmes posés par les duplications génomiques partielles ou totales qui sont fréquentes surtout pour des génomes d'espèces ayant subi plusieurs cycles de diploïdisation comme l'esturgeon (Ludwig A *et al.*, 2001).

II- Annotation des génomes de cichlidés et identification des gènes OR et TAAR

L'olfaction constitue l'un des sens majeurs qui permet aux animaux d'appréhender le monde extérieur. Tous les animaux ne présentent pas les mêmes capacités olfactives. Ainsi le chien, le rat, mais vraisemblablement beaucoup d'autres mammifères sauvages possèdent un sens olfactif très développé. L'homme au contraire a perdu en grande partie celui-ci au bénéfice d'autres fonctions cognitives vraisemblablement.

C'est en 1991 que pour la première fois des récepteurs olfactifs ont été isolés par des

expériences de clonage chez le rat (Buck et Axel Cell 1991). Depuis soit par clonage moléculaire mais le plus souvent par annotation des données de séquençage les répertoires de gènes de récepteurs olfactifs de nombreux animaux ont été établis (Kiefer *et al.*, 1996 ; Zhao *et al.*, 1998 ; Hatt *et al.*, 2001 ; Quignon *et al.*, 2005). Quelques années plus tard Fredriksson *et al.*, 2003; Gloriam *et al.*, 2005 ; puis Liberlès et Buck 2006 démontraient l'existence d'une autre catégorie de gènes s'exprimant dans l'épithélium olfactif comme les OR. Ceux-ci appelés TAAR, pour trace amine-associated receptors, sont impliqués dans la détection des amines à l'état de trace.

Ainsi que cela a été rappelé précédemment la carte RH de tilapia que nous avons construite a entre autre été utilisée par le BROAD Institute au cours de l'assemblage des données de séquence du génome de tilapia, un des cinq cichlidés à faire partie d'un vaste programme systématique de séquençage des génomes. Notre implication dans cette étape analytique nous a amené à participer à l'annotation des génomes de ces cinq poissons et plus particulièrement à l'identification des répertoires des gènes OR et TAAR.

Pour cette tâche nous avons utilisé plusieurs programmes informatiques de recherche, de comparaison, d'analyse des séquences d'ADN et de protéines. Nous avons par ailleurs pour faciliter et accélérer l'acquisition des données écrit plusieurs petits sous-programmes destinés à enchaîner les opérations. Comme cela a été décrit dans la partie des résultats, les répertoires des gènes OR et TAAR de ces cinq cichlidés sont composés d'une centaine de gènes par poisson avec une proportion faible de pseudogènes, proportion en tout cas plus faible que celle observée chez des poissons modèles comme le poisson zèbre ou medaka. L'exploitation des données de comparaison des séquences obtenues par Blast, ont apporté deux éléments majeurs pour la compréhension de l'histoire de l'évolution des gènes OR et TAAR. D'une part, des séquences présentes dans les génomes d'au moins deux des cinq poissons peuvent être très similaires voire identiques à 100% (BUR64 et ZEB 084, BUR48 et ZEB 069, BUR59 et NYE025, NYE009 et ZEB073) suggérant un événement de duplication précédant de peu la spéciation Burtoni / Nyererie, Burtoni / Zebra et Nyererie /Zebra ou une pression de sélection très forte à l'endroit de ces gènes. Par ailleurs, dans la séquence de tilapia, on a trouvé deux gènes totalement identiques, TIL 078 et TIL137 présents dans deux contigs différents. S'agit-il du signe d'une duplication récente ou d'une erreur d'assemblage ?

D'autre part de façon totalement inattendue nous avons identifié des gènes OR et des gènes TAAR présentant plusieurs exons codants. Ceci est nouveau puisque on a toujours considéré jusque là que les gènes OR et TAAR des mammifères et des poissons à la différence de ceux des invertébrés (nématodes et insectes) ne possédaient que deux exons dont un seul codant.

Certes on pourrait arguer que ces gènes résultent d'artefacts de séquençage ou ne sont pas fonctionnels. Si seule une analyse des transcrits exprimés par l'épithélium olfactif serait de nature à lever ce doute, deux éléments de structure de ces gènes suggèrent très fortement que ce ne sont pas des artefacts et qu'ils sont selon toute vraisemblance fonctionnels. Le premier argument tient dans la structure des séquences de jonction exon-intron qui sont très conservées. Le deuxième argument est apporté par la position des introns par rapport à la structure des gènes et de la structure bi-dimensionnelle des récepteurs. Pour les OR à deux exons codants, l'intron les séparant se situe le plus souvent dans la séquence définissant la première boucle interne. Dans le cas de gènes OR avec trois exons codants le deuxième intron se positionne dans la partie du gène codant la région transmembranaire 4. Pour les gènes TAAR, la situation est plus simple, tous les gènes épissés ont deux exons codants et l'intron les séparant est toujours (à une exception près) situé dans la région codant la première boucle interne.

Si ce fait était définitivement établi par l'analyse des transcrits et encore mieux une analyse fonctionnelle déterminant le ligand de certains d'entre eux, la question immédiate qui se poserait serait celle de l'existence de gènes OR et TAAR multi-exons dans d'autres espèces de poissons et autres vertébrés plus largement. A cet égard, il convient de souligner que l'absence d'intron dans la partie codante des gènes OR et TAAR de poissons comme dans ceux de tous les vertébrés, est encore maintenant fortement établie à l'image d'un dogme. C'est ainsi, que Alioto et Ngai dans un travail publié en 2005 visant à identifier le répertoire complet des gènes OR de zebrafish n'hésitaient pas à écrire « First, all coding sequences were required to be longer than 700 base pairs in length and show highest sequence similarity to previously characterized OR sequences ». De même, Lindermann et Hoener écrivaient en 2005 « The trace amine-associated receptors (TAARs) form a specific family of G protein-coupled receptors (GPCRs) that are single-exon encoded and have coding sequences of about 1 kb in length », propos repris par ailleurs par Yasuyuki Hashiguchi et Mutsumi Nishida (2007) dans un travail intitulé « Evolution of Trace Amine-Associated Receptor (TAAR) Gene Family in Vertebrates: Lineage-Specific Expansions and Degradations of a Second Class of Vertebrate Chemosensory Receptors Expressed in the Olfactory Epithelium ». La conséquence de cette prise de position quelque peu absolutiste est que les stratégies suivies pour extraire des séquences génomiques les répertoires des gènes OR ou TAAR excluaient d'emblée tout gène épissé et qu'une revisite de ces séquences génomiques serait souhaitable. Par ailleurs une recherche des gènes OR et TAAR dans d'autres phylums serait plus que souhaitable. Ceci étant posé, une analyse préliminaire des banques NCBI et ENSEMBL nous

a permis d'identifier plusieurs gènes OR de poissons (medaka, stickelback, zebrafish ou takifugu) formés de plus d'un exon codant, indiquant que ce problème devrait être analysé à nouveau sans *a priori*.

Le fractionnement des gènes eucaryotes en introns et exons est une source potentielle de diversité protéique. En effet, dès lors qu'un gène comporte au moins deux exons codants, un processus d'épissage alternatif augmentant grandement la diversité protéique peut être mis en place (Graveley, 2001; Smith and Valcarcel, 2000). Sans éliminer *a priori* cette possibilité pour les gènes OR ou TAAR épissés, il convient de noter que compte tenu de la structure tridimensionnelle très précise de ces récepteurs tout épissage alternatif éliminant un exon aboutirait à la production d'une protéine dépourvue des propriétés d'une GPCR. Pour autant ceci n'empêcherait nullement que le peptide produit suite à un épissage alternatif d'un gène OR ou TAAR n'ait une autre fonction. Par ailleurs étant donné que les gènes OR d'invertébrés comme ceux de la drosophile ou du nématode (Vosshall LB *et al.*, 1999 ; Robertson HM, 2000) et vraisemblablement ceux des poissons ou du moins de certains d'entre eux ont plusieurs exons, pourquoi les OR de mammifères seraient-ils seuls à n'être formés que d'un exon codant et de ce point de vue être une exception parmi les autres GPCR de classe A? En est-il réellement ainsi et à quel moment au cours de l'évolution cette transition se serait-elle produite ? Il est clair que ces questions et d'autres connexes méritent d'être approfondies et que pour commencer une analyse exhaustive des répertoires OR et TAAR des poissons modèles doit être menée avant d'envisager sérieusement le cas des mammifères et plus généralement des tétrapodes. D'un point de vue symétrique, la découverte en 1938 du premier cœlacanthe vivant, celui-ci ayant été jusque là considéré comme une espèce éteinte n'existant qu'à l'état de fossile à la jonction des poissons et des tétrapodes a été un événement considérable (Smith JLB. 1939). Les séquences des génomes de plusieurs espèces de cœlacanthe ont récemment été déterminées (Amemiya *et al.*, 2013; Nikaido *et al.*, 2013) et plus récemment encore une publication a fait état du répertoire des gènes OR, fondé sur l'analyse d'une de ces séquences (Picone *et al.*, 2013). Malheureusement en l'état aucune de ces trois publications ne permet d'aborder la question de la structure fine des récepteurs OR ou TAAR et de l'existence ou non de gènes multi-exons, ce qui serait du plus haut intérêt compte tenu de la position phylogénique des coelacanthe.

Les séquences des gènes OR comme les gènes TAAR épissés se distribuent dans des sous-familles particulières. Ceci suggère que leur genèse s'est faite par duplication successive d'un seul gène OR ou TAAR ancestral.

Le répertoire des gènes OR et TAAR identifiés dans cette étude montre que le nombre de

copies de gènes est comparable à celui trouvé chez d'autres poissons. Avec une taille de génome à peu près similaire, *O. niloticus* aurait le plus grand nombre de gènes OR fonctionnels (n=158) et *N. brichardi* le moins (n=69). Si l'idée que la capacité olfactive est en corrélation avec le nombre de gènes fonctionnels, l'olfaction serait peut-être plus essentielle à *O. niloticus* pour sa survie qu'à *N. brichardi*.

Comme cela a été rapporté dans l'introduction, le génome des poissons de l'ordre des téléostéens a été l'objet d'un troisième tour d'amplification globale dénommée WGD pour « whole genome duplication » en anglais. Les conséquences des cis-duplications des gènes OR et TAAR intervenues au cours de l'évolution sont évidentes lorsqu'on examine la composition des sous-familles et leur organisation génomique. Ainsi, la sous-famille A1 d' *O. niloticus* est composée de 13 gènes complets auxquels s'ajouteront vraisemblablement d'autres gènes parmi les OR actuellement incomplets et dénommés « edges », ou encore la sous-famille D1 dans laquelle à ce jour 11 gènes OR complets ont été identifiés. Ces gènes partagent une forte analogie de séquence. Ils sont localisés les uns à la suite des autres, conséquences des cis-duplications pour constituer les « scaffolds 77 et 37 » localisés sur le chromosome LG16-21. Ce mode d'organisation à l'échelle du génome n'a rien de particulier, on retrouve le même pour les répertoires d'OR de mammifères (Quignon *et al.*, 2005). Alors que la présence de familles ou sous-familles de gènes OR localisées sur plusieurs chromosomes rendrait facilement compte des deux premiers cycles d'amplification, les conséquences du troisième tour de WGD ayant affecté les poissons téléostéens ne sont pas très apparentes. Le nombre de gènes OR des poissons modèles et des cichlides est de l'ordre de la centaine comme celui des OR de classe 1 des mammifères, alors qu'on s'attendrait pour les poissons à retrouver des répertoires plus larges. D'un autre côté, l'organisation des gènes TAAR localisés sur plusieurs locus, rend bien compte de nombreux événements de cis- et trans-duplications résultants des WGD ainsi que leur nombre élevé supérieur à celui trouvé chez les mammifères (Liberles & Buck 2006).

L'Évolution de structure des gènes et relations phylogéniques

La nature des relations entre séquences protéiques entières orthologues et paralogues et de différentes régions protéiques (transmembranes, boucles extra et intracellulaires) a été examinée afin de déterminer les caractéristiques évolutives de ces gènes. Il apparaît que les gènes OR et TAAR dans leur ensemble ont évolué sous une pression de sélection positive c'est à dire favorisant la diversité protéique, propice à l'amélioration qualitative et

quantitative des capacités olfactives. Par ailleurs l'analyse de la fréquence des mutations synonymes et non synonymes faite avec une fenêtre glissante montre pour les gènes TAAR des valeurs relativement élevées de dN ciblant certaines régions comme TM4 et EL2. A l'inverse la région IL2 favoriserait l'accumulation des mutations synonymes. En revanche pour les gènes OR, nous n'avons pas trouvé de régions protéiques favorisant l'un ou l'autre type de mutations. Dans un travail similaire effectué sur les répertoires des gènes OR de medaka et stickelback, Chen Ming *et al.* (2010) ont identifié les régions transmembranaires TM4, TM5 et TM6 ainsi que la région EL2 comme ayant le rapport Ka/Ks le plus élevé, indiquant qu'une pression de sélection positive s'exercerait davantage sur ces parties des molécules réceptrices où plusieurs auteurs ont par ailleurs suggéré qu'elles participeraient au site de liaison des molécules odorantes (Chen Ming *et al.*, 2010 ; Man *et al.*, 2004). La sélection positive détectée chez les cinq espèces de cichlidés suggère pour celles-ci un besoin et ou une capacité d'adaptation aux environnements spécifiques qu'ils rencontrent.

Conclusion générale

Aujourd'hui, de grands défis se posent à la pisciculture, comme pour d'autres secteurs d'élevage animal. D'un côté, la limitation des ressources halieutiques disponibles, les menaces sur les milieux aquatiques continentaux et côtiers et la nécessité d'une croissance en raison de l'incapacité de la pêche à augmenter sa production vont nécessiter un accroissement important de l'efficacité des systèmes piscicoles. D'un autre côté, les changements que la pisciculture intensive fait subir au milieu externe risque d'induire en retour des modifications importantes sur les élevages. Contrairement aux espèces terrestres, les poissons sont des animaux à sang-froid, donc très sensibles au moindre changement de la température du milieu et très dépendants de leurs milieux d'élevages et de leurs caractéristiques biologiques (agents pathogènes) ou physicochimiques (polluants, oxygène).

Pour relever ces défis l'amélioration des connaissances de la génétique des espèces élevées est importante voire primordiale. Mon travail de thèse au cours duquel j'ai réalisé des panels RH de différentes espèces aquacoles avec ou sans cartographie montre la faisabilité et le bien-fondé de cette approche même si chaque nouvelle espèce reste un cas particulier. Par ailleurs l'identification des gènes codant pour les récepteurs olfactifs et les TAAR de cinq cichlidés ouvre la voie à de nouvelles études sur l'évolution de ces gènes particulièrement importants pour la physiologie et l'adaptation des poissons à un environnement en perpétuel changement.

BIBLIOGRAPHIE

- Abaffy T, Malhotra A, and Luetje C. W.** *The Molecular Basis for Ligand Specificity in a Mouse Olfactory Receptor : A network of functionally important residues* J Biol Chem 282, 1216-24. (2007)
- Alioto TS and Ngai J.** The odorant receptor repertoire of teleost fish. BMC Genomics. Dec 6; 6:173. (2005)
- Amemiya CT, Prohaska SJ, Hill-Force A, Cook A, Wasserscheid J, Ferrier DE, Pascual-Anaya J, Garcia-Fernández J, Dewar K, Stadler PF** The amphioxus Hox cluster : characterization, comparative genomics, and evolution. J Exp Zool B Mol Dev Evol 310:465–477 (2008)
- Andersson J.O.** Lateral gene transfer in eukaryotes. *Cell Mol Life Sci*, **62**, 1182-1197. (2005)
- Andersson L, Archibald A, Ashburner M, Audun S, Barendse W, Bitgood J, Bottema C, Broad T, Brown S, Burt D, Charlier C, Copeland N, Davis S, Davisson M, Edwards J, Eggen A, Elgar G, Eppig JT, Franklin I, Grewe P, Gill T 3rd, Graves JA, Hawken R, Hetzel J, Womack J, et al.** Comparative genome organization of vertebrates. The First International Workshop on Comparative Genome Organization. *Mamm Genome.*; **7**: 717. (1996)
- Barski. G, Sorieul. S, Cornefert F.** Production dans des cultures in vitro de deux souches cellulaires en association, de cellules de caractère "hybride". *CR Acad Sci (Paris)*, 251 18. (1960).
- Balakirev. E.S, and Ayala. F.J.** Pseudogenes : are they "junk" or functional DNA? *Annu Rev Genet*, **37**, 123-151. (2003)
- Barlow GW.** The Cichlid Fishes : Nature's Grand Experiment In Evolution. Basic Books, 15 déc. (2008)
- Bernardi. G, Wiley. EO, Mansour. H, Miller. MR, Orti. G, Haussler. D, O'Brien. SJ, Ryder. OA, Venkatesh. B.** The fishes of Genome 10K. *Mar Genomics*. Sep; **7** : 3-6. (2012)
- Ben-Arie N, Lancet D, Taylor C, Khen M, Walker N, Ledbetter DH, Carrozzo R, Patel K, Sheer D, Lehrach H, et al.** Olfactory receptor gene cluster on human chromosome 17: possible duplication of an ancestral receptor repertoire. *Human Molecular Genetics*; **3**, 229-235. (1994)
- Brennan. PA, Binns. EK.** Vomeronasal mechanisms of mate recognition in mice. *Chem Senses*. Jan; **30** Suppl 1 :148-9. (2005)

Brenan C, Morrison T, Roberts D, Hurley J. A nanofluidic system for massively parallel PCR Proc. SPIE 6886, Microfluidics, BioMEMS, and Medical Microsystems VI, 688609 (2008)

Brosius. J. The contribution of RNAs and retroposition to evolutionary novelties. *Genetica*, **118**, 99-116. (2003)

BROADHURST. C.L, CUNNANE. S.C, CRAWFORD. M.A. Rift Valley lake fish and shellfish provided brain-specific nutrition for early Homo. *Br. J. Nutr.*, **79**, 3-21. (1998)

Buck. L.B and Axel. R. A novel multigene family may encode receptors: a molecular basis for odor recognition. *Cell* **5**; 175-187. (1991)

Burt. D.W, Bruley. C, Dunn. I.C, Jones. C.T, Ramage. A, Law. A.S, Morrice. D.R, Paton. I.R, Smith. J, Windsor D, Sazanov. A, Fries. R, and Waddington. D. The dynamics of chromosome evolution in birds and mammals. *Nature*, **402**, 411-413. (1999)

Bucher HC, Hengstler P, Schindler C, Meier G. N-3 polyunsaturated fatty acids in coronary heart disease: a meta-analysis of randomized controlled trials. *Am J Med. Mar*; **112**(4):298-304. (2002)

Chen. M, Peng. Z, He S. Olfactory receptor gene family evolution in stickleback and medaka fishes. *Sci China Life Sci. Feb*; **53** (2):257-66. (2010)

Cox. DR, Burmeister. M, Price. ER, Kim. S, Myers. RM. Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. *Science. Oct* **12**; **250**(4978):245-50 (1990).

Crawford MA, Bloom M, Broadhurst CL, Schmidt WF, Cunnane SC, Galli C, Gehbremeskel K, Linseisen F, Lloyd-Smith J, Parkington J. Evidence for the unique function of docosahexaenoic acid during the evolution of the modern hominid brain. *Lipids*, **34** Suppl., S39-47. (1999)

CRIBIU. E.-P, SCHIBLER. L. Introduction à la cartographie des génomes complexes. *INRA Prod. Anim.*, 69-72. (2000)

Derrien T, André C, Galibert F, Hitte C. AutoGRAPH : an interactive web server for automating and visualizing comparative genome maps. *Bioinformatics. Feb* **15**; **23**(4):498-9. (2007)

De Robertis EM. Evo-devo : variations on ancestral themes. *Cell. 25*; **132**(2): 185–195. (2008)

Duboule D. The rise and fall of Hox gene clusters. *Development*; **134**: 2549–2560 (2007).

Dulac. C, Axel. R, A novel family of genes encoding putative pheromone receptors in mammals. *Cell*. Oct 20; 83(2):195-206. (1995)

Eichler. E.E, and Sankoff. D. Structural dynamics of eukaryotic chromosome evolution. *Science*, **301**, 793-797. (2003)

Emes. RD, Beatson. SA, Ponting. CP, Goodstadt. L. Evolution and comparative genomics of odorant- and pheromone-associated genes in rodents. *Genome Res*. Apr; 14 (4) :591-602. (2004)

Fitch W.M, Leiter J.M.E, Li X, Palese P. Positive Darwinian evolution in human influenza A viruses. *Proceedings of the National Academy of Sciences of the USA* **88** : 4270-4274. (1991).

Freitag. J, Ludwig. G, Andreini. I, Rössler. P. and Breer. H. Olfactory receptors in aquatic and terrestrial vertebrates. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology* Volume 183, Number 5 , 635-650 (1998).

Fredriksson R, Lagerström MC, Lundin LG, Schiöth HB. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol Pharmacol*. Jun; 63(6):1256-72. (2003)

Francino M P. An adaptive radiation model for the origin of new gene functions. *Nat Genet*, **37**: 573-577. (2005)

Genereux. D.P and Logsdon J.M. Much ado about bacteria-to-vertebrate lateral gene transfer. *Trends Genet*, **19**, 191-195. (2003)

GISSI. Dietary supplementation with n-3 polyunsaturated fatty acids and vitamin E after myocardial infarction : results of the GISSI-Prevenzione trial. *The Lancet*, **354** (9177), 447-455. (1999).

Gilad Y, Przeworski M, Lancet D. Loss of olfactory receptor genes coincides with the acquisition of full trichromatic vision in primates. *PLoS Biol*. Jun; 5 (6):e148. (2007)

Gilad Y, Bustamante CD, Lancet D, Pääbo S. Natural selection on the olfactory receptor gene family in humans and chimpanzees. *Am J Hum Genet*. Sep; 73(3) :489-501. (2003).

Glusman. G, Yanai. I, Rubin. I, Lancet D. The complete human olfactory subgenome. *Genome Res* **11**: 685–702. (2001).

Gloriam. DE, Bjarnadóttir. TK, Yan. YL, Postlethwait. JH, Schiöth. HB, Fredriksson. R. The repertoire of trace amine G-protein-coupled receptors : large expansion in zebrafish. *Mol Phylogenet Evol.* May; 35(2):470-82. (2005)

Goldman. N, Yang Z. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol.* Sep; 11(5) :725-36. (1994)

Grus WE and Zhang J. Rapid turnover and species-specificity of vomeronasal pheromone receptor genes in mice and rats. *Gene.* Oct 13; 340 (2):303-12. (2004)

Graveley. BR. Alternative splicing : increasing diversity in the proteomic world. *Trends Genet.* Feb; 17(2):100-7. (2001)

Gu X, Wang Y, and Gu J. Age distribution of human gene families shows significant roles of both large- and small-scale duplications in vertebrate evolution. *Nat Genet*, **31**, 205-209. (2002)

Guyon R, Senger F, Rakotomanga M, Sadequi N, Volckaert FA, et al. A radiation hybrid map of the European sea bass (*Dicentrarchus labrax*) based on 1581 markers : Synteny analysis with model fish genomes. *Genomics* 96 : 228-238. (2010)

Guyon R, Rakotomanga M, Azzouzi N, Coutanceau JP, Bonillo C, et al. A high-resolution map of the Nile tilapia genome : a resource for studying cichlids and other percomorphs. *BMC Genomics* 13 : 222. (2012)

Gyapay G, Schmitt K, Fizames C, Jones H, Vega-Czarny N, Spillett D, Muselet D, Prud'homme JF, Dib C, Auffray C, Morissette J, Weissenbach J, Goodfellow PN. A radiation hybrid map of the human genome. *Hum Mol Genet.* Mar;5(3):339-46 (1996).

Hayes. H, Petit. E, Lemieux. N, Dutrillaux. B. Chromosomal localization of the ovine beta-casein gene by nonisotopic in situ hybridization and R-banding. *Cytogenet Cell Genet.*; 61(4):286-8. (1992).

Harris. H, Watkins. JF, Campbell. GL, Evans. EP, Ford. CE. Mitosis in hybrid cells derived from mouse and man. *Nature*, 207 (997), 606-8. (1965)

Hale MC, McCormick CR, Jackson JR, Dewoody JA. Next-generation pyrosequencing of gonad transcriptomes in the polyploid lake sturgeon (*Acipenser fulvescens*): the relative merits of normalization and rarefaction in gene discovery. *BMC Genomics.* Apr 29; 10 :203 (2009).

Hatt. H, Lang. K, Gisselmann. G. Functional expression and characterization of odorant receptors using the Semliki Forest virus system. *Biol Chem.* Aug; 382(8):1207-14. (2001)

Hett.A.K, and Ludwig.A. SRY-related (Sox) genes in the genome of European Atlantic sturgeon (*Acipenser sturio*). *Genome*, 48 :181–186. (2005)

Hett.A.K, Pitra.C, Jenneckens.I, Ludwig. A. Characterization of Sox9 in European Atlantic sturgeon (*Acipenser sturio*). *J Hered.* Mar-Apr; 96(2):150-4. (2005)

Hett.AK, Ludwig. A. SRY-related (Sox) genes in the genome of European Atlantic sturgeon (*Acipenser sturio*). *Genome.* Apr; 48 (2) :181-6. (2005).

Herrada. G, Dulac. C. A novel family of putative pheromone receptors in mammals with a topographically organized and sexually dimorphic distribution. *Cell.* Aug 22; 90 (4):763-73. (1997)

Hitte C, Madeoy J, Kirkness EF, Priat C, Lorentzen TD, Senger F, Thomas D, Derrien T, Ramirez C, Scott C, Evanno G, Pullar B, Cadieu E, Oza V, Lourgant K, Jaffe DB, Tacher S, Dréano S, Berkova N, André C, Deloukas P, Fraser C, Hori. M. Frequency-dependent natural selection in the handedness of scale-eating cichlid fish. *Science.* Apr 9; 260(5105):216-9. (1993)

Hukriede N, Fisher D, Epstein J, Joly L, Tellis P, Zhou Y, Barbazuk B, Cox K, Fenton-Noriega L, Hersey C, Miles J, Sheng X, Song A, Waterman R, Johnson SL, Dawid IB, Chevrette M, Zon LI, McPherson J, Ekker M. The LN54 radiation hybrid map of zebrafish expressed sequences. *Genome Res.* Dec; 11(12):2127-32. (2001).

Jaillon O, Aury J-M, Brunet F, Petit J.L, Stange-Thomann N et al. (59 autres auteurs). Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* 431: 946–57 (2004)

Jones S, Martin R, and Pilbeam D. *The Cambridge Encyclopedia of Human evolution.* Press Syndicate of the University of Cambridge, Cambridge, UK. (1992)

Jukes T.H and Cantor C.R. Evolution of protein molecules. (Munro H N, eel.) *Mammalian protein mnet.abolism, III.* New York : Academic Press. p. 21-132. (1969)

Katada S, Hirokawa T, Oka Y, Suwa M, and Touhara K. *Structural basis for a broad but selective ligand spectrum of a mouse olfactory receptor : mapping the odorant-binding site J Neurosci* 25, 1806-15. (2005)

Keverne. EB. Brain evolution, chemosensory processing, and behavior. *Nutr Rev.* Nov;62(11 Pt 2):S218-23; discussion S224-41. (2004)

Kiefer. H, Krieger. J, Olszewski. JD, Von Heijne. G, Prestwich. GD, Breer. H. Expression of an olfactory receptor in *Escherichia coli* : purification, reconstitution, and ligand binding. *Biochemistry*. Dec 17; 35(50) : 16077-84. (1996)

Kimura. M. Evolutionary rate at the molecular level, *Nature*, **217**, 624-626. (1968)

Kishida T, Kubota S, Shirayama Y, Fukami H. The olfactory receptor gene repertoires in secondary-adapted marine vertebrates : evidence for reduction of the functional proportions in cetaceans. *Biol Lett*. Aug 22; 3(4) :428-30. (2007)

Lane. RP, Young. J, Newman. T, Trask. BJ. Species specificity in rodent pheromone receptor repertoires. *Genome Res*. Apr; 14(4):603-8. (2004)

Lane. RP, Cutforth. T, Axel. R, Hood. L, Trask. BJ. Sequence analysis of mouse vomeronasal receptor gene clusters reveals common promoter motifs and a history of recent expansion. *Proc Natl Acad Sci U S A*. Jan 8; 99(1) :291-6. (2002)

Laberge. F, Hara. TJ. Neurobiology of fish olfaction : a review. *Brain Res Brain Res Rev*. Aug; 36(1):46-59. (2001).

Lévêque. C and Didier P. Partie d'ouvrage : Les poissons des eaux continentales africaines : diversité, écologie, utilisation par l'homme
La diversité des poissons africains : l'héritage de l'évolution, Paris : IRD, p. 83-96. ISBN 2-7099-1432-8. (1999)

Lemieux. N, Dutrillaux. B, Viegas-Pequignot. E. A simple method for simultaneous R- or G-banding and fluorescence in situ hybridization of small single-copy genes. *Cytogenet Cell Genet.*; 59(4) :311-2. (1992)

Lemons D, McGinnis W. Genomic evolution of Hox gene clusters. *Science* ; 313: 1918–1922. (2006)

Ludwig A, Belfiore N M, Pitra C, Svirsky V and Jenneckens I. Genome Duplication Events and Functional Reduction of Ploidy Levels in Sturgeon (*Acipenser*, *Huso* and *Scaphirhynchus*). *Genetics* 158 : 1203–1215 (2001)

Lewin. HA, Larkin. DM, Pontius. J, O'Brien SJ. Every genome sequence needs a good map. *Genome Res*. Nov; 19(11): 1925-8. (2009)

Li WH, Wu CI, Luo CC. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol Biol Evol.* Mar; 2(2) :150-74. (1985)

Liberles SD, Buck LB. A second class of chemosensory receptors in the olfactory epithelium. *Nature.* Aug 10; 442 (7103) :645-50. (2006).

Liberles SD, Buck. LB. A second class of chemosensory receptors in the olfactory epithelium. *Nature.* Aug 10; 442(7103) :645-50. (2006)

Lindblad-Toh K, Ostrander EA, Galibert F. Facilitating genome navigation : survey sequencing and dense radiation-hybrid gene mapping. *Nat Rev Genet.* Aug; 6(8):643-8. (2005)

Liberles S.D and Buck. L.B. A second class of chemosensory receptors in the olfactory epithelium. *Nature,* 442; 645-650. (2006)

Liberles SD, Horowitz LF, Kuang D, Contos JJ, Wilson KL, Siltberg-Liberles J, Liberles DA, Buck LB. Formyl peptide receptors are candidate chemosensory receptors in the vomeronasal organ. *Proc Natl Acad Sci U S A.* Jun 16; 106 (24):9842-7. (2009).

Long. M. Evolution of novel genes. *Curr Opin Genet Dev,* 11, 673-680. (2001)

Luchetta. P, Maurel. MC, Higuet. D, and Vervoot. M. *Évolution moléculaire.* (2005)

Lyons. LA, Laughlin. TF, Copeland. NG, Jenkins. NA, Womack. JE, O'Brien. SJ.

Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. *Nat Genet.* Jan; 15(1):47-56 (1997)

Lynch, V.J. Wagner, G.P. Multiple chromosomal rearrangements structured the ancestral vertebrate Hox-bearing protochromosomes" *PLoS Genetics* 5(1) : e1000349. (2009)

Malnic. B, Gonzalez-Kristeller. D.C, and Gutiyama. L.M. The Neurobiology of Olfaction. Menini A, editor. Boca Raton (FL) : CRC Press; **Chapter 7 Odorant Receptors** PMID: 21882436 (pub med this book). (2010)

- Malnic. B, Godfrey. PA, Buck. LB.** The human olfactory receptor gene family. *Proc Natl Acad Sci USA*, 101 :2584-2589 (2004)
- Man. O, Gilad. Y, Lancet. D.** Prediction of the odorant binding site of olfactory receptor proteins by human-mouse comparisons. *Protein Sci*, 13 : 240-254. (2004)
- MARTIN. A.** *Apports nutritionnels conseillés pour la population française*. Editions Tec & Doc, Paris, 650 p. (2001)
- Marta TOMASZKIEWICZ.** Thèse : Molecular characterization of the Y chromosome-linked sex-determining region of the platyfish *Xiphophorus maculatus*. (2012)
- Makalowski W, Boguski MS.** Synonymous and nonsynonymous substitution distances are correlated in mouse and rat genes. *J Mol Evol.* Aug; 47(2) :119-21. (1998)
- Meyer A, Schartl M.** Gene and genome duplications in vertebrates : the one-to-four (to eight in fish) rule and the evolution of novel gene functions. *Curr Opin Cell Biol* 1: 699–704 (1999)
- Meyer A, Van de Peer Y.** From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). *Bioessays* 27: 937–45. (2005)
- Nadeau. JH and Taylor. BA.** Lengths of chromosomal segments conserved since divergence of man and mouse. *Proc Natl Acad Sci* ; 81: 814. (1984)
- Nadeau. JH and Sankoff D.** Counting on comparative maps. *Trends Genet.*, 14 (12), 495-501. (1998)
- Ngaib. J, Chess. A, Dowling. M. M, Necles. N, Macagno. E. R, and Axel R.** Coding of olfactory information : topography of odorant receptor expression in the catfish olfactory epithelium. *Cell*, 72, (5), 667-80. (1993)
- Niimura Y, Nei M,** Evolutionary dynamics of olfactory and other chemosensory receptor genes in vertebrates. *J Hum Genet.*; 51 (6) :505-17. (2006)
- Niimura Y, Nei M.** Evolutionary dynamics of olfactory receptor genes in fishes and tetrapods. *Proc Natl Acad Sci U S A.* Apr 26; 102(17):6039-44. (2005).
- O'Brien. SJ.** Human gene mapping. London Conference. Eleventh International Workshop on Human Gene Mapping. London, UK, August 18-22. *Cytogenet Cell Genet* ; 58 : 1124. (1991)
- O'Brien SJ, Womack JE, Lyons LA, Moore KJ, Jenkins NA, Copeland NG.** Anchored reference loci for comparative genome mapping in mammals. *Nat Genet.*; 3 : 103. (1993)

Ouyang Z, Liang J. Detecting positively selected sites from amino Acid sequences : an implicit codon model. *Conf Proc IEEE Eng Med Biol Soc.* 2007 :5302-6. (2007).

Parker HG, VonHoldt BM, Quignon P, Margulies EH, Shao S, Mosher DS, Spady TC, Elkahoulou A, Cargill M, Jones PG, Maslen CL, Acland GM, Sutter NB, Kuroki K, Bustamante CD, Wayne RK, Ostrander EA. An Expressed *Fgf4* Retrogene Is Associated with Breed-Defining Chondrodysplasia in Domestic Dogs. *Science.* Aug 21; 325(5943): 995-8. (2009)

Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M. *Crystal structure of rhodopsin : A G protein-coupled receptor.* *Science* 289, 739-745. (2000)

Patthy L. Modular assembly of genes and the evolution of new functions. *Genetica*, **118**, 217-231. (2003)

Pontecorvo. G. Production of mammalian somatic cell hybrids by means of polyethylene glyco (PEG) treatment. *Somatic Cell Genet*, 1 (4), 397-400. (1975)

Ptashne. M, Gilbert W. Genetic repressors. *Sci Am.* Jun; 222 (6):36-44. (1970)

Quignon P, Giraud M, Rimbault M, Lavigne P, Tacher S, Morin E, Retout E, Valin AS, Lindblad-Toh K, Nicolas J, Galibert F. The dog and rat olfactory receptor repertoires. *Genome Biol.* 6(10):R83. Epub Sep 28. (2005)

Robertson HM. The large *srh* family of chemoreceptor genes in *Caenorhabditis* nematodes reveals processes of genome evolution involving large duplications and deletions and intron gains and losses. *Genome Research.* 2000; 10 (2), 192-203).

Rouquier. S, Giorgi. D. Olfactory receptor gene repertoires in mammals. *Mutat Res.* Mar 1; 616 (1-2):95-102. (2007).

Rouquier S, Blancher A, Giorgi D. The olfactory receptor gene repertoire in primates and mouse : evidence for reduction of the functional fraction in primates. *Proc Natl Acad Sci U S A.* Mar 14; 97(6):2870-4. (2000)

Robinson. DM, Jones. DT, Kishino. H, Goldman. N, Thorne. JL. Protein evolution with dependence among codons due to tertiary structure. *Mol Biol Evol.* Oct; 20(10) : 1692-704. (2003)

Rogers. J.H. The origin and evolution of retroposons. *Int Rev Cytol*, **93**, 187-279. (1985)

Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes CA, Hutchison CA, Slocombe PM, Smith M. Nucleotide sequence of bacteriophage phi X174 DNA, *Nature*, vol. 265, p. 687-695 (1977)

Sarropoulou. E, Nousdili. D, Magoulas. A, Kotoulas. G. Linking the genomes of nonmodel teleosts through comparative genomics. *Mar Biotechnol (NY)*, **10**(3) :227-233. (2008)

SHEEL. J. J. *Rivulins Of The Old World*, TFH publ. 480 p. (1968)

Schneeberger. EE and Harris. H. An ultrastructural study of inter-specific cell fusion induced by inactivated Sendai virus. *J Cell Sci.* 1 : 401-406. (1966)

Schrock, E et al, Multicolor Spectral Karyotyping (SKY) of Human Chromosomes. *Science*, 1996. 273 (5274) : p 494-7.

Senger F, Priat C, Hitte C, Sarropoulou E, Franch R, Geisler R, Bargelloni L, Power D, Galibert F. The first radiation hybrid map of a perch-like fish : the gilthead seabream (*Sparus aurata* L). *Genomics.* Jun; 87 (6):793-800. (2006)

Shepherd. GM. The human sense of **smell** : are we better than we think? . *PLoS Biol.* May; 2 (5):E146. (2004)

SIMOPOULOS.A.P. Evolutionary aspects of diet and essential fatty acids. *World Review of Nutrition & Dietetics*, 88, 18-27. (2001)

SIMOPOULOS.A.P. The importance of the ratio of omega- 6/omega-3 essential fatty acids. *Biomed. Pharmacother*, 56, 365-379. (2002)

Smith. CW, Valcárcel. J. Alternative pre-mRNA splicing : the logic of combinatorial control. *Trends Biochem Sci.* Aug;25(8):381-8. (2000)

Smith JLB. A living fish of Mesozonic type. *Nature* 143 : 455–456 (1939)

Spehr M, Gisselmann G, Poplawski A, Riffell JA, Wetzel CH, Zimmer RK, Hatt H. Identification of a testicular odorant receptor mediating human sperm chemotaxis. *Science*. Mar 28; 299(5615):2054-8. (2003)

Sullivan SL, Adamson MC, Ressler KJ, Kozak CA, Buck LB. The chromosomal distribution of mouse odorant receptor genes. *Proceedings of the National Academy of Sciences of the United States of America*, 93, (2), 884-8. (1996)

Sullivan SL, Bohm S, Ressler K. J, Horowitz L. F, and Buck L. B. Target-independent pattern specification in the olfactory epithelium. *Neuron*, 15, (4), 779-89. (1995).

Venter. JC. E. coli sequencing. *Science*. Feb 3; 267(5198):60. (1995)

Viegas-Pequignot. E, Dutrillaux. B, Magdelenat. H, Coppey-Moisan. M. Mapping of single-copy DNA sequences on human chromosomes by in situ hybridization with biotinylated probes: enhancement of detection sensitivity by intensified-fluorescence digital-imaging microscopy. *Proc Natl Acad Sci U S A*. Jan; 86(2):582-6. (1989).

Volff. J-N. Evolution and biodiversity in teleost fish. *Heredity* **94**, 280–294 (2005).

Vosshall LB, Amrein H, Morozov PS, Rzhetsky A, Axel R. A spatial map of olfactory receptor expression in the *Drosophila* antenna. *Cell*; 96(5) 725–7365 (1999).

Walter MA, Spillett. DJ, Thomas. P, Weissenbach. J, Goodfellow. PN. A method for constructing radiation hybrid maps of whole genomes. *Nat Genet*; 7 : 22-8. (1994).

Wagner CE, Harmon LJ & Seehausen O. Ecological opportunity and sexual selection together predict adaptive radiation. *Nature* (2012)

Weth. F, Nadler. W, and Korsching S. Nested expression domains for odorant receptors in zebrafish olfactory epithelium. *Proceedings of the National Academy of Sciences of the United States of America*, 93, (23), 13321-6. (1996)

Woese CR, Kandler O, Wheelis ML, "Towards a natural system of organisms : Proposal for the domains Archaea, Bacteria, and Eucarya", *Proc. Natl. Acad. Sci. USA*, Vol.87, No.12, June 1, 1990, p. 4576-4579.

Yang Z, dos Reis M. Statistical properties of the branch-site test of positive selection. *Mol Biol Evol.* Mar; 28(3) : 1217-28. (2011)

Yang Z, Nielsen R. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *J Mol Evol.* Apr; 46(4):409-18. (1998)

Yang. H, Shi. P, Zhang. YP, Zhang. J. Composition and evolution of the V2r vomeronasal receptor gene repertoire in mice and rats. *Genomics.* Sep; 86 (3):306-15. (2005)

Yasuyuki. H and Mutsumi N. Evolution of trace amine associated receptor (TAAR) gene family in vertebrates : lineage-specific expansions and degradations of a second class of vertebrate chemosensory receptors expressed in the olfactory epithelium. *Mol Biol Evol*, 24 (9): 2099-2107. (2007)

Yerganian. G. and Nell MB. Hybridization of dwarf hamster cells by UV-inactivated Sendai virus. *Proc Natl Acad Sci U S A*, 55 (5), 1066-73. (1966)

Young. JM, Kambere. M, Trask. BJ, Lane. RP. Divergent V1R repertoires in five species: Amplification in rodents, decimation in primates, and a surprisingly small repertoire in dogs. *Genome Res.* Feb; 15 (2):231-40. (2005)

Yu F, Sabeti PC, Hardenbol P, Fu Q, Fry B, Lu X, Ghose S, Vega R, Perez A, Pasternak S, Leal SM, Willis TD, Nelson DL, Belmont J, Gibbs RA. Positive selection of a pre-expansion CAG repeat of the human SCA2 gene. *PLoS Genet.* Sep; 1 (3):e41. (2005)

Yukuto S & Mutsumi N. Teleost fish with specific genome duplication as unique models of vertebrate evolution. *Environ Biol Fish* 88:169–188 (2010)

Zdobnov. EM, Von Mering. C, Letunic I, Bork. P. Consistency of genome-based methods in measuring Metazoan evolution. *FEBS Lett.* Jun 13; 579 (15):3355-61. (2005)

Zhao. H, and Firestein. S. *Vertebrate odorant receptors.* *CMLS Cellular and Molecular Life Sciences* 56, 647-659. (1999)

Zhao. H, Ivic. L, Otaki. JM, Hashimoto. M, Mikoshiba. K, Firestein. S. Functional expression of a mammalian odorant receptor. *Science*. 1998 Jan 9; 279(5348): 237-42.

Zhang. J, Nielsen. R, Yang. Z. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol*. Dec; 22(12) : 2472-9. (2005)

Zhang X and Firestein S. Genomics of olfactory receptors. *S.Results Probl Cell Differ.*; 47:25-36. (2009).

Zhang X, Rodriguez I, Mombaerts P, Firestein S. Odorant and vomeronasal receptor genes in two mouse genome assemblies. *Genomics*. May; 83(5):802-11. (2004)

Zozulya S, Echeverri F, and Nguyen T. *The human olfactory receptor repertoire* *Genome Biology* 2, RESEARCH0018. (2001)