Structure, évolution et expression de gènes " chimériques " spécifiques des Primates

Fleur Toulemonde-Darre

▶ **To cite this version:**

Fleur Toulemonde-Darre. Structure, évolution et expression de gènes " chimériques " spécifiques des Primates. Biochimie [q-bio.BM]. Université Nice Sophia Antipolis, 2007. Français. NNT : . tel-00170281

HAL Id: tel-00170281
https://theses.hal.science/tel-00170281

Submitted on 10 Sep 2007

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

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

Université de Nice Sophia Antipolis - Faculté des Sciences

Ecole Doctorale des Sciences de la Vie

THÈSE

Pour obtenir le titre de

Docteur en Sciences
de l'Université de Nice Sophia Antipolis

Aspects de Biologie Moléculaire et Cellulaire

Présentée et soutenue publiquement par

Fleur TOULEMONDE – DARRÉ

Structure, évolution et expression de gènes « chimériques » spécifiques des Primates

Thèse dirigée par Jean-Louis NAHON

Soutenue Le 17 Janvier 2007

Jury

Dr. Philippe VERNIER	Président du Jury
Dr. Marie-Claude POTIER	Rapportrice
Dr. Hervé TOSTIVINT	Rapporteur
Dr. Véronique BARRIEL	Examinatrice
Dr. Richard CHRISTEN	Examinateur
Dr. Jean-Louis NAHON	Directeur de thèse

A Léane et Hanaé, les enfants de ma thèse,

A Pascal,

REMERCIEMENTS

J'adresse d'abord mes remerciements à Marie-Claude Potier et Hervé Tostivint, qui ont accepté d'être les rapporteurs de mes travaux de thèse, ainsi qu'à Véronique Barriel, Philippe Vernier et Richard Christen pour avoir accepté de lire et évaluer ce travail.

Mes remerciements s'adressent également à Jean-Louis Nahon, mon directeur de thèse ; je lui suis reconnaissante de m'avoir accueillie dans son équipe et de m'avoir laissé l'opportunité de réorienter mes travaux au gré des résultats mais aussi des envies. Cette liberté m'a beaucoup enseigné. Merci aussi pour ces discussions conceptuelles très stimulantes que tu as partagées avec moi, Jean-Louis.

Il me faut tout particulièrement remercier ceux et celles qui, de près ou de loin, par des manips ou à l'occasion de discussions fertiles, ont contribué à l'avancement de mes travaux de thèse. Merci à tous les membres de l'équipe, et en particulier à Christine qui a été d'un soutien technique et humain très précieux. Merci aussi à Alice, Carole et Antoine qui m'ont relue et soutenue au cours de la rédaction de ce manuscrit.

Je tiens à remercier avec force ceux qui ont cru en moi et en mes capacités, m'ont aidée à y croire moi-même et m'ont accordé de leur temps… Christine (encore !), Gérard et Bruno, je vous adresse toute ma reconnaissance.

Enfin, je remercie par-dessus tout ceux qui ont garanti mon indispensable équilibre, sans lequel je ne serais parvenue au terme de ces quatre années de thèse. Merci à mes filles, Léane et Hanaé, et à Pascal, qui tous trois m'ont accordé le meilleur. Merci aussi à tous les amis, Marion et Jérome, Sam, Fred, Fiorella, Olivier et leur petit Thao, toujours présents et prêts à partager de bons moments.

TABLE DES MATIERES

INTRODUCTION 9

1. Gènes, génomes et espèces : définitions générales en génétique 9

1.1. Espèces et spéciation 9

1.1.1. Evolution de la notion d'espèce 9

1.1.2. Problématique de la notion d'espèce 10

1.1.3. Modes de spéciation 11

1.2. Génome et épigénome 12

1.2.1. Taille et structure des génomes 12

1.2.1.1. Les mystères des tailles de génome *12*

1.2.1.2. Structure des génomes *14*

1.2.2. Epigénome 17

1.3. La notion de gène 17

1.3.1. Un terme à redéfinir 17

1.3.2. Organisation des gènes eucaryotes 19

2. Mécanismes d'évolution des génomes 21

2.1. Modes d'évolution des génomes 21

2.1.1. Polyploïdisation 21

2.1.1.1. Modes de polyploïdisation *21*

2.1.1.2. Occurrence et mise en évidence de la polyploïdisation *23*

2.1.1.3. Conséquences génomiques de la polyploïdisation *25*

2.1.2. Remaniements chromosomiques 25

2.1.3. Duplications 26

2.1.3.1. Origine des gènes dupliqués *27*

2.1.3.2. Terminologie des gènes dupliqués *30*

2.1.3.3. Devenir des gènes dupliqués *30*

2.1.4. Transfert latéral de gènes 32

2.1.5. Pseudogènes et rétropseudogènes 33
2.1.5.1. D'où viennent les pseudogènes et rétropseudogènes ? *33*
2.1.5.2. Où vont-ils ? *35*
2.1.6. Evolution de la structure des gènes par brassage d'exons 35
2.1.6.1. Domaines et modules protéiques *36*
2.1.6.2. Mécanismes à l'origine du brassage d'exons *36*
2.1.6.3. Rôle des introns dans le brassage d'exons *38*
2.1.7. Séquences répétées et éléments transposables 38
2.1.7.1. Les séquences répétées extragéniques *38*
2.1.7.2. Les transposons : des séquences répétées particulières *39*
2.1.7.3. Impact des éléments transposables sur le génome *40*
2.1.8. Mutations ponctuelles et petites insertions/délétions 42
2.1.8.1. Qu'est-ce qu'une mutation ? *43*
2.1.8.2. La sélection naturelle et les mutations *43*
2.1.8.3. La notion d'horloge moléculaire *45*
2.2. La création *de novo* : un générateur de nouveauté 46
2.2.1. Devenir de gènes, modules ou séquences dupliqués et acquisition de fonctions originales 46
2.2.2. Rétroposition, exaptation d'exons et création de nouveauté génique 47
3. Les Primates et leur évolution 49
3.1. Les Primates 49
3.1.1. Phylogénie des Primates actuels 49
3.1.1.1. Définition des Primates *49*
3.1.1.2. Phylogénie des Primates *49*
3.1.2. Biologie des Primates et impact de l'environnement 54
3.1.2.1. Répartition géographique des Primates : niches et évènements écologiques *54*
3.1.2.2. Diversité biologique et phénotypique *56*
3.2. Les Primates et leur évolution 62
3.2.1. Modifications génomiques des Primates et au sein des Primates 62
3.2.1.1. Organisation globale du génome et cytogénétique *63*

3.2.1.2. Création de nouveaux gènes *67*

3.2.1.3. ADN non codant, séquences répétées et éléments transposables des Primates *73*

3.2.1.4. Mutations ponctuelles, vitesses évolutives variables et polymorphisme *74*

3.2.2. Modifications transcriptomiques participant au renforcement ou au développement de propriétés des Primates 76

3.2.2.1. Epissages alternatifs *76*

3.2.2.2. Profils d'expression différents *79*

3.2.3. Phénotypes et pathologies spécifiques des Primates 82

4. Nos modèles d'étude et objectifs de la thèse 85

4.1. Modèles animaux 85

4.1.1. Présentation des espèces étudiées : relations phylogénétiques, caractéristiques phénotypiques, biotopes et types de données disponibles 85

4.1.1.1. Tarsius syrichta *85*

4.1.1.2. Saguinus Oedipus *87*

4.1.1.3. Cebus capucinus et Cebus appella *87*

4.1.1.4. Chlorocebus aethiops *88*

4.1.1.5. Macacca mulatta et Macacca fascicularis *89*

4.1.1.6. Hylobates lar *90*

4.1.1.7. Pan troglodytes et Pan paniscus *90*

4.1.1.8. Homo sapiens *92*

4.1.2. Les modèles animaux expérimentaux adaptés à l'étude de gènes et pathologies spécifiques des Primates 92

4.1.2.1. Description de la maladie de Parkinson *92*

4.1.2.2. Modèles expérimentaux *93*

4.2. Les gènes chimériques Primates-spécifiques 95

4.2.1. Le système MCH et les gènes *PMCHL* 95

4.2.1.1. Le système à MCH *95*

4.2.1.2. Apparition et évolution des gènes PMCHL1 et PMCHL2 ..... *100*

4.2.2. Données préliminaires sur les gènes dérivant de la béta-glucuronidase 105

4.3. Objectifs et plan de travail 109

4.3.1. Evolution du précurseur peptidique de la MCH chez les Mammifères et les Poissons 109

4.3.2. Mise en perspective des apports des séquences de génomes entiers de Primates 109

4.3.3. Etude de famille de gènes chimériques spécifiques des Primates 109

4.3.3.1. Les gènes PMCHL *109*

4.3.3.2. Les gènes GUSL *110*

RESULTATS 111

1. Structure et activité de la Melanin Concentrating Hormone des Mammifères et des poissons Téléostéens 111

1.1. Contexte bibliographique 111

1.2. Principaux résultats 113

1.2.1. Evolution de la pro-MCH 113

1.2.2. Activité de la MCH chez les Poissons 113

1.3. Article publié 114

2. Données de séquençages de génomes entiers de Primates 125

2.1. Principaux aspects abordés 125

2.2. Revue publiée 126

3. Etudes de familles de gènes chimériques spécifiques des Primates 137

3.1. Etude des gènes orthologues de la famille *PMCHL* 137

3.1.1. Principaux résultats 137

3.1.1.1. Analyse structurale des gènes PMCHL *137*

3.1.1.2. Expression des gènes PMCHL chez l'Homme et le Macaque crabier ... *138*

3.1.2. Article soumis 138

3.1.3. Données complémentaires 183

3.1.3.1. Tentatives de clonage du gène PMCHL1 humain 183

3.1.3.2. Expression de PMCHL dans un modèle de macaques « parkinsoniens » 187

3.2. Etude des gènes paralogues de la famille *GUSL* 194

3.2.1. Objectifs 194

3.2.2. Matériels et méthodes 195

3.2.3. Résultats 198

3.2.3.1. Structure et mise en place des gènes dérivés de la β-Glucuronidase 198

3.2.3.2. Données d'expression des gènes GUSL 5-9-10 208

3.2.4. Discussion 211

DISCUSSION 219

1. Caractéristiques générales des gènes « chimériques » étudiés : des modèles complémentaires 219

1.1. Aspects méthodologiques 220

1.1.1. Séquences analysées 220

1.1.2. Evolution : structure et relations phylogénétiques 221

1.1.3. Expression : caractérisation qualitative et quantitative 221

1.2. Histoires évolutives 221

1.3. Questions de spécificité 222

2. *PMCHL* et *GUSL* : éléments de duplications segmentaires 224

2.1. Duplications segmentaires simples et composées 224

2.2. Duplications segmentaires des Primates 225

2.2.1. Abondance des duplications récentes 225

2.2.2. Localisation des duplications 225

2.3. Duplications segmentaires et polymorphisme 226

2.3.1. Polymorphisme du nombre de copies 226

2.3.2. Polymorphisme nucléotidique 226

2.4. Duplications segmentaires et pathologies 227

2.4.1. Mécanistique 227

2.4.2. Exemples de pathologies associées aux duplications segmentaires 227

2.4.3. Qu'en est-il de *PMCHL* et *GUSL* ? 228

3. Nouveaux gènes et transcription : les épissages alternatifs 230

3.1. Diversifier le protéome… 230

3.2. … mais pas seulement ! 231

3.3. Qu'en est-il pour les transcrits PMCHL et GUSL ? 231

CONCLUSION GENERALE 233

BIBLIOGRAPHIE 235

TABLE DES FIGURES

INTRODUCTION

Figure 1.1.3 **Modalités de spéciation**....13
Figure 1.2.1.2.a **Le génome des procaryotes**....15
Figure 1.2.1.2.b **Organisation du génome eucaryote**....16
Figure 1.3.2 **Structure et avenir d'un gène**....20
Figure 2.1.1.1 **Mécanismes de la polyploïdisation**....22
Figure 2.1.1.2 **Diploïdie apparente après tétraploïdisation**....24
Figure 2.1.3.1 **Duplication en tandem d'un gène**....28
Figure 2.1.3.2 **Terminologie des gènes dupliqués**....29
Figure 2.1.3.3 **Non-fonction et néo-fonction**....31
Figure 2.1.5.1 **Origine des pseudogènes**....34
Figure 2.1.6.2 **Mécanismes moléculaires du brassage d'exons**....37
Figure 2.1.7.3 **Insertion d'un élément transposable**....41
Figure 3.1.1.2.a **Phylogénie des Primates**....50
Figure 3.1.1.2.b **Classification des principaux genres actuels et subfossiles* de l'ordre des Primates**....52
Figure 3.1.1.2.c **Relations de parenté entre les Hominoïdes actuels**....53
Figure 3.1.2.1 **Répartition mondiale des Primates actuels non-humains**....55
Figure 3.1.2.2.2 **Modes de locomotion des Primates**....58
Figure 3.2.1.1.a **Remaniements chromosomiques chez les Primates**....64
Figure 3.2.1.1.b **Mouvements des centromères au cours de l'évolution**....65
Figure 3.2.1.1.c **Modèle de dispersion des duplications segmentaires par déplacement du centromère**....66
Figure 3.2.1.2.a **Origine du gène ECP par néo-fonction**....68
Figure 3.2.1.2.b **Origine du gène *Kua-UEV1* par fusion transcriptionnelle**....70
Figure 3.2.1.2.c **Origine et évolution du gène *BC200 RNA* chez les Primates**....71
Figure 3.2.2.1 **Différentes formes alternatives**....77
Figure 3.2.2.2 **Variation d'expression génique des Primates**....80
Figure 4.1.1 **Les Primates étudiés**....84
Table 4.1.1 **Récapitulatif des caractéristiques des espèces étudiées**....86

Figure 4.2.1.1.1.a **Le système MCH : gène, ARNm et précurseur peptidique**.....96
Figure 4.2.1.1.1.b **Le locus *MCH*/AROM**.....96
Figure 4.2.1.1.2.a **Synthèse de l'hormone de mélanoconcentration**.....98
Figure 4.2.1.1.2.b **Expression de la MCH et de son récepteur**.....99
Figure 4.2.1.2.a **Localisations chromosomiques des gènes *MCH* et *PMCHL***.....101
Figure 4.2.1.2.b **Modèle de rétroposition du messager MCH antisens**.....102
Figure 4.2.1.2.c **Exaptation et création d'exons**.....103
Figure 4.2.2.a **Colocalisation de séquences dérivées de gènes sur le chromosome 5**.....106
Figure 4.2.2.b **Glu 5-10 chez les Primates**.....107

RESULTATS

Figure 3.1.3.1.2.a **Protocole de clonage de *PMCHL1* –loxP-IRES-eGFP-IRES-loxP par recombinaison homologue**.....184
Figure 3.1.3.1.2.b **Protocole de construction de *PMCHL1* –loxP-IRES-eGFP-IRES-loxP par PCR**.....185
Table 3.1.3.2.2 **Oligonucléotides utilisés en RT-PCR**.....189
Figure 3.1.3.2.3 **Expression de messagers PMCHL1 chez des macaques traités au MPTP**.....192
Table 3.2.2 **Oligonucléotides utilisés en RT-PCR**.....196
Figure 3.2.3.1.1.a **FISH sur caryotype numérique**.....199
Figure 3.2.3.1.1.b **FISH sur caryotype numérique : zoom sur les chromosomes 5, 6, 7 et 22**.....200
Figure 3.2.3.1.1.c **Gènes *GUSL* identifiés par BLAST « manuel »**.....201
Figure 3.2.3.1.1.d **Analyse visuelle des séquences du groupe A**.....203
Figure 3.2.3.1.2.a **Analyse phylogénétique des séquences du groupe A**.....205
Figure 3.2.3.1.2.b **Analyse phylogénétique des séquences du groupe B**.....207
Figure 3.2.3.1.3 **Etude des duplications *GUSL* du chromosome 5 humain**.....209
Figure 3.2.3.2.2.a **Expression transcriptionnelle des gènes *GUSL* chez l'Homme**.....210
Figure 3.2.3.2.2.b **Expression traductionnelle potentielle des gènes *GUSL* chez l'Homme**.....212
Figure 3.2.4 **Modèle de mise en place des gènes *GUSL* du groupe A chez l'Homme**.....214

INTRODUCTION

1. Gènes, génomes et espèces : définitions générales en génétique

1.1. Espèces et spéciation

1.1.1. Evolution de la notion d'espèce

Depuis l'avènement de la théorie de l'évolution, la notion d'espèce a connu beaucoup de développements mais aucun consensus n'a jamais pu être obtenu sur sa définition. Concept empirique, la notion d'espèce a évolué avec le temps et son histoire a été marquée par la pensée de grands naturalistes comme Linné, Buffon et Darwin.

On a considéré dans un premier temps, aux dix-huit et dix-neuvième siècles, les espèces comme des entités fixes définies par des critères morphologiques. Cette conception typologique a trouvé son apogée avec les travaux de Linné et l'établissement de collections d'individus « typiques » de l'espèce.

Cuvier introduit ensuite une notion de parenté, en proposant qu'«une espèce peut être définie comme la collection de tous les corps organisés nés les uns des autres ou de parents communs et de ceux qui leur ressemblent autant qu'ils se ressemblent entre eux ».

Puis cette conception a évolué vers la définition d'une espèce « taxonomique » pour laquelle l'analyse mathématique d'un grand nombre de critères suffirait à établir un seuil à partir duquel on pourrait dire que deux individus appartiennent à des espèces différentes. L'espèce serait alors plus un concept commode qu'une entité biologique réelle.

Les insuffisances de ces méthodes ont conduit à une autre approche qui est la notion d'espèce biologique fondée essentiellement sur les critères d'interfécondité et d'isolement (Mayr, 1942), avec là encore quelques difficultés pour différencier par exemple des espèces qui ne sont pas naturellement en contact.

Ceci a conduit à amender cette définition de l'espèce en y incluant une composante écologique. A compter de 1963, Ernst Mayr définit ainsi l'espèce comme une « communauté reproductive de populations, reproductivement isolée d'autres communautés et qui occupe une niche particulière dans la nature » (Mayr, 1963). Cette définition opérationnelle de l'espèce n'est toutefois pas exempte de problèmes (par exemple, la reconnaissance des niches).

Une grande partie de ces problèmes peut cependant être évitée si, se référant à Cuvier et à la notion de parenté qu'il a introduite, l'on considère les êtres vivants dans leur histoire. L'évolution est en effet un processus historique et les espèces sont le résultat de l'éclatement d'espèces qui les ont précédées, ce qui définit la spéciation.

1.1.2. Problématique de la notion d'espèce

Désormais, la définition la plus communément citée est celle d'Ernst Mayr. Il définit donc le concept d'espèce biologique (ou espèce isolée) en indiquant que les espèces sont des « groupes de populations naturelles, effectivement ou potentiellement interféconds, qui sont génétiquement isolés d'autres groupes similaires » (Mayr, 1963). De nombreuses autres définitions ont également cours : par exemple l'espèce peut être définie comme une population dont les membres peuvent se croiser sans difficultés dans des conditions naturelles. Une autre définition, taxonomique, repose sur la notion de ressemblance (ou au contraire de degré de différence), concept encore très utilisé en paléontologie, où il n'y a pas d'autre option.

La difficulté à admettre une définition unique de la notion d'espèce est également liée aux questions d'interfécondité présente ou absente. Celle-ci n'est en effet pas toujours tranchée : des populations A_1, A_2 peuvent être interfécondes, ainsi que A_2 et A_3 .... et A_{n-1} et

A_n et l'on peut pourtant avoir des populations A_1 et A_n qui ne le sont pas. Il est alors périlleux de limiter chacune des espèces. La théorie de l'évolution enrichit ainsi la notion d'espèce d'une dimension qui pose de grands problèmes d'individualisation, que seul le recours aux équilibres ponctués pourrait simplifier. En effet, la théorie de l'évolution repose sur le fait qu'une espèce peut évoluer, par le biais des phénomènes adaptatifs, vers la formation de nouvelles espèces. Ce processus est extrêmement lent, et l'évolution d'une espèce crée l'apparition de sous-espèces intermédiaires, chacune adaptée au milieu qui l'environne. Entre deux espèces, il existe un continuum de sous-espèces très proches. La notion d'espèce étant dès lors envisagée comme un ensemble continu, à quel moment peut-on dire que deux espèces sont différentes ?

Il apparaît nettement que le concept d'espèce fait encore l'objet de controverses, les différents domaines d'application ayant en outre chacun leurs exigences et contraintes. Pour notre part, nous nous en tiendrons à la définition d'Ernst Mayr qui est consensuelle et semble tout à fait pertinente.

1.1.3. Modes de spéciation

Comme souligné précédemment, une espèce n'apparaît pas instantanément par une mutation qui aboutirait à la formation d'un individu d'un type nouveau. Voilà bien l'origine d'une des difficultés à définir l'espèce. Les espèces s'individualisent à partir de populations appartenant à une espèce d'origine.

Le concept de spéciation a été essentiellement développé par Ernst Mayr. Celle-ci résulte des deux moteurs principaux de l'évolution : la sélection naturelle et/ou la dérive génétique, selon différentes modalités écologiques : on distingue la spéciation sympatrique de la spéciation allopatrique.

Quatre éléments sont donc retenus comme déterminants de la spéciation : les forces (sélection et dérive) et les modalités (sympatrique ou allopatrique). Il semble primordial d'en préciser le sens et la teneur. La sélection naturelle tend à ne retenir que les individus (et leurs phénotypes) présentant un avantage sélectif, ce qui se manifeste par une meilleure survie (liée à une meilleure adaptation au milieu) et la production d'un plus grand nombre de descendants. La sélection naturelle apparaît donc quand les conditions suivantes sont réunies : renouvellement d'une population d'individus par mortalité et reproduction, différences phénotypiques des individus d'une population à un instant donné, héritabilité de certains de ces caractères variables et variabilité du nombre de descendants. La dérive génétique est la

modification aléatoire des fréquences des différentes caractéristiques des individus. Elle a donc, par le processus de stochasticité, un impact d'autant plus important que la population est de petite taille. En outre, dans un certain nombre de cas, une corrélation positive peut être établie entre vitesse d'évolution et spéciation (Webster et al., 2003). La spéciation est dite allopatrique lorsque des populations initialement interfécondes évoluent en espèces distinctes du fait d'un isolement géographique. C'est le mode de spéciation de loin le plus fréquent chez les animaux. La spéciation sympatrique caractérise la distinction d'espèces issues de populations non isolées géographiquement. Dans ce mode de spéciation, la sélection naturelle joue un rôle crucial dans la divergence des populations et un échange génétique se poursuit pendant un temps au début de la séparation des espèces (voir Figure 1.1.3).

La spéciation est entrée au cours de ces dernières années dans l'ère de la biologie moléculaire (Wu and Ting, 2004). Des éléments géniques sont ainsi identifiés qui concourent à l'isolement reproductif des populations. Les gènes dits de spéciation affecteraient la valeur adaptative des individus hybrides, créant ainsi une barrière de spéciation. Un tel gène est exceptionnellement bien décrit chez des espèces de Drosophile : il code, chez *Drosophila simulans* pour deux nucléoporines Nup96 et Nup98 qui interagissent avec un gène inconnu du chromosome X de *Drosophila melanogaster* et cause la mort de la descendance hybride mâle (Presgraves et al., 2003).

1.2. Génome et épigénome

Les caractéristiques des espèces, populations et individus, celles qui les rassemblent et celles qui les différencient, sont donc en premier lieu portées et transmises par le génome, substrat de notre identité génétique. Le génome est en effet l'ensemble du matériel génétique d'un individu ou d'une espèce. Le terme génome est du à Hans Winkler qui le définit pour la première fois en 1920 : « the haploid chromosome set, which, together with the pertinent protoplasm, specifies the material foudations of the species » (traduction en anglais de Lederberg et McCray) (Luchetta et al., 2005).

1.2.1. Taille et structure des génomes

1.2.1.1. Les mystères des tailles de génome

Les tailles des différents génomes peuvent être déterminées par des méthodes physico-chimiques et sont éventuellement confirmées par les données de séquençage dans les espèces concernées. La taille mentionnée correspond à celle du génome haploïde. Un génome peut

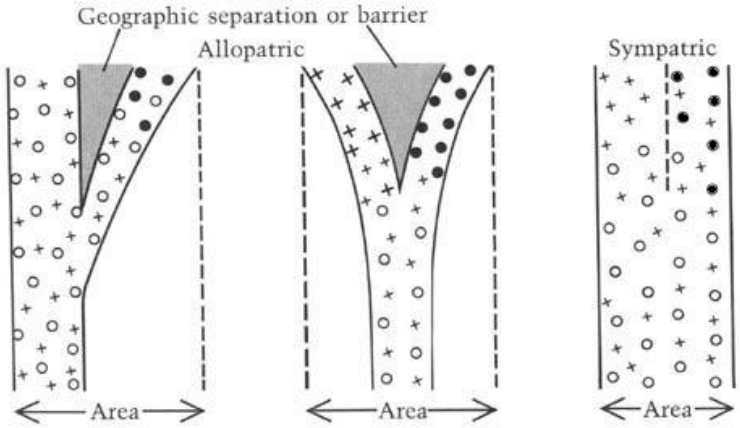


Figure 1.1.3 **Modalités de spéciation**

Représentation des divergences de populations lors de spéciations sympatrique et allopatrique. L'axe vertical représente la progression temporelle, du plus ancien (en bas) au plus récent (en haut). Les cercles et croix représentent les différents génotypes. La coloration noire des symboles illustre l'apparition des nouvelles espèces. Modifiée à partir d'une illustration de "Fossils and the History of Life", de GG Simpson (1983).

ainsi couvrir de quelques kilobases (par exemple celui du HIV) à des centaines de gigabases (cas de certaines amibes). Les génomes de mammifères, quant à eux, s'étendent sur quelques gigabases. On note une très grande disparité dans les tailles de génomes, y compris entre espèces apparentées, et sans relation à la complexité apparente de l'organisme. Cette absence de corrélation constitue le « paradoxe de la valeur C », proposé par CA Thomas en 1971. Il est cependant primordial de noter qu'en première approximation (compte tenu, entre autre, du débat sur la notion de gène) le nombre de gènes est en relation avec ladite complexité des organismes. Mais cette relation est elle-même contestée par Claverie en 2001 (Claverie, 2001; Comeron, 2001; Gregory, 2001), qui introduit le paradoxe de la valeur N du nombre de gènes. La différence de taille entre génomes apparentés est alors portée par des séquences dites non codantes, dont les rôles et fonctions restent à définir. Ces questions relatives à l'ADN dit non codant constituent « l'énigme de la valeur C » (terme du à TR Grégory) (Comeron, 2001; Doolittle and Sapienza, 1980; Gregory, 2001).

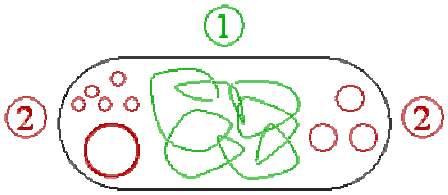
1.2.1.2. Structure des génomes

Les pièces élémentaires des génomes sont les acides nucléiques (principalement l'ADN mais parfois également l'ARN) qui s'organisent en grandes molécules : les chromosomes.

Les génomes de procaryotes se présentent sous la forme d'un chromosome circulaire (rarement plusieurs) auquel sont associées d'autres petites molécules d'ADN circulaire appelées plasmides. On identifie également des épisomes, molécules d'ADN circulaire, extrachromosomiques, qui peuvent se répliquer de manière autonome, à l'instar des plasmides. Les épisomes possèdent cependant certains gènes supplémentaires codant pour la synthèse d'enzymes de restriction qui permettent son intégration aux chromosomes cellulaires ou bactériens par une recombinaison épisomale. Les génomes procaryotes sont très riches en gènes et comportent peu de régions intergéniques, qui sont essentiellement régulatrices (voir Figure 1.2.1.2.a).

Les génomes eucaryotes sont généralement constitués de plusieurs molécules d'ADN linéaires de très grande taille situées dans le noyau. S'y ajoutent un nombre variable de plasmides circulaires (chez certains champignons, notamment) et surtout les chromosomes des organites (mitochondries et chloroplastes) (voir Figure 1.2.1.2.b). Le génome nucléaire se scinde en autant de chromosomes qu'en compte l'espèce, dont la densité génique et la complexité structurale sont très variables selon les régions. On identifie une grande variété de composants des génomes eucaryotes : outre les séquences codantes à proprement parler, et

A



B

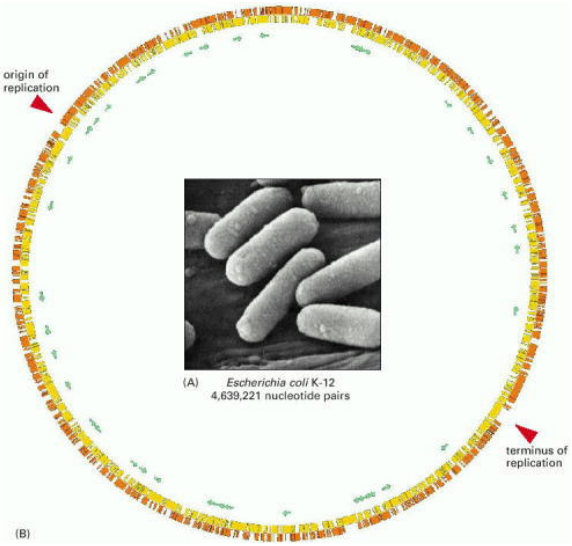


Figure 1.2.1.2.a **Le génome des procaryotes**

A.Illustration de l'organisation du genome des parocaryotes. En vert (1), le chromosome circulaire, en rouge (2), les plasmides et épisomes. **B**. L'exemple représentatif du génome. d'*E. Coli*. En (A), un amas de bactéries E. Coli. En (B), un diagramme du génome circulaire d'E. Coli. Les gènes codant pour des protéines sont représentés par les barres oranges et jaunes en fonction du brin d'ADN à partir duquel ils sont transcrits. Les gènes ne codant que pour des molécules d'ARN sont figurés par des flèches vertes.((A) tiré de Tony Brain and Science Photo Library, (B), d'après FR Blattner et al, Science, 1997 © AAAS)

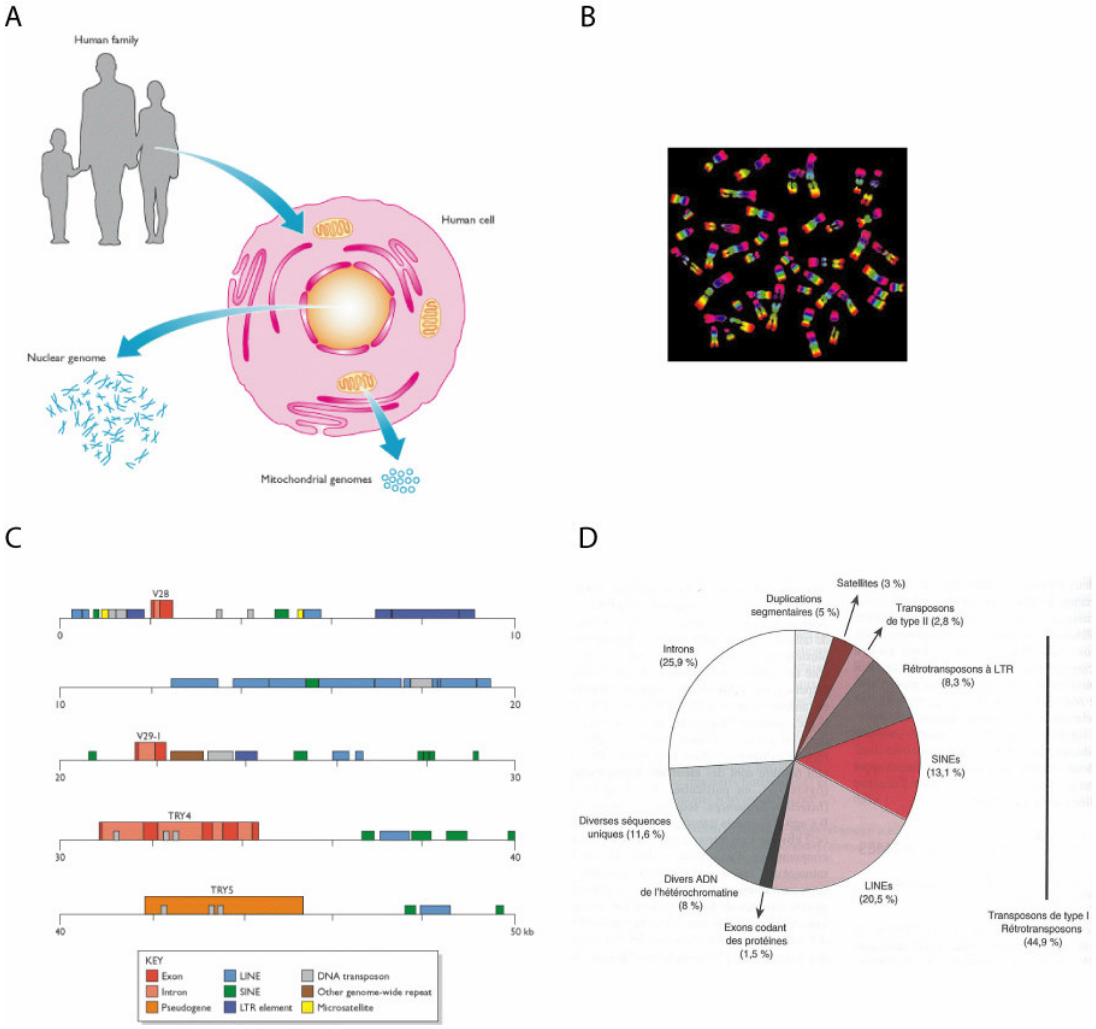


Figure 1.2.1.2.b **Organisation du génome eucaryote**

A, Le matériel génétique eucaryote est localisé dans le noyau et les mitochondries. **B**, caryotype : chromosmes eucaryotes en métaphase de mitose. On distingue les deux chromatides de chaque chromosome. **C**, segment du génome humain. Cette carte indique la position des gènes, segments de gènes, pseudogènes, séquences répétées et microsatellites, sur un fragment de 50kb du locus du récepteur β aux cellules T (chromosome 7). La faible densité génique est ainsi illustrée. LINE/SINE : Long/Short Interspered Nuclear Repeat, LTR : Long Terminal Repeat. Extrait de (Rowen et al., 1996). **D**, résumé des différentes composantes du génome humain. Près de 50 % du génome sont constitués de séquences répétées, parmi lesquelles 45% sont des éléments transposables. Extrait de (Luchetta et al., 2005), d'après les données de séquençage du génome humain.

bien qu'elles restent encore à définir, s'y trouvent des introns, diverses séquences uniques, des séquences variées de l'hétérochromatine (régions fortement condensées de l'ADN, le plus souvent inaccessible à la machinerie cellulaire) et des satellites (séquences répétées localement) et enfin des séquences répétées d'origines diverses (transposons, notamment) (voir Figure 1.2.1.2.b) (Batzer and Deininger, 2002; Hoskins et al., 2002).

1.2.2. Epigénome

Si le génome détermine, comme décrit précédemment, l'identité de l'individu, il ne suffit pour autant à expliquer le phénotype de chaque cellule. Deux cellules contenant la même séquence d'ADN peuvent en effet présenter des phénotypes différents, différences qui sont, comme le matériel génétique, transmises d'une cellule mère à ses cellules filles au cours de la mitose (Govin, 2006).

Ainsi, une information non codée par la séquence de l'ADN existe, et participe à la détermination du phénotype. Cette information est dite épigénétique et est définie par l'ensemble des changements stables et héréditaires qui ne sont pas associés à une modification de la séquence d'ADN, autrement dit, du génome (Wu and Morris, 2001). L'épigénome regroupe ainsi les modifications de la structure chromatinienne (méthylations, modifications d'histones et présence de variants d'histones…) induisant des variations d'expression génique et les phénomènes d'empreinte parentale. L'épigénèse inclut également, outre l'épigénome, les éléments ne codant pas pour des protéines mais néanmoins fonctionnels que sont les ARN dits non codants.

1.3. La notion de gène

L'un des éléments structurels du génome apparaît être le gène. Mais comment définir clairement cette entité : gène, porteur d'un caractère phénotypique de Mendel et Morgan et des biologistes de l'évolution ou gène, portion d'ADN codant pour une protéine des débuts de la biologie moléculaire ?

1.3.1. Un terme à redéfinir

Il est tout à fait significatif d'apprendre que lors de l'élaboration d'un projet (non abouti) de lexique de la génétique par la société française de génétique, l'entrée « gène » avait été très problématique. La question « qu'est-ce qu'un gène ? » posée à dix-huit spécialistes

français du gène (La recherche, décembre 2001) suscite majoritairement deux types de réponses, conformément aux deux définitions « classiques ».

Des biophysiciens, spécialistes de génétique moléculaire ou virologues se réfèrent au gène dans le contexte des théories de l'évolution, comme l'élément objet de la sélection naturelle. Le gène est selon cette définition une unité abstraite, porteuse d'information, transmise par un individu à sa descendance et déterminant un caractère phénotypique. Mais cela implique un lien unique et systématique entre le gène et le caractère, ce qui est incompatible avec les connaissances relatives aux interactions entre plusieurs éléments géniques pour la manifestation d'un caractère.

Pour d'autres (ou pour les mêmes, mesurant la difficile conjugaison des deux définitions « classiques » du gène), biophysicien, biologistes moléculaires, spécialistes en biotechnologies, virologues ou généticien des populations, le gène est sommairement un fragment d'ADN codant pour une protéine par l'intermédiaire d'un ARN, dit messager. Cette deuxième définition présente deux écueils majeurs qu'il nous faut développer :

- Comment borner cette séquence codante, en dresser la géométrie ? Introns et régions promotrices et régulatrices en font-ils partie ? Quid des séquences possédant plusieurs sites de terminaison de la transcription, des épissages alternatifs et donc permettant potentiellement la synthèse de protéines différentes ? Quid de fragments d'ADN présentant des séquences codantes sur leurs deux brins : un gène ou deux gènes ? Quid des gènes situés dans les introns d'autres gènes ?
- Que faire des séquences ne codant pas pour la synthèse d'une protéine ? N'ont-elles donc ni fonction ni devenir ? N'est-ce point réducteur de résumer l'information portée par le génome à celle codant pour des protéines ?

Le philosophe et biophysicien Henri Atlan ajoute en outre qu'il ne faut oublier « que les molécules d'ADN ne font rien par elles-mêmes, puisqu'elles sont inertes, contrairement aux protéines et aux ARN » (La recherche, décembre 2001).

Ne parvenant pas à une définition unique et consensuelle, il semble néanmoins raisonnable d'admettre, dans le contexte de l'étude des génomes et de leur évolution, comme définition du gène la notion d'unité transcriptionnelle informative, tout en mesurant les limites de ce concept. Il s'agit alors d'un enchaînement de nucléotides, c'est-à-dire une portion d'acide désoxyribonucléique (ADN), destiné à être transcrite en acide ribonucléique (ARN).

1.3.2. Organisation des gènes eucaryotes

Les gènes eucaryotes sont éminemment divers : par leur fonction, leur taille, leur structure.

La plupart du temps, un gène commence par une séquence de nucléotides appelée région promotrice, dont le rôle est de permettre l'initiation mais surtout la régulation (tous les gènes ne sont pas exprimés dans toutes les cellules) de la transcription de l'ADN en ARN. Elle se situe majoritairement en 5' du premier exon et s'étend parfois dans le premier intron. Un gène eucaryote comprend également souvent des séquences introniques absentes de l'ARNm et qui régulent au minimum l'épissage, et se termine par une séquence terminatrice, qui marque la fin de la transcription (voir Figure 1.3.2).

Les gènes eucaryotes s'étendent sur quelques centaines de bases à plusieurs mégabases, et comportent également un nombre très variable d'exons, d'un à plusieurs dizaines. Il est intéressant de souligner la corrélation négative manifeste entre la longueur du gène et la fraction du gène qui est présente dans l'ARN messager mature. Ainsi, les messagers sont, eux, de tailles équivalentes et l'accroissement de la taille des gènes repose sur l'allongement des séquences non transcrites, qui peuvent être régulatrices, par exemple.

Comme souligné précédemment, les gènes spécifient des ARN qui peuvent coder pour des chaînes polypeptidiques, ou protéines, mais également des ARN ne codant pas pour des protéines. C'est le cas pour les ARN ribosomaux ou les ARN de transfert, mais également de la famille des ARN dit non codants qui portent en réalité l'information, sans nécessité que celle-ci soit traduite. Ces derniers participent entre autre aux fonctions de régulation de la transcription, de réplication, de maturation, modification, translocation et stabilité des ARN et/ou de translocation et dégradation des protéines (Storz, 2002). L'étendue des fonctions des gènes non codants pour une protéine est certainement très vaste et reste encore méconnue.

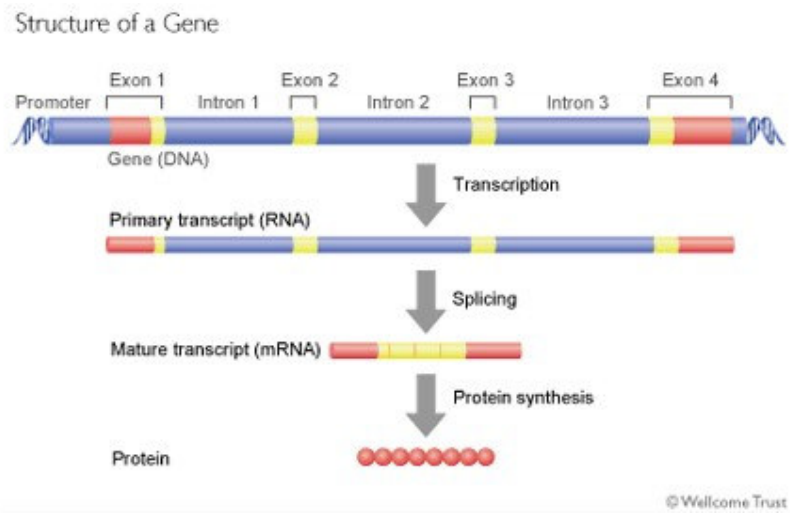


Figure 1.3.2 **Structure et avenir d'un gène**

Le gène eucaryote est composé de séquences promotrice, régulatrices, d'introns et d'exons. Exons et Introns sont transcrits pour former l'ARN pré messager, puis les introns sont éliminés par la machinerie d'épissage. La traduction s'opère alors, au cours de laquelle la protéine est produite. © Wellcome Trust.

2. Mécanismes d'évolution des génomes

2.1. Modes d'évolution des génomes

Une grande part de l'histoire de l'évolution est enregistrée dans les génomes des organismes actuels, et peut être déchiffrée à partir d'une analyse attentive de leurs séquences d'ADN. Les séquençages à grande échelle procurent en effet un aperçu complet des forces évolutives qui ont façonné la structure des chromosomes eucaryotes. Les principes généraux de l'évolution des génomes qui se dégagent ainsi lors d'études de génétique moléculaire font l'objet de ce sous-chapitre. Ils sont présentés par décroissance de la taille des fragments d'ADN affectés.

2.1.1. Polyploïdisation

Le mécanisme majeur de l'augmentation de la taille des génomes eucaryotes est la polyploïdisation, qui correspond à une duplication globale de l'ensemble du génome. Les génomes eucaryotes sont majoritairement haploïdes (possédant une seule copie de chaque chromosome) ou diploïdes (en possédant deux) ; ces deux niveaux de ploïdie coexistent en outre chez les espèces à reproduction sexuée (Luchetta et al., 2005).

2.1.1.1. Modes de polyploïdisation

Autopolyploïdie et allopolyploïdie s'opposent, qui associent les génomes de la même espèce ou de deux espèces distinctes mais proches, respectivement. On reconnaît trois types principaux d'autopolyploïdisation : la non-réduction gamétique, la fertilisation multiple et l'endoréplication. Tous trois conduisent à un doublement de la ploïdie (voir Figure 2.1.1.1).

La non-réduction gamétique découle d'un accident méïotique qui donne naissance à un gamète dont la ploïdie n'a pas été réduite. Lors de la fécondation ultérieure, le zygote est ainsi polyploïde. Cette polyploïdie est cependant généralement impaire, de sorte que si le zygote est viable (ce qui est loin d'être systématique), il est en général infertile, en raison de l'impossibilité d'une répartition homogène des chromosomes lors de la méïose.

Lors de la fertilisation multiple, un gamète femelle est fécondé par plusieurs gamètes mâles distincts, donnant ainsi naissance à un zygote polyploïde. De même que la non réduction gamétique, ce mode de polyploïdisation produit généralement des individus de

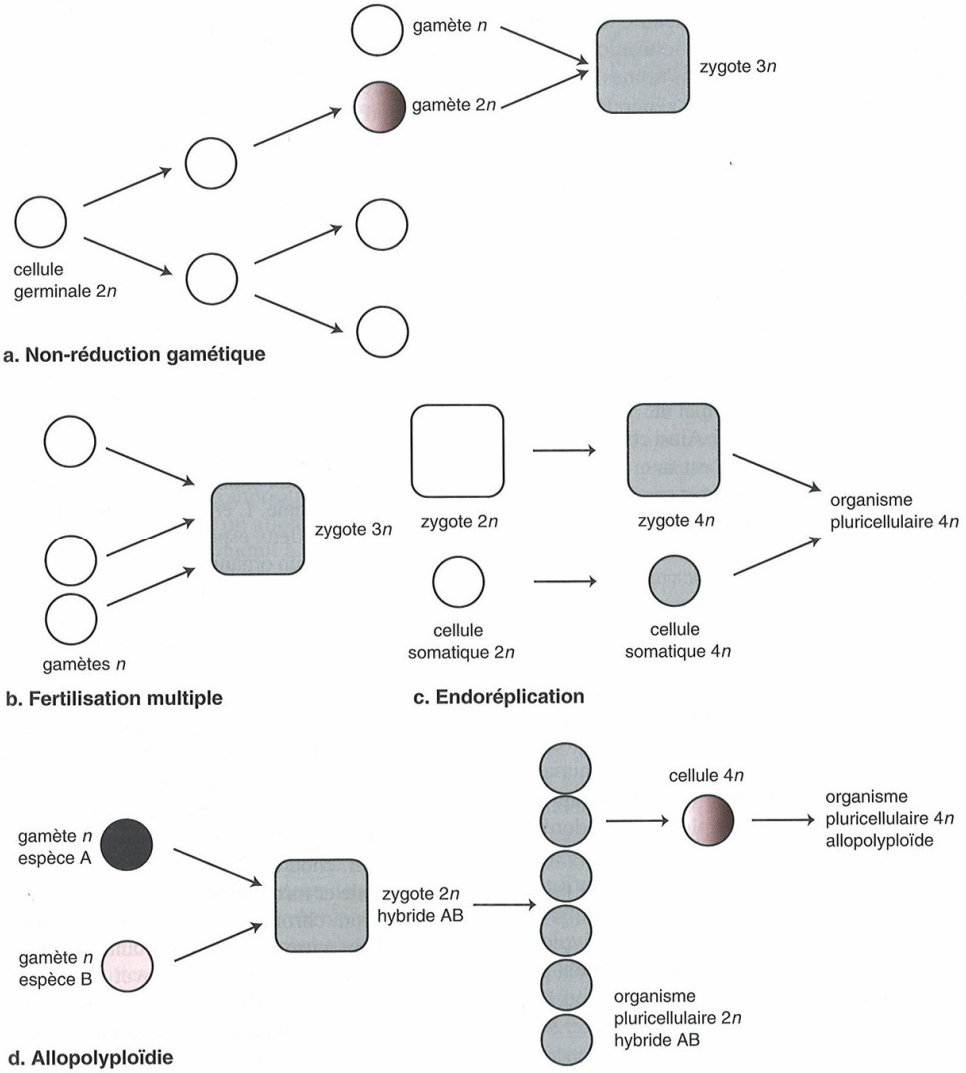


Figure 2.1.1.1 **Mécanismes de la polyploïdisation**

Non-réduction gamétique (**a**), Fertilisation multiple (**b**) et endoréplication (**c**) sont les trois modes d'autopolyploïdisation. L'allopolyploïdisation (**d**) joint les génomes de deux espèces proches mais distinctes. Extrait de (Luchetta et al., 2005).

ploïdie impaire. Leur stérilité, particulièrement fréquente chez les espèces animales, empêche le maintien évolutif de cette forme de polyploïdie.

L'endoréplication consiste en des mitoses anormales lors desquelles il y a duplication du matériel génétique mais non suivie de la séparation en deux cellules filles (cytocinèse). Elle peut affecter les cellules zygotiques ou somatiques et concourt à la création de populations polyploïdes. La parité de cette polyploïdie assure en effet la capacité reproductive des individus.

L'allopolyploïdiation résulte de la fusion des génomes de deux espèces proches mais distinctes. Elle concerne essentiellement les espèces végétales et produit des hybrides qui peuvent exceptionnellement être fertiles, si la fusion est suivie d'une étape d'endoréplication. L'organisme produit est alors tétraploïde, mais il est plus rigoureux de considérer ces nouvelles espèces comme amphidiploïde, leur formule chromosomique relevant davantage du *2n+2n* que du *4n*.

2.1.1.2. Occurrence et mise en évidence de la polyploïdisation

L'étude des caryotypes et l'observation de multivalents lors de la métaphase méïotique permettent de mettre en évidence les polyploïdisations les plus récentes. Mais il est nettement plus délicat de décrire avec certitude des évènements de polyploïdisation anciens. En effet, de nombreux réarrangements chromosomiques surviennent ensuite au cours du temps qui masquent la duplication génomique (voir Figure 2.1.1.2) (Strachan and Read, 1996). La polyploïdie ne peut alors qu'être déduite de l'observation de deux espèces proches possédant l'une un nombre donné de chromosomes, l'autre une quantité double.

Les études génomiques ont néanmoins permis d'identifier un certain nombre de polyploïdisations anciennes. De nombreuses espèces portent les traces de ces évènements au cours de leur histoire évolutive (Seoighe, 2003). Ainsi Susumu OHNO (Ohno, 1970) proposa dès les années 1970 qu' « un ou plusieurs épisodes de duplications de génomes se soient déroulés lors de l'évolution des Vertébrés ». Cette affirmation a fait et fait encore l'objet d'un débat : les séquences dupliquées des génomes de Vertébrés pourraient également découler de duplications segmentaires. Il est cependant désormais relativement consensuel d'affirmer qu'un évènement au moins de polyploïdisation a façonné les génomes de Vertébrés. Cet évènements serait survenu il y a 750 millions d'années environ. Un deuxième « tour » de polyploïdisation, plus débattu, pourrait avoir eu lieu il y a 450 millions d'années (Gu et al., 2002).

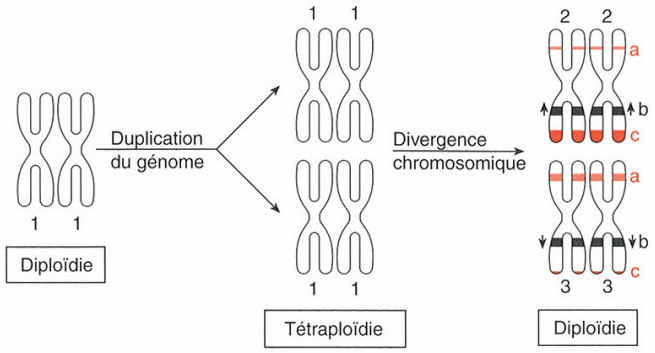


Figure 2.1.1.2 **Diploïdie apparente après tétraploïdisation**

La duplication du génome peut conduire à un état tétraploïde transitoire avant que la divergence des chromosomes restaure la diploïdie. Délétions interstitielles (en a, en haut), ou terminales (c, en bas) et inversions (b) sont quelques uns des remaniements concourant à la divergence chromosomique. Extrait de (Strachan and Read, 1996).

Il faut noter enfin que la polyploïdisation est fréquente chez les plantes et rare chez les animaux. Elle n'est cependant pas absente de leur histoire évolutive, et ce dans des lignées très diverses. Différentes hypothèses sont proposées pour expliquer cette différence. La reproduction sexuée, entre deux individus morphologiquement et génétiquement distincts, chez les animaux constitue une limite à la polyploïdisation dans la mesure où cette duplication pourrait perturber la détermination du sexe. En outre l'allopolyploïdie n'est favorisée chez les animaux ni au stade du croisement interspécifique ni à celui de l'endoréplication. Enfin la létalité embryonnaire des animaux polyploïdes en réduit également l'occurrence.

2.1.1.3.Conséquences génomiques de la polyploïdisation

Le premier effet de la polyploïdisation est l'accroissement substantiel de la taille des génomes. La duplication du génome constitue une duplication de chacun des gènes, permettant la divergence rapide d'une des copies de chaque gène. La duplication du génome est ainsi « une source importante de nouveautés génétiques et de complexification des organismes vivants » (d'après S. Ohno). En outre, les généticiens des populations témoignent de la plus grande variabilité génotypique (et par là-même phénotypique) possible dans les espèces polyploïdes. Le nombre et la variété des hétérozygotes sont également accrus chez les autopolyploïdes et les réarrangements chromosomiques favorisés chez les allopolyploïdes (Soltis and Soltis, 1999). Enfin, la polyploïdisation constitue naturellement un facteur majeur de spéciation (Otto and Whitton, 2000; Taylor et al., 2001).

Wolfe a cependant récemment suggéré que les évènements de polyploïdisation du génome dans la lignée menant aux vertébrés, qu'ils aient effectivement existé ou non, n'auraient pas eu d'impact majeur sur le développement de leur protéome, sa diversification tenant essentiellement aux épissages alternatifs des gènes et aux protéines variantes ainsi produites (Adams et al., 2003; Wolfe, 2001)**.**

2.1.2. Remaniements chromosomiques

L'observation des caryotypes types d'espèces plus ou moins proches révèle parfois l'existence de différences cytogénétiques importantes. Les réarrangements génomiques dont ils témoignent concernent de grands fragments d'ADN et sont relativement fréquents chez les mammifères, jusque dans leur histoire évolutive récente. En terme de modification génomique, on distingue en particulier les remaniements affectant la taille du génome

(amplification ou réduction) de ceux qui ne la changent pas (on parle de réarrangements équilibrés) (Jones et al., 1992).

Crossing-overs inégaux lors de la méïose (ce qui implique la recombinaison entre deux séquences non-alléliques entre deux chromatides non sœurs de chromosome homologues), échanges inégaux entre chromatides sœurs (c'est-à-dire entre chacune des molécules d'ADN identiques constituant le chromosome pré-mitotique), translocations chromosomiques et transpositions de multiples copies d'ADN (aussi appelées duplications) sont les principaux facteurs de l'amplification génomique et constituent de véritables duplications sub-génomiques. Ces processus peuvent cependant également 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 de phénomènes de translocation équilibrée (on évoque ainsi le transfert de régions chromosomiques entre chromosomes non homologues), d'exceptionnelle tranposition de simple copie d'ADN (déplacement d'un fragment d'ADN d'un locus à un autre), d'inversions chromosomiques (de part et d'autre du centromère), mais également de fusion ou fission de chromosomes.

Par l'étude des relations synténiques, c'est-à-dire des bornes de ces remaniements, Nadeau et Taylor ont suggéré dès le milieu des années 1980 que les points de coupures (breakpoints) entre segments homologues devaient être répartis de façon uniforme et aléatoire (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 évènements de recombinaison homologue (Eichler and Sankoff, 2003).

Les taux de réarrangements chromosomiques diffèrent largement d'un génome à un autre. Parmi les espèces de vertébrés, par exemple, ils varient de 0,2 à 1 ou 2 remaniement par million d'années (Burt et al., 1999). L'ensemble de ces réarrangements chromosomiques est ensuite transmis à la descendance à une fréquence faible, en raison de l'infertilité conséquente. En effet, ils induisent des difficultés d'appariements des paires de chromosomes lors de la méïose.

2.1.3. Duplications

La duplication de fragments d'ADN est un phénomène très général, observé dans la plupart des génomes. Elle concerne des fragments de quelques kilobases à plusieurs centaines

de kilobases et constitue, en réalité une forme de remaniement chromosomique à détailler. Ces duplications sont dites segmentaires par opposition aux duplications de génome entier des évènements de polyploïdisation. Leur abondance varie considérablement entre les différents génomes eucaryotes analysés, probablement en raison de variations temporelles. On identifie en effet deux pics de duplications segmentaires : l'un a suivi la vague de radiation des mammifères (et pourrait en réalité être spécifique des Primates), l'autre a probablement eu lieu aux premiers temps de l'évolution des métazoaires (Eichler and Sankoff, 2003; Gu et al., 2002).

2.1.3.1. Origine des gènes dupliqués

Différents mécanismes peuvent être à l'origine des duplications ; l'analyse du type de répétition observé permet d'en déduire l'origine.

Translocation et inversion (si elle est péricentromérique) concourent dans certains cas à la duplication d'une partie du génome dans la descendance.

Mais l'évènement de duplication le plus fréquent survient à l'occasion d'un crossing-over inégal entre chromosomes homologues lors de la méiose (voir Figure 2.1.3.1) (Smith, 1974). Un glissement lors de l'appariement des chromatides non-sœurs répartit de façon inégale les allèles. Au cours de ce processus, l'un des chromosomes appariés perd l'une de ses copies et l'autre en acquiert une en tandem (Elemento et al., 2002). En fonction du caractère essentiel ou non du gène exclu/dupliqué, la pression de sélection permettra ou non le maintien de chacune des formes. La répétition de ce processus est ensuite favorisée par la présence de gènes dupliqués en tandem, de sorte que leur nombre tend à s'accroître, puis à fluctuer.

Des mécanismes réplicatifs sont également impliqués dans la duplication de fragments d'ADN. Lorsque l'initiation de la réplication se répète localement, différentes copies du même fragment sont produites. Ce processus de sur-réplication favorise ainsi, chez les eucaryotes, des duplications en tandem. De même, si des mésappariements décalés surviennent lors de la synthèse du brin complémentaire, des duplications en tandem peuvent être engendrées (Elemento et al., 2002). Ces deux mécanismes, assez peu connus, ne permettent la duplication que de fragments d'ADN de taille restreinte (quelques centaines de paires de bases).

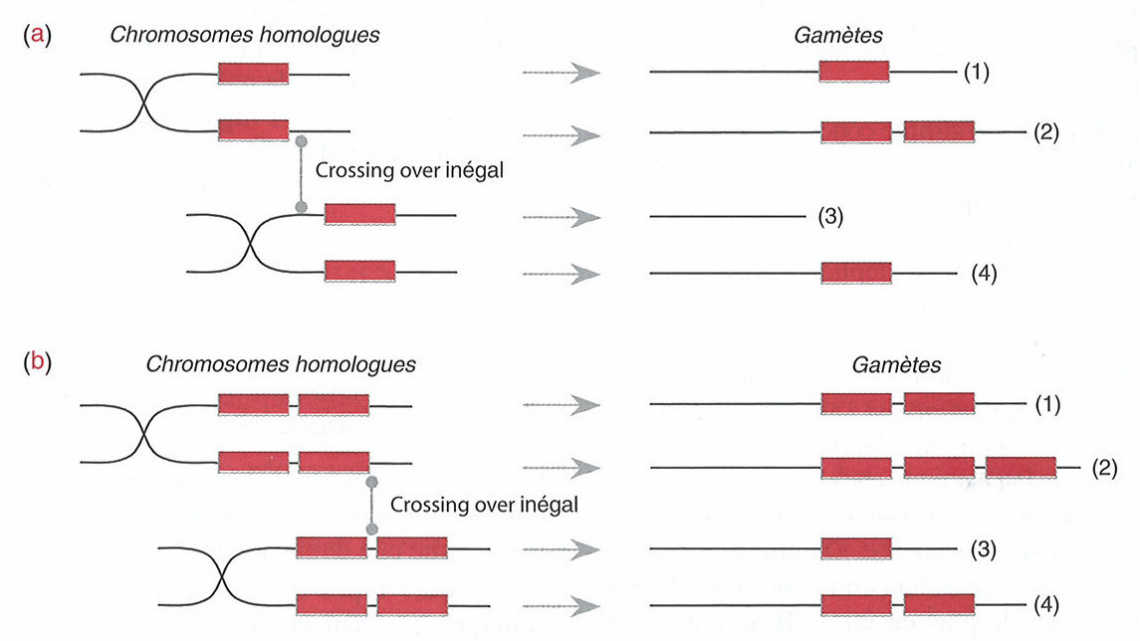


Figure 2.1.3.1 **Duplication en tandem d'un gène**

(a) Une duplication simple peut être obtenue au cours de la méiose par crossing over inégal entre les chromatides de deux chromosomes homologues. Après la méiose, les gamètes (1) et (4) possèderont toujours le gène d'origine, tandis que le gamète (2) en possèdera deux copies et le gamète (3) aucune. **(b)** Lors des méioses ultérieures, les similitudes de séquence dans cette région auront tendance à favoriser ces crossing over inégaux. On observera alors un accroissement (2) ou une diminution (3) du nombre de copies du gènes dans les futurs gamètes. Extrait de (Luchetta et al., 2005).

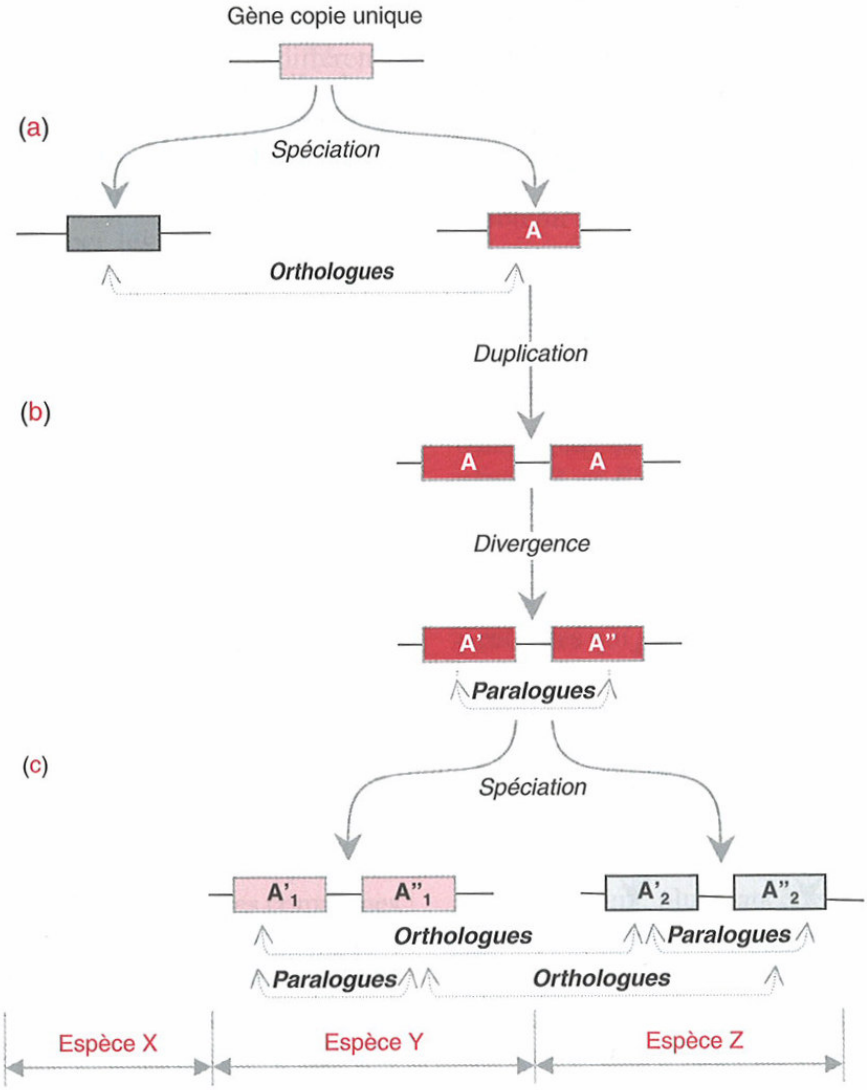


Figure 2.1.3.2 **Terminologie des gènes dupliqués**

(a) Deux gènes orthologues dérivent d'un gène unique par spéciation. **(b)** Deux gènes paralogues dérivent d'un gène unique par duplication et sont présents au sein d'une même espèce. **(c)** Suite à un nouvel évènement de spéciation, ces deux gènes (A' et A'') se trouvent dans deux espèces différentes et divergent pour donner les gènes A'_1, A''_1, A'_2 et A''_2. Extrait de (Luchetta et al., 2005).

2.1.3.2. Terminologie des gènes dupliqués

L'acquisition de nouvelles copies de fragments d'ADN a amené la mise en place d'une terminologie permettant de définir de façon précise et unique les relations de parenté entre séquences (voir Figure 2.1.3.2). L'ensemble des gènes possédant une origine commune, c'est-à-dire provenant d'un même gène ancestral, que ce soit par duplication ou spéciation, sont dits homologues. On distingue ensuite les orthologues des paralogues selon qu'ils émergent suite à un évènement de spéciation ou de duplication, respectivement. Les orthologues possèdent en conséquence généralement une même fonction, tandis que les paralogues peuvent diverger d'un point de vue fonctionnel.

2.1.3.3. Devenir des gènes dupliqués

L'implication des duplications dans les processus évolutifs n'est plus à démontrer, mais selon des modalités qui restent à définir. Le maintien de plusieurs copies redondantes d'un même gène pose en effet des problèmes d'instabilité à la cellule, d'autant qu'une copie est suffisante à assurer la fonction ancestrale. S. Ohno proposa donc dès 1970 le double modèle de non-fonction / néo-fonction (Ohno, 1970). Il considère qu'une copie du gène étant suffisante, la deuxième est exempte de pression de sélection et peut évoluer librement. La relaxation des contraintes sélectives permet l'accumulation de mutations, qui concourent généralement à la perte de fonction de la copie génique, qui devient pseudogène. Ainsi est définie la non-fonction. Cette copie est ensuite éventuellement perdue. Mais de nouvelles fonctions, au sens large, peuvent également émerger à la faveur de cette relaxation de contrainte, voire d'une pression de sélection positive, favorisant la divergence. Leur occurrence est cependant statistiquement moins probable (voir Figure 2.1.3.3) (Eichler, 2001; Holland, 1999; Prince and Pickett, 2002).

Selon ce modèle d'évolution des séquences dupliquées, seul un petit nombre de copies devraient subsister, du fait des contraintes évolutives (Ohta, 2000). Mais les observations des génomes vont à l'encontre de cette hypothèse. Il apparaît en réalité que l'essentiel des gènes dupliqués dont les différentes copies sont maintenues sont les produits d'évènements de polyploïdisation. Ainsi, chez les Vertébrés, les gènes dupliqués identifiés dans le génome seraient apparus (en terme de duplication) à la faveur du second épisode de polyploïdisation, il y a 450 Millions d'années. A. Force (Force et al., 1999) proposa en outre en 1999 le modèle de Duplication – Dégénérescence – Complémentation (DDC) qui permet d'expliquer le maintien de l'ensemble des copies des gènes dupliqués. Ce modèle fait appel à la notion de

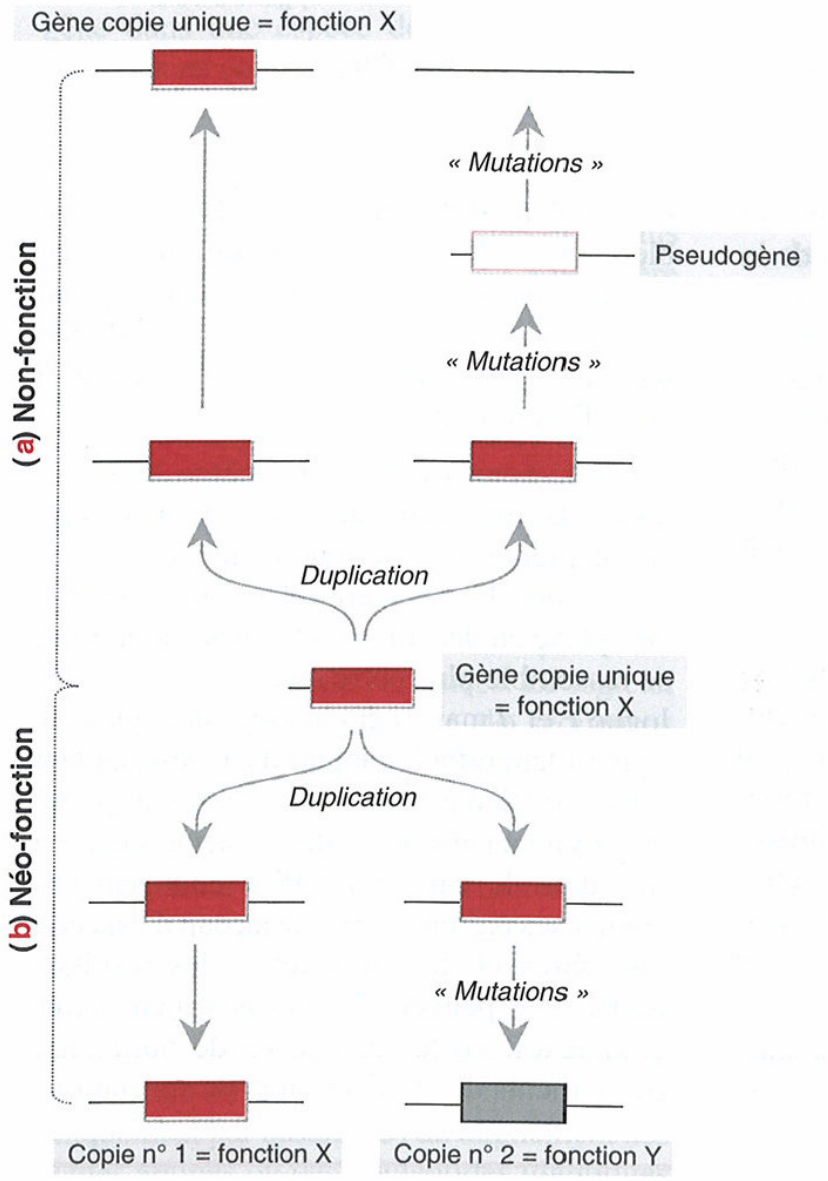


Figure 2.1.3.3 **Non-fonction et néo-fonction**

Le modèle d'Ohno (1970) propose deux voies pour l'évolution des copies de gènes dupliqués **(a)** La voie de la non-fonction, où l'une des copies conserve sa fonction d'origine, alors que l'autre devient un pseudogène. **(b)** La voie de la néo-fonction, où l'une des copies conserve sa fonction initiale tandis que l'autre peut évoluer librement et éventuellement acquérir une nouvelle fonction. Extrait de (Luchetta et al., 2005).

sous-fonction, le terme fonction étant entendu dans son acception la plus large. La fonction d'origine du gène, ou son profil d'expression, serait répartie entre les gènes dupliqués. Selon que les mutations affectent l'une des copies dans sa région codante ou régulatrice, on observe un partage des fonctions *stricto sensu* ou des sites d'expression, respectivement. Au total, les différentes copies assument ainsi la fonction initiale. Divers exemples de cette complémentation fonctionnelle sont observés chez les poissons et les vertébrés, notamment. Ce maintien des différentes copies du gène dupliqué favorise en outre à long terme l'apparition de fonctions nouvelles.

Les deux modèles complémentaires présentés précédemment (celui d'Ohno et celui de Force) impliquent une divergence élévée entre paralogues, de sorte que les gènes paralogues sont rapidement plus différents entre eux que les gènes orthologues. Un certain nombre d'observations vont cependant à l'encontre de ce principe. Ainsi, les membres de la famille multigénique des ARN ribosomaux 16S sont très proches en terme de séquences. On parle, dans ce cas, d'évolution concertée, ou coordonnée, des paralogues. Ceci implique l'existence de conversion génique entre paralogues. La conversion génique maintient en effet la similarité des séquences par l'échange régulier de fragments d'ADN entre copies des gènes dupliqués. Cette homogénéïsation est curieusement restreinte aux séquences répétées non codantes et à un petit nombre de familles multigéniques (Liao, 1999).

2.1.4. Transfert latéral de gènes

Le processus de transfert latéral de gène, relativement méconnu et également appelé transfert horizontal de gène, survient essentiellement entre organismes procaryotes, mais peut également concerner les génomes eucaryotes. Il définit un échange de matériel génétique entre espèces distinctes et favorise l'acquisition de fonctions de l'organisme donneur, presque systématiquement procaryote, par l'organisme receveur qui peut être eucaryote. Les transferts latéraux de gènes sont des évènements rares, mis en évidence par la présence de gènes similaires dans des espèces très distantes (Genereux and Logsdon, 2003).

Les impératifs mécanistiques des transferts latéraux des gènes sont les suivants : le matériel génétique étranger doit d'abord pénétrer dans la cellule receveuse, sous forme d'ADN nu ou au sein de la cellule donneuse. Cela se produit généralement entre un procaryote et un organisme phagotrophe. Il faut ensuite que le gène soit 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).

L'incorporation endosymbiotique des génomes des organites (mitochondrie et chloroplaste) par les cellules eucaryotes constitue un transfert latéral de gènes. Il semble clair aujourd'hui que les transferts horizontaux ont joué un rôle au cours de l'évolution (notamment au moment de la séparation des trois domaines du vivant : archae, bacteriae, eucaryotes) (Brown, 2003). Le transfert latéral de gènes reste quoiqu'il en soit minoritaire comparé aux grands processus d'évolution des génomes qui sont à l'origine de la grande diversité des eucaryotes (Luchetta et al., 2005). Le transfert latéral de gène est en outre exceptionnel chez les animaux (Andersson, 2005).

2.1.5. Pseudogènes et rétropseudogènes

2.1.5.1. D'où viennent les pseudogènes et rétropseudogènes ?

Les pseudogènes sont des gènes non fonctionnels qui présentent des similitudes avec un ou plusieurs gènes fonctionnels ; en d'autres termes ce sont, initialement, des copies inactives de gènes fonctionnels. Cette absence de fonction est souvent due à l'absence de promoteur ou de séquence de régulation. De ce fait, le pseudogène est exempt de toute contrainte sélective, ce qui permet une accumulation de mutations diverses. On le reconnaît en général par la présence aberrante de codons stop dans la séquence codante.

On distingue deux types de pseudogènes selon les modalités de leur mise en place : les pseudogènes dupliqués et les rétropseudogènes. Le pseudogène dupliqué, plus souvent mentionné dans la littérature, provient de la duplication conforme d'un gène actif telle que décrite précédemment (cf 2.1.4). La copie évoluant vers la non-fonction, en conséquence de mutations dans le promoteur, les régions de régulation ou les exons, est dite pseudogène. Le rétropseudogène n'existe que chez les Métazoaires et provient d'une rétroposition, c'est-à-dire de la transcription inverse d'un ARN messager suivi d'une intégration dans l'ADN génomique (voir Figure 2.1.5.1). Le mécanisme de cette rétroposition est particulièrement bien décrit pour les séquences de Primates et sera donc rapporté ultérieurement (cf 3.1.3.2) (Dewannieux et al., 2003). Le rétropseudogène 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). Ces marques sont parfois perdues au cours de l'évolution.

Les pseudogènes ont été longtemps considérés comme des éléments inactifs et sans intérêt. La découverte de leur abondance au cours des programmes de séquençages des

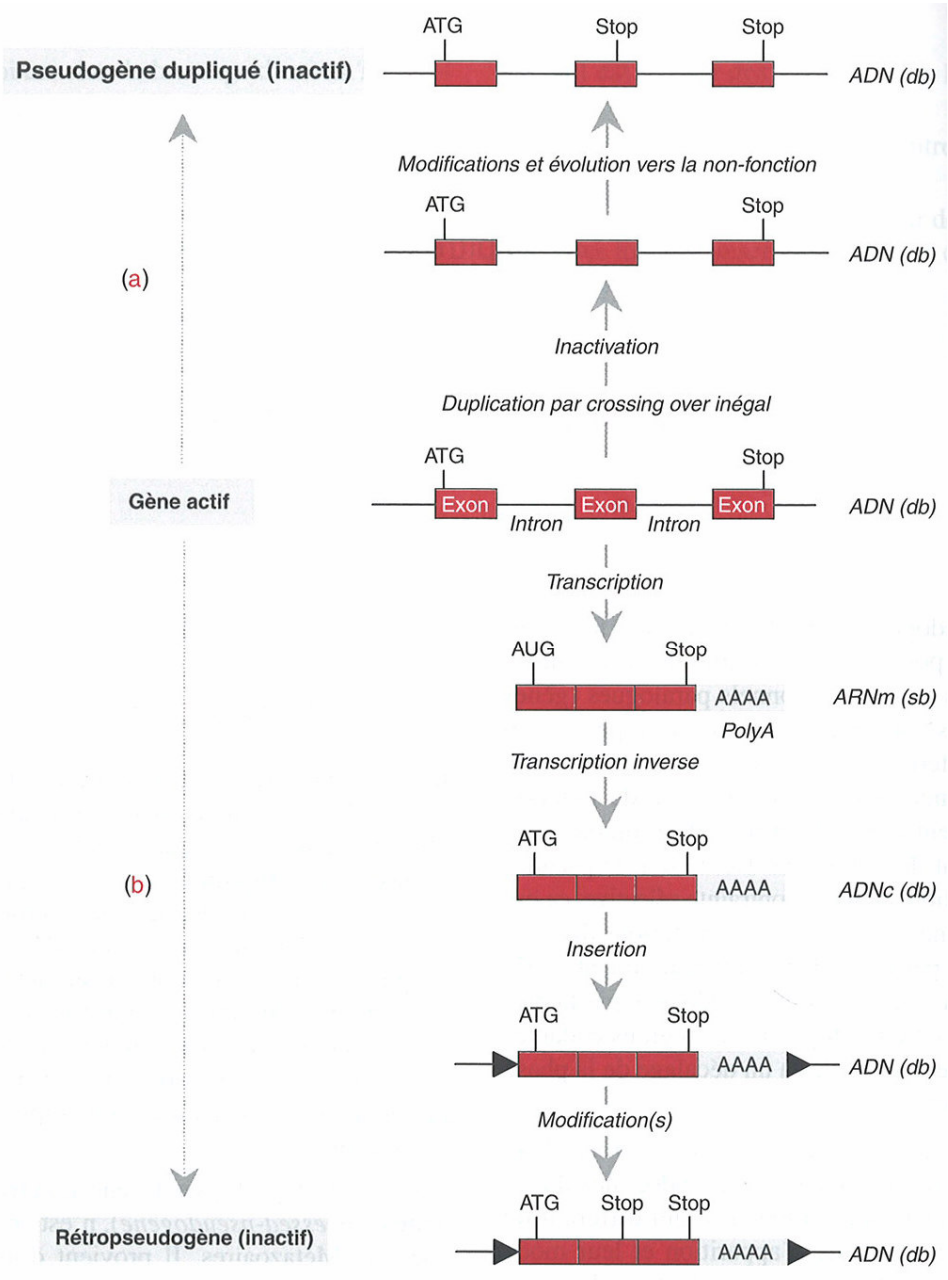


Figure 2.1.5.1 **Origine des pseudogènes**

Deux types de pseudogènes peuvent être produits à partir d'un gène actif. **(a)** Les pseudogènes dupliqués. **(b)** Les rétropseudogènes. Les triangles noirs indiquent la présence de séquences répétées directes de part et d'autre du rétrotransposon inséré dans le génome. Extrait de (Luchetta et al., 2005). (sb) simple brin, (db) double brin.

génomes en a fait l'objet de recherches. Ils peuvent ainsi être considérés comme des « traces » de l'évolution, ce qui les rend essentiels à la compréhension des mécanismes moléculaires de l'évolution (Harrison et al., 2002).

2.1.5.2. Où vont-ils ?

Lorsqu'un gène a subi une duplication et que l'une des copies devient pseudogène, on peut observer une inactivation partielle ou complète de celle-ci. Ce sont en général des mutations qui agissent sur la transcription, la traduction ou les deux à la fois. Au cours du temps, on peut envisager une reprise d'activité du pseudogène due à différentes modifications de type mutations, inversions, délétions… (voir Figure 2.1.5.1). Cette succession d'évènements est assez rare et doit se faire dans un temps limité : on estime, selon la loi de Dollo, qu'un pseudogène qui n'a pas repris d'activité (généralement du même ordre que celle du gène ancestral) au bout de quelques millions d'années est définitivement éteint, car il a accumulé trop de mutations et que cette évolution est irréversible (Gould, 1970; Marshall et al., 1994). On observe en outre chez les Eucaryotes que cette reprise d'activité est restreinte à certaines familles de protéines.

Pseudogènes dupliqués et rétropseudogènes peuvent aussi par la suite acquérir de nouvelles fonctions. Les rétropseudogènes sont particulièrement associés à la création de nouveautés géniques. Ces devenirs et caractéristiques feront l'objet d'un développement ultérieur (cf 2.2) (Balakirev and Ayala, 2003).

2.1.6. Evolution de la structure des gènes par brassage d'exons

Le brassage d'exons a été proposé dès la fin des années 1970 par W. Gilbert 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 (essentiellement métazoaires) et on dénombre aujourd'hui une centaine de familles de gènes ainsi constituées, principalement des gènes de mammifères. Elles répondent en particulier aux fonctions de signalisation entre cellules de l'organisme (propres aux métazoaires) (description exhaustive dans (Patthy, 2003)). Le brassage d'exons est le plus couramment détaillé, mais il existe également, par exemple, des processus de brassage de promoteurs (Luchetta et al., 2005).

2.1.6.1. Domaines et modules protéiques

Une protéine est constituée d'un ensemble de régions appelées domaines. Un domaine est défini comme une partie de la protéine qui possède une « autonomie » fonctionnelle et structurale. Un domaine peut être codé par un ou plusieurs exons. Le terme module, souvent employé dans la littérature, correspond également à une unité protéique, mais en ne considérant que la structure. Un ou plusieurs modules peuvent ainsi être nécessaires pour former un domaine (Apic et al., 2001).

Lorsqu'un gène subit un brassage d'exons, le maintien de la nouvelle structure est dépendant de la stabilité de l'ensemble. Des échanges peuvent se faire entre différents modules à la seule condition que l'insertion d'une séquence ne modifie pas la structure et la fonction globale de la protéine. On admet donc par définition qu'une protéine n'a pu évoluer par brassage d'exons que lorsqu'elle possède plusieurs modules, elle est alors dite protéine mosaïque.

2.1.6.2. Mécanismes à l'origine du brassage d'exons

Le brassage d'exons peut se faire de différentes manières, mais fait toujours intervenir des échanges de séquences composées d'introns et d'exons. Chez les Eucaryotes, les introns ont joué un rôle essentiel en permettant, comme nous le préciserons ultérieurement, d'insérer ou de supprimer des séquences nucléotidiques sans changer la phase ouverte de lecture, et donc de conserver la structure protéique (voir Figure 2.1.6.2).

L'un des mécanismes principaux à l'origine du brassage d'exons est le crossing-over inégal qui, comme indiqué précédemment, permet de dupliquer des exons. Cet échange se produit généralement lors d'une recombinaison homologue au cours d'une meïose et peut s'appliquer à un ou plusieurs exons. Ce mécanisme résulte alors en une duplication de module ou de domaine.

L'autre mécanisme principal, qui représente une part importante du brassage d'exon, est la rétrotransposition, déjà mentionnée précédemment (Brosius, 1991). Différentes expériences ont en effet montré qu'une séquence répétée de type LINE-1 pouvait, lors de sa retrotransposition, « emporter » une séquence en aval, lorsque le signal de polyadénylation de l'élément transposable n'est pas reconnu et que la transcription se poursuit. La conséquence n'est alors pas une duplication d'exon(s), mais une insertion d'un nouvel exon, d'un nouveau module ou d'un nouveau domaine (Moran et al., 1999).

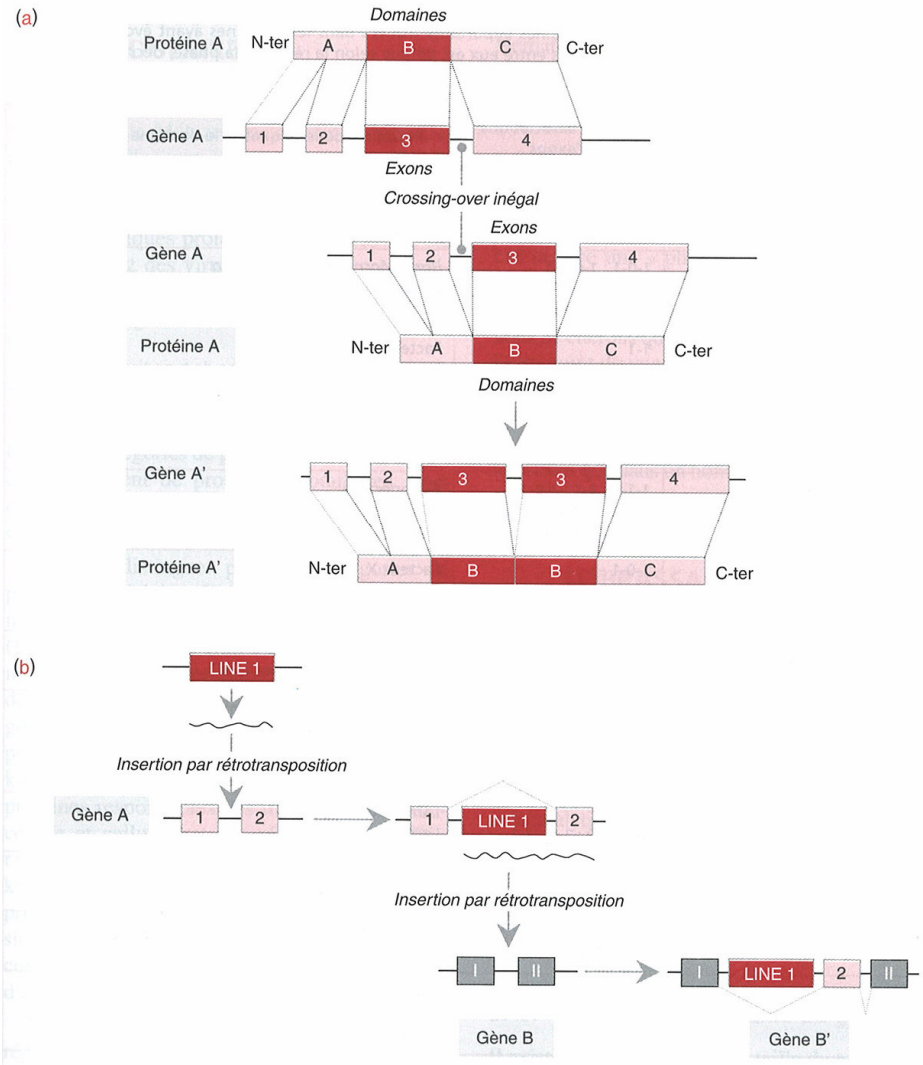


Figure 2.1.6.2 **Mécanismes moléculaires du brassage d'exons**

Deux mécanismes principaux sont à l'origine du brassage d'exons. **(a)** La duplication d'une portion de gène peut se produire par un crossing-over inégal. Il en résulte la formation d'un nouveau gène (A'), et la duplication d'un domaine ou la création d'un nouveau domaine. **(b)** L'acquisition d'un nouvel exon peut se faire par rétroposition. Le nouveau gène formé (B') possède alors un (rarement des) exon supplémentaire. Extrait de (Luchetta et al., 2005).

2.1.6.3. Rôle des introns dans le brassage d'exons

L'évolution par brassage d'exons est directement liée à l'apparition des introns « spilcéosome-dépendants » que l'on trouve dans le noyau des Eucaryotes. Ces introns de grande taille possèdent une machinerie d'épissage qui ne dépend pas de leur séquence nucléotidique interne. Ceci permet des insertions d'exons (par crossing-over inégaux) à l'intérieur même des introns. La machine d'épissage peut ensuite reconnaître les bornes de ces nouveaux exons et les conserver pour créer une nouvelle protéine.

Lorsqu'on analyse de manière approfondie la structure des gènes qui ont évolué par brassage d'exons, on constate que la jonction intron-exon entre les modules ne se fait pas de façon aléatoire, mais se soumet à la « règle de la phase », selon laquelle, quel que soit le mécanisme utilisé, il faut toujours conserver la phase de lecture présente sur l'ARN messager de façon à ne pas modifier la séquence et la structure de la protéine. Ceci impose des contraintes fortes sur les phases des exons insérés et des introns dans lesquels ils s'insèrent. Pour que le système soit fonctionnel, il faut que cette insertion se produise dans une région non-codante et que la phase de lecture de la protéine en aval de l'insertion soit conservée (Luchetta et al., 2005).

Réciproquement, l'observation même que la distribution des phases d'introns dans le génome n'est pas aléatoire suggère qu'un grand nombre de gènes eucaryotes sont apparus par brassage d'exons (Kaessmann et al., 2002; Long, 2001; Long et al., 2003a; Wilhelm and Wilhelm, 2001).

Le brassage d'exons peut également être impliqué dans la formation de gènes possédant des caractéristiques inédites, qui sera abordée ultérieurement (cf 2.2).

2.1.7. Séquences répétées et éléments transposables

2.1.7.1. Les séquences répétées extragéniques

Les génomes eucaryotes contiennent de nombreuses familles de séquences hautement répétées. Elles présentent deux types d'organisation : 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, qui peut-être isolé par centrifugation en gradient de densité, se compose de répétitions d'une séquence plus ou moins

complexe et constitue l'essentiel de l'hétérochromatine. 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. Enfin, l'ADN microsatellite est défini par des séries d'unités simples répétées en tandem et dispersés dans le génome. L'unité de base de l'ADN microsatellite ne compte que quelques nucléotides (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 ou ectopique.

2.1.7.2. Les transposons : des séquences répétées particulières

Les familles d'ADN répétitifs dispersés se composent de séquences transposables. Ces éléments ont été découverts dans les années 1940 par B. McClintock, découverte pour laquelle elle reçut le prix Nobel en 1983 (McClintock, 1956; McClintock, 1984). Il s'agit, en première approximation, de séquences d'ADN capables de se déplacer de façon autonome et de se multiplier dans les génomes. Mais cette définition pose le problème de la perte de mobilité de séquences au cours de l'évolution. Ne pouvant s'en satisfaire, il semble utile de caractériser ces éléments d'un point de vue moléculaire.

On distingue les éléments transposables par leur séquence et leur mode de transposition. Les rétrotransposons font intervenir un intermédiaire ARN pour leur insertion dans le génome, et possèdent, ou non, de longues séquences répétées terminales (LTR Long Terminal Repeat). D'autres éléments transposables utilisent un modèle conservatif médié par une molécule d'ADN (Capy et al., 1997).

Les rétrotransposons à LTR ont une structure proche des rétrovirus et contiennent l'information nécessaire à la synthèse des protéines impliquées dans la rétrotransposition (protéines de capside, enzyme de la transcription et de l'intégration). Les rétrotransposons sans LTR sont subdivisés selon leur taille en LINEs et SINEs (Long et Short Interspersed Nuclear Elements). Les premiers peuvent mesurer plusieurs kilobases et possèdent les séquences nécessaires à une reverse-transcription et une transposition couplées. Les seconds comptent quelques centaines de nucléotides et ne possèdent pas de séquences codantes. Le mécanisme de leur rétrotransposition reste donc énigmatique, mais il semblerait que certains

SINEs soient réverse-transcrits par l'enzyme du LINE auquel ils sont apparentés. Les analyses des génomes séquencés ont en outre montré que des espèces même proches pouvaient présenter des taux d'activité de rétrotransposition très différents (Makalowski, 2000; Wilhelm and Wilhelm, 2001).

Les « simples » transposons se caractérisent par la présence à leurs extrémités de séquences inversement répétées, entourant une ou plusieurs phases de lecture codant pour une transposase. Le transposon est excisé du site donneur et inséré au site accepteur. Au site donneur subsiste, après réparation, la signature du transposon excisé, un fragment de l'élément excisé ou l'élément complet.

Les séquençages de génomes entiers, essentiellement d'organismes modèles, révèlent une grande disparité des éléments transposables, tant du point de vue quantitatif que qualitatif. Le nombre d'éléments transposables est néanmoins très probablement sous-estimé, dans la mesure où ils représentent un obstacle au séquençage et où l'hétérochromatine, qu'ils occupent majoritairement, a été négligée du point de vue du séquençage des génomes. On considère néanmoins que le nombre d'éléments transposables est régulé. La nécessité d'une relative stabilité chromosomique limite leur expansion et cette pression de sélection semble avoir favorisé le maintien d'éléments ayant une capacité de transposition modérée. Différents mécanismes sont proposés pour cette régulation, qui font intervenir les uns les transposons eux-mêmes, d'autres le génome « hôte » (Luchetta et al., 2005).

2.1.7.3. Impact des éléments transposables sur le génome

Du fait de leur présence en grand nombre dans les génomes, les éléments transposables peuvent être à l'origine de remaniements chromosomiques tels que précédemment décrits. Les recombinaisons entre séquences d'un même élément situées à des sites non homologues en sont à l'origine. Si les recombinaisons se produisent entre éléments d'un même chromosome, on peut observer des inversions ou des délétions. S'il s'agit d'éléments situés sur des chromosomes différents, on constate des translocations, équilibrées ou non (Anxolabéhère et al., 2000).

L'insertion d'un élément transposable dans un gène peut en modifier l'expression, rarement de façon avantageuse (c'est bel et bien une réalité probabilistique et statistique). L'impact des transposons au niveau des gènes dépend essentiellement de leur site d'insertion, dans les séquences codante ou régulatrice (voir Fig 2.1.7.3). L'insertion dans une séquence transcrite favorise l'apparition de transcrits composites, dont l'épissage peut être perturbé.

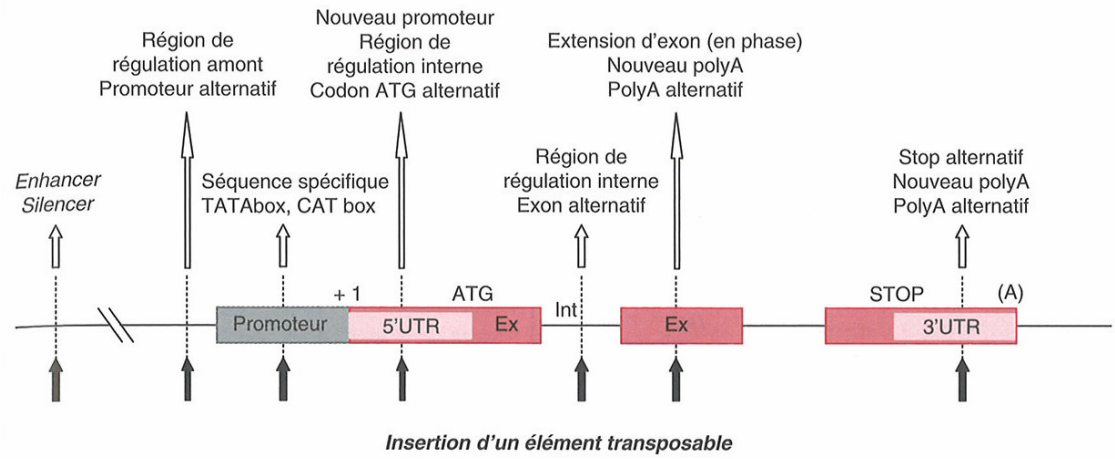


Figure 2.1.7.3 **Insertion d'un élément transposable**

Conséquences possibles (flèches blanches) de l'insertion d'un élément transposable dans différents éléments d'un gène (flèches noires). Extrait de (Luchetta et al., 2005).

Une protéine tronquée est généralement obtenue. Il apparaît néanmoins que des éléments transposables peuvent être identifiés comme codant pour une portion de la protéine dans plus de deux cents cas au sein des génomes de vertébrés (Makalowski, 2000; Zdobnov et al., 2005). L'insertion d'éléments transposables dans des introns peut également favoriser leur recrutement comme exon, aussi appelé exaptation, comme cela a été mis en évidence dans le génome humain. L'arrêt de la transcription est généralement précipité, de sorte qu'on obtient encore une protéine tronquée. En outre, la présence d'un élément transposable dans un gène peut résulter d'une cascade d'évènements et impliquer plusieurs gènes. Ces processus d'insertion d'éléments transposables dans les séquences codantes des gènes se révèlent donc très fréquemment délétères. Ils ne sont toutefois pas systématiquement éliminés par la sélection naturelle, et d'autant moins que les gènes atteints possèdent des paralogues, et sont donc à même de modifier substantiellement le génome. Les éléments transposables sont présents en plus grand nombre dans les régions non transcrites et non traduites des gènes, où la pression de sélection s'exerce de façon moins drastique. Leur insertion dans ces régions peut créer un nouveau promoteur, une région de régulation, un nouveau site (alternatif) d'initiation de la traduction, de polyadénylation ou de terminaison de la traduction. La modification de l'expression d'un gène (niveaux d'expression, spécificité développementale ou tissulaire) par l'insertion d'un élément transposable peut s'effectuer à distance, par création d'une nouvelle région cis-régulatrice.

Enfin, le génome lui-même recrute des éléments transposables pour leurs fonctions intrinsèques. Il utilise pour cela leur première caractéristique : la mobilité. Ainsi, par exemple, la fonction structurelle télomérase, d'élongation des extrémités des chromosomes, est assurée chez la drosophile par des LINEs. De même, la variabilité extraordinaire du système immunitaire des Vertébrés est conférée par une capacité de recombinaison inédite, médiée par des transposases dérivant d'anciens transposons simples et ayant perdu leur mobilité par délétion de leurs séquences terminales inversement répétées (Jones and Gellert, 2004).

Du fait de leur impact important sur les génomes, les éléments transposables participent donc largement à leur plasticité.

2.1.8. Mutations ponctuelles et petites insertions/délétions

Les génomes sont radicalement remodelés et agrandis par recombinaison génétique, nous l'avons vu. Ils sont également ajustés finement par mutation (Luchetta et al., 2005).

2.1.8.1. Qu'est-ce qu'une mutation ?

Les mutations résultent d'erreurs de réplication et de lacunes du système de réparation de l'ADN. Elles peuvent être induites par exposition à différents types de mutagènes présents dans notre environnement ou produits par l'environnement des cellules. La principale source de mutation est cependant représentée par les mutations endogènes, notamment les erreurs spontanées de la réplication et de la réparation de l'ADN.

On distingue, parmi les substitutions, les transitions des transversions, qui sont respectivement des substitutions entre purines (adénine ou guanine) ou entre pyrimidines (thymine ou cytosine) et des substitutions entre purine et pyrimidine. On observe davantage de transitions que de transversions, ce qui s'explique probablement par de moindres différences d'encombrement stérique et le moindre impact fonctionnel lors des transitions. Il faut également mentionner les insertions et délétions de quelques nucléotides, qui lorsqu'elles surviennent dans une région codante provoquent généralement un décalage du cadre de lecture et une modification conséquente de la séquence protéique. Insertions et délétions sont très rares dans les régions codant pour des protéines.

Les mutations affectent aussi bien les régions codantes que non-codantes de notre génome, bien qu'à des taux différents. Lorsqu'elles se situent dans une séquence codant pour une protéine, les mutations sont dites non-synonymes ou synonymes, selon qu'elles affectent, ou non, la séquence polypeptidique. 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.

2.1.8.2. La sélection naturelle et les mutations

L'évolution des séquences d'ADN peut donc être considérée comme étant sous le contrôle combinatoire de deux types de forces : d'une part des forces mutationelles qui produisent aléatoirement dans le génome des mutations elles-mêmes aléatoires quant à leur

utilité éventuelle ; d'autre part des forces sélectives qui vont tendre à éliminer les mutations ou, au contraire, à en augmenter la fréquence dans la population. La sélection négative ou purificatrice permet l'élimination progressive des mutations défavorables pour les individus de la population, tandis que la sélection positive, ou darwinienne, permet la fixation de mutations favorables.

On observe essentiellement les effets de pressions de sélection négative, qui éliminent les mutations des régions codantes, mutations qui occasionneraient des changements phénotypiques significatifs. C'est la raison pour laquelle les régions codantes sont exceptionnellement bien conservées entre espèces, même distantes. La sélection positive serait, elle, rare, et la fixation de mutations est généralement due à la dérive génétique des populations, processus stochastique et aléatoire qui permet la fixation de mutations neutres ou quasiment neutres. E. Zdobnov et ses co-auteurs suggèrent que la complexité des génomes survient à l'occasion d'accumulations de « déchets » génomiques approximativement neutres lorsque la sélection purificatrice est relachée (Zdobnov et al., 2005). Ainsi, si l'on considère un gène, il coexistera généralement de nombreuses variantes de ce gène différant par une ou plusieurs mutations (des allèles) fonctionnellement équivalents dans la population. La population est alors polymorphe pour de nombreux loci, mais ce polymorphisme est fondamentalement instable : les allèles sont en réalité en cours de fixation ou d'élimination. La théorie mettant en avant l'importance d'évènements aléatoires, de mutation et de dérive, dans l'évolution moléculaire a été appelée neutralisme et doit beaucoup à M. Kimura. Notons qu'il est à présent admis que neutralisme et sélectionnisme constituent deux moteurs importants de l'évolution des génomes.

La sélection s'exerçant sur une séquence codante peut être déterminée par le calcul des taux de substitutions synonymes et non-synonymes, notés respectivement K_S et K_A. Ces valeurs sont calculées en comparant les séquences orthologues dans deux espèces différentes en utilisant les formules empiriques suivantes :

$$K_S = -3/4 \ln [1-(4M_S/3N_S)]$$

$$K_A = -3/4 \ln [1-(4M_A/3N_A)]$$

Où M_S et M_A sont les nombres de substitutions, respectivement synonymes et non-synonymes, et N_S et N_A les nombres de sites, respectivement synonymes et non synonymes (selon la dégénérescence ou non du code génétique au locus analysé). Ces formules sont établies selon la méthode de Jukes et Cantor à un paramètre, qui considère que les taux de transitions et transversions sont égaux. On considère généralement qu'un rapport K_A/K_S à

peu près égal à 1 indique l'absence de sélection, $K_A/K_S > 1$ une sélection positive et $K_A/K_S < 1$ une sélection négative. $K_A/K_S > 1$ indique en effet que des substitutions changeant des acides aminés s'accumulent davantage (et plus vite) que les mutations neutres, suggérant qu'elles sont positivement sélectionnées. L'analyse du rapport K_A/K_S permet donc d'obtenir des indications sur les pressions de sélection s'exerçant sur les séquences codantes, mais ces informations restent partielles et ne sont pas toujours conclusives. En particulier, des phénomènes tels que la conversion génique sont susceptibles de réduire ou d'accentuer les effets d'une éventuelle pression de sélection. La sélection positive a néanmoins pu être ainsi confirmée, aussi bien chez les procaryotes que chez les Primates. Les gènes pour lesquels une sélection positive a été mise en évidence sont essentiellement impliqués dans trois grandes fonctions : les systèmes de défense et immunité, l'évitement des systèmes de défense et immunité (chez des organismes pathogènes) et la reproduction.

2.1.8.3. La notion d'horloge moléculaire

La notion d'horloges moléculaires est basée sur l'observation, faite dans les années 1960, que les taux de substitution des acides aminés de certaines protéines étaient à peu près les mêmes dans différentes lignées de Mammifères. E. Zuckerlandl et L. Pauling proposèrent en 1965 le postulat selon lequel le taux d'évolution est constant au cours du temps, pour toute protéine et toute espèce, ce qui implique, en d'autre terme, l'existence d'une horloge moléculaire (Benton and Ayala, 2003). Ce modèle fait également appel au neutralisme, dans la mesure où il requiert d'admettre que les processus façonnant le génome sont majoritairement liés à des évènements neutres du point de vue de la sélection naturelle (Ohta, 1987; Zdobnov et al., 2005). Ainsi, le niveau de divergence d'une protéine ou d'un gène entre deux espèces permettrait la mesure du temps écoulé depuis leur dernier ancêtre commun.

Cette idée de taux d'évolution contraint a été contestée à trois niveaux (Douzery et al., 2004).

- Tout d'abord, il est à présent clairement démontré que les vitesses d'évolution varient au gré des gènes ou protéines. Ces différences sont vraisemblablement dues à des différences de pression de sélection, mais aussi à des inégalités lors de la réplication et de la réparation des séquences d'ADN. Les séquences de l'hétérochromatine sont en effet répliquées plus tardivement au cours du cycle cellulaire, à un moment ou les enzymes sont mobilisées par d'autres sites, ce qui affecte la fidélité de la réplication. En outre, ces séquences sont moins riches en GC que celles de l'euchromatine, et le système de

réparation semble y être moins performant. Il n'y aurait donc pas d'horloge moléculaire unique et universelle, mais la possibilité d'horloges différant d'un gène à l'autre subsiste.

- En outre, il apparaît distinctement lors d'analyses phylogénétiques que les vitesses d'évolution dépendent également de l'espèce. Britten a montré dès 1986 que l'évolution moléculaire avait ralenti chez les espèces à durée de génération plus longue. En effet, les mutations transmises doivent apparaître lors de la gamétogénèse, puis ne peuvent envahir la population qu'après reproduction. La limitation de la fréquence des générations réduit donc la probabilité de transmettre une mutation. Ces espèces possèdent également un système de réparation de l'ADN plus performant. Ainsi, la vitesse d'évolution des Primates est inférieure à celle des rongeurs, les autres mammifères adoptant une position intermédiaire. De même, le rythme apparent de l'horloge moléculaire est d'autant ralenti que la taille effective de la population est grande, en raison d'une sélection purificatrice plus prononcée (Zdobnov et al., 2005). Ces différences de vitesse évolutive entre lignées peuvent être testées et quantifiées.
- Enfin, il semblerait qu'il existe des variations de la vitesse d'évolution d'un gène dans une lignée au cours du temps (Alba and Castresana, 2005). Ce dernier point est le plus problématique, mais aussi le plus contestable, bien qu'un nombre croissant d'études indique un ralentissement de la vitesse d'évolution d'un gène au cours de son histoire évolutive.

La notion d'horloge est donc à prendre avec mesure et précautions. Elle permet néanmoins, après une étape de calibration indispensable, de dater les évènements, et ce aussi bien à partir de données phylogénétiques relatives aux organismes qu'aux gènes (Lopez et al., 2002). On précise ainsi respectivement des évènements de spéciation et de duplication. Ces deux types d'approches seront développés au cours de ma thèse.

2.2. La création *de novo* : un générateur de nouveauté

2.2.1. Devenir de gènes, modules ou séquences dupliqués et acquisition de fonctions originales

Comme décrit précédemment (cf 2.1.4), les gènes dupliqués subissent bien plus qu'une perte de fonction. La relaxation des contraintes pesant sur leur évolution favorise l'apparition de fonctions nouvelles, éventuellement complémentaires de celles du gène

ancestral (Balakirev and Ayala, 2003). De même, l'organisation en modules et le brassage d'exons (cf 2.1.6) est très favorable à l'acquisition de nouvelles fonctions protéiques, ou plus exactement de nouvelles combinaisons de fonctions protéiques. Néanmoins, dans un cas comme dans l'autre, la parenté des séquences se traduit par des proximités fonctionnelles.

On observe cependant des pseudogènes dupliqués ayant acquis des fonctions véritablement distinctes de celle du gène dont ils dérivent. Ainsi, le pseudogène *Makorin 1-p1* est le premier exemple connu de pseudogène jouant un rôle sur la production des ARN de sa propre copie active *Makorin 1* (Hirotsune et al., 2003; Lee, 2003). Il s'agirait d'un titrage d'agents de dégradation de l'ARN ou de répression de l'activité génique, de sorte que l'existence du pseudogène *Makorin 1-p1* favorise une élévation du niveau d'expression du gène *Makorin 1*. D'autre rôles et fonctions inédits des pseudogènes pourront être révélés par le développement d'observations de ce type.

2.2.2. Rétroposition, exaptation d'exons et création de nouveauté génique

Contrairement aux pseudogènes dupliqués, le rétropseudogène s'insère dans un nouvel environnement génétique, et peut donc jouer un rôle primordial dans la création de nouveautés géniques. Les étapes de duplication et d'inactivation se font de manière concomitante pour le rétropseudogène, dans la mesure où le promoteur n'est pas impliqué. Ceci élimine tout risque d'interférence avec le produit de synthèse d'origine.

Il faut donc envisager plusieurs possibilités quant à l'évolution des rétropseudogènes en fonction du lieu où ils s'insèrent. Ils peuvent soit redevenir actifs, s'ils sont insérés dans une région qui possède fortuitement un promoteur (Long, 2001), soit rester « en attente » d'une nouvelle fonction s'ils sont dans une région qui permettra plus tard de recréer un promoteur. 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 rétropseudogène qui sera co-transcrit avec ces nouveaux exons (Long, 2001).

De nombreux exemples de gènes actifs provenant de rétropseudogènes ont déjà été décrits et, parmi les mieux caractérisés, on peut citer celui de la super-famille des récepteurs couplés aux protéines G chez les Vertébrés, qui inclut des récepteurs olfactifs, des récepteurs de neurotransmetteurs, de cytokines et de neuropeptides... Plusieurs des gènes de cette famille n'ont pas d'introns dans leur région codante. Il apparaît distinctement que ces rétropseudogènes constituent donc un « réservoir » pour la création de nouveaux récepteurs,

fixant d'autres ligands, exerçant d'autres fonctions ou exprimés de façon différentielle (Brosius, 2003).

En outre, et de façon particulièrement notoire, les rétropseudogènes peuvent aussi créer directement un « nouveau gène » s'ils s'associent, par le processus d'exaptation précédemment décrit, à des séquences codantes préexistantes ou à des exons créés *in situ* et *de novo* par le processus d' « exonisation ». Plusieurs cas de telles créations ont été décrits, l'un des plus exemplaires étant celui des gènes *PMCHL*, qui fait l'objet de mon travail de doctorat. Il faut également citer le cas du gène jingwei de la Drosophile, qui a fait l'objet d'une analyse particulièrement fine (Long et al., 2003b).

Dans l'analyse des processus de l'évolution, nous nous sommes particulièrement intéressés aux évènements récents d'apparition de familles géniques chez l'Homme. La lignée des Primates s'impose dès lors comme un bon modèle d'étude, qui nous permet de surcroît d'examiner également quelques spécificités humaines.

3. Les Primates et leur évolution

3.1. Les Primates

3.1.1. Phylogénie des Primates actuels

3.1.1.1. Définition des Primates

L'ordre des Primates fut introduit par Charles Linné en 1758 dans la dixième édition de son *Systema Naturae*. L'Homme figurait déjà aux côtés des singes et des paresseux dans l'édition de 1735 pour former l'ordre des anthropomorphes, dont le nom fut donc ensuite changé en Primates, terme qui signifie premiers. Ils sont alors définis aux moyens des caractères partagés suivants : « *Quatre incisives inférieures parallèles, une seule canine* [à chaque demi-mâchoire]*, une paire de mamelles pectorales, extrémités des membres antérieurs propres à saisir, clavicules, se déplacent sur quatre pattes, grimpent aux arbres et se nourrissent de fruits* ». Le groupe ainsi constitué par Linné comprenait l'homme, les Prosimiens, les Simiens, mais aussi les Dermoptères et Chiroptères (grande famille des chauves-souris) qui en ont depuis été exclu. Les auteurs ont par la suite tous souligné la difficulté à définir les Primates, quand bien même on se limiterait aux seules espèces actuelles. Cela tient au faible nombre de caractères effectivement partagés par tous. On a ainsi pu dire qu'aucun trait distinctif ne caractérise tous les Primates à l'exception d'un caractère négatif : l'absence de spécialisation. En d'autres termes, tous les Primates vivants partageraient un certain nombre de caractères primitifs qui auraient été perdus par tous les autres mammifères placentaires. L'ordre des Primates n'a finalement pu être délimité qu'en utilisant un agrégat de caractères définis par des tendances évolutives, liées pour la plupart à des fonctions adaptatives majeures telles que la vision, la structure de l'oreille moyenne, le mode de locomotion et la reproduction (Beard, 1993; Thomas, 1999) (cf. 3.1.1.2). Naturellement, aucun de ces caractères pris isolément n'est exclusif aux Primates.

3.1.1.2. Phylogénie des Primates

L'établissement de cladogrammes illustrant les relations entre espèces, en l'occurrence entre Primates, impose la reconnaissance des similarités. On en distingue de trois sortes. D'abord, les similarités convergentes ont été développées indépendamment, tandis que les

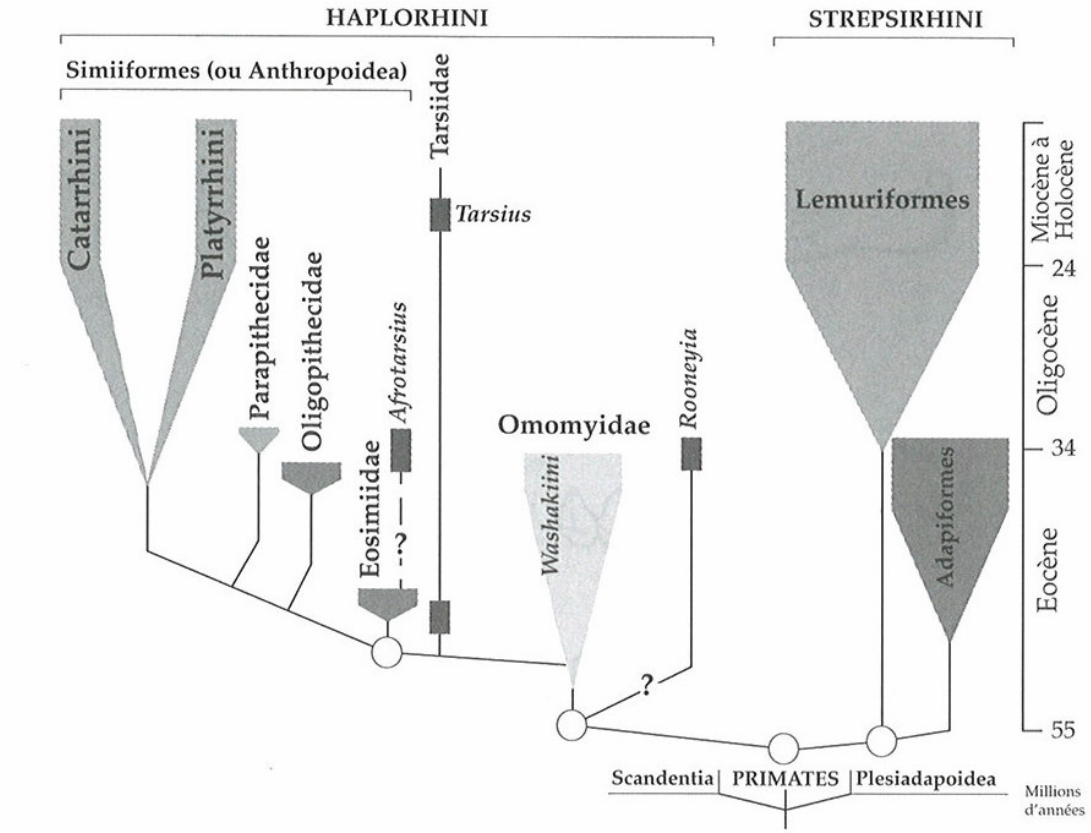


Figure 3.1.1.2.a **Phylogénie des Primates**

Illustration des temps d'apparition des espèces actuelles et fossiles de Primates et de leurs relations de parenté. Les grands groupes consensuels de primates y figurent. Seul le positionnement des tarsiers, non-conforme au consensus actuel pose problème. Tiré de (Thomas, 1999), d'après (Kay et al., 1997). Les dates de divergence approximatives sont les suivantes : Strepsirhiniens/Haplorhiniens : 65 à 85MA, Tarsiers : 50-65MA, Platyrhiniens/Catarhiniens : 35-40MA. MA : millions d'années.

similarités par homologie sont héritées d'un ancêtre commun. Parmi elles, les traits primitifs sont partagés par toutes les espèces étudiées et hérités de l'ancêtre commun à toutes, tandis que les traits dérivés sont apparus chez un ancêtre plus tardif au sein de l'arbre. Seules ces dernières similarités sont indicatives des relations phylogénétiques entre les espèces. Ce type de raisonnement a d'abord été appliqué aux traits phénotypiques des espèces, mais convient également à l'analyse de leurs séquences nucléotidiques (Jones et al., 1992).

A la base des Primates, la subdivision entre Strepsirhiniens (aussi appelés Prosimiens) et les Haplorhiniens (les Simiens) fait à présent consensus. La forme de leur face et le développement d'un rhinarium les distingue. Les tarsiers ont longtemps posé un problème de positionnement du fait de leur ressemblance aux Prosimiens mais de leur absence de rhinarium ; on les classe désormais parmi les Haplorhiniens. Parmi les Haplorhiniens, l'infraordre des Platyrhiniens est caractérisé notamment par des narines écrasées et une longue queue préhensile. Les Catarhiniens, qui regroupent cercopithécoïdes et Hominoïdes, ne présentent pas ces traits. L'une des difficultés de la définition de l'organisation des Primates résulte de la seule observation des espèces actuelles. Une chronologie de leur apparition peut néanmoins être établie (cf Fig 3.1.1.2.a), ainsi qu'une classification des principaux groupes actuels et sub-fossiles (cf Fig 3.1.1.2.b). La classification illustrée présente néanmoins un écueil au regard des controverses actuelles : les Pongidés y regroupent les chimpanzés (*Pan*), les gorilles (*Gorilla*) et les orangs-outans (*Pongo*) et en excluent l'homme (*Homo*).

Les relations phylogénétiques entre Hominoïdes (les *Hylobates*, *Pongo*, *Gorilla*, *Pan* et *Homo*) font en effet débat et quatre possibilités majeures s'affrontent (cf Fig 3.1.1.2.c), la plus probable et raisonnable, compte tenu des connaissances actuelles, étant celle d'un ancêtre commun aux gorilles, chimpanzés et hommes et d'une divergence simultanée des lignées de ces trois genres il y a 5 à 7 millions d'années (Bailey, 1993; Dutrillaux, 1980; Goodman et al., 1994). L'hypothèse d'un phylum regroupant les chimpanzés et l'homme retient également tout particulièrement notre attention (Barriel, 2004; Gagneux and Varki, 2001; Goodman, 1999). Elle assure que le dernier ancêtre commun à l'homme et au chimpanzé vivait il y à 4,6 à 6,2 millions d'années, tandis que le gorille avait divergé il y a 6,2 à 8,4 millions d'années (Chen and Li, 2001).

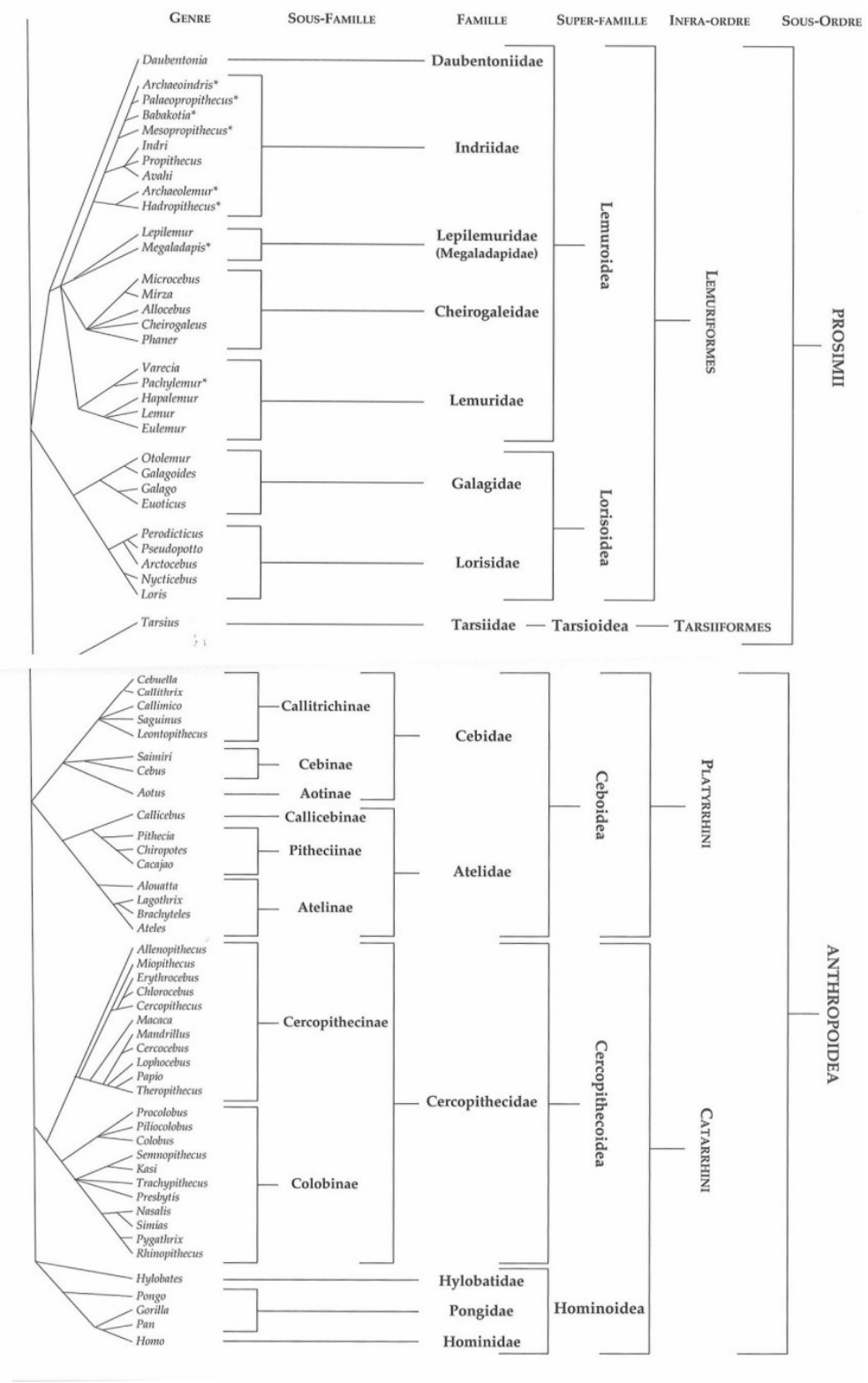


Figure 3.1.1.2.b **Classification des principaux genres actuels et subfossiles* de l'ordre des Primates**

Tiré de (Thomas, 1999).

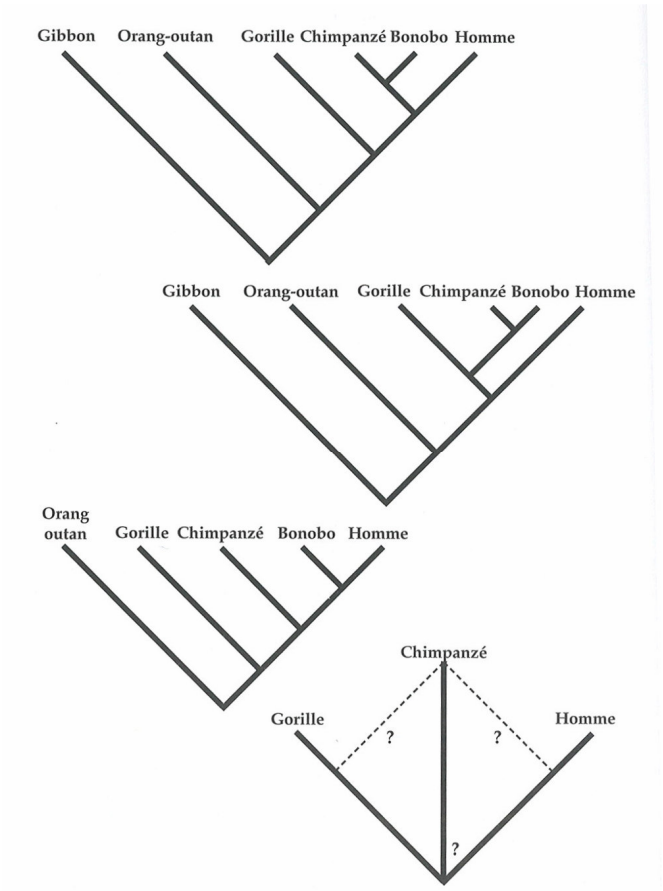


Figure 3.1.1.2.c **Relations de parenté entre les Hominoïdes actuels**

Les différents travaux écartent l'orang-outan (*Pongo*) d'une parenté étroite avec l'homme. Il reste néanmoins trois possibilités principales (présentées de haut en bas) :

- L'ensemble des Chimpanzé (*Pan*) serait le plus proche parent de l'homme,
- l'ensemble Gorille-Chimpanzé (*Hylo –Pan*) serait le groupe le plus proche de l'homme,
- Seul le Bonobo (*Pan paniscus*) aurait droit au titre de plus proche parent.

Finalement, il semble plus prudent et plus logique au vu des différentes approches, de proposer une trichotomie Chimpanzé-Homme-Gorille, tant que nous n'avons pas de meilleurs arguments.Tiré de (Senut, 1999).

3.1.2. Biologie des Primates et impact de l'environnement

3.1.2.1. Répartition géographique des Primates : niches et évènements écologiques

La grande diversité des Primates, qui rend l'établissement de leur classification si ardue, a déjà été soulignée. On compte d'ailleurs 200 à 235 espèces de Primates vivants (Corbet and Hill, 1991; Groves, 1993), constituant un ordre très varié, dont la diversité, rapportée à leur aire de distribution actuelle (essentiellement les zones tropicales et sub-tropicales, si l'on excepte l'homme), est considérable. Les différents genres de Primates occupent en effet des régions et niches écologiques bien distinctes, comme le souligne la figure 3.1.2.1. Les Primates ont ainsi colonisé, sous ces latitudes, aussi bien des milieux forestiers, que des savanes et steppes et même des milieux semi-arides et désertiques (Thomas, 1999).

Davantage de données relatives à la géographie et à l'écologie sont disponibles pour les Hominoïdes. Leur habitat, à l'exclusion de l'homme, est désormais restreint aux zones intertropicales d'Asie et d'Afrique. Les Hylobatidés, de petite taille, ne sont qu'asiatiques, de même que les Pongidés. Gorilles et Chimpanzés ne sont eux présents qu'en Afrique. Enfin, l'homme, exception notable, a colonisé tous les continents et toutes les latitudes. Cette répartition, et son évolution, est conditionnée par le contexte chronologique et paléontologique. La connaissance et la prise en compte des variations du climat, de la géologie et par conséquent de l'environnement permettent de mieux comprendre cette répartition actuelle.

L'ouverture de l'Atlantique sud, qui a commencé il y a quelque 120 millions d'années, a progressivement séparé l'Amérique du Sud de l'Afrique, coupant ainsi les flux géniques entre les individus des espèces qui vivaient sur cette aire géographique jadis commune. Cette séparation a naturellement été très lente, s'étalant sur plus de 30 millions d'années et, pendant très longtemps, des échanges restaient possibles entre les deux nouveaux continents, mais ces échanges se sont peu à peu réduits, jusqu'à un isolement complet des faunes et des flores sud-américaines d'une part et africaines d'autre part.

Mais c'est par d'autres phénomènes que se sont séparés les Platyrhiniens et les Cathariniens, également appelés singes du Nouveau et de l'Ancien Monde, respectivement. Probablement transportés par des radeaux naturels, des primates et des rongeurs africains ont pu coloniser l'Amérique du Sud il y a environ 40 millions d'années, alors que l'Atlantique Sud était déja ouvert.

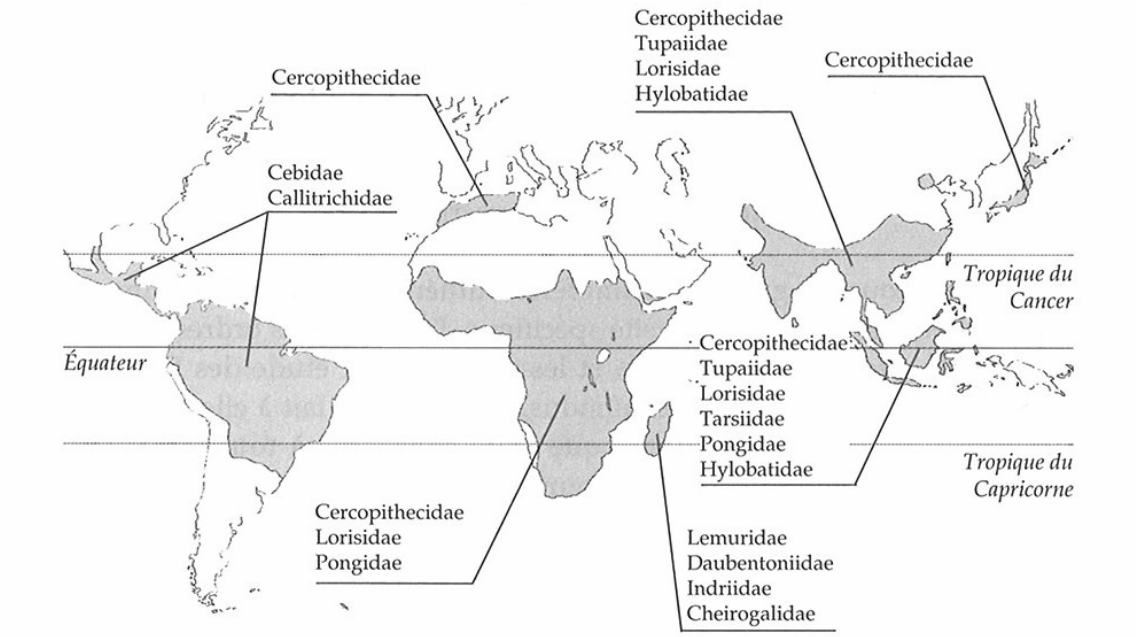


Figure 3.1.2.1 **Répartition mondiale des Primates actuels non-humains**

On observe une répartition presque exclusivement intertropicale. Tiré de (Thomas, 1999).

Il y a 19 millions d'années, la plaque tectonique d'Afrique heurte la plaque eurasiatique, établissant une connexion entre elles qui favorise la dispersion de la faune africaine, et en particulier des grands singes panafricains du Miocène. Il y a 15 millions d'années, la fermeture de la mer Téthys modifie la circulation des eaux marines, engendrant d'importants changements des climats locaux et globaux. C'est vraisemblablement à ce moment que les Pongidés envahissent l'Eurasie. Parallèlement, la diversité des grands singes s'amoindrit nettement en Afrique, peut-être en raison d'une compétition avec les cercopithècoïdes très présents. Cette dernière hypothèse n'est cependant pas suffisante, puisque le déclin des Hominoïdes d'Afrique aurait débuté avant l'expansion des cercopithècoïdes et qu'ils n'occupent pas exactement les mêmes niches écologiques. Beaucoup plus récemment, il y a 6 à 8 millions d'années, les grands singes déclinent en Eurasie, ou seuls subsistent les orangs-outans et gibbons. On assiste ainsi à une concentration géographique des Hominoïdes, alors qu'ils étaient dispersés au Miocène (il y a 15 à 20 millions d'années). Cette brève histoire de la répartition géographique des grands singes illustre avec force que l'évolution des Hominoïdes est profondément liée aux facteurs géographiques et géologiques (Andrews, 1992; Gibbons, 2002; Senut, 1999).

3.1.2.2. Diversité biologique et phénotypique

Les Primates, comme cela a été mentionné précédemment, sont très divers à de multiples égards. Leurs caractéristiques physiques, mais également leurs biologie et vie sociale (voire sociétale) sont extrêmement variées.

3.1.2.2.1. Caractéristiques physiques

Le poids des Primates varie, pour les seules espèces actuelles, d'un facteur proche de 8000, entre le plus petit des microcèbes qui ne dépasse pas 30 grammes et les gorilles, qui atteignent 250 kilogrammes. Quant à leur taille, elle est, selon les espèces, de 13 à 180 centimètres.

Un certain nombre de caractères physiques sont en revanche partagés par les Primates, mais à des degrés variables en fonction de leurs degrés d'évolution et en corrélation avec leurs modes de vie, qui seront détaillés ultérieurement. Le premier de ces traits est l'organisation caractéristique du pied et de la main avec pouces en opposition, et semblerait être lié à la vie arboricole s'il était partagé avec d'autres mammifères arboricoles non-Primates, ce qui n'est pas le cas. Un autre trait phénotypique manifeste repose sur la position centrale « en avant »

des yeux de la plupart des Primates, dont on a supposé qu'elle est liée au mode de vie arboricole, mais aussi au développement de la vision au détriment de l'odorat, notamment pour ce qui concerne la prédation. Ce placement des yeux permet en effet une vision binoculaire.

D'autres traits morphologiques, enfin, servent précisément à la classification des Primates et signent leur diversité, comme cela a été précisé précédemment (cf 3.1.1.2). La face des Strepsirhiniens est allongée en museau, avec formation d'un rhinarium, zone cutanée sans poils entourant les narines, avec absence de soudure de la lèvre supérieure ; celle des Haplorhiniens n'est plus allongée en museau, le rhinarium a disparu et la lèvre supérieure est soudée. Ces deux sous-ordres des Primates sont d'ailleurs définis par ces différences morphologiques. On distingue ensuite deux infra-ordres des Haplorhiniens : celui des Platyrhiniens, à narines écartées et à queue longue et préhensile, celui des Catarhiniens, à narines rapprochées, à queue parfois absente et de toute façon jamais préhensile. En plus de ces deux caractères distinctifs, Platyrhiniens et Catarhiniens possèdent des nombres de dents différents.

3.1.2.2.2. Biologie des individus

Les caractéristiques biologiques des différents sous-ordres, genres, espèces de Primates sont également variables, et ce, à plusieurs égards.

Si la majorité des Primates est diurne, d'autres sont cependant demeurés nocturnes. A l'exception d'une espèce, tous les Haplorhiniens sont actifs le jour. En revanche, l'essentiel des Strepsirhiniens est nocturne, seuls les lémuridés et quelques indriidés ayant accédé à la vie crépusculaire ou diurne. Cette différence de mode de vie semble être étroitement liée au poids corporel. En effet, à de rares exceptions près, les Primates de plus d'un kilogramme sont diurnes, les autres sont nocturnes (Thomas, 1999).

Cette variété de rythme d'activité s'accompagne de différences plus ou moins marquées du développement des sens de l'ouïe, de l'olfaction et de la vision. La tendance générale des Primates est à l'augmentation de l'acuité visuelle et à l'acquisition de la vision binoculaire, au détriment de l'olfaction et de l'audition. Cette évolution est cependant nettement plus marquée chez les espèces diurnes que chez les espèces nocturnes.

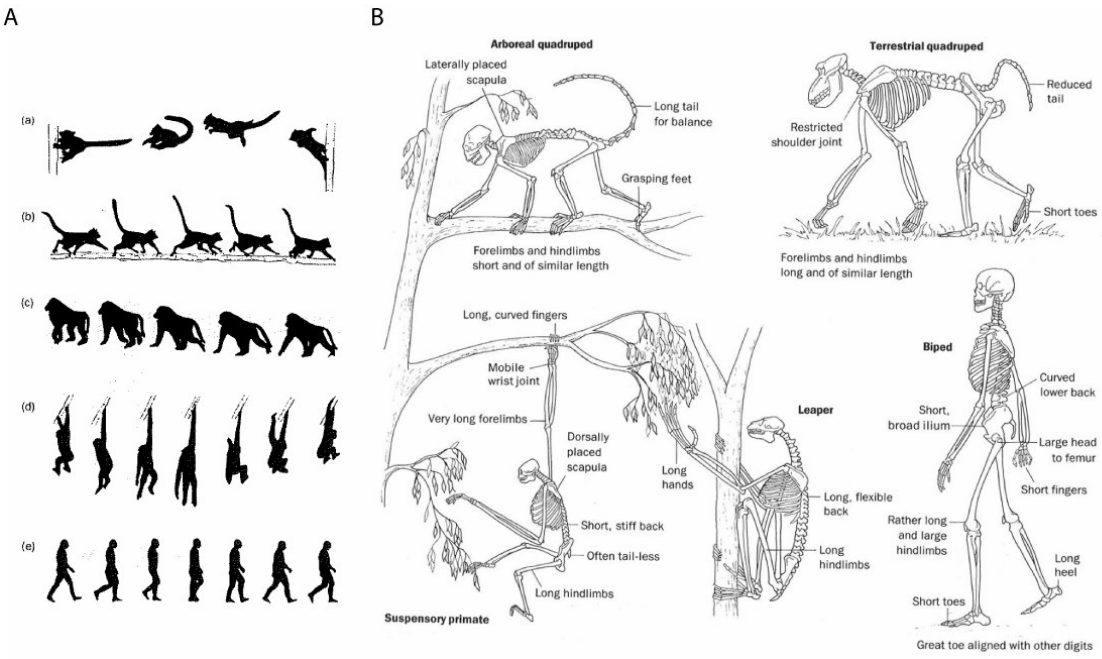


Figure 3.1.2.2.2 **Modes de locomotion des Primates**

A, Les primates se déplacent selon cinq modes distincts : le saut (a), la quadrupédie arboricole (b), la quadrupédie terrestre (c), la suspension (d) et la bipédie (e). **B**, ces différents modes de locomotion s'accompagnent de changements anatomiques (Jones et al., 1992).

- Ainsi, au sein des Primates, la communication par l'odeur est prévalente chez les Prosimiens, bien qu'elle soit également relativement développée chez les Platyrhiniens. L'olfaction est en revanche nettement moins développée chez les Catarhiniens, et n'est presque pas utilisée pour communiquer, bien que ces signaux puissent être importants dans certains contextes sociaux particuliers. La communication olfactive permet aux mammifères de distinguer les espèces, individus, sexes, conditions sexuelles, identités de groupes, dominances relatives et états émotionnels. Chez les Primates, les odeurs sont essentiellement utilisées pour l'espacement interindividuel et intergroupe, pour l'établissement et l'affirmation d'une dominance et pour l'indication de l'état sexuel (Jones et al., 1992).
- L'ensemble des Primates utilise ses traits caractéristiques et muscles faciaux pour exprimer des émotions. Les expressions faciales participent ainsi aux relations sociales. Le répertoire des expressions faciales est partagé par tous les Primates, mais l'intensité et la fréquence de ces expressions sont très variables, de même que les couleurs et morphologies des faces de Primates. L'ensemble de ces caractéristiques faciales est plus nettement associé au comportement et à l'écologie de l'espèce chez les Catarhiniens, dont la vue est également plus performante. La vision est en effet chez ces espèces le sens primordial et l'ensemble du système visuel y est surdéveloppé. La vision est stéréoscopique et trichromatique, les aires visuelles cérébrales occupent davantage d'espace (Jones et al., 1992).

Les Primates recourent également à une extraordinaire diversité de modes de locomotion. Certaines espèces courent le long des branches, d'autres sur le sol. Certains battent des bras, d'autres bondissent d'arbre en arbre. Certains posent leurs pattes antérieures à plat sur le sol, d'autres marchent sur la pointe des doigts, d'autres encore sur leurs articulations. Enfin, seule l'espèce humaine a évolué vers la bipédie. Cette grande variété de modes de locomotion et de postures permet à chaque espèce de s'adapter au mieux à son environnement et requiert le développement d'adaptations anatomiques particulières. On distingue cinq grandes catégories de comportement locomoteur : la quadrupédie arboricole, la quadrupédie terrestre, le saut, la suspension et la bipédie (cf Fig 3.1.2.2.2). La quadrupédie arboricole est le mode de locomotion le plus fréquemment employé par les espèces de Primates, aussi bien Strepsirhiniens qu'Haplorhiniens, qui utilisent alors souvent leur queue

en guise de balancier. La quadrupédie terrestre est communément développée par des singes Catarhiniens, tels que les Babouins (*Papio*). Les lémuriens et tarsiers, de petite taille, emploient essentiellement le saut pour se déplacer, tandis que d'autres espèces, appartenant à divers ordres, se suspendent pour grimper et avancer. Enfin, la bipédie est propre à l'espèce humaine (Jones et al., 1992).

Dans leur habitat naturel, les Primates sélectionnent leurs aliments en fonction de leurs besoins nutritifs. Les aliments les plus abondants, mais aussi les moins digestes, sont les feuilles, que seuls les plus gros mammifères sont capables d'assimiler sans troubles. D'autres mammifères, carnivores ou insectivores, sont adaptés à l'exploitation des ressources animales, plus rares et plus mobiles, mais aussi plus nourrissantes. Seuls les Primates, et les Haplorhiniens en particulier, ont acquis la capacité d'exploiter les ressources intermédiaires : les fruits. Cependant, en raison de l'incapacité à digérer à la fois de grandes quantités de ressources animales et végétales, aucune espèce (à l'exception notable de l'espèce humaine, qui cuisine…) ne peut réellement être qualifiée d'omnivore. On peut caractériser les régimes alimentaires des Primates par l'apport relatif des trois types de ressources alimentaires. Pour l'essentiel des Primates, les fruits représentent 55 à 80% des apports. Les ressources animales et foliaires constituent alors des apports d'appoint. Il est difficile d'établir une corrélation entre le positionnement phylogénétique d'une espèce et son régime alimentaire (Jones et al., 1992).

3.1.2.2.3. Biologie des espèces

La biologie de l'espèce, en terme de reproduction et de vie sociale notamment, présente également quelques particularités et différences au sein des Primates.

Les Primates possèdent une caractéristique reproductive distinctive : ils se reproduisent lentement et produisent peu de descendants, privilégiant la qualité à la quantité. Certaines Primates femelles ont plusieurs petits simultanément, mais la majorité n'en porte qu'un à la fois et le développement est toujours relativement lent, bien que les petits soient bien développés à la naissance (on parle de descendance précoce). La maturité sexuelle est atteinte tardivement, la période de gestation est souvent longue et les naissances espacées. Les Primates ne consacrent qu'une petite portion de leurs ressources à la reproduction. En revanche, ils s'investissent tous dans l'élevage des petits, et ce, pour un temps relativement

long après la naissance. Au sein de l'ordre des Primates, on observe des différences majeures entre Strepsirhiniens et Haplorhiniens, différences de forme de placenta et de taille relative du nouveau-né (à taille égale de la mère, les nouveaux-nés Haplorhiniens sont en moyenne trois fois plus gros que les nouveaux-nés Strepsirhiniens). Pour ces deux aspects, l'espèce humaine est absolument représentative des espèces d'Haplorhiniens, à l'exception d'une particularité notoire : le cerveau humain se développe encore dans l'année qui suit la naissance comme celui des autres Haplorhiniens le fait au cours de la vie fœtale. Ceci explique la grande dépendance du petit d'Homme aux soins parentaux.

En terme de relations sociales, la reproduction des Primates se définit également par un moindre investissement (en terme de ressources allouées) des mâles relativement aux femelles. La polygynie constitue la norme et les cas de monogamie restent anecdotiques. Les comportements sociaux associés à la reproduction (rencontre, dominance, soins parentaux, lactation…) ne peuvent être corrélés à la classification des Primates (Jones et al., 1992).

Les traits d'histoires de vie sont donc éminemment variables entre Primates. La taille corporelle est en général un bon élément prédictif du mode de vie, comme cela a d'ailleurs été mis en évidence précédemment. Mais elle ne fait pas tout. En comparaison aux autres mammifères, à gabarit équivalent, les Primates vivent lentement et possèdent un gros cerveau. La taille du cerveau serait d'ailleurs un meilleur élément de corrélation aux différents traits d'histoire de vie. Cependant, certains Primates tendent à se développer plus lentement que les autres, même après correction des effets de la taille. Ainsi, les espèces dont la gestation est plus courte sont également celles dont la maturité sexuelle est la plus précoce. Ces différences du « rythme » de vie pourraient être liées à l'habitat, comme l'indique le constat que, toutes choses égales par ailleurs, les espèces vivant dans la forêt humide tropicale se développent bien plus lentement que les espèces occupant la savane ou la forêt secondaire. La stabilité environnementale de la forêt humide tropicale, où la sélection naturelle favorise l'attribution de ressources à la production de jeunes compétitifs, qui se développent lentement, pourrait expliquer ce constat (Jones et al., 1992).

Un certain nombre d'espèce de Primates présentent une véritable organisation sociale. D'autres cependant ne forment pas de groupe social ; il s'agit essentiellement des Primates nocturnes qui restent solitaires. Certaines espèces nocturnes vivent en famille monogame, mais jamais dans de plus grands groupes. Ces espèces solitaires ont naturellement des interactions sociales. La rencontre entre individus se produit la nuit et dépend largement des

odeurs et vocalisations. Les signaux visuels sont davantage utilisés par les Primates diurnes. Ces derniers vivent en général dans de grands groupes stables (d'une famille à une horde). Deux pressions de sélection particulières peuvent expliquer la formation de groupes sociaux chez les Primates. La première est la prédation, le groupe assurant une meilleure éviction et une meilleure défense contre le prédateur. Le deuxième facteur repose sur la mutualisation des efforts pour l'accès à la nourriture, notamment pour les arbres fruitiers dispersés. La vie en groupe accroît également la compétition pour l'accès aux ressources locales, il est donc nécessaire, pour le maintien de cette structure sociale, que les bénéfices soient supérieurs aux coûts. On classifie les Primates dans quatre groupes sur les bases de leur organisation sociale : les individus solitaires, les couples, les harems et les groupes contenant au moins deux mâles et plusieurs femelles. Pour toutes les espèces de Primates, le premier lien social est celui qui s'établit entre la mère et son petit, et les femelles et leur descendance femelle (autrement dit les matrilignages) forment le noyau des groupes sociaux. Il est plus ou moins aisé de déterminer le profil social de chacune des espèces de Primates, la plus grande hétérogénéité étant observée dans l'espèce humaine, ou coexistent, dans différentes populations, monogamie, polygynie et polyandrie (Jones et al., 1992).

3.2. Les Primates et leur évolution

L'évolution des Primates peut être analysée suivant deux composantes : ce qui rend les Primates différents des autres Mammifères, par exemple de la souris, et ce qui diffère entre espèces de Primates. On s'intéressera notamment aux différences génomiques, transcriptomiques et enfin pathologiques.

3.2.1. Modifications génomiques des Primates et au sein des Primates

Un certain nombre de paramètres génomiques sont identifiés qui distinguent les Primates de la Souris, avec laquelle le dernier ancêtre commun date de 80 millions d'années. D'autres différences isolent l'une ou l'autre des espèces ou lignées de primates. L'analyse de ces différences semble d'autant plus intéressante que si l'on admet que « ce qui est conservé entre espèces distantes doit être fonctionnellement important », on assure aussi qu'entre espèces proches « ce qui diffère pourrait être important pour expliquer les différences phénotypiques » (Ruvolo, 2004). Ces changements génomiques sous-tendent en outre les capacités uniques au cours de l'évolution de l'Homme et des autres Primates. Il s'agit

essentiellement de réarrangements chromosomiques, de duplications géniques (essentiellement par duplication segmentaire), d'amplification et de dispersion d'éléments transposables et enfin de délétions, insertions et mutations ponctuelles (Gagneux and Varki, 2001).

3.2.1.1. Organisation globale du génome et cytogénétique

La synténie définit les correspondances d'organisation de portions du génome entre espèces différentes, ou encore la conservation de l'ordre des gènes entre deux espèces apparentées. On constate une conservation partielle de synténie entre les espèces de Rongeurs et de Primates, mais l'ordre des gènes est rarement bien conservé, à l'exception notoire des gènes du chromosome X. En revanche, la synténie est extrêmement bien conservée au sein des Primates, si l'on excepte les variations du nombre de modifications structurales de type inversions (péri- et paracentriques), fusions chromosomiques et translocations, toutes ces variations étant équilibrées chez les Primates. On note également des différences de positionnement des bandes C, essentiellement composées de séquences répétées, et une redistribution générale de l'hétérochromatine sur les différents chromosomes (données rassemblées dans (Gagneux and Varki, 2001)) (Jones et al., 1992; Strachan and Read, 1996). L'exception cytogénétique du gibbon (*Hylobates lar*) doit être soulignée : son caryotype est difficilement comparable à celui des autres Primates, autres espèces d'*Hylobates* comprises. Cette particularité, qui n'est en revanche pas présente au niveau moléculaire, pourrait être liée à l'écologie et la structure sociale des gibbons, qui sont monogames, territoriaux et arboricoles. Ce mode de vie favorise en effet la fixation de variants par dérive génétique. L'ensemble des remaniements chromosomiques de grande ampleur ayant affecté la lignée des Primates est représenté dans la Figure 3.2.1.1.a. (Jones et al., 1992).

Des analyses cytogénétiques moléculaires des chromosomes de primates ont démontré que les centromères se déplacent souvent, sans pour autant altérer l'ordre des gènes environnants, ni leurs structures (Fig 3.2.1.1.b). Les régions centromériques semblent en outre fréquemment associées aux duplications segmentaires. Ainsi, l'élévation locale des taux de recombinaison au niveau de l'hétérochromatine riche en duplication des centromères récemment inactivés constitue une clé de l'évolution des duplications segmentaires euchromatiques (Fig 3.2.1.1.c) (Jackson, 2003). Réciproquement, il est intéressant de noter que les duplications segmentaires elles-mêmes favorisent les remaniements chromosomiques

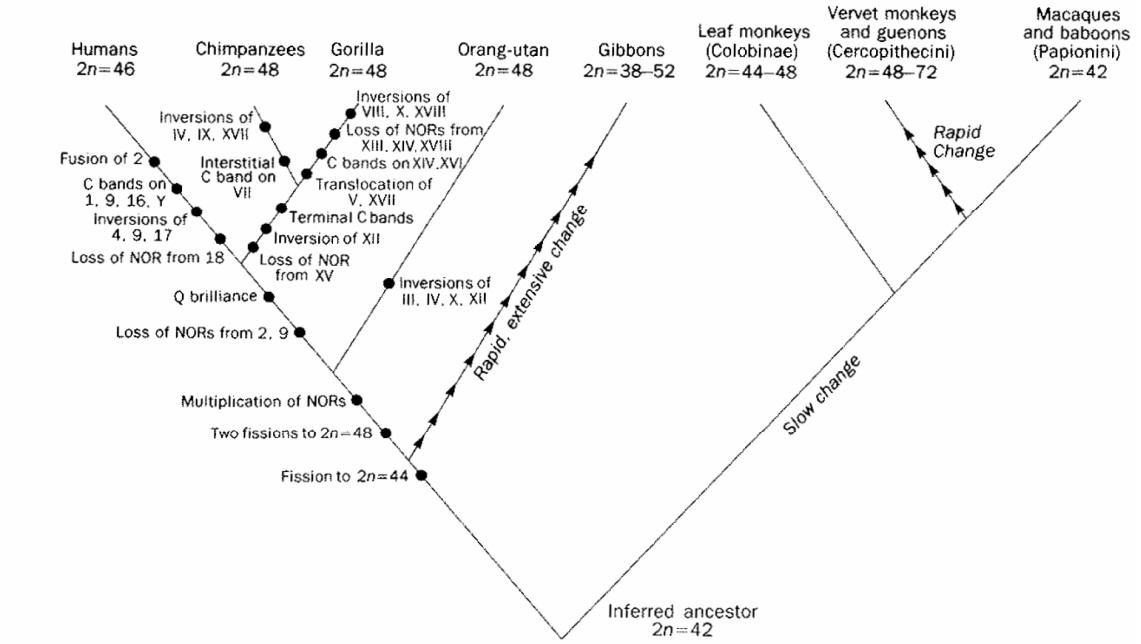


Figure 3.2.1.1.a **Remaniements chromosomiques chez les Primates**

Chaque point représente un évènement déduit de la comparaison des caryotypes des espèces actuelles. Au sein d'une branche, l'enchaînement des points est arbitraire. L'accent est mis sur les espèces de Catarhiniens, dont les chromosomes ont été étudiés de façon plus exhaustive. On identifie ainsi des évènements de fusion, de fission, d'inversion, de translocation, de modification des bandes C, et de perte d'organiseurs nucléolaires (NORs). Humans : *Homo*, Chimpanzee : *Pan*, Gorilla : *Gorilla*, Orang-utan : *Pongo*, Gibbon : *Hylobates.* (Jones et al., 1992).

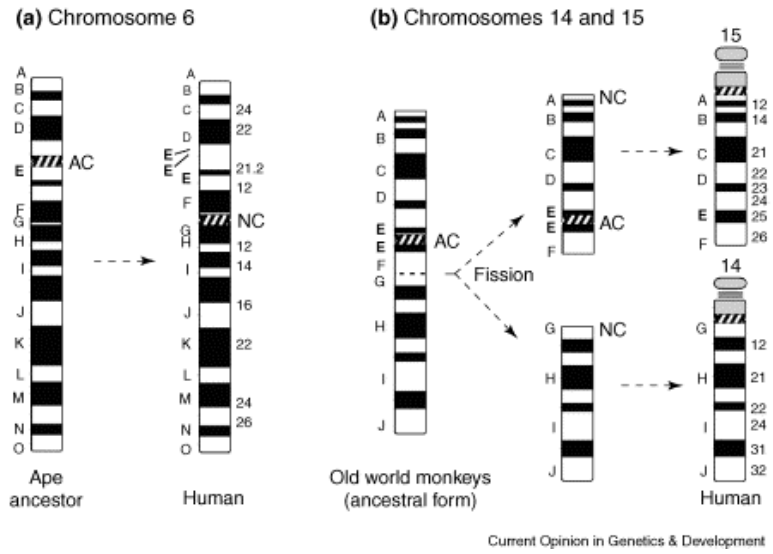


Figure 3.2.1.1.b **Mouvements des centromères au cours de l'évolution**

Les lettres à gauche de chaque représentation de chromosome indiquent la position de signaux obtenus par hybridation fluorescente in situ avec des sondes humaines. Dans les deux cas (chromosome 6 en **(a)**, chromosomes 14 et 15 en **(b)**), l'ordre des marqueurs est conservé entre l'ancêtre commun des hominidés et l'homme, seule la position des centromères varie. L'explication la plus parcimonieuse est celle d'un repositionnement des centromères. La multiplicité des signaux obtenus avec les sondes les plus proches des anciens centromères révèle l'existence de duplications segmentaires. AC : centromère ancestral, NC : néocentromère. Extrait de (Jackson, 2003).

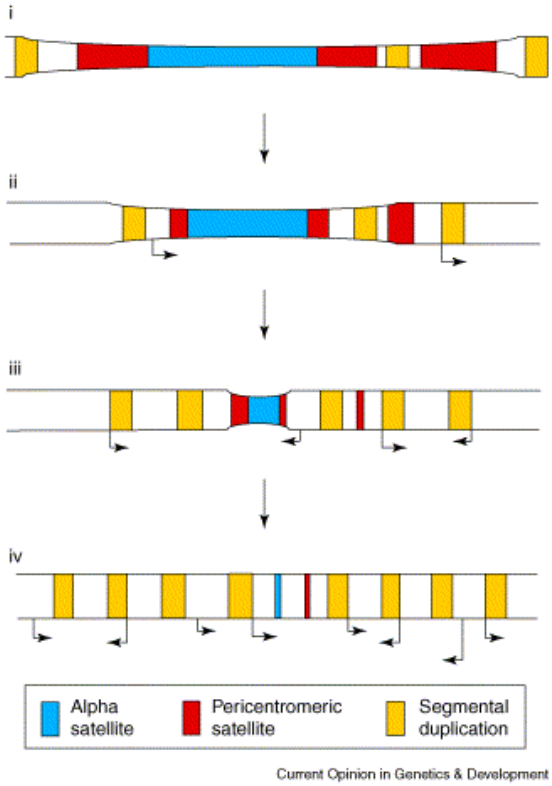


Figure 3.2.1.1.c **Modèle de dispersion des duplications segmentaires par déplacement du centromère**

L'environnement hétérochromatique du centromère ancien est représenté par la constriction entourant le domaine satellite central. Les flèches horizontales signent la transcription. **(i)** Duplications segmentaires et satellites au sein de l'hétérochromatine non-recombinante du centromère récemment inactivé. **(ii)** La perte stochastique (aléatoire) de satellites induit une relaxation de la chromatine. Les niveaux de transcription et de recombinaison augmentent. **(iii)** Les duplications segmentaires se répandent grâce à la recombinaison homologue non-allélique au niveau de l'ancien péricentromère. **(iv)** Le nombre de copies des duplications segmentaires est maintenu par l'équilibre entre les évènements stochastiques de perte et de gain (par recombinaison et conversion) de copies. Extrait de (Jackson, 2003).

(Samonte and Eichler, 2002). Ces néocentromères possèdent enfin un rôle fondamental dans l'apparition de pathologies humaines (Amor and Choo, 2002).

L'analyse des délétions et insertions de l'ordre de la kilobase (de 0,2 à 8kb), qui constituent donc de petits remaniements chromosomiques, a révélé que ces réarrangements étaient distribués de façon aléatoire, occupant aussi bien des régions géniques qu'intergéniques, et aucune caractéristique particulière n'a pu être détectée aux bornes de ces insertions/délétions. Ce type de réarrangements semble en outre avoir été fréquent au cours de l'évolution des Primates et avoir contribué de façon conséquente à la divergence génétique de ces espèces (Frazer et al., 2003).

L'ensemble des réarrangements chromosomiques constituerait ainsi une fraction tout à fait significative des différences de séquences entre l'Homme et le Chimpanzé (Frazer et al., 2003).

Enfin, bien que la taille des génomes et le nombre de gènes soient similaires dans les différentes espèces, un certain nombre de gènes spécifiques des Primates (totalement absents des génomes des autres mammifères), voire de l'homme, peuvent être identifiés (Strachan and Read, 1996).

3.2.1.2. Création de nouveaux gènes

Bien qu'en nombre restreint, quelques gènes fonctionnels ont été identifiés comme étant spécifiques des Primates ou même de la seule espèce humaine. Certains de ces gènes ont évolué et acquis une nouvelle fonction suite à une duplication simple, selon le modèle de la néo-fonction décrit précédemment (cf 2.1.4). Ils participent le plus souvent aux duplications segmentaires, très abondantes dans le génome humain, qui sont organisées en mosaïques de modules unitaires et localisées dans des régions hetéro- et euchromatiques spécifiques (Bailey et al., 2002b; Eichler, 2001; Stankiewicz et al., 2004) (cf 3.2.1.1). La fusion de gènes semble également avoir été opérante dans la lignée des Primates et a ainsi créé de la nouveauté génétique. D'autres gènes, enfin, ont été véritablement créés *de novo* par rétroposition, brassage d'exons, duplications et recrutement de séquences adjacentes au site d'insertion. Les duplications segmentaires sériées pourraient d'ailleurs avoir orchestré l'évolution du génome des Primates en engendrant ces nouveaux gènes par fusion et fission (Bailey et al., 2002b; Courseaux and Nahon, 2001; Eichler, 2001; Stankiewicz et al., 2004). Quelques uns des exemples de tels gènes finement analysés méritent d'être décrits et deux d'entre ces gènes chimériques feront l'objet d'un développement plus approfondi (cf 4.2).

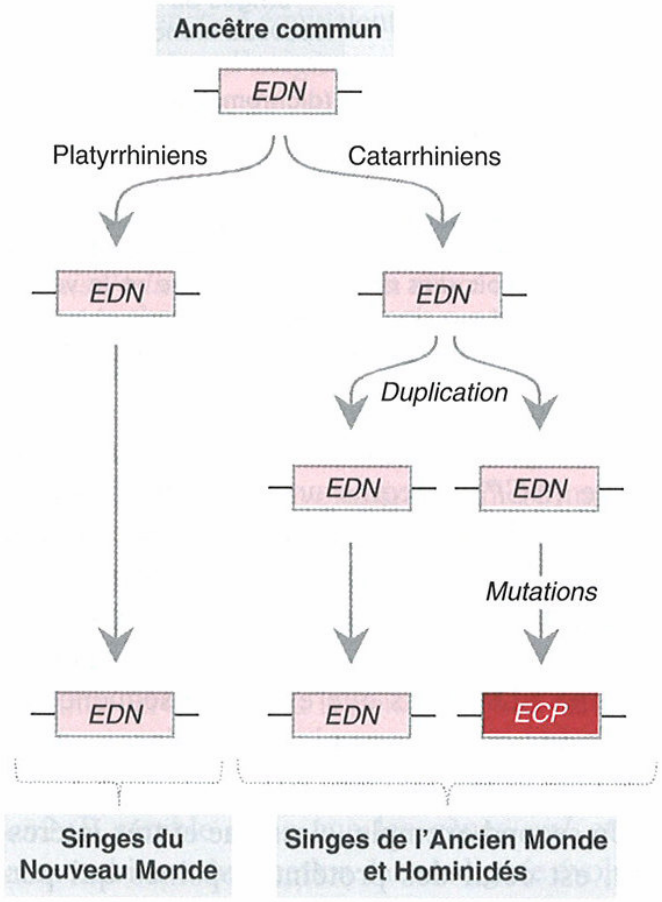


Figure 3.2.1.2.a **Origine du gène ECP par néo-fonction**

Dans la lignée des Catarhiniens, le gène *EDN* (*Eosinophilic-Derived Neurotoxin*) a subi une duplication. L'une des copies a conservé sa fonction d'origine, tandis que l'autre, appelée *ECP* (*Eosinophil Cationic Protein*) a subi des mutations lui permettant d'acquérir de nouvelles propriétés. Extrait de (Luchetta et al., 2005)

L'exemple de la **duplication** du gène de l'EDN (*Eosinophil-Derived Neurotoxin*) dans la lignée des Catarhiniens est tout à fait représentatif de l'acquisition d'une nouvelle fonction dans une lignée particulière (Fig 3.2.1.2.a). Dans la lignée des Platyrhiniens, le gène EDN, qui fait partie de la famille des RNAse A, est resté sous forme de copie unique et a conservé la structure et la fonction du gène d'origine (rôle antiviral). Dans la lignée des Catarhiniens, ce gène a subi une duplication. L'une des copies (*EDN*) a conservé la fonction d'origine, tandis que l'autre (*ECP* pour *Eosinophil Cationic Protein*) a acquis une nouvelle fonction, et code pour une protéine antibactérienne. Cette activité antibactérienne n'est pas liée à l'activité RNAse, mais est la conséquence de l'accumulation d'arginines en remplacement d'autres acides aminés. Cette accumulation relève, en outre, de la sélection darwinienne positive (Zhang et al., 1998). Plus généralement, les duplications segmentaires ont été largement étudiées chez les Primates et sont particulièrement abondantes chez l'Homme où elles représentent près de 14% des séquences génomiques (contre 6,6% chez la Souris et moins de 5% chez le Chimpanzé) (Bailey and Eichler, 2006). Elles n'auraient pas seulement contribué à la formation de nouveaux gènes, mais pourraient également avoir favorisé les variations génétiques et phénotypiques propres à l'homme ou aux primates, sans que le nombre de copies en soit le seul déterminant. De même, les duplications segmentaires pourraient être l'un des déterminants des fréquents réarrangements du génome du Gibbon (*Hylobates lar*) (Bailey and Eichler, 2006). Ces évènements et les variations génétiques qu'ils impliquent sont associés à la susceptibilité à un certain nombre de pathologies (cf 3.2.3).

La **fusion** de gènes est décrite chez les procaryotes et quelques eucaryotes, mais impose en général des contraintes structurales fortes. Un exemple unique de fusion de gènes, affranchi de ces contraintes, est décrit chez l'Homme. Dans le cas des gènes *UEV* et *Kua*, il s'agit, en effet, d'une fusion transcriptionnelle. Le gène *UEV* (*Ubiquitin-conjugating Enzyme Variant*) produit un variant inactif des enzymes E2 responsables de la régulation de l'élongation des chaînes d'ubiquitine. Ce gène est présent chez tous les Eucaryotes et est dupliqué chez l'Homme. La copie d'origine *UEV2* a conservé sa fonction d'origine, alors que la copie dupliquée (sur un autre chromosome) *UEV1* a fusionné avec le gène *Kua* adjacent, jusqu'alors inconnu. La fusion est de type transcriptionnelle et réalisée par un mécanisme d'épissage alternatif, lui-même probablement réalisé sous la contrainte de modifications génomiques. Trois types d'ARN messagers sont alors produits : les deux ARN messagers d'origine (produits respectivement par *UEV1* et *Kua*) et un troisième ARN messager Kua-UEV1 permettant la synthèse d'une protéine chimérique (cf Fig 3.2.1.2.b). Cette protéine

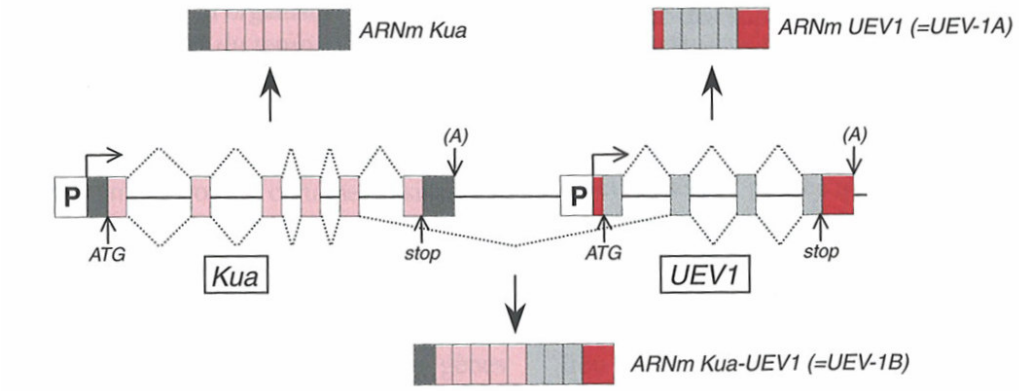


Figure 3.2.1.2.b **Origine du gène *Kua-UEV1* par fusion transcriptionnelle**

Chez l'Homme, la duplication du gène *UEV* et la fusion transcriptionnelle de l'une de ses copies avec le gène *Kua* conduit à la formation d'une protéine chimèrique qui n'a pas encore acquis de fonction. Extrait de (Luchetta et al., 2005).

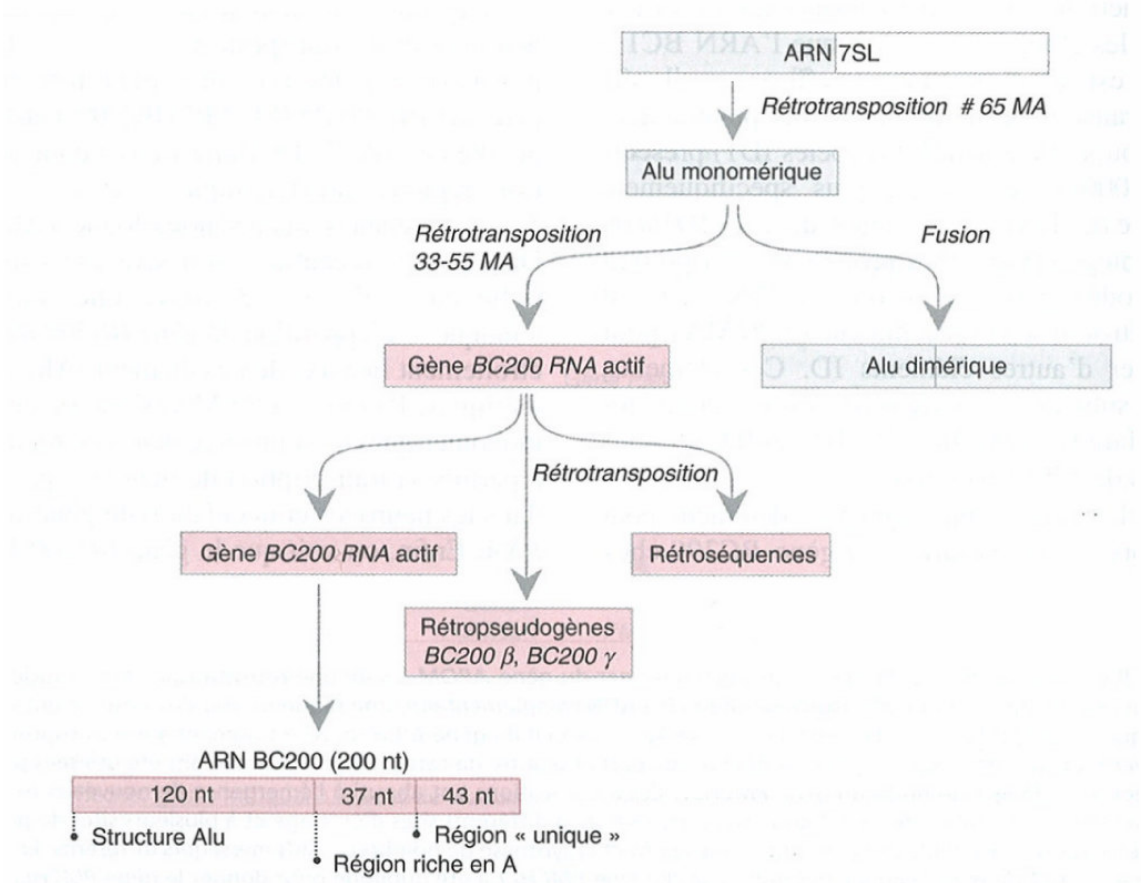


Figure 3.2.1.2.c **Origine et évolution du gène *BC200 RNA* chez les Primates**

Ce gène est apparu dans la lignée des Primates il y a 33 à 55 millions d'années par rétrotransposition d'une séquence Alu monomérique. Ces séquences Alu sont spécifiques aux Primates et dérivent de la rétrotransposition d'un ARN 7SL. Extrait de (Luchetta et al., 2005).

possède, pour le moment, les mêmes propriétés que la protéine UEV mais le domaine Kua permet son adressage aux endomembranes. Cette jeune protéine chimérique spécifique de l'espèce humaine n'a donc pas encore acquis de nouvelle fonction, mais il s'agit sans doute d'une période de transition (Long, 2000; Thomson et al., 2000).

Les exemples parmi les plus pertinents de gènes chimériques issus d'évènements de **rétroposition** et spécifiques des Primates sont les gènes *PMCHL1* et *PMCHL2* qui font l'objet de mon travail de thèse et feront, à ce titre, l'objet d'un développement approfondi ultérieur. Mais le cas des gènes *BC1 RNA* et *BC200 RNA* (pour *Brain Cytoplasmic RNA*) mérite également que l'on s'y attarde. Il constitue un exemple de rétroposition que l'on pourrait qualifier de convergente. En effet, si *BC200 RNA* est spécifique des Haplorhiniens, *BC1 RNA* est, lui, propre aux Rongeurs et tous deux ont des fonctions proches, d'ARN non codants pour une protéine, produits dans les noyaux des cellules nerveuses et dirigés ensuite dans le cytoplasme dendritique où ils vont s'associer à des protéines pour former des particules ribonucléoprotéiques. *BC1 RNA* et *BC200 RNA* dérivent tous deux, mais de manière indépendante, de la rétrotransposition de petits ARN et pourraient être liés à l'origine de l'apparition de certains éléments répétitifs spécifiques (séquences Alu dans le cas des Primates). Dans le génome des Primates, et plus particulièrement de l'Homme, on trouve en effet près de 50 000 séquences Alu (SINEs). Les séquences Alu monomériques sont d'abord apparues par rétrotransposition d'une partie de l'ARN 7SL, qui est impliqué dans la synthèse protéique au niveau du réticulum endoplasmique. Les séquences Alu « actuelles » ont une séquence dimérique constituée de deux régions similaires qui proviendraient de deux monomères. Le gène BC200 RNA est composé de trois régions caractéristiques : une séquence homologue à Alu lm (left monomer) en 5', une séquence riche en adénosine centrale et une séquence « unique » en 3'. Il y a 33 à 55 millions d'années, un Alu monomérique s'est intégré dans une région qui a permis sa transcription de manière spécifique dans les neurones et la création du gène *BC200 RNA*. Enfin, ce gène possède lui-même une forte capacité à la rétrotransposition, comme en témoigne l'existence de deux pseudogènes *BC200 RNAβ* et *BC200 RNAγ*, apparus après la divergence Homme/Chimpanzé (Fig 3.2.1.2.c) (Kuryshev et al., 2001; Martignetti and Brosius, 1993; Skryabin et al., 1998). De manière plus générale, il semblerait qu'un pic de rétroposition chez les Primates (il y a 5 à 10 millions d'années) ait favorisé l'émergence de jeunes gènes spécifiques de la lignée humaine. Ces gènes sont préférentiellement exprimés dans les testicules, à la différence des gènes dont ils dérivent. Ces

rétrogènes acquièrent ensuite majoritairement des fonctions spermatogénétiques (Marques et al., 2005).

3.2.1.3. ADN non codant, séquences répétées et éléments transposables des Primates

Un certain nombre des composantes du génome des Primates se distingue nettement de celles des autres mammifères, et en particulier des Rongeurs. En effet, si les séquences codant pour des protéines sont relativement bien conservées, les introns sont eux presque incomparables. Si l'ADN minisatellite télomérique est bien conservé en raison d'une pression de sélection assurant une reconnaissance continue par la télomérase, le reste de l'ADN hautement répétitif (ADN satellite, minisatellite et microsatellite) est lui très rapidement divergent. De même, les éléments dispersés hautement répétitifs sont en général mal conservés. On identifie ainsi des séquences répétées transposables spécifiques des Primates, telles que les séquences Alu. Cependant, ces pseudoéléments matures dérivés de l'ARN 7SL possèdent des équivalents convergents dans le génome des Rongeurs : la répétition B1 déjà mentionnée précédemment (cf 3.2.1.2) (Strachan and Read, 1996). On note, dans les génomes de Primates, une richesse particulière en éléments transposables anciens (et une relative pauvreté en éléments récents), qui indique que, depuis la radiation des Mammifères, l'activité globale des transposons a décliné progressivement dans la lignée évolutive conduisant aux Primates. En outre, ces séquences, pourtant fréquemment non fonctionnelles, sont donc éliminées à une vitesse très lente dans la lignée des Primates. Le déclin des éléments transposables est très marqué dans les 35 à 50 derniers millions d'années, à l'exception d'un pic d'activité des éléments Alu il y a environ 40 millions d'années. Un tel déclin n'est pas observé chez les Rongeurs (Strachan and Read, 1996).

Ces éléments transposables affectent en outre les propriétés d'autres éléments du génome, et en particulier des gènes. L'importance et l'impact des éléments transposables dans l'évolution des gènes des Primates et de leurs produits peuvent être déduits de l'analyse systématique du génome humain et de sa comparaison avec les séquences génomiques des Rongeurs. Il semblerait que pas moins de 4% des gènes humains contiennent dans leur région codante des séquences présentant des similitudes avec des éléments transposables (Nekrutenko and Li, 2001). Près de 90% de ces éléments correspondraient à des insertions à l'intérieur d'introns qui ont ensuite été recrutés en tant qu'exons. Par ailleurs et à titre d'exemple, l'insertion, dans le génome de l'ancêtre des Primates actuels, d'un élément

transposable en amont de l'unité de transcription du gène *CYP19*, codant pour l'aromatase P450, enzyme clé de la synthèse des oestrogènes, a permis, chez les Primates, la création d'un nouveau promoteur spécifique du placenta (Luchetta et al., 2005).

Au sein des Primates, l'ADN non codant peut présenter un degré élevé de similitudes des séquences. L'ADN hautement répété présente, lui, une évolution extrêmement rapide en dépit de cette très grande identité nucléotidique globale. Des séquences mini- et microsatellites peuvent ainsi différer de façon conséquente entre l'homme et d'autres Primates. Les séquences télomériques sont conservées, mais des séquences minisatellites hypervariables montrent une évolution transitoire chez les Primates (Gray and Jeffreys, 1991). Les microsatellites présentent également des différences au niveau des positions orthologues chez l'homme et les autres Primates (Rubinsztein et al., 1995a; Rubinsztein et al., 1995b). De même, l'ADN dispersé hautement répétitif varie au gré des espèces. Bien que la sous-famille de répétitions Alu soit retrouvée chez tous les Primates, plusieurs sous-familles ont été individualisées (Jurka, 1991) et semblent s'être développées à différentes périodes de l'histoire des Primates. La plus ancienne sous-famille, la répétition AluJ, a été datée à environ 55 millions d'années. Cette famille, comme d'autres, est caractérisée par une divergence considérable entre ses membres, mais avec toujours une relativement étroite ressemblance avec la séquence consensus de l'ARN 7SL. Enfin, un petit nombre de séquences Alu appartient à des familles dont l'origine au cours de l'évolution est extrêmement récente et contient des membres activement transposés. Cela inclut les familles Sb1 et Sb2, cette dernière semblant majoritairement spécifique à l'homme et datant de moins de 2 millions d'années (Strachan and Read, 1996; Zietkiewicz et al., 1994). Ces répétitions Alu sont ensuite des points chauds de recombinaison, qui peuvent engendrer des délétions de grande taille, des insertions, des duplications. Ainsi, ces éléments transposables spécifiques des Primates affectent de façon conséquente l'évolution des génomes et l'apparition de pathologies à substrat génétique (Strachan and Read, 1996).

3.2.1.4. Mutations ponctuelles, vitesses évolutives variables et polymorphisme

Le calcul des pourcentages d'identité entre les séquences de Primates est à l'origine des estimations de dates de divergence des espèces, par l'utilisation des horloges moléculaires. Cependant les relations de parenté ainsi établies ne sont parfois pas corroborées par les données fossiles. Il semble donc essentiel pour dater les divergences successives de faire appel aux données paléontologiques en sus des données moléculaires, d'autant que

lesdites horloges moléculaires battent, comme nous l'avons détaillé précédemment, à des rythmes différents selon les lignées (Senut, 1999). La comparaison des taux de substitutions nucléotidiques de l'homme, des singes de l'ancien monde, des singes du nouveau monde et des rongeurs indique en effet un ralentissement possible chez l'homme, le chimpanzé et le gorille (Hacia, 2001).

Beaucoup se sont étonnés des seuls 1 à 2% de différences entre les génomes humain et de chimpanzés, 1 à 2% qui font pourtant toute la différence (Barriel, 2004) ! Cette divergence est principalement portée par les régions introniques, intergéniques et 5'UTR (Ebersberger et al., 2002; Hellmann et al., 2003), mais certaines mutations affectent plus directement le phénotype. Les mécanismes moléculaires spécifiques de l'évolution des Primates méritent donc d'être détaillés. Le cas du gène *FOXP2* (*Forkhead-box P2 transcription factor*) est particulièrement marquant. Ce gène serait impliqué dans la possibilité physique qu'a l'Homme, à la différence d'autres Primates, de développer un langage articulé. En effet, une mutation ponctuelle de ce gène co-ségrège avec un grave déficit du langage dans une famille humaine où la moitié des membres est affectée. Il semble qu'il soit nécessaire d'avoir deux copies fonctionnelles du gène *FOXP2* pour développer un langage normal. Des orthologues de ce gène sont connus chez divers Primates ainsi que chez la Souris. Seuls trois acides aminés (sur 715) diffèrent entre la protéine de l'homme et celle de la souris. Deux de ces différences sont liées à des évènements ayant eu lieu après la divergence de l'Homme et du Chimpanzé. Cela suggère une accélération de l'évolution de cette protéine dans la lignée des hominidés. L'absence de polymorphisme en ces loci au sein de l'espèce humaine suggère qu'il s'agit bien de sélection positive, et non simplement de relaxation des contraintes. Ainsi, deux petites mutations suffiraient à conférer une propriété unique à une espèce de primates, l'Homme en l'occurrence (Enard et al., 2002b; Luchetta et al., 2005). D'autres exemples de gènes modifiés par mutation ou petites insertions/délétions et conférant probablement des différences phénotypiques entre l'Homme et les autres Primates sont maintenant disponibles (Varki and Altheide, 2005). Les exemples de l'*ASPM* (*abnormal spindle-like microcephaly* associate) et de la *microcéphaline* (Dorus et al., 2004; Evans et al., 2004), celui de la chaîne lourde de myosine 16 *MYH16* (Stedman et al., 2004), ainsi que celui de la *CMP-Neu5Ac hydroxylase* (Chou et al., 2002; Chou et al., 1998) sont particulièrement documentés et la fonctionnalité de ces gènes (respectivement impliqués dans l'augmentation de la taille du cerveau humain, des modifications de son anatomie et des changements de résistance aux pathogènes) est avérée.

Il apparaît également que les ratios Ka/Ks sont fréquemment supérieurs dans les séquences de primates, et encore d'avantage dans les séquences humaines, relativement à celles des Rongeurs, et cette différence est nettement accrue lorsqu'il s'agit de gènes spécifiques des systèmes nerveux (Dorus et al., 2004) ou olfactif (Clark et al., 2003), par opposition aux gènes dit « de ménage ».

Enfin, la moindre variabilité génétique de l'espèce humaine, comparée à celle des chimpanzés par exemple, cette variabilité étant en outre supérieure à l'intérieur d'une population humaine qu'entre populations différentes, suggère que l'espèce humaine est relativement uniforme. L'hypothèse du « bottleneck » (goulot de bouteille), qui consiste en une réduction drastique du nombre des individus à un moment de l'histoire évolutive, justifie cette plus grande proximité génétique entre individus de l'espèce par perte de la variabilité préalablement existante (Jones et al., 1992; Strachan and Read, 1996). Cette réduction drastique du nombre des êtres humains serait survenue il y a 70 ou 80 000 ans, à l'époque de Néandertal, quelques dizaines de milliers d'années après l'apparition des premiers *Homo sapiens.* La taille critique de population alors atteinte est estimée à 5 à 10 000 individus (Harpending et al., 1998; Jorde et al., 1997).

3.2.2. Modifications transcriptomiques participant au renforcement ou au développement de propriétés des Primates

Au-delà des modifications génomiques, les spécificités et différences phénotypiques des Primates tiennent à des modifications d'expression de gènes communs aux différentes espèces, ces modifications étant elles-mêmes consécutives aux changements génomiques. Il peut s'agir de modifications du processus d'épissage et de maturation de l'ARN messager ou de l'activité promotrice (déterminant le profil tissulaire et temporel et les niveaux d'expression). Néanmoins, compte tenu de l'étendue et de la variabilité qualitative et quantitative du transcriptome, la comparaison entre espèces distinctes s'avère beaucoup plus fastidieuse que celle des génomes, en particulier pour ce qui est des données qualitatives.

3.2.2.1. Epissages alternatifs

Il convient en premier lieu de préciser les modalités de l'épissage alternatif, avant d'examiner les exemples de gènes pour lesquels un épissage alternatif spécifique des Primates est décrit et confère éventuellement des propriétés inédites. L'épissage alternatif est particulièrement fréquent chez les Métazoaires et permet d'obtenir plusieurs ARN messagers

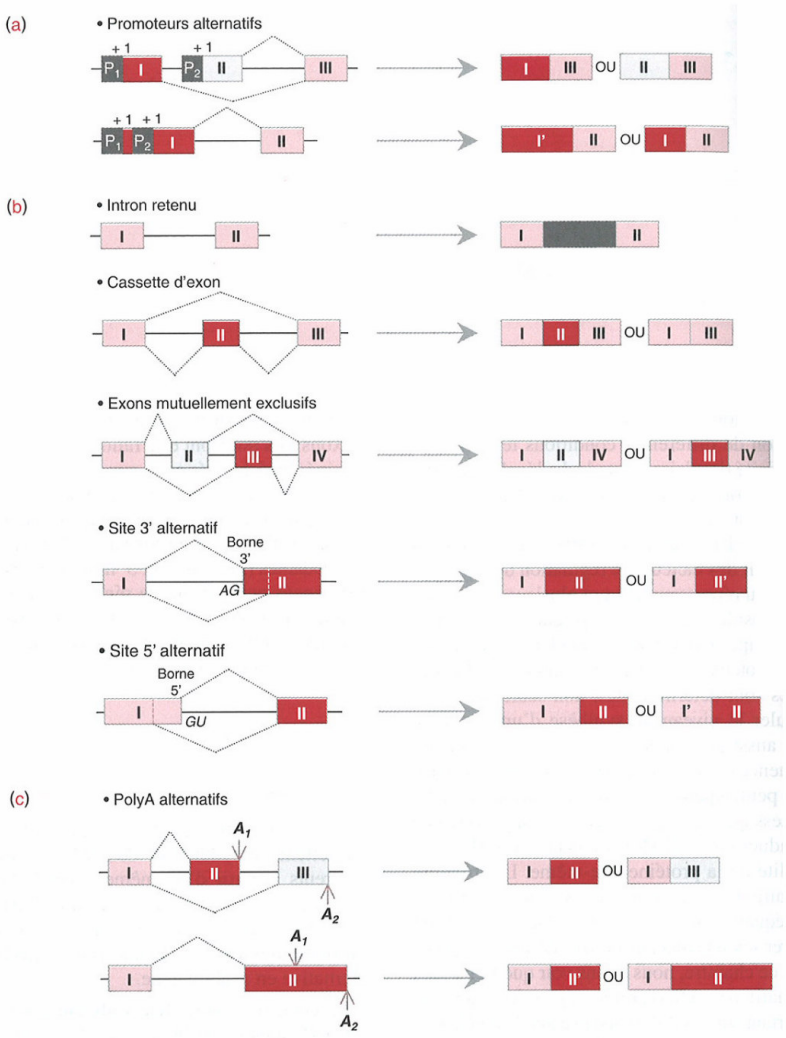


Figure 3.2.2.1 **Différentes formes alternatives**

(a) Dans un système de promoteurs alternatifs, le choix entre deux promoteurs P1 et P2 va définir un site d'initiation de la transcription différent. **(b)** Plusieurs cas sont à envisager pour un épissage alternatif. Un intron peut être retenu, un exon éliminé (exon cassette), des exons peuvent être mutuellement exclusifs. Enfin, les bornes d'épissage 3' et 5' des introns peuvent être mlultiples. **(c)** Dans un processus de polyadénylation alternative, il peut y avoir un chois alternatif du dernier exon, ou bien un même exon peut posséder deux sites de polyadénylation alternatifs. Extrait de (Luchetta et al., 2005).

différents à partir d'un seul type d'ARN pré-messager. Le choix des sites d'épissages, qui dépend de l'environnement cellulaire conduira à différentes associations d'exons.

L'épissage alternatif dépend donc du contexte cellulaire (spécificité tissulaire, stade de développement, réponse à un stimulus externe) par l'interaction de différents éléments cellulaires avec des régions activatrices et inhibitrices d'exons et d'introns. Les modifications des séquences génomiques de ces régions au cours de l'évolution, au même titre que les modifications de contexte cellulaire, peuvent donc favoriser de substantiels changements d'expression des gènes.

Lors du processus d'épissage alternatif, les exons peuvent être rallongés, raccourcis, éliminés ou échangés entre eux. Ces choix des sites d'épissage auront pour conséquence directe la création d'ARN messagers différents, soit au niveau de leur région codante (ORF), soit au niveau de leurs régions non codantes pour une protéine (régions 5' et 3' non traduites). Différents profils « alternatifs » permettent de réaliser cet épissage alternatif (Thanaraj and Clark, 2001). L'utilisation de promoteurs alternatifs peut conditionner le site d'initiation de la transcription et ainsi modifier le premier exon et potentiellement la partie N-terminale de l'éventuelle protéine (Figure 3.2.2.1 **(a)**). Le choix de différentes bornes d'introns sur un même ARN prémessager pour former des ARN messagers différents, permettant ainsi éventuellement la synthèse de plusieurs isoformes protéiques, constitue le second profil d'épissage alternatif. Il peut en résulter la rétention d'un intron, l'élimination d'un exon (dit exon cassette) ou l'exclusion mutuelle de deux exons. Dans d'autres cas, les exons peuvent être tronqués ou rallongés (Modrek et al., 2001) (Figure 3.2.2.1 **(b)**). Enfin, il existe des mécanismes de polyadénylation alternative susceptibles de modifier le dernier exon (Figure 3.2.2.1 **(c)**). Les différentes possibilités précédemment énumérées peuvent être associées pour donner des systèmes complexes. L'épissage alternatif a ainsi joué un rôle essentiel dans l'expansion du protéome et la diversification des Métazoaires. 35 à 50% des gènes humains feraient l'objet d'épissages alternatifs (Luchetta et al., 2005; Modrek et al., 2001; Roberts and Smith, 2002).

La grande majorité des messagers alternativement épissés détectés par analyse du transcriptome humain semble être spécifique. Ils proviennent aussi bien de gènes humains connus, que de nouveaux gènes spécifiques et participent à des fonctions biologiques fondamentales (Modrek et al., 2001).

L'exemple de la Talanin est probablement le plus pertinent pour démontrer l'apparition d'un épissage alternatif au cours de l'évolution des Primates et son implication

dans une pathologie spécifique. L'identification récente d'un locus de susceptibilité à la néphrolithiase à l'acide urique (UAN, Uric acid Nephrolithiasis) sur le chromosome 10 humain et la démonstration que le gène *ZNF365* de ce locus produisait différents transcrits par épissage alternatif a retenu l'attention de l'équipe de Mario Piratsu en Italie. Il a ensuite été démontré par analyse des mutations que l'un de ces transcrits (codant pour une isoforme protéique appelé Talanin) est associé à la néphrolithiase UAN. L'un des transcrits ZNF365 (la forme A) est retrouvée par homologie chez la Souris et y est spécifiquement exprimé dans le cerveau. En revanche, aucun homologue du transcrit ZNF365D n'a pu être identifié chez la Souris, et la région génomique correspondante n'est pas organisée en structure génique canonique chez les Rongeurs. Cela suggère que le transcrit ZNF365D est apparu après la divergence de la lignée des Hominoïdes de celle des Rongeurs. De façon très intéressante, l'acitivité uricase, présente chez les Rongeurs et la plupart des Mammifères, a été perdue dans la lignée des Hominoïdes à l'époque du Miocène. L'analyse du locus de la Talanin chez les Primates non-humains démontre que de nombreux codons stop prématurés empêchent la production d'une protéine fonctionnelle, malgré une structure exon/intron canonique. Chez l'homme, une protéine fonctionnelle est produite, suggérant que le transcrit alternatif ZNF365D a émergé au cours de l'évolution des Primates à partir d'une séquence génomique non-codante, restaurant ainsi une fonction uricase (Gianfrancesco et al., 2004).

3.2.2.2. Profils d'expression différents

L'avènement des puces à ADN de grande densité a permis l'étude de l'expression d'un grand nombre, voire de la totalité, des gènes codant pour des protéines des génomes de différentes espèces de primates et dans différents tissus. Cette technique permet de détecter aussi bien des changements de « site » (tissus, types cellulaires…) d'expression des gènes que de niveaux d'expression dans un même tissu (Varki and Altheide, 2005). La comparaison du niveau d'expression d'un grand nombre de gènes dans le cerveau, le foie et le sang de trois hommes, trois chimpanzés et un orang-outan indique d'abord que les variations inter individuelles intraspécifiques sont considérables relativement aux différences observées entre homme et chimpanzé. Néanmoins, l'utilisation de l'orang-outan comme groupe externe permet de positionner l'homme et le chimpanzé dans deux groupes mutuellement exclusifs. Il apparaît alors une nette accélération des variations d'expression chez l'Homme (comparativement au Chimpanzé), et de façon plus marquée dans le cerveau que dans le foie. Cela suggère que les profils d'expression génique ont changé davantage dans le cerveau que

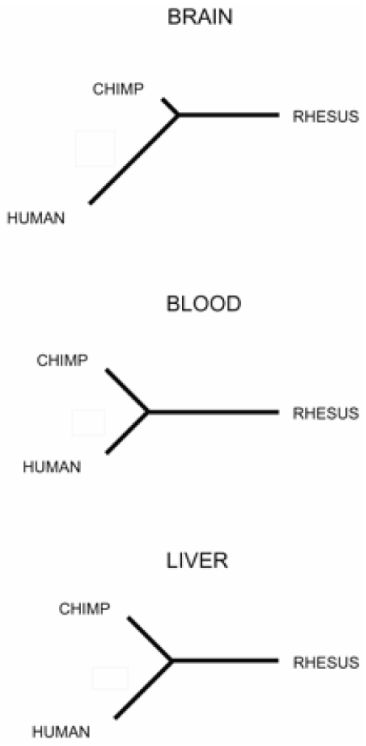


Figure 3.2.2.2 **Variation d'expression génique des Primates**

Arbres de distance représentant les valeurs relatives de variations d'expression entre trois espèces de Primates (Homme, Chimpanzé, Macaque) et dans trois tissus (cerveau, sang et foie). Les données sont déduites d'analyse de puces à ADN. Tiré de (Enard et al., 2002a).

dans d'autres tissus au cours de l'évolution récente de l'homme, ce qui est attesté par des expériences complémentaires utilisant le macaque comme groupe externe (Enard et al., 2002a; Khaitovich et al., 2004a) (Fig 3.2.2.2). Cette plus grande variabilité résulte en outre très fréquemment d'une augmentation des niveaux d'expression des gènes dans le cerveau humain, alors que les différences des niveaux d'expression dans le foie et le cœur de l'homme et du chimpanzé résultent autant d'augmentation que de diminution du niveau d'expression (Caceres et al., 2003; Gu and Gu, 2003; Khaitovich et al., 2004a). Il faut noter que si l'ensemble de ces études a clairement mis en évidence des variations de niveau d'expression des gènes, les patrons d'expression des gènes sont eux très conservés dans le cerveau de l'Homme et du chimpanzé (Khaitovich et al., 2004a; Varki and Altheide, 2005). Ceci rejoint la notion selon laquelle l'apparition des capacités cognitives spécifiques de l'homme s'est faite par recrutement de structures cérébrales existantes et de leurs fonctions moléculaires sous-jacentes (Khaitovich et al., 2004a).

Les gènes pour lesquels on observe des modifications d'expression cérébrale dans l'une ou l'autre des espèces appartiennent principalement aux classes de la régulation transcriptionnelle, de la transduction du signal, du métabolisme des lipides, de l'adhésion cellulaire (Caceres et al., 2003; Sikela, 2006) et de la transmission synaptique (Caceres et al., 2003). Ce type de données, d'ailleurs controversées, doit cependant être ensuite validé par d'autres outils de biologie moléculaire, pour que la question de l'importance des gènes identifiés pour l'établissement des phénotypes spécifiques puisse ensuite être examinée.

Certains gènes identifiés par puce à ADN ont été précisément caractérisés du point de vue de leur expression, ce qui permet d'en affirmer le rôle dans certains phénotypes. Ainsi, l'expression de la transthyretin est amoindrie dans le sang et le cerveau de l'homme (comparativement aux autres Primates) et cette différence pourrait être associée à un métabolisme des hormones thyroïdienne modifié chez l'Homme (Gagneux et al., 2001). La régulation positive spécifique de l'alpha 2-6 sialytransférase dans les cellules épithéliales humaines pourrait expliquer la résistance relative des chimpanzés au virus influenza A humain (Gagneux et al., 2003). Enfin, l'apparition, chez les Primates, d'un promoteur additionnel alternatif pour le gène *CYP 19* codant pour l'aromatase modifie le profil d'expression de ce gène. Cette enzyme clé de la synthèse des oestrogènes est exprimée dans les gonades, le cerveau et le tissu adipeux de la plupart des Mammifères. L'apparition du promoteur alternatif chez les Primates favorise l'expression additionnelle du gène *CYP19*

dans le placenta de ces espèces, contrôlant ainsi le niveau d'oestrogènes lors de la grossesse (Luchetta et al., 2005).

L'ensemble de ces modifications transcriptomiques affecte naturellement le protéome, notamment en terme de diversité.

3.2.3. Phénotypes et pathologies spécifiques des Primates

Les modifications phénotypiques, et en particulier cognitives, qui définissent les Primates ou l'Homme tiennent donc à des modifications génomiques, transcriptomiques et en conséquence protéomiques, mais aussi à l'interaction de ces sphères avec l'environnement. Les pathologies affectant spécifiquement les primates, ou l'homme seulement, constituent des indicateurs indirects de ces spécificités et de leurs substrats moléculaires.

Les bases génétiques des différences de susceptibilité de l'homme et du chimpanzé à différentes pathologies, telles que la malaria, les cancers de la peau, la maladie d'Alzheimer, le SIDA… sont encore relativement méconnues. L'exploration du système du CMH (Complexe Majeur d'Histocompatibilité) du Chimpanzé pourrait apporter quelques éléments de réponse (Gagneux and Varki, 2001). L'observation d'une susceptibilité réduite à l'infection par le virus HIV et à l'évolution vers le SIDA lorsque, par duplication segmentaire, le nombre de copies du récepteur CCL3L1 est accru constitue une piste pour la compréhension de la différence entre Homme et Chimpanzé (Bailey and Eichler, 2006; Gonzalez et al., 2005). L'exemple du gène *FOXP2* (Forkhead box P2) découle du même type d'approche : il a initialement été identifié dans le cadre de troubles du langage chez l'Homme (Lai et al., 2001), puis sa comparaison avec ses orthologues chez d'autres Primates (Chimpanzé, Gorille, Orang-Outan, Macaque) et chez la Souris a démontré que ce gène avait subi une sélection positive chez l'Homme, ce qui le rend significativement différent entre l'Homme et le Chimpanzé. L'ensemble de ce modèle est ainsi susceptible d'expliquer les différences d'articulation du langage entre ces espèces (Enard et al., 2002b). Enfin, des mutations dans un grand nombre d'autres membres de la famille forkhead sont responsables de pathologies spécifiquement humaines (glaucome congénital, agénèse thyroïdienne, syndromes de lymphe-demadistichiase et de blepharophimosis/ptosis/epicanthus inversus) (Lai et al., 2001).

De manière plus générale, les duplications segmentaires ont été largement impliquées dans l'émergence de pathologies spécifiques de l'Homme ou des Primates. Ainsi, le locus

SMA (Spinal Muscular Atrophy) sur le chromosome 5q13 humain (Melki et al., 1994) correspond à l'étendue d'une duplication segmentaire majeure ou siègent les gènes *SMN1* et *SMN2* chez l'Homme (Schmutz et al., 2004). Les recombinaisons favorisées par cette duplication segmentaire sont susceptibles d'affecter le gène *SMN1* et ainsi de provoquer l'atrophie spinale musculaire et la susceptibilité à cette pathologie, bien que les cas les plus connus relèvent d'une mutation ponctuelle du gène *SMN1* (Monani et al., 1999).

Figure 4.1.1 **Les Primates étudiés**

(a) *Tarsius syrichta*, **(b)** *Saguinus oedipus*, **(c)** *Cebus apella*, **(d)** *Chlorocebus aethiops*, **(e)** *Macacca mulatta*, **(f)** *Hylobates lar*, **(g)** *Pan troglodytes*, **(h)** *Pan paniscus*, **(i)** *Homo sapiens*.

4. Nos modèles d'étude et objectifs de la thèse

4.1. Modèles animaux

L'étude de modèles de familles de gènes spécifiques humains ou de primates permet la compréhension de l'évolution de ces espèces et de leurs génomes. Cette étude nécessite de pouvoir comparer les données phénotypiques (physiologiques et pathologiques) et moléculaires d'un certain nombre d'espèces, afin de déterminer si l'un ou l'autre des mécanismes d'évolution des génomes intervient et avec quelles conséquences sur l'expression génique et le phénotype.

4.1.1. Présentation des espèces étudiées : relations phylogénétiques, caractéristiques phénotypiques, biotopes et types de données disponibles

Les principales caractéristiques des espèces de Primates étudiées sont reprises dans la Table 4.1.1.

4.1.1.1. Tarsius syrichta

Le positionnement phylogénétique des Tarsiers, dont *Tarsius syrichta* (cf Figure 4.1.1 **(a)**), fait, comme nous l'avons déjà vu, l'objet de controverses. Il est néanmoins désormais consensuel de l'associer aux Haplorhiniens (cf Figure 3.1.1.2.b). Leur date de divergence d'avec l'ancêtre des autres Haplorhiniens est estimée à 60 millions d'années environ.

Les Tarsiers ne mesurent que 10 à 15 cm. Leurs grands yeux leur permettent une bonne vision de nuit adaptée à un mode de vie nocturne, et leurs grands pavillons leur assurent une bonne audition. Ce sont d'excellents sauteurs grâce à l'allongement de 2 os du pied. Cette aptitude, leur bonne vision de nuit et leur vivacité leur permettent de compléter leur alimentation végétale par des insectes ou de petits vertébrés tels que les lézards. Ils vivent exclusivement en Malaisie et aux Philippines et sont menacés d'extinction. De ce fait, seul leur ADN est véritablement disponible.

Nom commun	Nom latin	classification	date de divergence d'avec la lignée humaine	taille / poids	longévité	répartition géographique	habitat	alimentation	nocturne / diurne	matériel génétique disponible
Tarsier des Philippines	*Tarsius syrichta*	Strepsirhinien	65 MA	10 à 15 cm, 80 à 160 g	12 ans	Malaisie et Philippines	forêts côtières	végétaux, insectes, lézards	nocturne	ADNg
Tamarin pincné	*Saguinus Oedipus*	Plathyrhinien	35 MA	25 à 30 cm, 350 à 550 g	15 ans	nord de la Colombie	forêt dense	fruits, légumes, insectes et œufs	diurne	ADNg
Sapajou appelle	*Cebus apella*	Plathyrhinien	35 MA	35 à 45 cm, 2,5 et 4 kg.	35 - 40 ans	Amérique sud tropicale et subtropicale	forêts andines humides et chaudes	fruits, germes, feuilles, insectes	diurne	ADNg
Capucin moine	*Cebus capucinus*				30 ans	Sud et Nord-Ouest de l'Amérique du Sud	depuis les mangroves jusqu'aux forêts tropicales	fruits, feuilles, matières végétales, invertébrés et vertèbres de petite taille	diurne	ADNg
Singe Vert	*Chlorocebus aethiops*	Cercopithécoïdés	25 MA	45 à 60 cm, 3,5 à 5 kg	30 ans	du Sénégal à la Somalie, et jusqu'en Afrique du Sud	zones boisées et savanes	fruits, feuilles, racines, graines et œufs, petits reptiles, oiseaux	diurne	ADNg
Macaque rhésus	*Macacca mulatta*	Cercopithécoïdés	25 MA	jusqu'à 65 cm, 3 à 6 kg	30 ans	Afghanistan, Inde septentrionale, Chine méridionale	zones arides et terrains découverts	feuilles, aiguilles de pins, racines, insectes	diurne	ADNg
Macaque crabier	*Macacca fascicularis*			40 à 60 cm, 3 à 7 kg	30 ans	Asie du Sud-Est	forêts et mangroves	végétaux, mollusques, crabes	diurne	ADNg, tissus (animaux sains et traités au MPTP)
Gibbon à mains blanches	*Hylobates lar*	Hylobatidés	18 MA	40 à 90 cm, 5 à 10 kg	40 ans	Inde, Indochine, Malaisie	forêts	feuilles, fruits, fleurs, insectes, araignées, oiseaux, œufs	diurne	ADNg
Chimpanzé	*Pan troglodytes*	Panidés	5 MA	70 à 120 cm, 40 à 50 kg	40 à 50 ans	zone équatoriale africaine	forêts	fruits, feuilles,insectes, viande	diurne	ADNg
Bonobo	*Pan paniscus*			70 à 100 cm, 30 à 45 kg	40 ans	nord de la République Démocratique du Congo	forêts tropicales humides	fruits, feuilles	diurne	ADNg

Table 4.1.1 **Récapitulatif des caractéristiques des espèces étudiées**

4.1.1.2. Saguinus oedipus

Le Tamarin pinché *Saguinus oedipus* (cf Figure 4.1.1 **(b)**) est un primate Plathyrhinien de la famille de Cebidés et de la sous-famille des Challitrichidés (cf Figure 3.1.1.2.b). Son dernier ancêtre commun avec les Catarhiniens, dont l'Homme, date de 35 millions d'années environ (Enard and Paabo, 2004; Glazko and Nei, 2003).

Il mesure 25 à 30 centimètres, pèse 350 à 550 grammes et vit en moyenne 15 ans. Son unique habitat est la forêt dense du nord de la Colombie, où il s'alimente de fruits et légumes, et parfois aussi d'insectes ou d'œufs. Du fait de sa petite taille, le tamarin pinché est menacé par de nombreux prédateurs tels que les rapaces ou les serpents mais son plus terrible ennemi est l'homme qui détruit son habitat. Il fait en conséquence partie d'un programme européen d'élevage (EEP).

La seule ressource biologique disponible pour cette espèce est par conséquent son ADN génomique. Des tissus seront cependant accessibles prochainement dans le cadre du projet APES et d'une collaboration avec Lutz Walter du German Primate Center.

4.1.1.3. Cebus capucinus et Cebus appella

Les ADN génomiques de deux espèces de Cebus ont été utilisées au cours de nos travaux : le Sapajou apelle *Cebus apella* et le Capucin moine *Cebus capucinus* (cf Figure 4.1.1 **(c)**). Ils appartiennent à l'infra-ordre des Plathyrhiniens, à la famille de Cebidés et à la sous-famille des Cébinés (cf Figure 3.1.1.2.b). Leur dernier ancêtre commun avec les Catarhiniens, dont l'Homme, date de 35 millions d'années environ (Enard and Paabo, 2004; Glazko and Nei, 2003).

Cebus apella et *Cebus capucinus* mesurent tous deux 35 à 45 cm, et presque autant pour leur queue semi-préhensile, et pèsent entre 2,5 et 4 kg.

Le **Sapajou apelle** vit en Amérique sud tropicale et subtropicale à l'est des Andes, à l'exception du Chili et de l'Uruguay. Il habite les forêts andines humides et chaudes, jusqu'à 2700m d'altitude et s'y nourrit de fruits, germes, feuilles, insectes et petites proies. La femelle donne naissance à un petit par an, au terme d'une gestation de 5 mois environ.

L'aire de répartition du **capucin moine** se situe dans la partie Sud de l'Amérique Centrale et au Nord-Ouest de l'Amérique du Sud. Il occupe différents niveaux d'altitude depuis le niveau de la mer jusqu'aux contreforts nord de la cordillère des Andes. Il vit dans différents types de milieux végétaux depuis les mangroves jusqu'aux forêts tropicales

primaires et secondaires, qu'elles soient humides à feuilles caduques ou semi-sèches à feuilles persistantes. C'est un singe diurne et principalement arboricole, dont le régime alimentaire comprend 65 % de fruits, 15 % de feuilles et 20 % de matières végétales diverses (graines, écorces, bourgeons, gommes, fleurs), d'invertébrés (insectes, araignées, crabes, huîtres) et de vertébrés de petite taille (lézards, oiseaux, jeunes écureuils et coatis). En ce qui concerne le mode locomotion, le capucin moine est essentiellement quadrupède. La bipédie n'est qu'occasionnelle. La durée moyenne de gestation de *Cebus capucinus* est de 5 mois. En milieu naturel, le capucin moine ne forme pas de groupe mixte avec une autre espèce de Primates. Ses principaux compétiteurs en matière de nourriture sont le singe araignée à mains noires (*Ateles geoffroyi*), le singe hurleur à manteau (*Alouatta palliata*) et le coati (*Nasua narica*). Ses principaux prédateurs sont le boa constricteur, le serpent à sonnette, le caïman, le grand faucon noir, le jaguar, le puma, le coyote et les chasseurs.

4.1.1.4. Chlorocebus aethiops

Cette espèce était autrefois classée dans le genre *Cercopithecus* et nommée *Cercopithecus aethiops* ; elle est aujourd'hui nommée *Chlorocebus aethiops* (Linnaeus, 1758) (cf Figure 3.1.1.2.b). Elle appartient à l'infra-ordre des Catarhiniens, à la super-famille des Cercopithécoïdés, à la famille des Cercopithècidés et enfin à la sous-famille des Cercopithécinés. On estime sa divergence d'avec les Hominoïdes à il y a approximativement 25 millions d'années (Goodman, 1999).

Le Singe vert *Chlorocebus aethiops* (cf Figure 4.1.1 **(d)**) mesure 45 à 60 cm sans la queue et pèse 3,5 à 5 kg. Il vit en lisière des zones boisées et dans les savanes africaines, du Sénégal à la Somalie, et jusqu'en Afrique du Sud. Il est abondant dans le parc du Niokolo-Koba au Sénégal. Sa population y est estimée à 30 000 individus. Il se nourrit de fruits, feuilles, racines, graines et à l'occasion d'œufs, petits reptiles ou oiseaux. Il vit en groupe de 20 à 50 individus, le plus souvent au sol, mais trouve refuge dans les arbres. Il est également bon nageur. La maturité sexuelle est atteinte à 4 ans, la gestation dure plus de 5 mois et la longévité des individus est estimée à 30 ans au maximum.

La seule ressource biologique accessible pour cette espèce est son ADN génomique. Des tissus seront cependant disponibles dans le cadre du projet APES et d'une collaboration avec Lutz Walter du German Primate Center.

4.1.1.5. Macacca mulatta et Macacca fascicularis

Les macaques (cf Figure 4.1.1 **(e)**) appartiennent à la même sous-famille (des Cercopithécinés) que le *Chlorocebus aethiops* et se sont donc également séparés de la lignée des Hominoïdes il y a environ 25 millions d'années (cf Figure 3.1.1.2.b) (Goodman, 1999).

Les **macaques rhésus** *Macacca mulatta* atteignent une longueur de 65 centimètres, avec des queues atteignant 30 centimètres. Les mâles peuvent peser jusqu'à 6 kilogrammes, alors que les femelles sont autour de la moitié de ce poids. Ils passent la majorité de leur temps dans les arbres et leur longue queue leur sert de balancier pour sauter de branche en branche. Le macaque rhésus est commun dans tout l'Afghanistan jusqu'en Inde septentrionale et en Chine méridionale. Habitant les zones arides et les terrains découverts, les macaques rhésus peuvent aussi se trouver dans les prairies, les régions boisées et dans des régions montagneuses jusqu'à 2.500 mètres d'altitude. On dit qu'ils sont de bons nageurs et apprécient cette activité. Les rhésus sont connus pour leur tendance à se déplacer des espaces ruraux vers les secteurs urbains, venant chercher les aumônes ou les ordures des humains. Ils sont devenus de vrais parasites dans certaines zones, perçus comme un risque possible pour la sécurité et la santé publiques. Vivant le jour, les macaques rhésus sont aussi bien des animaux terrestres qu'arboricoles. Ils sont pour la plupart herbivores et se nourrissent de feuilles, d'aiguilles de pins, de racines et occasionnellement d'insectes ou de petits animaux. Comme les autres macaques, les groupes de singes rhésus comprennent des mâles et des femelles. Le groupe peut compter jusqu'à 180 individus mais la moyenne se situe à une vingtaine. Les femelles sont 4 fois plus nombreuses que les males. La hiérarchie sociale est matriarcale, le rang de chacun dépend de sa lignée avec la femelle dominante. Cette vie en groupe implique à la fois coopération et compétition. Les macaques sont capables de grandes performances, utiliser des outils ou échanger des informations, par exemple. Les femelles macaques n'ont qu'un petit tous les un à deux ans. La gestation dure près de six mois. Les macaques peuvent vivre une trentaine d'années.

Les **macaques crabier**s (ou macaques à longue queue) *Macacca fascicularis* vivent en groupes de plusieurs dizaines d'individus. D'un poids de 5 à 7 kg pour les mâles, celui-ci n'atteint que 3 ou 4 kg pour les femelles. La taille oscille de 40 à 60 cm. Les bandes sont très hiérarchisées. Elles se composent d'un mâle dominant, de plusieurs femelles et de leurs petits. On rencontre *Macacca fascicularis* dans la cime des arbres des jungles de montagne, mais également dans les forêts qui bordent les marais ou les fleuves, et les mangroves qui croissent

en Asie du Sud-Est, en Indonésie et dans les îles de la Sonde. Sur les traces du macaque à longue queue qui se régale de crabes, se profilent également les silhouettes de ses prédateurs : le tigre, le crocodile marin ou encore le python. Avantagé par son faible poids, le macaque crabier étend progressivement son territoire à la canopée pour échapper à la concurrence des autres espèces plus agressives. Il est également capable de nager.

Le génome entier du macaque rhésus est en cours de séquençage et les premières données sont d'ores et déjà disponibles sur www.ensembl.org. Nous disposons par ailleurs de son ADN génomique. Nous avons également accès à des ADN génomiques et des tissus, et donc des ARN, de *Macaca fascicularis* (sains et traités au MPTP, mimant la pathologie de Parkinson) issus d'un élevage chinois et prélevés dans des conditions normalisées.

4.1.1.6. Hylobates lar

Le Gibbon à mains blanches *Hylobates lar* appartient à la famille des Hylobatidés, à la super-famille des Hominoïdes (cf Figure 3.1.1.2.b). Son dernier ancêtre commun avec les autres Hominoïdes vivait il y a environ 18 millions d'années (Goodman, 1999).

Le Gibbon, singe sans queue, vit dans les forêts d'Inde, d'Indochine et de la Malaisie (cf Figure 4.1.1 **(f)**). Sa caractéristique la plus remarquable réside dans ses longs bras, grâce auxquels il se balance de branche en branche avec une grande agilité, en utilisant ses mains comme des crochets. Ce mode de déplacement s'appelle brachiation. Un *Hylobates lar* peut mesurer jusqu'à 90 cm de haut et l'envergure de ses bras fait environ le double. Le gibbon est le seul singe Haplorhinien, en dehors de l'Homme, qui marche uniquement sur ses pattes arrières, en levant les bras pour garder l'équilibre. Les gibbons sont monogames et l'unique petit du couple reste avec le groupe familial jusqu'à l'âge de cinq ou six ans. Ils mangent des feuilles, des fruits, des fleurs, des insectes, des araignées, des oiseaux et leurs œufs. Ils sont en général silencieux pendant la journée, mais hurlent au lever du soleil et en fin d'après-midi. Les populations de gibbons sauvages ont subi une sérieuse réduction du fait de la chasse et de la déforestation.

Nous ne disposons que de l'ADN génomique pour cette espèce.

4.1.1.7. Pan troglodytes et Pan paniscus

Deux espèces de chimpanzés retiennent particulièrement l'attention des biologistes moléculaires : *Pan paniscus* et *Pan troglodytes*. Tous deux sont des Hominidés et leurs

lignées se seraient séparése de la lignée humaine il y a environ 5 millions d'années (Chen and Li, 2001; Goodman et al., 1994).

Les **chimpanzés** mâles *Pan troglodytes* pèsent 50 kg et mesurent 120 cm, les femelles, 40kg et 70cm (cf Figure 4.1.1 **(g)**). On les rencontre dans 21 pays africains, de l'est à l'ouest de l'Afrique, au niveau de la zone équatoriale. Ils vivent surtout dans la forêt, tropicale humide ou sèche, et s'aventurent parfois dans la savane. Les chimpanzés vivent en communautés, regroupant jusqu'à plus de cent individus. La journée, ils se séparent en petits groupes pour aller chercher à manger, des fruits, des feuilles, des insectes et parfois aussi de la viande, dans la forêt. Décrits pour la première fois par Jane Goodall, les outils et inventions des chimpanzés se révèlent multiples et complexes, empreints d'un caractère traditionnel. L'espérance de vie des *Pan troglodytes* est de 40 à 50 ans, et la gestation dure près de huit mois. Malheureusement, aujourd'hui, les chimpanzés, comme les autres grands singes sont menacés de disparition. D'une population estimée à 1 million dans les années 1960, il n'en reste aujourd'hui probablement moins de 100 000.

Le **bonobo** *Pan paniscus* (cf Figure 4.1.1 **(h)**) mesure 70 à 100 cm et pèse 30 et 45 kg en moyenne, pour les femelles et les mâles respectivement. Il est aussi appelé chimpanzé pygmée, mais de pygmée il n'a que le nom étant simplement un peu plus élancé mais nullement plus petit que le chimpanzé commun. C'est sans doute des grands singes, celui qui nous ressemble le plus, d'un point de vue génétique mais aussi comportemental. Les bonobos ne vivent que dans les forêts tropicales humides du nord de la République Démocratique du Congo, souvent inondées. Là, ils vivent en groupes, au sein d'une société de type égalitaire à tendances matriarcales. Frugivores et folivores, les Bonobos passent beaucoup de temps dans la forêt, au sol ou dans les arbres à rechercher leur nourriture, toujours en groupe. Leur espérance de vie est d'environ 40 ans, et la gestation dure près de huit mois. En République Démocratique du Congo, les guerres civiles et la situation politique toujours instable ont mis les bonobos en péril. De 100 000 individus en 1980, les bonobos sont passés à 10 000 en 1990 et la pression de la chasse ne fait qu'augmenter suite à l'implantation massive des industries forestières et minières dans les forêts qui créent à tout va moultes pistes à travers la sylve jusqu'alors réputée impénétrable.

Nous ne disposons que d'ADN génomique pour *Pan paniscus*. Le génome de *Pan troglodytes* est actuellement séquencé en quasi-totalité et disponible sur www.ensembl.org. Nous disposons également de son ADN génomique, mais ne pouvons avoir accès à des tissus et ARN de bonne qualité. Les tissus de ces deux espèces devraient cependant être accessibles

à terme par l'intermédiaire de la Biothèque, en collaboration avec Erwan Bézard (Biothèque Primates, Valbonne & Bordeaux).

4.1.1.8. Homo sapiens

Nous avons accès à des ADN génomiques humains, ainsi qu'à des tissus, dont nous pouvons extraire les ARN, par l'intermédiaire de l'hôpital Edouard Herriot de Lyon et le National Research Specimen Bank (Los Angeles, USA). Nous pouvons également obtenir des tissus centraux de patients parkinsoniens et sains de la cérébrothèque de la Pitié-Salpétrière à Paris.

4.1.2. Les modèles animaux expérimentaux adaptés à l'étude de gènes et pathologies spécifiques des Primates

On dénombre un certain nombre de pathologies spécifiques des Primates, ou même de l'Homme seul. Il s'agit en particulier de maladies neurodégénératives. Parmi elles, la maladie de Parkinson n'affecte à l'état naturel que l'espèce humaine et ne présente aucun équivalent dans d'autres espèces. L'étude de cette pathologie, dont la prévalence augmente avec l'espérance de vie, requiert le développement de modèles animaux appropriés.

4.1.2.1. Description de la maladie de Parkinson

La maladie de Parkinson est la deuxième pathologie neurodégénérative la plus fréquente et affecte, selon les estimations, 1 à 2% de la population des plus de 65 ans. Elle se manifeste par une bradykinésie, des tremblements, une rigidité et des réflexes posturaux anormaux. Lorsque les symptômes apparaissent, une large proportion (50 à 70%) des neurones dopaminergiques de la partie compacte de la substance noire est déjà détruite. La synthèse et la libération de dopamine sont alors nettement réduites. La formation de corps de Lewy dans les neurones dopaminergiques survivants signe la pathologie de Parkinson. Il s'agit d'inclusions intracytoplasmiques composées essentiellement d'ubiquitine et d'α-synucléine, mais aussi de parkine (Orth and Tabrizi, 2003).

Des facteurs génétiques et environnementaux sont identifiés qui pourraient participer au développement de la maladie. Différentes mutations, affectant les gènes de l'ubiquitine, de l'α-synucléine, de la parkine et DJ-1, ont été isolées dans des cas familiaux de la maladie de Parkinson. Les toxines environnementales impliquées dans le développement de la maladie sont le monoxyde de carbone, le manganèse et de façon très importante le 1-méthyl-4-phényl-

1,2,3,6 tétrahydropyrimidine (MPTP). Il apparaît clairement que la maladie de Parkinson est une pathologie hétérogène, dont l'étiologie est complexe et multifactorielle. L'administration de L-dopa ou d'agonistes dopaminergiques en substitution de la dopamine permet de faire disparaître les symptômes mais ne freine pas la progression de la maladie.

L'utilisation de modèles animaux est nécessaire à la compréhension de la pathologie afin de pouvoir la dépister plus précocement et la traiter de façon appropriée. De tels modèles doivent reproduire les paramètres cliniques et pathologiques caractéristiques de la maladie de Parkinson, et en particulier des troubles de la locomotion sensibles à la L-dopa, une perte de neurones dopaminergiques progressive et chronique, et des inclusions semblables aux corps de Lewy (Orth and Tabrizi, 2003).

4.1.2.2. Modèles expérimentaux

4.1.2.2.1. Modèles génétiques

Des souris déficientes (knock-out) pour, respectivement, les gènes de l'α-synucléine et de la parkine ont été développées. Les premières ne présentent aucun symptôme objectif de la maladie de Parkinson. Leur développement est normal et elles sont absolument fertiles. Seul le contenu en dopamine dans le striatum distingue ces souris déficientes de souris sauvages, ainsi que le comportement locomoteur en réponse à des stimuli chimiques (administration d'amphétamine) et électriques. Ceci suggère que l'α-synucléine n'est indispensable ni pour le développement du cerveau murin, ni pour le fonctionnement des neurones dopaminergiques. Elle pourrait cependant intervenir dans la libération de dopamine. Des souris déficientes pour la parkine ont été récemment décrites. Elles sont viables et ne présentent, de façon surprenante, aucun phénotype pathologique. Cela pourrait résulter d'adaptations développementales compensatrices, si bien qu'il a été proposé que des souris pour lesquelles le gène de la parkine ne serait éteint qu'après la naissance pourraient présenter un phénotype plus sévère (Orth and Tabrizi, 2003).

Des lignées de souris transgéniques ont été établies pour tenter de modéliser le développement de la maladie de Parkinson. Les souris exprimant le gène sauvage de l'α-synucléine humain développent des inclusions cytoplasmiques et nucléaires dans les neurones du néocortex, du bulbe olfactif, de l'hippocampe et de la substance noire. Les analyses cliniques ont révélé des déficits de performance dans les tests d'équilibre chez ces souris âgées d'un an et surexprimant l'α-synucléine. Dans ce modèle, certaines caractéristiques de la pathologie sont donc reproduites (inclusions, défauts locomoteurs), mais d'autres sont

absentes (mort de neurones dopaminergiques de la substance noire, fibrilles des corps de Lewy, localisation atypique des inclusions dans le noyau cellulaire) (Orth and Tabrizi, 2003).

Ces modèles génétiques de la maladie de Parkinson permettent donc d'inférer le rôle des gènes dont l'expression est modifiée dans le développement de la pathologie. Ils ne constituent en revanche pas de bons modèles pour l'étude de la maladie dans son ensemble.

4.1.2.2.2. Modèles toxicologiques

Il existe trois modèles toxicologiques principaux de la maladie de Parkinson : l'administration de 6-hydroxydopamine (6-OHDA), de MPTP ou de roténone. Tous trois impliquent une mort des neurones dopaminergiques.

Dans le premier modèle toxicologique, l'injection directe, souvent unilatérale, de 6-OHDA dans la substance noire, essentiellement à des Rongeurs, induit un déficit moteur quantifiable, mais les corps de Lewy ne sont pas observables et on constate une atteinte non-spécifique, induite par la toxine, des neurones voisins. Ce modèle a été très utile et utilisé pour les études pharmacologiques, mais ne présente pas certaines des caractéristiques fondamentales de la maladie de Parkinson (Orth and Tabrizi, 2003).

Des rats perfusés quotidiennement avec de la roténone, un inhibiteur potentiel du complexe I de la chaîne respiratoire mitochondriale, développent une hyperkinésie, une instabilité posturale, une démarche hésitante et des tremblements des membres. Ces symptômes sont tous corrélés avec le degré de dégénérescence des neurones dopaminergiques nigro-striataux. On observe en microscopie des inclusions cytoplasmiques semblables aux corps de Lewy. En revanche, moins de la moitié des animaux traités à la roténone développent effectivement ces lésions (Orth and Tabrizi, 2003).

Des cas de maladie de Parkinson ont été rapportés chez des toxicomanes dont la drogue aurait été contaminée par du MPTP, un inhibiteur du complexe I de la chaîne respiratoire mitochondriale. Des modèles murin et de Primate de la maladie utilisant cette neurotoxine ont alors été développés. Chez les Primates non-humains, le MPTP induit des tremblements, une rigidité, une akinésie et une instabilité posturale, symptômes qui disparaissent tous lors de l'administration de L-dopa ou d'agonistes dopaminergiques. Les neurones dopaminergique sont détruits sélectivement dans certaines aires cérébrales, dont la substance noire, conformément à l'idiopathie de la maladie de Parkinson. Néanmoins, la pathologie déclenchée par un traitement aigu au MPTP n'induit pas le développement de corps de Lewy caractéristiques. Une administration chronique de MPTP pourrait cependant

pallier à ce défaut. Le modèle MPTP représente ainsi le modèle de maladie de Parkinson le mieux caractérisé et répond à l'ensemble des critères d'un bon modèle de cette pathologie, à l'exception de la formation d'inclusions (Orth and Tabrizi, 2003).

Nous avons accès, dans le cadre de nos études, et en collaboration avec l'équipe d'Erwan Bézard (CNRS, UMR 5543, Bordeaux), à des macaques crabiers traités au MPTP de façon chronique, pendant 6, 12 ou 15 jours et sacrifiés après 6, 12 et 25 jours respectivement. Ceci permet d'observer les modifications phénotypiques et moléculaires se produisant au cours de la maladie, et notamment lors de la phase asymptomatique.

Ce type de pathologies, spécifiques des Primates ou de l'Homme, pourrait en outre être déterminé, au moins partiellement, par des gènes jeunes, présents uniquement dans ces espèces. L'étude de leurs structures, de leur évolution et éventuellement de leurs fonctions et implication dans des pathologies spécifiques pourrait ainsi constituer une clé de la compréhension des spécificités des Primates et de l'Homme.

4.2. Les gènes chimériques Primates-spécifiques

Je me suis intéressée au cours de ma thèse à deux familles de gènes spécifiques des Primates. Pour l'une d'entre elle, la famille des gènes *PMCHL*, le gène ancestral (de la MCH ou Melanin Concentrating Hormone) était largement étudié au laboratoire et un modèle d'apparition des gènes chimériques était proposé. La deuxième famille de gènes à avoir retenu mon attention est la famille des gènes *GUSL*, dérivés du gène de la β-Glucuronidase. Leur caractérisation était alors tout à fait préliminaire.

4.2.1. Le système MCH et les gènes *PMCHL*

4.2.1.1. Le système à MCH

En 1983, Kawachi, Baker et leurs collaborateurs identifiaient chez le saumon la Melanin Concentrating Hormone (MCH), un peptide cyclique de 17 acides aminés impliqué dans la pigmentation de la peau des poissons téléostéens. La structure génique en déterminant la séquence a ensuite été finement disséquée chez les poissons et les mammifères, avant qu'un retour aux fonctions du neuropeptide soit opéré. Seules les données relatives au système à MCH des mammifères seront présentées ici.

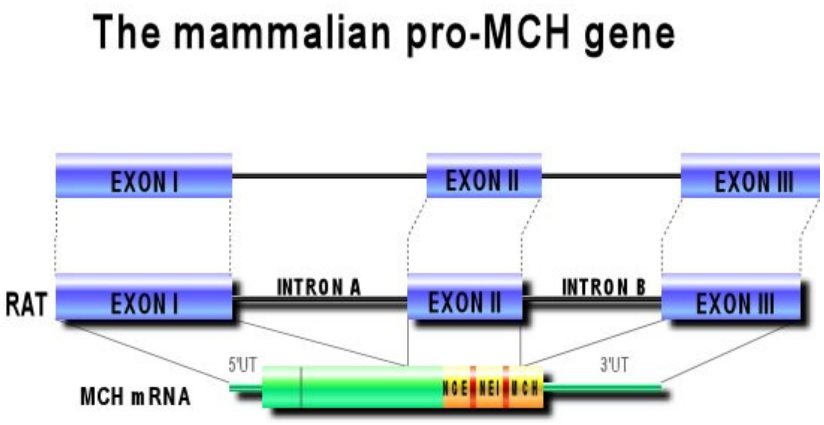


Figure 4.2.1.1.1.a **Le système MCH : gène, ARNm et précurseur peptidique**

Les structures géniques de l'homme (en haut) et du rat sont présentées, ainsi que les correspondances entre exons et peptides synthétisés : NGE Neuropeptide Glycine(G)-Glutamic Acid (E), NEI Neuropeptide Glutamic acid (E)-Isoleucine (I) et MCH Melanin Concentrating Hormone.

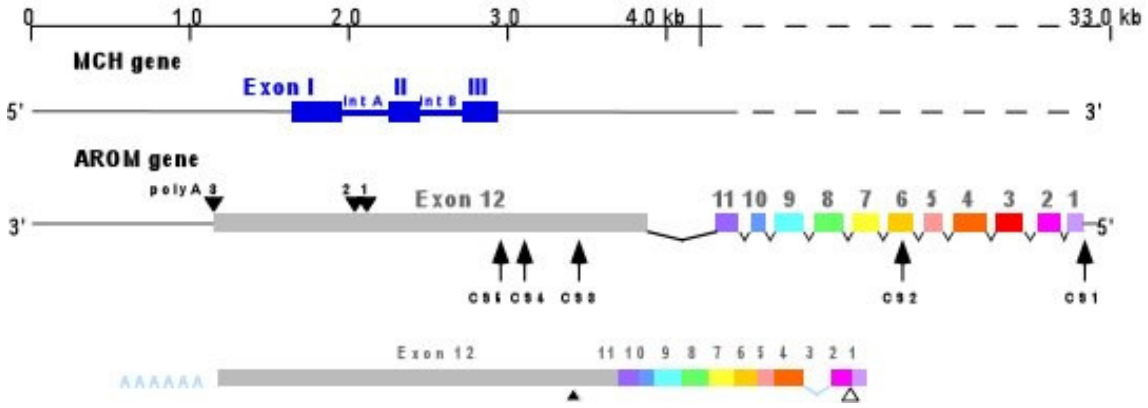


Figure 4.2.1.1.1.b **Le locus *MCH*/AROM**

Le gène *MCH* compte trois exons (notés I à III). Sur le brin opposé se trouve le gène *AROM Antisense RNA Overlapping MCH gene*, dont le douzième exon, parfois transcrit, couvre l'ensemble de la région trancrite du gène MCH. CS1 à CS5 : Cap sites 1 à 5, polyA 1 à poly A 3 : sites de polyadénylation 1 à 3. L'un des transcrits principaux est présenté. Il est. non épissé et recouvre les exons II et III ainsi que les régions introniques et régulatrices flanquantes du gène *MCH*. D'après (Borsu et al., 2000).

4.2.1.1.1. Le locus *MCH* : gènes en présence et messagers produits

Le gène codant pour la MCH est composée de trois exons et deux introns. Chez l'Homme, le gène est localisé sur le chromosome 12q24.1 (Pedeutour et al., 1994; Viale et al., 1997) et s'étend sur environ 1,4kb. Le premier exon couvre la région 5' non traduite de l'ARN messager et la partie N-terminale du précurseur peptidique, incluant le peptide signal qui permet l'adressage à la voie de sécrétion. Le deuxième exon comprend les séquences codant pour les peptides putatifs NGE (pour Neuropeptide Glycine(G)-Glutamic Acid (E)) et NEI (Neuropeptide Glutamic acid (E)-Isoleucine (I)), ainsi que pour les trois premiers acides aminés de la MCH. Le reste de la séquence codant pour le neuropeptide MCH et la région 3' non traduite sont portées par l'exon III (cf Figure 4.2.1.1.1.a).

Au cours de l'étude de la séquence et de l'expression ribonucléique du gène de la Melanin Concentrating Hormone, une nouvelle unité de transcription associée a été mise en évidence. La caractérisation fonctionnelle d'ARNm longs reconnus par une sonde radioactive correspondant au cDNA MCH a permis la découverte de l'existence de messagers antisens « naturels » transcrits à partir de la région génomique complémentaire au gène *MCH*. Ce nouveau gène a été nommé *AROM* pour *Antisense-RNA-Overlapping-MCH* (Borsu et al., 2000). Deux classes principales d'ARN messagers sont exprimées : des messagers non épissés recouvrant les exons II et III ainsi que les régions introniques et régulatrices flanquantes du gène *MCH*, et des messagers alternativement épissés codant pour des protéines liant les acides nucléiques et complémentaires de la région 3' flanquante du gène *MCH* (cf Figure 4.2.1.1.1.b).

4.2.1.1.2. L'hormone de mélanoconcentration : structure, expression et fonctions

La traduction du gène *MCH* donne naissance à la pré-pro-MCH, qui devient pro-MCH après clivage du peptide signal. Les peptides issus du précurseur pro-MCH par maturation et clivages au niveau de doublets d'acides aminés basiques sont essentiellement, dans le cerveau de mammifère, la MCH mature et le peptide NEI amidé (Bittencourt et al., 1992; Nahon et al., 1989; Viale et al., 1999) (cf Figure 4.2.1.1.2.a). La MCH est un peptide cyclique très conservé chez les Vertébrés. Il compte 19 acides aminés chez les Mammifères et présente une forte similarité structurale avec le peptide des Poissons, à l'exception d'une extension N-terminale et de quatre substitutions (Vaughan et al., 1989).

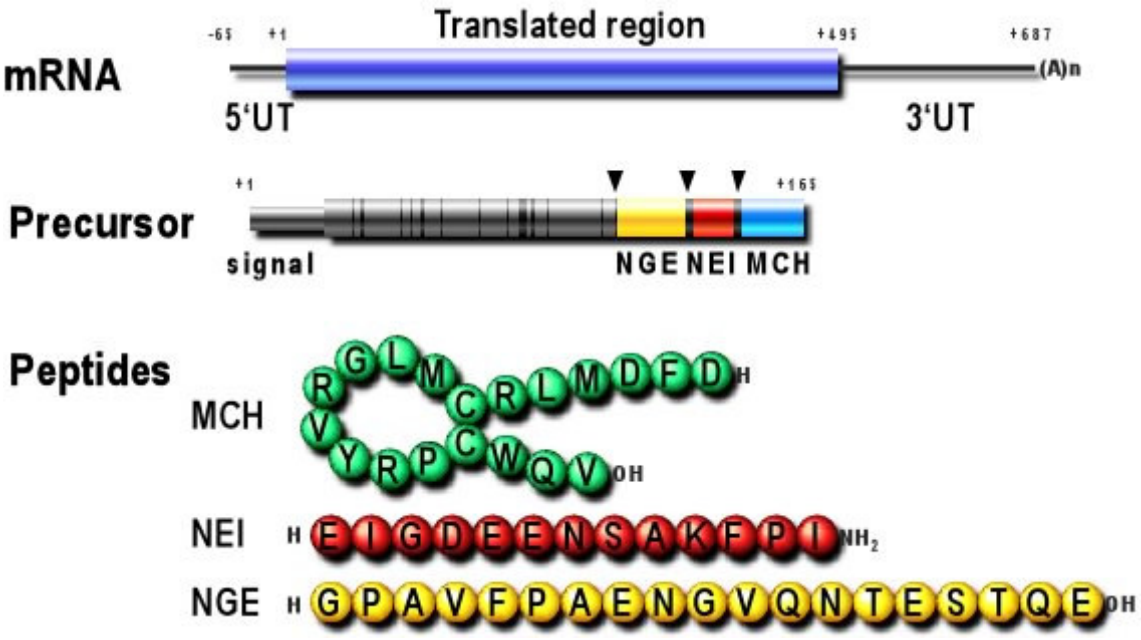


Figure 4.2.1.1.2.a **Synthèse de l'hormone de mélanoconcentration**

Le messager du gène MCH compte 752 nucléotides, dont 495 sont traduits. Le Précurseur peptidique ainsi synthétisé comporte un peptide signal, un prodomaine et trois peptides libérés par clivage. Le peptide principal est la Melanin Concentrating Hormone, un petit peptide cyclique.

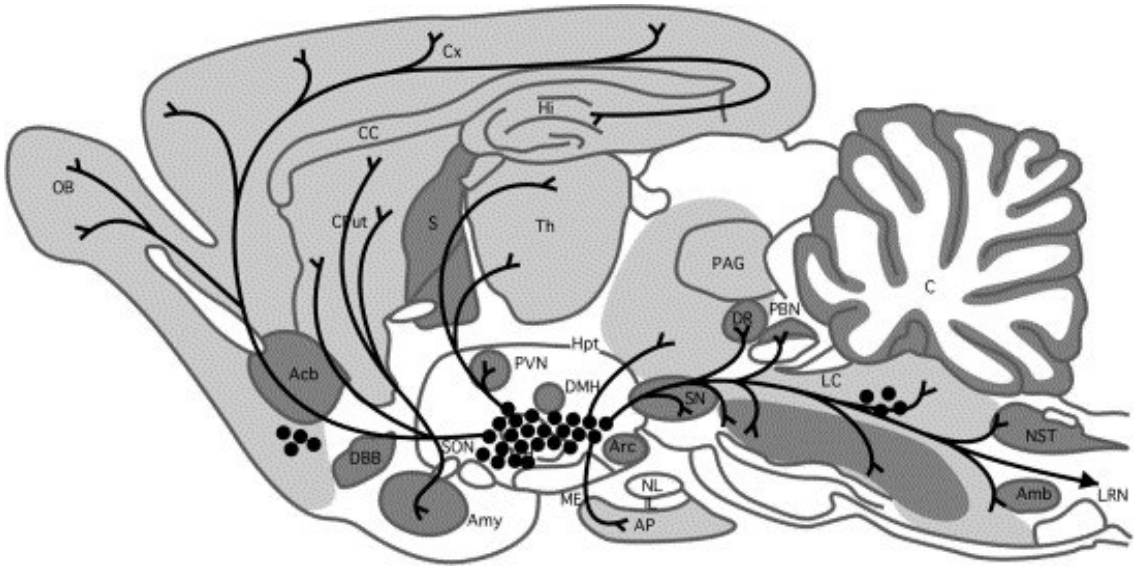


Figure 4.2.1.1.2.b **Expression de la MCH et de son récepteur**

Section sagittale de cerveau de rat figurant la distribution des neurones à MCH ainsi que leurs projections. L'expression transcriptionnelle du récepteur MCH-R1 est également indiquée. Les points noirs représentent les corps cellulaires des neurones à MCH et les lignes correspondent aux fibres. De forts, moyens et faibles niveaux d'expression du récepteur MCH-R1 sont figurés en gris foncé et clair et en blanc, respectivement.
Acb noyau accumbens, Amy Amygdale, Amb Noyau ambigu, AP Glande adéno-pituitaire, Arc Noyau arqué, C cervelet, Cx Cortex, CC corps calleux, CPut Caudé Putamen, DMH Hypothalamus dorso-médian, Hi Hippocampe, LC Locus coeruleus, ME Eminence médiane, NST Noyau de la strie terminale, OB Bulbe olfactif, PAG Substance grise péri-aqueducale, PBN Noyau para-brachial, PVN Noyau péri-ventriculaire, S Striatum, SN Substance noire, Th Thalamus.

La population neuronale exprimant la MCH chez les Mammifères est principalement confinée à l'hypothalamus dorso-latéral et à la zona incerta, tandis que les projections touchent de multiples aires cérébrales, du bulbe olfactif à la mœlle épinière, à l'exception notoire du cervelet, de l'éminence médiane externe et de la glande pituitaire. De petits groupes de neurones exprimant la MCH sont cependant également identifiés dans des aires extrahypothalamiques telles que le tubercule olfactif et le tegmentum du pont (Bittencourt et al., 1992) (cf Figure 4.2.1.1.2.b). Enfin, la MCH est également sythétisée dans certains tissus périphériques (Hervieu and Nahon, 1995).

Une grande variété de fonctions semble être sous le contrôle de ce neuropeptide, dont principalement le comportement alimentaire, l'homéostasie énergétique et la régulation de l'axe du stress et des émotions. On observe en effet une nette surexpression des messagers MCH chez des souris génétiquement obèses ou après une privation de nourriture (Qu et al., 1996). Les effets orexigéniques de la Melanin Concentrating Hormone ont ensuite été confirmés dans diverses expériences. Ce sont les effets les plus documentés pour ce neuropeptide et les données physiologiques (expériences d'injection de peptide), pharmacologiques (utilisation d'antagonistes) et génétiques (modèles de souris invalidées pour les gènes *MCH* ou *MCHR1*, ou surexprimant le gène *MCH*) concordent parfaitement. Ils sont médiés par l'interaction avec le récepteur MCH-R1 (cf Figure 4.2.1.1.2.b). La MCH est cependant impliquée dans un certain nombre d'autres fonctions centrales et périphériques, telles que la reproduction, la prise hydrique, les soins de toilettage, le comportement agressif, la locomotion, le sommeil, l'apprentissage et la constitution de la mémoire (Nahon, 1994; Nahon, 2006).

4.2.1.2. Apparition et évolution des gènes PMCHL1 et PMCHL2

L'émergence de deux gènes chimériques spécifiques des primates a été mise en évidence dans le laboratoire de Jean-Louis Nahon (Breton et al., 1993; Pedeutour et al., 1994). Nommés *PMCHL1 (Pro Melanin Concentrating Hormone Like 1)* et *PMCHL2 (Pro Melanin Concentrating Hormone Like 2)* du fait de leur homologie partielle avec le gène *MCH* authentique, ils se localisent respectivement sur les chromosomes 5p14.3 et 5q13.3 humains (Viale et al., 2000) (cf Figure 4.2.1.2.a). Ils sont exclusivement présents dans des génomes de primates (Viale et al., 1998a).

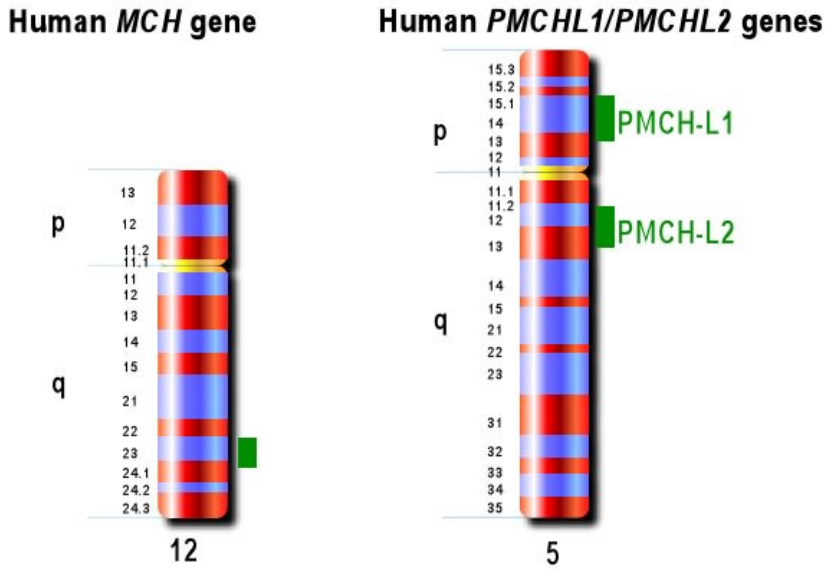


Figure 4.2.1.2.a **Localisations chromosomiques des gènes *MCH* et *PMCHL***

Le gène *MCH* se situe sur le chromosome 12q23 chez l'Homme, tandis que les gènes qui en sont dérivés, *PMCHL1* et *PMCHL2*, se trouvent sur un autre chromosome. On les détecte en effet de part et d'autre du centromère du chromosome 5 humain.

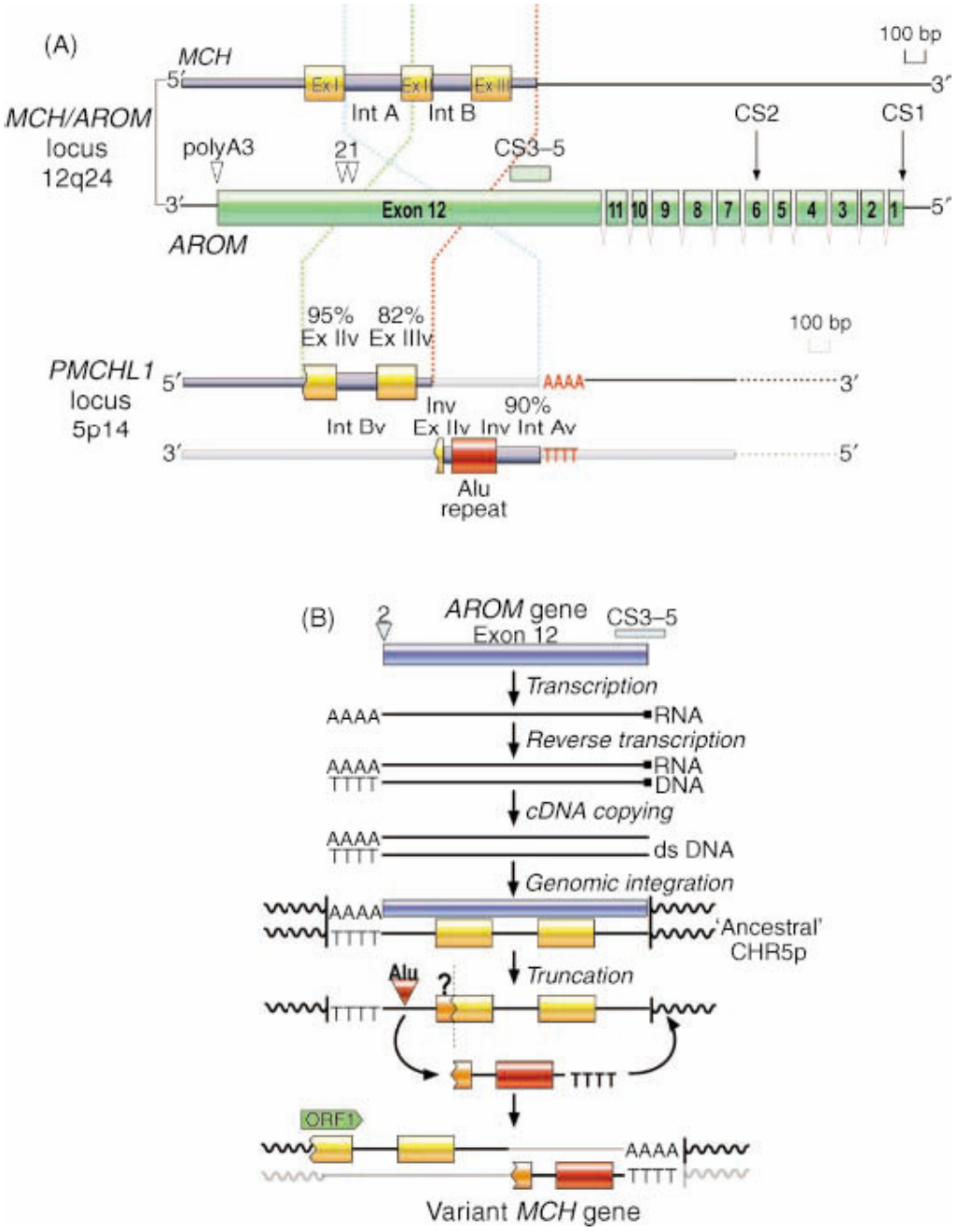


Figure 4.2.1.2.b **Modèle de rétroposition du messager MCH antisens**

(A) Structures exon/intron des gènes *MCH*, *AROM* et *PMCHL1*. Les brins d'ADN sens et antisens sont figurés pour les loci *MCH/AROM* et *PMCHL*. CS 1-5 Cap sites 1-5 du gène *AROM*, polyA 1-3 sites de polyadénylation 1-3 du gène *AROM*. D'après (Borsu et al., 2000). **(B)** Modèles de création des gènes *MCH variants* par rétroposition d'un ARNm AROM. D'après (Courseaux and Nahon, 2001). Tiré de (Nahon, 2003).

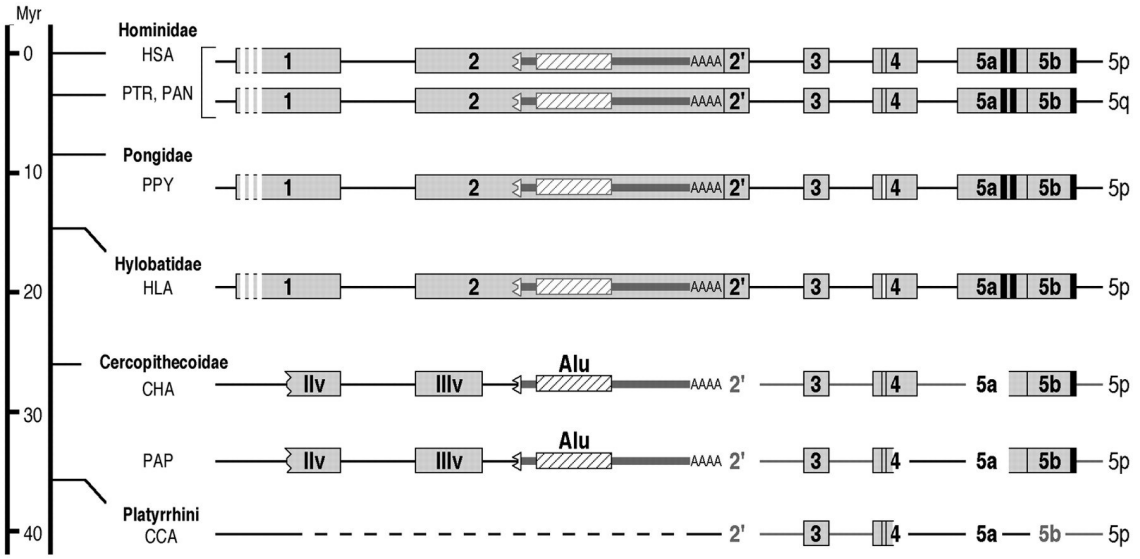


Figure 4.2.1.2.c **Exaptation et création d'exons**

Modèle proposé d'émergence des gènes chimériques *PMCHL1* et *PMCHL2* au cours de l'évolution des Primates. Une séquence dérivée de la MCH a été rétrotransposée sur le chromosome 5p au moment de la divergence des Catarhiniens. Les bornes exon/intron et les signaux de polyadénylation ont été créés ensuite par mutation avant la divergence des Hylobatidés. Un évènement de duplication a ensuite conduit à la distribution de *PMCHL1* et *PMCHL2* sur les chromosomes 5p et 5q. Les sites de polyadénylation sont indiqués par de petites barres noires verticales. Tiré de (Courseaux and Nahon, 2001).

Des études de séquençage et d'hybridation fluorescente in situ ont été réalisées afin de comprendre les modalités de la création et de l'évolution des gènes *PMCHL1* et *PMCHL2*. Il a ainsi été établi que trois types d'évènements génétiques étaient associés à l'émergence de ces gènes (cf Figure 4.2.1.2.b).

D'abord s'est produite la transposition d'une copie des exons II et III, et de l'intron les séparant, du gène *MCH* authentique sur l'équivalent du chromosome 5p chez l'ancêtre des Catarhiniens, il y a environ 30 millions d'années. L'identification d'une série d'adénosines à la fin de la séquence dérivée du gène *MCH*, qui coïncide exactement avec l'un des sites de polyadénylation des ARN messagers du gène *AROM*, antisens naturel du gène *MCH*, ainsi que la coïncidence entre l'extrémité 5' de la séquence insérée et un cap site du gène *AROM*, suggèrent que cette séquence a probablement été insérée sur l'ancêtre du chromosome 5p lors d'un évènement de rétrotransposition d'un ARNm AROM (Courseaux and Nahon, 2001). Puis il y aurait eu troncation de l'exon II, transposition de sa partie 5' en aval de l'exon III variant et en orientation inverse (Viale et al., 2000), et accumulation de mutations avant la divergence des Hominoïdes. C'est à la même période qu'une séquence Alu aurait été insérée en aval des séquences dérivées de la *MCH* (Viale et al., 2000). Enfin, le gène variant *PMCHL1* a été dupliqué pour créer *PMCHL2* chez les Hominidés et Pongidés, il y a environ 10 millions d'années (Viale et al., 1998a).

Chez l'Homme des transcrits des gènes *PMCHL* ont été identifiés dans le cervelet, l'hippocampe et le cortex et semblent absents de l'hypothalamus d'individus adultes humains, où le gène ancestral *MCH* est, lui, fortement exprimé (Viale et al., 2000; Viale et al., 1998b). Des études ultérieures ont mis en évidence la présence de messagers sens, parfois épissés de multiples façons alternatives, mais également d'ARNm antisens du gène *PMCHL1*, dont les extrémités 5' et 3' ont été caractérisées dans le cerveau humain (Courseaux and Nahon, 2001; Viale et al., 2000). Ces données d'expression indiquent en outre que les séquences rétroposées dérivées de la *MCH* ont recruté des séquences présentes au site d'insertion, et possédant, ou ayant acquis de façon fortuite, une capacité transcriptionelle ou de régulation. Une structure chimérique a ainsi été créée par rétroposition, exaptation de séquences et brassage d'exons (Courseaux and Nahon, 2001; Nahon, 2003) (cf Figure 4.2.1.2.c).

Ainsi, les gènes *PMCHL* présentent toutes les caractéristiques attendues de gènes « en recherche de fonction » : ils sont apparus par duplication au cours de l'évolution des

Catarhiniens, ils sont faiblement exprimés, limitant ainsi les effets délétères de mutations dominantes et enfin, leur profil d'expression n'est pas ubiquitaire, mais clairement distinct de celui du gène ancestral *MCH* (Viale et al., 2000; Viale et al., 1998a). L'accumulation de mutations, dont certaines ont notamment favorisé l'apparition de séquences reconnues par la machinerie d'épissage, renforce cette hypothèse (Courseaux and Nahon, 2001). Il semblerait néanmoins que les gènes *PMCHL* soient mieux conservés dans la lignée des Hominidés, ce qui pourrait indiquer l'apparition d'une contrainte sélective et donc probablement d'un rôle émergent de ces gènes (Courseaux and Nahon, 2001).

4.2.2. Données préliminaires sur les gènes dérivant de la béta-glucuronidase

Au cours de l'analyse du locus SMA (Spinal Muscular Atrophy), sur le chromosome 5q13 humain, de nombreux pseudogènes exprimés ont été identifiés. Ces éléments génomiques contiennent des cassettes d'exons et introns très similaires à celles de gènes fonctionnels distribués sur d'autres chromosomes (Courseaux et al., 2003; Nahon, 2003) (cf Figure 4.2.2.a). L'une de ces structures génomiques atypiques a retenu l'attention du laboratoire de Jean-Louis Nahon et sa structure, son organisation chromosomique et son évolution ont été examinées. Il s'agit des gènes initialement dénommés *Glu 5-10* (Courseaux et al., 2003).

Les gènes *Glu 5-10*, à présent nommés *GUSL* (*GUS-like*), présentent de grandes similarités avec les exons 5, 9, 10 et les séquences introniques adjacentes, du gène de la β-Glucuronidase (*GUSB*), qui est un gène ubiquitaire et très conservé au cours de l'évolution, localisé sur le chromosome 7q21 humain (Sargent et al., 1994; Theodosiou et al., 1994). En combinant des analyses *in silico* et d'hybridation fluorescente *in situ* (FISH), il a été montré qu'un certain nombre d'évènements de transposition et de duplication inter- et intrachromosomiques a permis la création et l'expansion de cette nouvelle famille de gènes (Courseaux et al., 2003).

Le modèle proposé stipule qu'un évènement de transposition à partir de la région chromosomique ancestrale correspondant au chromosome humain 7q21 a conduit à l'insertion de Glu 5-10 sur l'équivalent du bras court du chromosome 5 après la divergence des Catarhiniens, il y a environ 30 millions d'années. Puis une série de duplications d'un large fragment incluant l'élément *Glu 5-10* et d'autres exons, provenant d'autres régions progénitrices, a eu lieu dans la lignée des Hominidés dans les 10 derniers millions d'années,

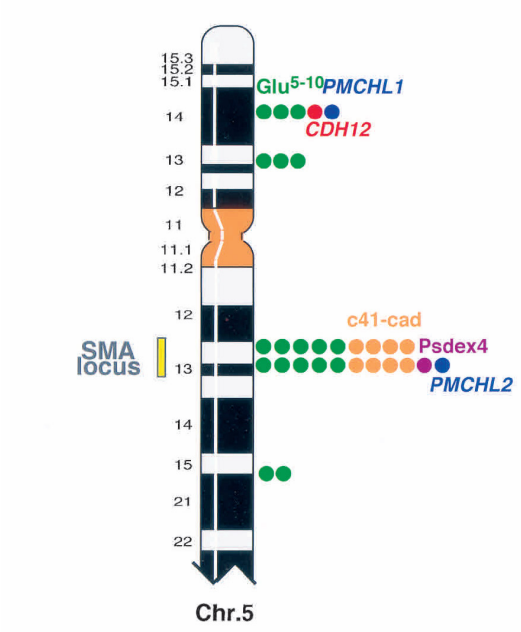


Figure 4.2.2.a **Colocalisation de séquences dérivées de gènes sur le chromosome 5**

Les signaux d'hybridation obtenus en FISH, avec des sondes spécifiques, sur le chromosome 5 humains sont représentés par des points colorés. (bleu) *PMCHL*, (vert) *Glu 5-10*. Le nombre de points est proportionnel à l'intensité du signal. Tiré de (Courseaux et al., 2003).

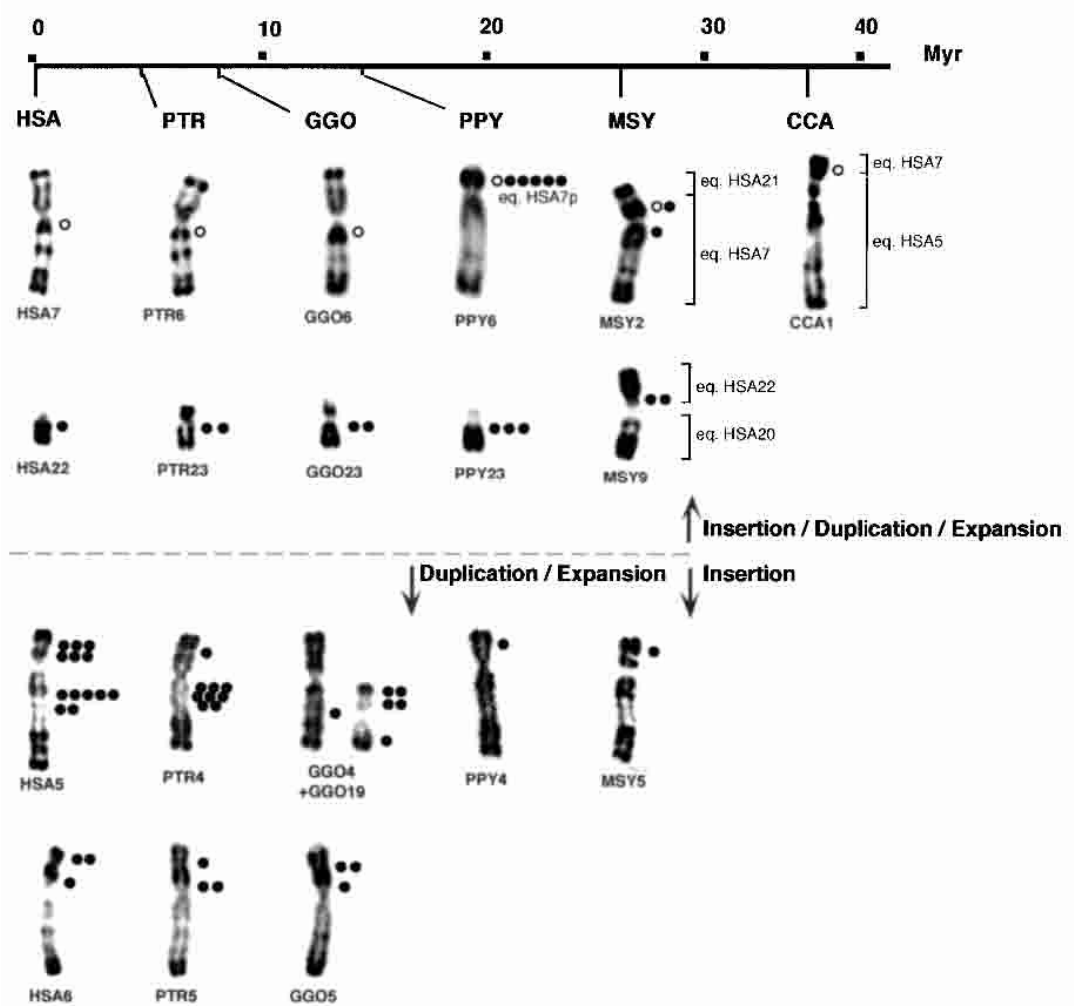


Figure 4.2.2.b **Glu 5-10 chez les Primates**

Résumé des résultats d'hybridation *in situ* obtenus avec une sonde Glu 5-10 chez les Primates. Le nombre de points (en blanc, le gène ancestral, en noir, les paralogues dérivés) indique l'intensité moyenne du signal observé en un certain locus. Les crochets illustrent les synténies entre chromosomes de l'Homme, du Macaque et du *Cebus*. On présuppose que l'homologue du chromosome 7 humain était séparé en deux dans le caryotype de l'ancêtre des Mammifères placentaires. Ces deux éléments ont fusionné avant la séparation des cercopithécoïdés et des Hominidés (il y a environ 25 millions d'années), puis des inversions péri- et paracentriques sont survenues dans la lignée humaine. Les différences de la chronologie de l'expansion des séquences paralogues dérivées de GUSB sont matérialisées par les pointillés horizontaux. En haut, la première insertion a lieu dans des régions péricentromériques des ancêtres des chromosomes 7 et 22. Les phénomènes de duplication/expansion sont simultanés de cette insertion et prennent plus ou moins d'ampleur selon les lignées. En bas, La première insertion s'effectue dans la région euchromatique de l'ancêtre du chromosome 5p. L'expansion est retardée et affecte les homologues des chromosomes 5 et 6 humains. HSA *Homo sapiens*, PTR *Pan troglodytes*, GGO *Gorilla gorilla*, PPY *Pongo pygmaeus*, MSY *Macacca sylvana*, CCA *Cebus capucinus*. Tiré de (Courseaux et al., 2003).

ce qui a conduit à la distribution actuelle de multiples copies des gènes *Glu 5-10* sur différents chromosomes. Les différentes copies présentent une grande identité de séquence, et sont également très semblables au gène ancestral de la β-Glucuronidase. Elles sont positionnées sur le chromosome 5 humain (5p14, 5p13, 5q13, 5q15), mais également sur les chromosomes 6 (6p11 et 6p21), 22q11 et 7q11.

Une ébauche de chronologie de l'émergence et de la dispersion des séquences dérivées de *GUSB* peut être proposée suite à l'analyse de la distribution chromosomique dans d'autres espèces de primates (cf Figure 4.2.2.b). L'analyse, parcellaire pour des raisons techniques, des structures actuelles des paralogues dérivés de *GUSB* semblerait indiquer que le ou les évènement(s) de transposition initiaux a/ont impliqué une grande portion du gène *GUSB*, dont une/des partie(s) a/ont ensuite été délétée(s). Une copie ainsi transposée dans une région péricentromérique a ensuite rapidement été utilisée comme source pour d'autres évènements de transposition/duplication chez les Primates. La copie initiale dérivée de *GUSB* et positionnée sur l'ancêtre du chromosome 5p humain a en revanche connu une période de stabilité (sans duplication) de près de 20 millions d'années, suivie de séries de duplications dans la lignée des Hominidés dans les 10 derniers millions d'années (Courseaux et al., 2003) (cf Figure 4.2.2.b).

Des transcrits chimériques contenant les exons 5, 9 et 10 dérivés du gène de la β-Glucuronidase sont identifiés chez l'Homme. Certains comportent également l'exon 11 de *GUSB*. Des exons exogènes au gène de la β-Glucuronidase, uniques ou eux-mêmes répétés dans le génome, et mentionnés précédemment pour leur présence au sein du même duplicon, sont co-transcrits, ce qui permet de définir ces ARN messagers comme mosaïques (Courseaux et al., 2003).

Les mécanismes de transposition et duplication des gènes dérivés de *GUSB*, ainsi que leurs structures exactes et leurs potentiels transcriptionnels précis restent à définir. Cette famille de gènes, dont un certain nombre de membres sont apparus très récemment au cours de l'histoire des Primates (Courseaux et al., 2003), constitue un modèle tout à fait approprié à l'étude des premiers temps de la création de nouveaux gènes.

4.3. Objectifs et plan de travail

Je me suis attachée au cours de ma thèse à étudier les processus de l'évolution des génomes, particulièrement des génomes de primates. Mon travail s'est alors organisé en trois temps, le dernier d'entre eux constituant la thématique centrale de ma thèse. Cette étude thématique des gènes chimériques spécifiques des Primates s'est elle-même concentrée sur l'étude de deux exemples majeurs.

4.3.1. Evolution du précurseur peptidique de la MCH chez les Mammifères et les Poissons

L'occasion m'a été donnée d'examiner les séquences du précurseur peptidique de la MCH dans deux lignées de Vertébrés et d'en déduire une caractéristique évolutive imprévue et inédite chez les Mammifères. Ce travail a été intégré dans un article, paru en 2004, relatant une analyse comparative de la structure et de l'activité de la MCH des poissons Téléostéens et des Mammifères.

4.3.2. Mise en perspective des apports des séquences de génomes entiers de Primates

Afin d'aborder ensuite, et de façon plus précise, les processus évolutifs des génomes de Primates, j'ai participé au cours de ma thèse à une étude bibliographique présentant un état des lieux des données de séquences génomiques disponibles pour les Primates et des connaissances pouvant en être déduites. Ce travail de revue a été publié en 2004.

4.3.3. Etude de famille de gènes chimériques spécifiques des Primates

Je me suis enfin intéressée à l'étude de la structure, de l'évolution et de l'expression de deux familles de gènes spécifiques des Primates.

4.3.3.1. Les gènes PMCHL

La première d'entre elles est la famille des gènes *PMCHL*, qui a été étudiée par une approche de comparaison d'orthologues. Les résultats obtenus font l'objet d'un article soumis en décembre 2006.

4.3.3.2. Les gènes GUSL

Les gènes *GUSL*, qui dérivent du gène authentique *GUSB*, ont été identifiés et caractérisés par l'analyse des paralogues humains. Ces travaux sont réunis dans un article en préparation.

RESULTATS

1. Structure et activité de la Melanin Concentrating Hormone des Mammifères et des poissons Téléostéens

J'ai participé, au cours de ma thèse, à une analyse comparative de la structure et de l'activité de la MCH des poissons et des mammifères. L'évolution du précurseur peptidique de la MCH y a été examinée. Cette étude constitue ainsi, pour cet aspect, une première approche de l'évolution des gènes du système à MCH.

1.1. Contexte bibliographique

Le peptide MCH, comme cela a déjà été indiqué, a été initialement découvert par ses propriétés d'agrégation des pigments de la peau de saumon (Kawauchi et al., 1983). Il n'y a cependant pas d'éléments indiquant que la MCH possède cette propriété chez d'autres espèces que les Néoptérygiens, ce qui suggère que cette fonction est apparue et a évolué uniquement chez un ancêtre des poissons Téléostéens et Holostéens (Griffond and Baker, 2002). La MCH

des Mammifères possède une forte similarité de séquence avec celle du saumon (Vaughan et al., 1989). Elle est largement distribuée dans le cerveau et régule un grand nombre de fonctions comportementales, notamment de prise alimentaire et de balance énergétique.

Les efforts initiaux pour caractériser le/les récepteur(s) à la MCH, que ce soit chez les Poissons ou chez les Mammifères, ont pâti du manque de tests de liaison appropriés, ainsi que de la forte hydrophobicité du peptide MCH et de sa haute sensibilité à la dégradation (Schlumberger et al., 2002b).

Le premier récepteur à la MCH, le MCH-R1, a été découvert par une stratégie de pharmacologie inverse (Chambers et al., 1999; Saito et al., 1999). Il s'agit d'un variant d'épissage d'un récepteur couplé à une protéine G (GPCR) nommé SLC1/GPR24 (Lakaye et al., 1998). Les ARNm et protéines MCH-R1 sont exprimés dans le cerveau de rat et de souris, de façon concordante avec les projections des neurones à MCH (Hervieu et al., 2000; Kilduff and de Lecea, 2001). De récentes expériences d'inactivation du gène du récepteur MCH-R1 chez la souris (Chen et al., 2002; Marsh et al., 2002) et d'administration d'antagonistes non-peptidiques bloquant spécifiquement le récepteur MCH-R1 chez le rat (Borowsky et al., 2002; Takekawa et al., 2002) confirment l'implication de ce récepteur dans la régulation de l'homéostasie énergétique. Le couplage de ce récepteur à de multiples protéines G (Gi, Go, Gq) est désormais bien établi dans les modèles de cellules transfectées. L'inhibition subséquente de l'AMP cyclique et l'activation des MAP Kinases ont également été proposées. Un second récepteur à la MCH, nommé MCH-R2, a été identifié chez l'homme par des études *in silico* (Mori et al., 2001). MCH-R2 semble également exister chez les Poissons (Logan et al., 2003), mais est absent chez les Rongeurs (Tan et al., 2002). Le récepteur MCH-R2 ne possède que 38% d'identité de séquence protéique avec MCH-R1 et leurs distributions cérébrales diffèrent partiellement. MCH-R2 ne s'associerait en outre qu'avec la protéine Gαq, ce qui conduit à une augmentation des niveaux de Ca^{2+} intracellulaire. L'importance fonctionnelle de ce deuxième récepteur reste méconnue.

De façon paradoxale, la structure et la signalisation du/des récepteur(s) de Poissons n'ont été que peu étudiées. Les récepteurs orthologues à MCH-R1 et MCH-R2 sont nommés mchr1 et mchr2, respectivement, chez le zebrafish et le *Fugu*, où ils ont été identifiés par exploitation des données de génomes. Très peu d'informations relatives aux voies de signalisation activées par les récepteurs à la MCH de Poissons sont disponibles. L'activation d'une protéine kinase C (Abrao et al., 1991) et l'inhibition de l'activité de l'adénylate cyclase (Oshima et al., 2001; Svensson et al., 1991) sont ponctuellement décrites.

1.2. Principaux résultats

1.2.1. Evolution de la pro-MCH

L'amplification et le séquençage des gènes *MCH* de Rongeurs et de Primates met en évidence une totale conservation de la structure intron/exon.

Au sein des Mammifères, les peptides MCH et NEI sont parfaitement conservés, ce qui suggère une contrainte fonctionnelle importante. De façon plus surprenante, le prodomaine est lui-même hautement conservé. Chez les Poissons téléostéens, la conservation du peptide MCH est également quasi totale, quelques acides aminés des extrémités différant cependant. En revanche, aucune conservation du Mgrp, homologue au NEI des mammifères, n'est observée. Il en va de même pour le propeptide.

Ainsi, bien que possédant des dates de divergence similaires, les phylums des poissons Clupéocéphales et des Mammifères présentent de grandes disparités de conservation du prodomaine du peptide MCH, ce qui suggère une contrainte structurale largement plus importante chez les Mammifères. Ce phénomène peut en outre, d'après nos données, également être généralisé à d'autres neuropeptides. Dans la mesure où aucune différence fondamentale entre poissons et mammifères n'est décrite pour l'adressage et la maturation des peptides en général, il n'existe pas d'explication simple pour une telle différence des taux d'évolution. Cependant, les différences fonctionnelles entre la MCH des Poissons et celle des Mammifères, accompagnant le passage d'un rôle principalement endocrine à un système essentiellement neuronal, pourraient s'accompagner de différences, lignée-spécifiques, de régulation du pro-domaine.

1.2.2. Activité de la MCH chez les Poissons

Un test permettant de mesurer l'activité d'agrégation de pigments de la MCH a été développé par Bruno Cardinaud au laboratoire. Il s'agit de mesurer la densité moyenne (en pixel) d'un patch de peau d'anguille dans différentes conditions (avant et après traitement à la MCH ou d'autres agents).

On observe ainsi une réponse maximale pour 10 nM de MCH et la valeur EC50 est obtenue à 0,5 nM de MCH. L'allure de la courbe dose-réponse exclut en outre, chez l'anguille, l'existence d'un récepteur de basse affinité responsable d'un effet dispersant (cf Figure 3B de (Cardinaud et al., 2004)).

De précédentes études basées sur l'analyse des mélanophores de poisson et l'utilisation d'analogues de la MCH de poisson indiquent que l'activité peptidique dépend essentiellement du noyau cyclique. Chez les mammifères, le rôle crucial de l'Arg^6 (correspondant à l'Arg^4 de la MCH de poisson) a été démontré dans des études de structure-activité et de liaison aux récepteurs MCH-R1 et MCH-R2. Les capacités de la MCH de mammifère, de la MCH de saumon, de la MCH [Phe^{13}, Tyr^{19}] et de la MCH_{6-17} à induire l'agrégation des mélanosomes ont été évaluées dans notre test. Il apparaît à l'issue de ces tests et de l'analyse de leurs résultats qu'il est encore impossible de savoir si le récepteur à la MCH de la peau d'anguille est un homologue fonctionnel de MCH-R1 ou de MCH-R2. La structure minimale pour l'activité de la MCH sur son/ses récepteur(s) est cependant la même chez les poissons et les mammifères (il s'agit de la MCH_{6-17} cyclique), en dépit des centaines de millions d'années d'évolution qui les séparent.

Différentes voies de signalisation sont impliquées dans le mouvement des chromatophores. Les voies de l'AMP cyclique, de l'IP3/diacylglycérol, des MAP kinases et des phosphoinositide 3-kinases semblent cruciales pour le contrôle de ce mouvement de pigments. Nous confirmons que la MCH est capable de réduire les niveaux d'AMPc dans les patchs de peau d'anguille. Cette régulation serait indépendante des protéines de type Gαi canoniques, mais pourrait impliquer l'activation de phosphodiestérases.

1.3. Article publié

Ce travail, présenté page suivante, a été publié dans Peptides en Septembre 2004.

Mon implication personnelle dans ce travail s'est concentrée sur les analyses des séquences des précurseurs peptidiques de la MCH dans les différentes espèces et leur évolution. Enfin, j'ai participé à la rédaction du manuscrit.

Article peptides

2. Données de séquençages de génomes entiers de Primates

Afin de mieux appréhender les connaissances disponibles sur les génomes de primates pour les utiliser de façon pertinente et rigoureuse dans nos études, j'ai participé à la rédaction d'un travail de revue du génome de Chimpanzé commun *Pan troglodytes*. Expériences du laboratoire et données bibliographiques ont ainsi été réunies, détaillant l'état du séquençage du génome de Chimpanzé, mais également les relations phylogénétiques le liant/distançant de l'Homme et les modifications phénotypiques correspondantes.

2.1. Principaux aspects abordés

Le séquençage de génomes de Primates, et notamment celui du Chimpanzé, devrait fournir des données comparables aux séquences humaines. L'émergence de fonctions propres aux Hominidés, telles que les fonctions cognitives les plus élaborées, pourra ainsi être examinée d'un point de vue génétique.

L'analyse bibliographique que nous avons réalisée permet de mettre en évidence la nécessité de pouvoir comparer les séquences et chromosomes homologues. Le moment de la divergence des Hominidés, désormais estimée à moins de 6,2 millions d'années, peut ainsi être précisé. Les contraintes fonctionnelles, portant sur les séquences codant pour une protéine, mais aussi sur les autres types de séquences, ont été passées en revue. D'éventuelles pressions de sélection positive ont été décrites. D'autres contraintes ou accélérations de l'évolution devraient être mises en lumière au cours de l'analyse comparative des génomes de Mammifères et en particulier de Primates.

L'impact des duplications segmentaires a été décrit précédemment. On peut penser que leur implication éventuelle dans la création de gènes spécifiques de l'espèce humaine pourra être déduite de la comparaison des génomes. De même, la connaissance du rôle de ces duplications dans l'établissement de barrières de spéciation en est à ses prémices mais prendra de l'ampleur avec ces nouvelles données.

Les comparaisons de transcriptomes devraient également être facilitées par ces données de séquençage. Un modèle finalement « neutraliste » de l'expression génique relative dans les cerveaux de l'Homme et des autres Primates a ainsi d'ores et déjà pu être proposé.

2.2. Revue publiée

Ce travail de revue à été publié dans le Vivant, journal français en ligne abordant « l'actualité des sciences et [les] débats sur le vivant », en décembre 2004. Il concourt à une meilleure compréhension des génomes de primates et de leur évolution.

Revue le vivant

3. Etudes de familles de gènes chimériques spécifiques des Primates

3.1. Etude des gènes orthologues de la famille *PMCHL*

La première des familles de gènes spécifiques des Primates à avoir retenu notre attention a été la famille des gènes *PMCHL*. Leurs séquences et profils d'expression ont été analysés chez différentes espèces ; les modèles d'évolution ont ainsi été déduits de l'étude de gènes orthologues.

3.1.1. Principaux résultats

3.1.1.1. Analyse structurale des gènes PMCHL

Au cours de ce travail, nous avons établi la structure des gènes *PMCHL1* et *PMCHL2* ainsi que l'organisation régionale de leurs loci sur le chromosome 5. Les limites du duplicon ont été précisées. Il correspond à un fragment de 92 kb comprenant essentiellement la partie 5' du troisième intron de la Brain-Cadhérine, sans qu'aucune séquence particulière ne puisse être identifiée à ses bornes. La structure fine des gènes *PMCHL* a ensuite été analysée. Il est apparu que la séquence Alu présente dans l'exon 2 est de type Alu Sg, contrairement à ce qui était supposé jusqu'alors. Un noyau dense en fragments de séquences répétées a pu être mis en évidence dans le troisième intron, et il a été proposé qu'il contribue à l'instabilité de ces gènes. On observe également l'apparition de mutations spécifiques de lignées particulières.

L'obtention et l'analyse des séquences des gènes *PMCHL* de sept espèces de Primates ont permis de préciser la structure fine de ces gènes et leur évolution. La concomitance de l'insertion de l'élément Alu Sg et de l'évènement de rétroposition a été ainsi documentée : ces deux processus remontent à il y a au plus 35 millions d'années. Des analyses phylogénétiques indiquent plus précisément que la rétrotransposition a eu lieu très rapidement après la divergence des Platyrhiniens et des Catarhiniens, soit il y à 30 à 35 millions d'années.

Un algorithme spécifique a ensuite été développé afin de mesurer la variabilité nucléotidique de séquences non codantes pour des protéines. Ce nouvel outil permet de mettre en évidence des régions particulières du gène *PMCHL1* présentant une hypervariabilité ou une contrainte forte apparentes. Des taux de mutations variables le long des séquences et entre

les différentes espèces ont ainsi été documentés. La majorité des régions concernées contient des insertions/délétions.

3.1.1.2. Expression des gènes PMCHL chez l'Homme et le Macaque crabier

L'étude des transcrits épissés des gènes *PMCHL* humains a permis de les caractériser et d'en estimer l'abondance relative. Un nouvel exon et de nouveaux épissages alternatifs ont alors été mis en évidence.

L'expression des transcrits des gènes *PMCHL* a ensuite été comparée dans les cerveaux de l'Homme et du Macaque. Des messagers non épissés couvrant les exons 1-2 ou 4-5 ont été détectés dans les cortex des deux espèces, tandis que seuls les messagers comprenant les exons 4-5 sont identifiés dans leurs cervelets (cf Figure 5 de « Evolutionnary variations in structure and brain expression of the *PMCHL1* gene in macaque and human », Fleur DARRE-TOULEMONDE, Alain CORINUS, Audrey DELERUE-AUDEGOND, Richard CHRISTEN and Jean-Louis NAHON, soumis à Molecular Biology and Evolution).

Cela suggère que deux unités de transcription distinctes produisent ces deux types de messagers, mais de façon identique chez l'Homme et le Macaque. Ceci renforce le choix du macaque comme modèle *in vivo* de primate non humain pour nos études ultérieures.

Au cours de la recherche des sites d'initiation de la transcription des messagers comprenant les exons 1 et 2, il est apparu que ces sites différaient dans les deux espèces. D'après des expériences de RACE-PCR, réalisées précédemment (Viale et al., 2000), et de RT-PCR, présentées dans l'article, un cap site majeur du gène *PMCHL1* se positionne chez l'Homme près du site d'insertion de la séquence dérivée du gène *MCH*. Chez le Macaque crabier, des ARN messagers comprenant les exons 1 et 2 débutent au moins 2,5 kb (voire jusqu'à 7 kb) en amont du site d'insertion de la séquence rétroposée. Les régions promotrices de l'unité de transcription *PMCHL1* ont donc divergé entre ces deux espèces.

3.1.2. Article soumis

Ce travail a été effectué en collaboration avec Richard Christen (Laboratoire de Biologie Virtuelle, CNRS, Nice), qui a réalisé les analyses phylogénétiques et les mesures de variabilité nucléotidique. Il est actuellement soumis au journal Molecular Biology and Evolution.

J'ai personnellement participé au développement de ce projet ainsi qu'à la conception et à la rédaction du manuscrit. J'ai réalisé l'ensemble des expériences nécessaires à l'étude structurale des gènes *PMCHL*, à l'exception des analyses réalisées en collaboration, pour

lesquelles j'ai cependant fourni l'ensemble des données de séquence. Enfin, j'ai contribué aux expériences d'analyse des profils d'expression des gènes *PMCHL*. J'en ai réalisé la partie qualitative (portant sur les messagers épissés) et ai défini les conditions de l'étude quantitative (portant sur les ARNm non épissés), que j'ai ensuite partiellement menée.

3.1.3. Données complémentaires

3.1.3.1.Tentatives de clonage du gène PMCHL1 humain

3.1.3.1.1. Objectifs

J'ai initié un projet de souris Knock-in *PMCHL1*, dont le génome aurait été ainsi enrichi d'un gène spécifique des Primates. Il s'agissait d'insérer, de façon ciblée (au locus hprt), dans le génome murin, le gène *PMCHL1* humain comprenant une séquence IRES-eGFP (pour Internal Ribosome Entry Site – ectopic Green Fluorescent Protein). La souris ainsi produite exprimerait la protéine GFP dans les cellules possédant le transcrit PMCHL1. Des analyses transcriptionnelle et traductionnelle pourraient alors être menées chez la Souris, ainsi que d'éventuelles études comportementales. En outre, la présence d'éléments de type loxP aux bornes de l'IRES-eGFP garantirait la possibilité d'exciser cet élément, exogène à *PMCHL1,* par croisement avec une souris exprimant la recombinase Cre.

3.1.3.1.2. Stratégies et Résultats

Il s'est d'abord agit de produire le fragment d'ADN comportant le gène *PMCHL1* dans lequel l'IRES-eGFP soit inséré. Ce fragment de grande taille (environ 20 kb initialement) devait ensuite être expédié à une société privée (Nucléis) réalisant la transgénèse ciblée, par recombinaison homologue au site *hprt*.

Au gré des difficultés, différentes approches ont été tentées pour obtenir ce fragment :

- La première approche consiste à insérer l'IRES-eGFP (de façon ciblée) dans un chromosome artificiel bactérien (BAC) contenant *PMCHL1*, et ce par recombinaison homologue en bactérie DY 380 (Copeland et al., 2001). Le fragment d'intérêt sera ensuite excisé par l'action d'enzymes de restriction ou par PCR, si les enzymes compatibles font défaut.

  Les bactéries comportant le BAC (incluant le gène *PMCHL1*) ont d'abord été cultivées, et le BAC en a été extrait en quantité par de multiples MiniPrep. Les bactéries recombinantes DY380 ont alors été rendues compétentes et ont été transformées avec le BAC par électroporation. En parallèle, un fragment loxP-IRES-eGFP-loxP a été amplifié par PCR à partir d'un plasmide contenant l'IRES-eGFP et d'amorces flanquées des séquences loxP. Ce produit a ensuite été lui-même amplifié avec des oligonucléotides présentant des extrémités homologues au site d'insertion choisi au sein de l'exon 2' de *PMCHL1*. Après avoir été rendues compétentes, les bactéries DY380 contenant le BAC ont été rendues recombinantes par un choc thermique, et le fragment loxP-IRES-eGFP-

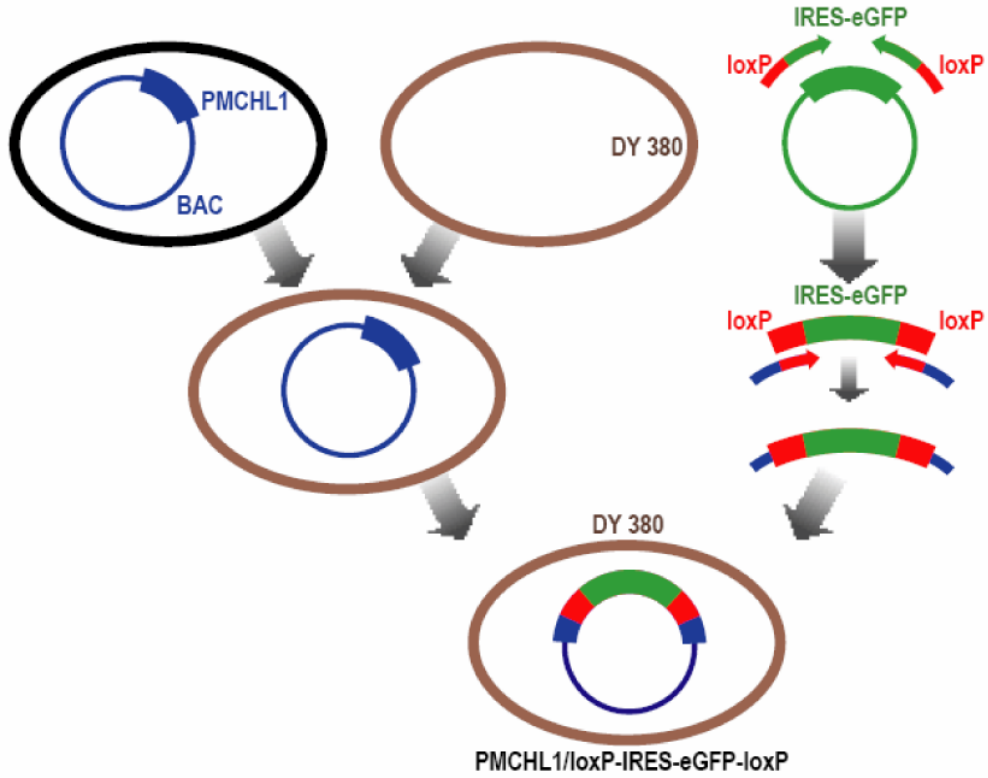


Figure 3.1.3.1.2.a **Protocole de clonage de *PMCHL1* –loxP-IRES-eGFP-IRES-loxP par recombinaison homologue**

Résumé du protocole mis en place pour l'obtention d'un transgène *PMCHL1* comportant un IRES-eGFP excisable par la Cre-recombinase. Il s'agit d'insérer le BAC contenant *PMCHL1* dans une souche bactérienne favorisant la recombinaison homologue, puis de la transformer avec un fragment recombinant portant l'IRES-eGFP.

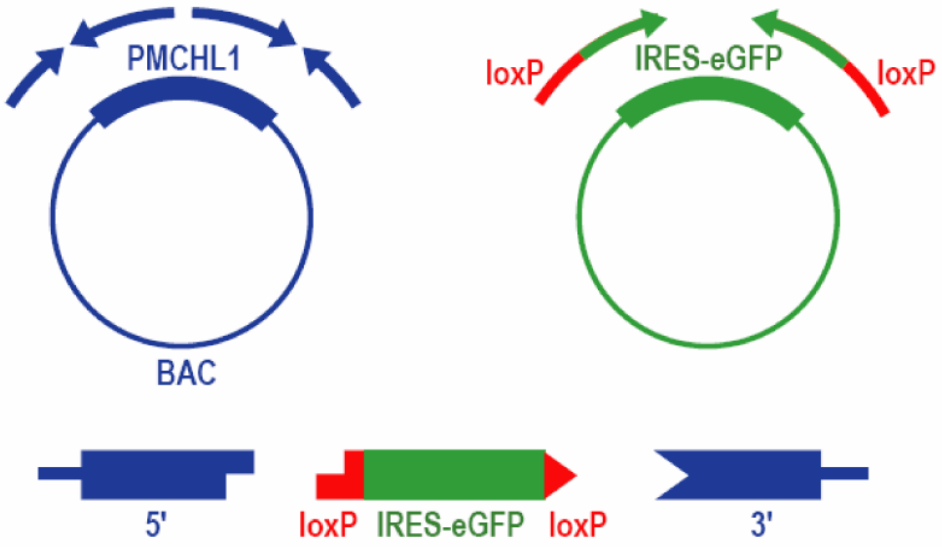


Figure 3.1.3.1.2.b **Protocole de construction de *PMCHL1* –loxP-IRES-eGFP-IRES-loxP par PCR**

Résumé du deuxième protocole mis en place pour l'obtention d'un transgène *PMCHL1* comportant un IRES-eGFP excisable par la Cre-recombinase. Il s'agit d'amplifier trois fragments par PCR, de les soumettre à une digestion enzymatique puis à une ligation.

loxP a été intégré par recombinaison homologue après électroporation (voir Figure 3.1.3.1.2.a). Après étalement, les clones ont été criblés par hybridation radioactive à partir de répliques des boîtes de bactéries. Les clones semblant positifs ont ensuite été remis en culture, l'ADN en a été extrait par Mini Prep et leurs séquences ont été analysées par de multiples PCR, couvrant différents fragments du gène *PMCHL1* mais également le site supposé d'insertion de l'IRES-eGFP. Il est malheureusement apparu que dans tous les cas, les propriétés recombinantes des bactéries DY380 avaient favorisé des évènements de recombinaison non homologue tels que le BAC était totalement remanié. Une autre approche, n'impliquant plus cette souche de bactérie et ne reposant plus sur un évènement de recombinaison homologue, a donc du être envisagée.

- La stratégie alors choisie a été celle de la ligation de produits de PCR afin d'obtenir le fragment attendu. Trois fragments ont donc été d'abord amplifiés par PCR à partir du BAC, qui correspondent respectivement à la région 5' du gène *PMCHL1* (de 2 à 10 kb en amont du site d'insertion de la séquence rétroposée dérivée de *MCH* à l'exon 2'), à l'IRES-eGFP flanqué de séquences loxP, et à la région 3' du gène (de l'exon 2' jusqu'à 1 à 7kb en aval de la fin de l'exon 5). Les amorces utilisées présentaient des extrémités comportant des sites de restriction, de sorte que chaque fragment a ensuite été digéré par des enzymes de restriction telles que le seul produit pouvant apparaître par ligation était le produit attendu (gène PMCHL1 incluant le loxP-IRES-eGFP-loxP). La ligation simultanée des trois produits s'avérant trop peu efficace, voire inefficace, il a été choisi de procéder à une première étape de digestion et de ligation de l'extrémité 3' du gène *PMCHL1* avec l'IRES-eGFP (voir Figure 3.1.3.1.2.b). Le produit obtenu devait ensuite subir une nouvelle étape de digestion enzymatique, de même que le fragment comprenant l'extrémité 5' de *PMCHL1* et une deuxième ligation aurait été réalisée. Mais le produit obtenu par ligation l'étant en quantité minime (et d'autant plus après l'indispensable étape de purification), l'insertion dans un vecteur de clonage (par TA cloning dans le vecteur TOPO XL adapté à l'insertion de produits de PCR de grande taille) et l'électroporation en bactéries afin de produire de grandes quantités de plasmides et d'en extraire le fragment attendu était incontournable. Les bactéries Top10 puis SURE2 (réputées pour leur très faible taux de recombinaison) ont été utilisées, mais dans les deux cas, le fragment de gène *PMCHL1* a été largement remanié, comme l'ont indiqué les expériences de contrôle par PCR et séquençage. Il est ainsi apparu que le gène *PMCHL1* était hautement instable et remanié en bactéries, et ce quelle que soit la souche bactérienne utilisée.

- Une nouvelle approche a alors été définie, qui évite ce passage en bactérie. Elle impose de renoncer à l'insertion d'un IRES-eGFP, choix qui a été conforté par la mise en évidence de l'abondance de transcrits antisens de *PMCHL1*, qui n'auraient donc pas conduit à la production de la GFP. Il s'agit donc d'amplifier directement par PCR le gène *PMCHL1* (de 3kb en amont de l'exon 1 à la fin de l'exon 5) à partir du BAC et d'en obtenir de grandes quantités (1 à 15 µg sont nécessaires pour la transgénèse ciblée). Ce fragment de PCR, un fragment obtenu par digestion du BAC (ces deux fragments ayant été préalablement manipulés de sorte à présenter des extrémités franches) et le BAC lui-même ont été fournis à la société Nucleis pour la réalisation de la transgénèse ciblée. Cependant, en dépit de nombreuses tentatives (certaines en collaboration avec d'autres laboratoires académiques), le remaniement des séquences n'a pu être évité, le passage en bactérie étant incontournable dans les phases ultérieures du projet.

Nous avons donc du nous résoudre à l'abandon de ce projet et de l'idée de cloner le gène *PMCHL1* en bactérie.

3.1.3.1.3. Discussion

L'ensemble des approches utilisées a donc démontré l'impossibilité d'amplifier ou de modifier les séquences *PMCHL* en bactéries. Les remaniements semblent incontournables. Deux questions se posent alors :

- Comment ces gènes ont-il pu être séquencés à partir de chromosomes artificiels bactériens dans le cadre du programme de séquençage du génome humain ?
- A quoi tiennent ces remaniements ?

Cette question a été évoquée dans l'article soumis à Molecular Biology and Evolution. Nous proposons que la haute densité en fragments de séquences répétées des gènes *PMCHL* facilite les évènements de recombinaison et rend ainsi cette séquence instable.

3.1.3.2. Expression de PMCHL dans un modèle de macaques « parkinsoniens »

3.1.3.2.1. Objectifs

Les macaques constituent, comme nous l'avons montré précédemment (dans l'article soumis à Molecular Biology and Evolution) un modèle pertinent et disponible pour l'étude des gènes spécifiques des primates en général, et de *PMCHL1* en particulier, dans la mesure où des expériences de distribution des ARNm de ce gène chez l'Homme et le Macaque révèlent une certaine conservation du profil d'expression.

Parallèlement, le développement d'un modèle de Macaques crabiers (*Macaca fascicularis*) traités de façon chronique au MPTP (en collaboration avec E. Bézard, CNRS, UMR 5543) et développant tous les signes de la maladie de Parkinson, constitue une opportunité formidable d'étudier les stades précoces de la maladie. Chez l'Homme, la pathologie est dépistée en phase terminale, lorsque 80% des neurones dopaminergiques sont déjà détruits ; la découverte de marqueurs précoces de la maladie constituerait donc une avancée clinique substantielle. Les gènes spécifiques des Primates constituent de bons gènes candidats pour la suceptibilité, l'initiation, le développement, la régulation ou l'indication de pathologies spécifiques de ces espèces.

Nous nous sommes donc intéressés à l'expression du gène *PMCHL1* chez le macaque traité au MPTP aux différents stades (asymptomatique ou symptomatique) de la maladie neurodégénérative de Parkinson. J'ai personnellement débuté ce projet, ai mis au point les premières conditions expérimentales (PCR et RT spécifiques) et ai obtenu des données préliminaires significatives dans le cerveau des Macaques traités et non traités. Ce travail a été, depuis, poursuivi au laboratoire. Je présenterai deux séries d'expériences attestant de la complexité des régulations du gène *PMCHL1*.

3.1.3.2.2. Matériels et méthodes

- Traitement chronique au MPTP et tissus disponibles

Nous disposons, dans le cadre d'une collaboration avec le Dr Erwan Bezard (Bordeaux) de différents tissus provenant de macaques (*Macaca fascicularis*) traités de façon chronique au MPTP (Bezard et al., 1997). Les individus traités pendant 6 et 12 jours sont en phase présymptomatique du développement de la maladie de Parkinson. Les macaques traités pendant 15 jours et prélevés 10 jours après la fin du traitement présentent tous les symptomes de la pathologie et en sont sévèrement atteints.

- Extraction des ARN

Les ARN ont été extraits à partir de ces tissus, conformément à la méthode standard guanidium phénol chloroforme (Chomczynski) ou en utilisant l'appareil fastprep (FP220A Thermo Instrument, QBiogene). Ils ont ensuite été traités à la RQ1 DNAse selon le protocole du fabricant (Promega), afin d'éliminer d'éventuelles contaminations par de l'ADN génomique.

Oligo	orientation	séquence	Température d'hybridation
PMCHL1A	F	CAATGGGATTATGCTGTCACAA	57°C
PMCHL4C	F	TGAGAATGTAAAGAGACCACCTT	58°C
P1mac	F	CCTCATTGTCTAAGGAGGAAATGCTAGGG	60°C
hMCH7	R	TTGGTCTAGAGTGTATG	45°C
PMCHL2B	R	TGATACATCTTAACACTACTTTCTTTT	56°C
PMCHL2D	R	TTTACTGGTTGACTGCACTCTCCTT	62°C
PMCHL5C	R	GATGGAGGTAAACCAAGGAGG	60°C

Table 3.1.3.2.2 **Oligonucléotides utilisés en RT-PCR**

Lorsque les températures d'hybridation diffèrent entre deux oligonucléotides utilisés pour la même amplification, l'hybridation est effectuée à la plus faible d'entre elles.

- Reverse-transcription

Les ADN complémentaires ont été synthétisés au cours de la réverse-transcription de 2 à 3 µg d'ARN total. La reverse-transcriptase Superscript II a été utilisée selon le protocole du fabricant. Les amorces nucléotidiques sont des 15-mer d'oligo-dT ou des amorces spécifiques permettant de distinguer les messagers sens et antisens. Les séquences de ces oligonucléotides figurent dans la Table 3.1.3.2.2.

Afin d'être à même de révéler une contamination par l'ADN génomique, la reverse transcription est systématiquement également réalisée en absence d'enzyme.

- PCR

Les oligonucléotides ont été commadés à Eurogentec (Belgique). Leurs séquences, Tm et les conditions de PCR figurent dans la Table 3.1.3.2.2.

2 à 4 µL de reverse-transcription sont utilisés pour la réaction de PCR selon le protocole du fabricant (Hotmaster, Eppendorf), dans un volume total de 25 µL. Trente-cinq cycles de PCR sont effectués, de la façon suivante : 30 s à 96°C (dénaturation), 30 s à une température variable dépendant du Tm des olignucléotides (hybridation), 30 s à 2 min à 72°C (extension). Une étape finale de 7 min d'extension à 72°C termine la réaction.

- Southern blot

Les produits de PCR ont été déposés sur gel d'agarose (1%) contenant du bromure d'éthidium (BEt), ont subi une électrophorèse et ont été visualisés sous les UV. Le gel et les échantillons déposés sont alors traités et transferrés sur une membrane de cellulose (Biodyne A, Polylabo), selon un protocole décrit précédemment (Viale et al., 1998b). Des fragments spécifiques marqués au ^{32}P sont préparés selon le protocole du fabricant (Prime-a-gene labelling system, Promega) et appliqués à la membrane (5.10^{-5} dpm.mL^{-1}). Les conditions de pré-hybridation, d'hybridation et de lavage ont été décrites par ailleurs (Pedeutour et al., 1994). L'analyse densitométrique est ensuite effectuée à l'aide du phosphoimager Fujifilm (BAS-1500).

3.1.3.2.3. Résultats

L'expression des ARNm de *PMCHL1* a été analysée par RT-PCR sur ARN extraits de tissus de macaques traités au MPTP. Des expériences préliminaires, réalisées par Alexandra Cervantes, Audrey Delerue et moi-même au laboratoire, ont permis de définir les couples

d'amorces. Les résultats obtenus par Alexandra Cervantes, sont présentés dans la Figure 3.1.3.2.3.

Les messagers contenant les exons 1 et 2 (couple PMCHL1A/hMCH7) sont exprimés dans le cortex préfrontal de macaque quel que soit le traitement administré (aucun, injection de saline, ou divers traitements au MPTP). Ils sont en revanche absents, dans nos expériences, de l'aire V1 de macaques sains, et présents lorsque les animaux sont traités. Les macaques ayant subi des injection de MPTP pendant 12 jours présentent cependant un moindre niveau d'expression des messagers 1-2 de *PMCHL1*, ce qui est mis en évidence par un marquage plus faible, voire absent, selon les échantillons, en Southern Blot (cf Figure 3.1.3.2.3). Les messagers détectés sont sytématiquement non épissés, et comportent donc l'intron A en plus des exons 1 et 2.

Les messagers contenant les exons 4 et 5 (oligonucléotides PMCHL4C/PMCHL5C) sont présents dans toutes les conditions dans l'aire V1 des macaques. Ils ne sont en revanche, et de façon notoire, détectés dans le cortex préfrontal que chez les individus non traités au MPTP (contrôles ou salins). Une faible expression est également détectée dans cette aire cérébrale chez deux des trois macaques traités pendant 12 jours au MPTP (cf Figure 3.1.3.2.3). La taille des fragments amplifiés en PCR est compatible avec des messagers non épissés, comprenant les exons 4 et 5, mais également l'intron D.

Les études précédentes ayant mis en évidence l'existence de messagers PMCHL1 sens et antisens chez l'Homme, nous avons examiné, dans le cortex préfrontal de macaques, l'orientation des messagers comportant les exons du gène *PMCHL1* détectés précédemment. J'ai, pour cela, réalisé des expériences de reverse-transcription spécifique, suivies de PCR et Southern Blot.

Aucun messager sens (RT spécifique avec l'ologonucléotide PMCHL2B dans l'exon 2') n'est détecté dans nos conditions (trois couples de PCR différents) et dans aucun tissu (données non présentées).

Une RT spécifique des antisens a été effectuée avec l'oligonucléotide P1mac (situé en amont du site d'insertion du rétroposon *MCH*) sur les ARN extraits de cortex préfrontal de macaque. Elle permet de mettre en évidence des messagers antisens non épissés comportant les exons 1 et 2 (PCR PMCHL1A/hMCH7) dans des individus traités au MPTP pendant 6 jours (2 individus sur 5), des individus traités 12 jours (2/5) et des individus traités 15 jours et sacrifiés 10 jours après la fin du traitement (4/5). Ces ARN messagers ne sont en revanche pas détectés chez les individus non traités au MPTP (contrôles et salins) (données non présentées).

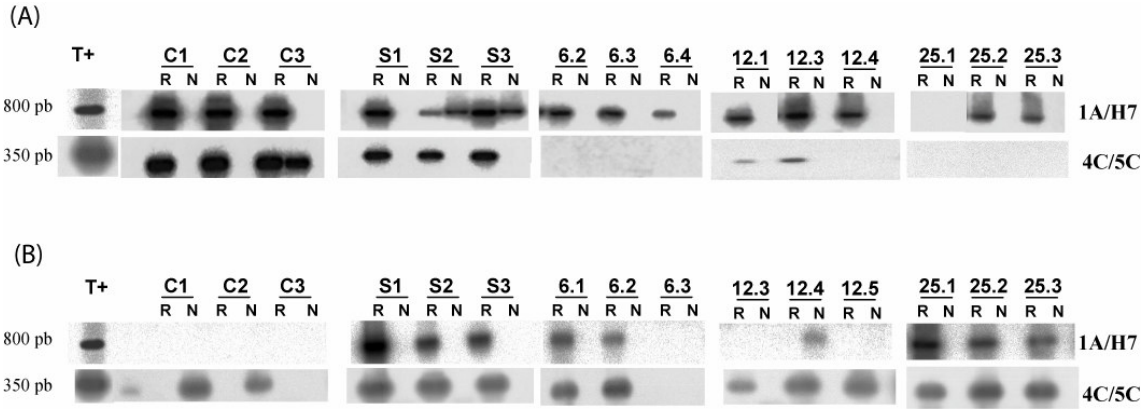


Figure 3.1.3.2.3 **Expression de messagers PMCHL1 chez des macaques traités au MPTP**

L'expression des messagers PMCHL1 est analysée par RT-PCR sur ARN extraits de cortex préfrontal ((**A**)) ou d'aire visuelle V1 ((**B**)) de macaques contrôles (C), traités par une solution saline (S) ou traités au MPTP pendant 6 ou 12 jours. Enfin, les derniers macaques sont traités au MPTP pendant 15 jours et sacrifiés au bout de 25 jours, alors qu'ils présentent tous les traits cliniques de la maladie de Parkinson. L'expression est révélée en Southern Blot. Les couples d'amplification et la taille du produit obtenu figurent à droite et à gauche des blots, respectivement. R : ARN reverse transcrits, N : ARN non reverse transcrits.

3.1.3.2.4. Discussion

Ces résultats très préliminaires, tant du fait du faible nombre d'individus et d'aires étudiés que du choix initial de couples de PCR, mettent en évidence une expression du gène *PMCHL1* dans le cortex préfrontal (PFC) et l'aire visuelle V1 de macaques traités, ou non, au MPTP. Cette expression apparaît en outre être régulée, dans la mesure où nous observons des différences selon les aires analysées, les traitements administrés et les fragments amplifiés. L'analyse des messagers comprenant respectivement les exons 1 et 2, d'une part, et 4 et 5, d'autre part, est particulièrement intéressante. Ces messagers présentent une tissus-sélectivité, étant exprimés différentiellement dans le cortex préfrontal et dans l'aire visuelle V1. L'expression des messagers comprenant les exons 4 et 5 semble en outre être affectée par le traitement au MPTP, puisqu'elle diminue fortement dans le cortex préfrontal des individus traités au MPTP, et ce dès 6 jours. Les macaques crabiers (*Macaca fascicularis*) ayant subi ces 6 jours d'injection sont en phase présymptomatique du développement de la maladie de Parkinson (Bezard et al., 1997). On peut dès lors spéculer que la diminution de l'expression des messagers 1-2 et surtout 4-5 constitue l'un des marqueurs précoces du développement de la maladie. Cependant, les causes et conséquences de ce défaut localisé d'expression sont absolument inconnues. De plus, la régulation de l'expression, des messagers 1-2 en particulier, est similaire chez les macaques traités au MPTP et injectés de solution saline, ce qui suggère un effet important du stress. Enfin, ces nouvelles données ne représentent malheureusement pas une piste vers une avancée clinique, dans la mesure ou ce marqueur putatif ne peut-être détecté qu'à partir de biopsies tissulaires.

Les données d'expression des messagers sens et antisens représentent elles aussi une énigme. L'ensemble de ces données ne permet en effet pas d'expliquer le signal obtenu avec les ARN polyadénylés totaux (reverse-transcrits avec un oligo dT) dans le cortex préfrontal de macaque. Il faut également signaler que les données extraites des expériences de RT spécifique ne sont pas non plus toujours compatibles avec les expériences d'hybridation in situ menées sur tranches de cerveau de macaque par Christine Payré et Alain Corinus au laboratoire (données non présentées). L'incohérence de ces résultats tient probablement à des limitations techniques compromettant la détection des ARN messagers dans l'une ou l'autre des conditions. De manière générale, les messagers du gène *PMCHL1* sont en effet faiblement exprimés et de faibles variations d'efficacité suffisent peut-être à se situer sous le seuil de détection. D'autres expériences devront donc être menées afin d'élucider ces différences de résultats. Il est cependant intéressant de noter que dans nos conditions, les messagers

PMCHL1 les plus aisément détectés dans le cortex préfrontal de macaques traités au MPTP sont synthétisés en antisens de la structure classiquement proposée pour le gène. Ceci est à mettre en regard du grand nombre de messagers antisens identifiés lors d'une recherche systématique de messagers humains. Il est ainsi proposé que la régulation de l'expression des gènes par un messager antisens pourrait être un phénomène commun dans les cellules de mammifères (Lehner et al., 2002). Comme cela a déjà été suggéré pour les messagers PMCHL humains, une régulation complexe du système *MCH/AROM/PMCHL* par des ARNm partiellement complémentaires serait susceptible d'intervenir chez le macaque.

Ces données préliminaires relatives à l'expression d'un gène chimérique spécifique des Primates confortent l'intérêt du modèle des macaques *Macaca fascicularis* traités au MPTP pour l'étude d'une pathologie neurodégénérative spécifique des Primates.

3.2. Etude des gènes paralogues de la famille *GUSL*

3.2.1. Objectifs

La deuxième famille de gènes spécifiques des primates à avoir retenu l'attention du laboratoire a été la famille *GUSL* (*GUS-like*). Seules les données de FISH et de séquences de BAC (chromosome artificiel bactérien) non assemblés étaient alors disponibles (Courseaux et al., 2003) (voir 4.2.2 de l'Introduction).

Les données préalablement établies d'hybridation *in situ* fluorescente sur chromosomes métaphasiques de diverses espèces de Primates indiquaient que des séquences dérivées des exons 5, 9 et 10 de la β-Glucuronidase étaient présentes en de nombreux exemplaires dans leurs génomes, et en particulier dans le génome humain (Courseaux et al., 2003). Le seul génome de Primates séquencé, assemblé et disponible dans son intégralité étant le génome humain, nous nous sommes concentrés sur l'analyse de ces éléments dans l'espèce humaine.

Il s'est donc agit, pour ma part, de caractériser précisément l'étendue et la structure de la famille des gènes *GUSL* à partir de données de séquençage du génome entier humain. Les propriétés transcriptionnelles des paralogues *GUSL* ont également été examinées, ainsi que leur éventuelle spécificité, développementale ou tissulaire, d'expression.

3.2.2. Matériels et méthodes

- Tissus

Différents tissus d'individus adultes humains nous ont été fournis par le National Research Specimen Bank (Los Angeles, Californie). Les tissus fœtaux et de nouveaux-nés proviennent de la faculté de médecine Laennec à Lyon. Les références, âge, sexe, cause du décès et délai postmortem de prélèvement ont été précisés par ailleurs (Viale et al., 2000). Les tissus de macaques adultes ont été obtenus dans le cadre d'une collaboration avec le Dr E. Bézard (Biothèque primates, Valbonne & Bordeaux).

- Préparation des ARN, reverse-transcription et PCR

Les protocoles utilisés sont ceux décrits précédemment (cf 3.1.3.2.2). Les oligonucléotides servant aux amplifications et leurs caractéristiques figurent dans la Table 3.2.2.

- Séquençage

Les séquences nucléotidiques de fragments amplifiés par PCR ont été déterminées en utilisant la polymérase Ampli Taq FS et le kit de séquençage Big Dye Terminator (Applera), ainsi qu'un séquenceur Perkin Elmer ABI PRISM 377.

- Alignements de séquences

Les séquences issues des génomes entiers de l'Homme et du Chimpanzé (sur www.ensembl.org) et les fragments séquencés ont été d'abord alignés en utilisant ClustalW. Les alignements sont ensuite vérifiés et éventuellement modifiés manuellement dans SEAVIEW (Galtier et al., 1996).

- BLAST

Les recherches de séquences par similarité afin d'identifier les gènes paralogues ont d'abord été réalisées par BLAST (Bedell et al., 2003) sur génome entier (via le serveur Ensembl), selon la même méthode que Crosier et al. dans l'étude des paralogues *KIAA0187* (Crosier et al., 2002).

Oligo	orientation	séquence	Température d'hybridation
5F	F	GTACCTGATGCACGAACACCCCG	69°C
9F	F	TGACCAACTCCACCTACGCAG	61°C
hSMA4F1 (X1Δ)	F	CTTTGGTGTAGTCCTGAAGCTTCC	64°C
hSMA3F1 (Rp1)	F	GGTGGGTACTTGGAGAACTTACTACG	62°C
10R	R	AGTAGCTGTTCACACGGATCA	57°C

Table 3.2.2 **Oligonucléotides utilisés en RT-PCR**

Lorsque les températures d'hybridation diffèrent entre deux oligonucléotides utilisés pour la même amplification, l'hybridation est effectuée à la plus faible d'entre elles.

Elles ont également été menées par Nicolas Pantel et Richard Christen avec le logiciel BLAT, et seules les séquences répondant à des contraintes de similarité et de colocalisation de plusieurs exons ont été extraites.

- WLAST

Le logiciel WLAST, développé au laboratoire de biologie virtuelle à Nice (Dr R. Christen), permet de comparer de très longues séquences. L'opération se déroule en deux temps : d'abord, la séquence choisie comme référence est divisée en plusieurs zones de longueur égale, chacune se voyant attribuer une couleur différente. Dans un deuxième temps, les mots de 11 nucléotides de la séquence de référence sont extraits. A chaque mot est affectée la couleur de la zone dans laquelle il se trouve. La présence de chacun de ces mots est successivement recherchée dans chacune des autres séquences, et toutes les positions où se retrouve exactement ce mot sont alors peintes de la couleur qui lui a été attribuée. On peut considérer que WLAST est une extension graphique de BLAT dans laquelle la partie recherche de séquence similaire est traitée par la carte graphique. Le résultat est une série de bandes de couleur sur chaque séquence, indiquant les zones identiques entre elles. Il permet de repérer facilement les délétions, et les duplications.

- FISH numérique

Un outil de visualisation des loci des séquences identifiées a été développé au laboratoire de Biologie Virtuelle par Richard Christen. Il permet de placer les séquences sur un schéma des chromosomes.

- Etudes phylogénétiques

Les dendogrammes phylogénétiques ont été reconstruits selon trois méthodes différentes : le Neighbour Joining (BIONJ), le maximum de vraissemblance (ML), et le maximum de parcimonie (MP). Pour l'analyse de Neighbour Joining (NJ), une matrice de distance est calculée avec la correction de Kimura à deux paramètres. Les valeurs de bootstraps sont obtenues en réalisant 1000 répétitions, le programme BIONJ et la correction de Kimura à deux paramètres. BIONJ est conforme à la publication *princeps* de Gascuel (Gascuel, 1997), ML et MP appartiennent au logiciel PHYLIP (Phylogeny Inference Package, version 3.573c, distribué par J. Felsenstein, Seattle, USA). Les dendogrammes phylogénétiques ont été mis en forme en utilisant NJPLOT (Perrière & Gouy, 1996).

- Calcul de Ka/Ks

Les calculs des rapports du nombre de substitutions non synonymes et synonymes ont été effectués en ligne sur http://www.bioinfo.no/tools/docs/kaksdoc à partir des séquences alignées du domaine codant et d'un arbre phylogénétique obtenu pour ces séquences, au format Newick.

3.2.3. Résultats

3.2.3.1. Structure et mise en place des gènes dérivés de la β-Glucuronidase

3.2.3.1.1. Les gènes *GUSL* et leur structure

Chacun des douze exons du gène *GUSB* a été recherché par similarité de séquence (en utilisant le logiciel BLAT) dans le génome humain, et sa position, ainsi que son orientation, ont été repérées. Après optimisation des paramètres par Nicolas Pantel et Richard Christen au laboratoire de Biologie Virtuelle, à Nice, 111 séquences humaines ont été identifiées, qui sont distribuées dans 24 structures comportant au moins trois unités en tandem. Un caryotype numérique illustrant ces positions et orientations a ensuite été généré par Nicolas Pantel. Il apparaît très clairement (cf Figure 3.2.3.1.1.a) que ces exons sont répandus dans le génome et se trouvent associés et regroupés de manière plus prononcée sur les chromosomes 5, 6, 7 et 22 humains. Tous les chromosomes, à l'exception des chromosomes 8, 16, 21 et Y, présentent au moins une copie d'un des exons de la *β-Glucuronidase*. Une vue rapprochée (cf Figure 3.2.3.1.1.b) des chromosomes 5, 6, 7 et 22 favorise une identification plus précise de ces duplications. Le chromosome 5 comporte l'essentiel des duplications, particulièrement présentes en 5q13.2. Les exons 5, 6, 9, 10 et 11 de *GUSB* y sont nombreux. Le chromosome 6 comprend deux duplications des exons 5, 6, 9, 10, 11. La copie ancestrale sur le chromosome 7 est identifiée par l'alignement des douze exons recherchés en 7q11.21. D'autres duplications sont identifiées sur ce chromosome. Enfin, le chromosome 22 présente une duplication des exons 1 à 4 de *GUSB*.

L'ensemble de ces éléments dupliqués peuvent donc être divisés en deux groupes : le groupe A regroupe les séquences comprenant les exons 5 à 11, le groupe B est composé des séquences portant les exons 1 à 4. Les séquences similaires à l'exon 12 de *GUSB* sont systématiquement isolées et n'entrent donc pas dans le cadre de notre étude.

Afin de caractériser précisément la structure des gènes comprenant ces exons dérivés de la *β-Glucuronidase* et localisés sur les chromosomes 5, 6, 7 et 22, une analyse manuelle

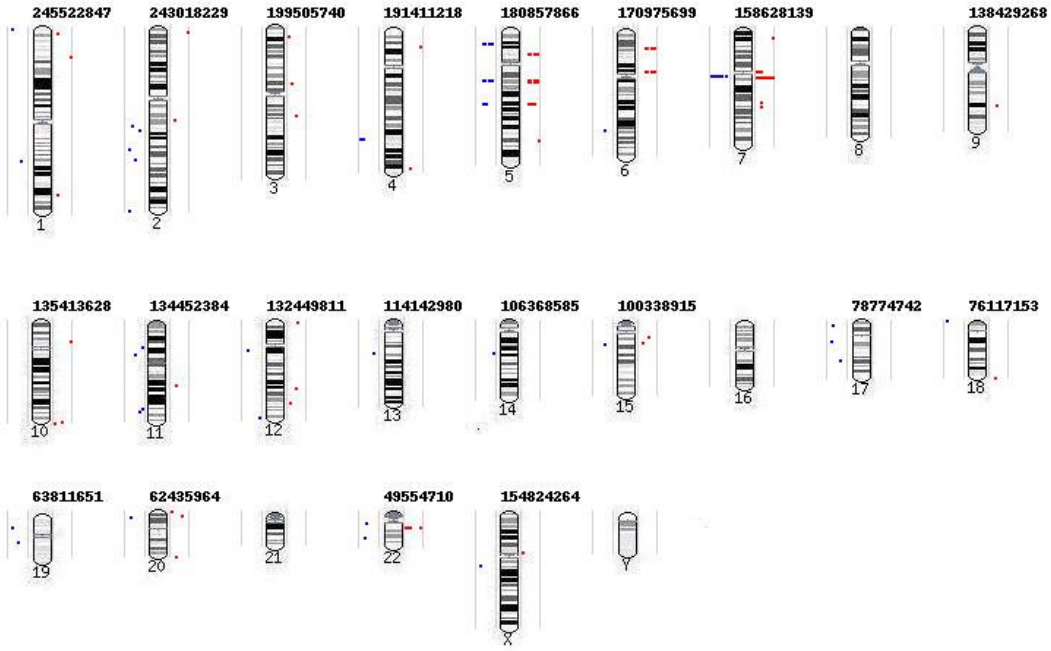


Figure 3.2.3.1.1.a **FISH sur caryotype numérique**

Les 24 chromosomes d'un haplotype humain sont figurés, leur taille (en paire de bases) est indiquée au dessus de chacun d'entre eux. A gauche et à droite de chaque chromosome, en bleu et en rouge, sont représentées les différentes copies de chaque exon de *GUSB*, orientées en antisens et en sens, respectivement. Au sein de chaque colonne de part et d'autre de chaque chromosome se trouvent douze colonnes virtuelles représentant chacun des exons de *GUSB*. Ainsi, un point bleu à l'extrême gauche représente une copie antisens de l'exon 1 de *GUSB* ; un point rouge à l'extrême droite, une copie sens de l'exon 12 du gène. Figure réalisée par Nicolas Pantel au cours de son M1 au laboratoire de biologie virtuelle à Nice.

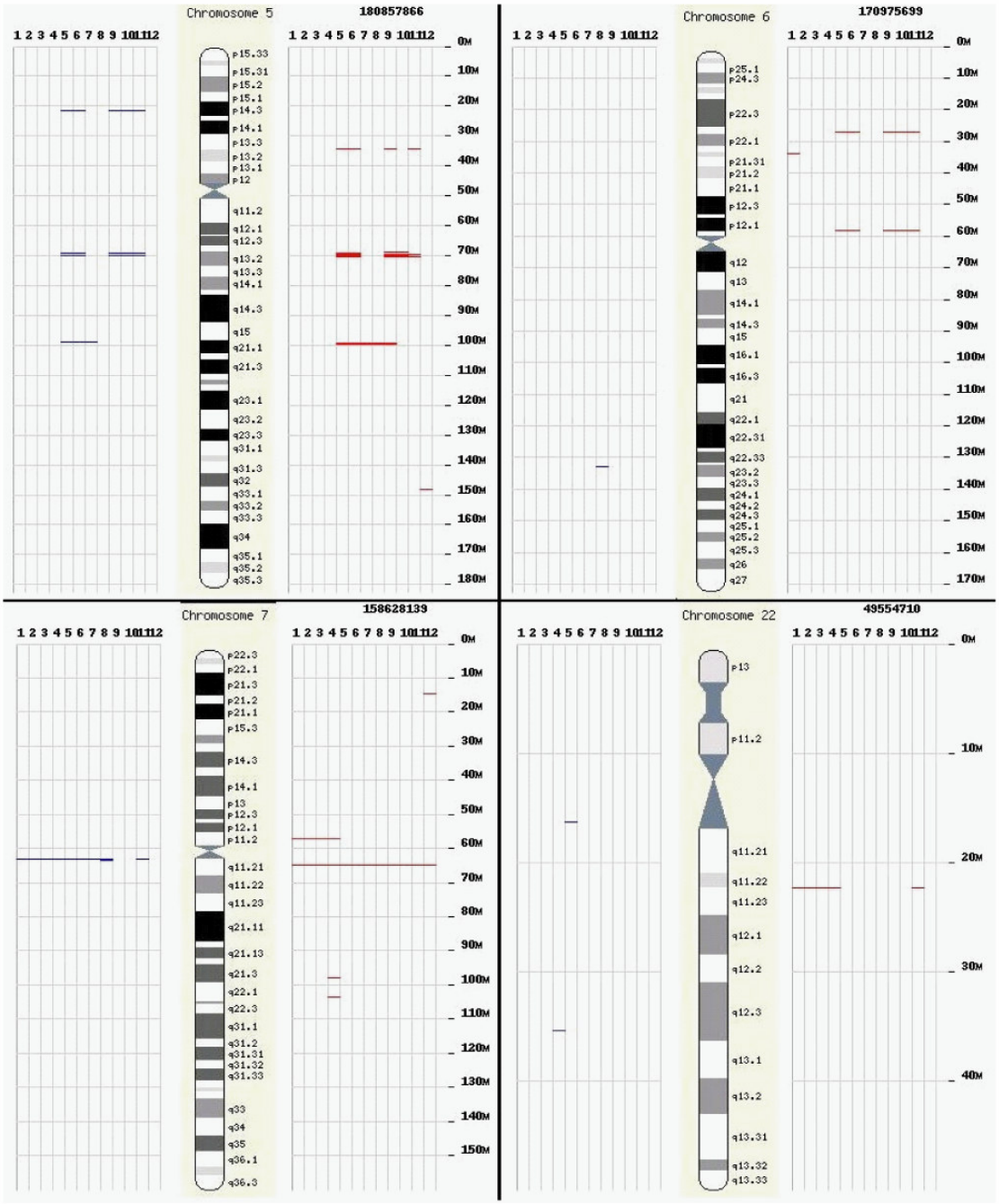


Figure 3.2.3.1.1.b **FISH sur caryotype numérique : zoom sur les chromosomes 5, 6, 7 et 22**

Figure réalisée par Nicolas Pantel au cours de son M1 au laboratoire de biologie virtuelle à Nice.

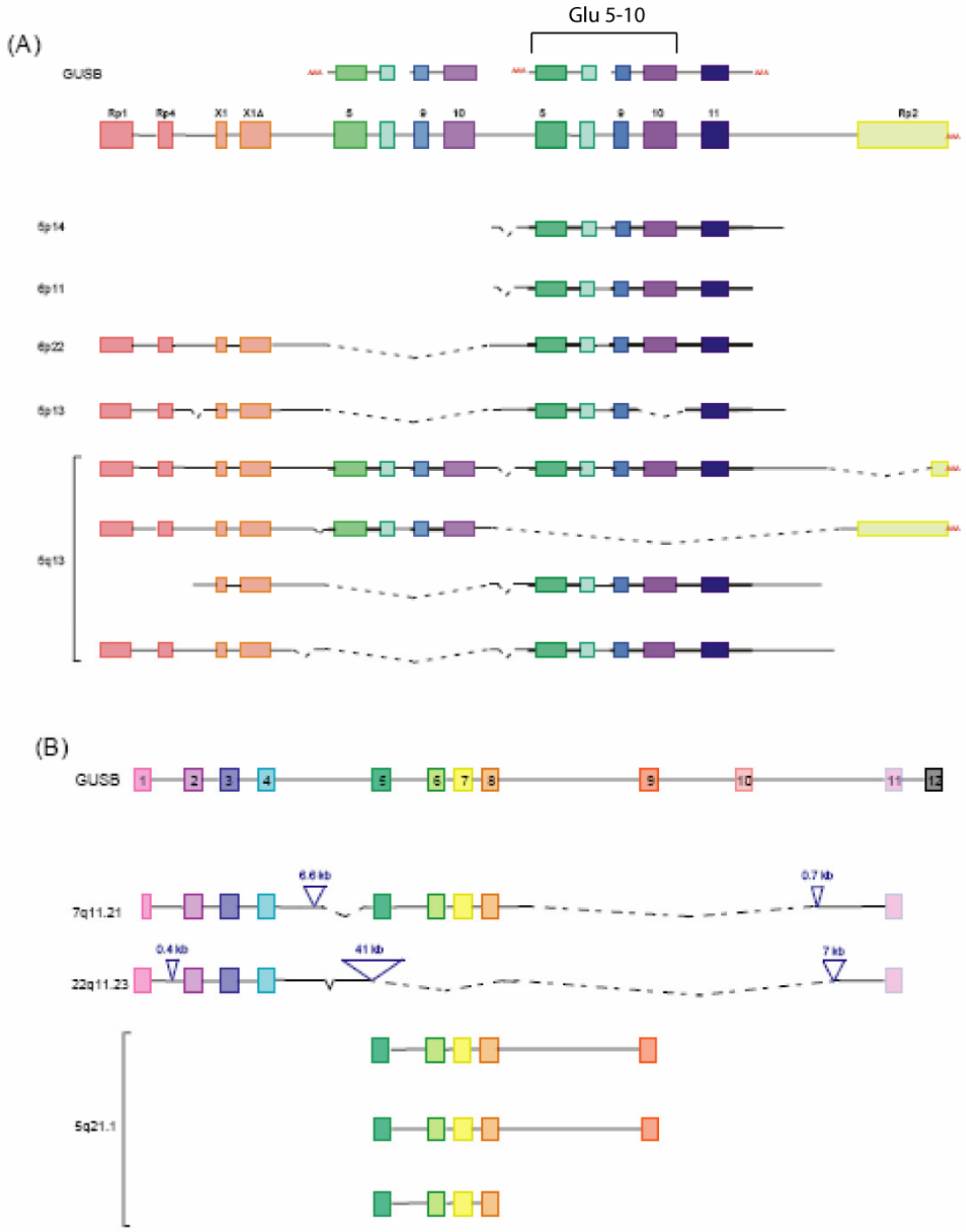


Figure 3.2.3.1.1.c **Gènes *GUSL* identifiés par BLAST « manuel »**

Chacun des douze exons de *GUSB* est recherché par similarité dans le génome entier humain. Les exons dérivés de *GUSB* associés en un même locus sont rassemblés en une unité génique. Elles sont alignées et leurs structures sont précisées. Deux classes de gènes *GUSL* (GUS-like) apparaissent alors, selon qu'ils comportent **(A)**, ou non **(B)** l'élément Glu 5-10. Les numéros au dessus des gènes correspondent aux numéros des exons de la *β-Glucuronidase*. Les loci figurent à gauche de chaque gène.

par BLAST a été menée en parallèle, au cours de laquelle chaque exon de *GUSB* a été recherché. Puis les exons dérivés de ce gène ont été localisés. Lorsque différents exons sont identifiés dans une même région (moins de 50 kb entre chaque exon) et dans la même orientation, ils sont regroupés dans une même unité de séquence. Les séquences environnantes et intervenantes sont également extraites et les « gènes » ainsi identifiés sont alignés dans Seaview. La longueur totale ainsi couverte par chaque gène est d'environ 100 kb. Leur organisation structurale peut alors être déduite. Les exons dérivés de *GUSB*, mais également des exons répétés dans le génome et identifiés dans une étude précédente (Courseaux et al., 2003), sont alors positionnés. En 2005, treize gènes dérivés de la *β-Glucuronidase* ont ainsi été identifiés par cette approche manuelle, et leurs structures ont été précisées (cf Figure 3.2.3.1.1.c). Ils sont séparés en deux sous-familles (correspondant aux groupes A et B), selon qu'ils comportent, ou non, l'élément Glu 5-10 initialement étudié. Cet élément est en outre parfois lui-même dupliqué au sein d'un gène *GUSL*, qui présente alors une structure exonique du type Rp1-Rp4-X1-X1Δ-5-9-10-5-9-10-11.

Cette analyse manuelle permet donc l'identification et la caractérisation des membres de la famille de gènes *GUSL*. Elle est cependant longue et fastidieuse, d'autant que les réassemblages constants des séquences du génome humain (et notamment des loci comportant des duplications segmentaires, particulièrement difficiles à assembler, dont le locus SMA) imposent de renouveler l'analyse de façon régulière.

Un outil logiciel automatique a alors été mis en place et développé par Richard Christen et Nicolas Pantel, au laboratoire de Biologie Virtuelle de Nice. Les duplications des groupes A et B sont d'abord identifiées par BLAT puis visuellement examinée grâce à l'outil WLAST.

- **Duplications du groupe A**

Les séquences identifiées par BLAT débutent à l'extrémité 5' du premier exon dupliqué et terminent à l'extrémité 3' du dernier exon. Mais la duplication peut, comme l'indiquent l'analyse manuelle et les alignements de séquences subséquents, s'étendre au-delà de ces bornes. Nicolas Pantel a donc procédé à d'autres analyses afin d'identifier les duplications entières, qui indiquent que les séquences obtenues par BLAT doivent être allongées de quelques kb en amont et en aval. L'analyse par Wlast de ces séquences étendues démontre leur similarité avec le gène *GUSB* authentique (cf Figure 3.2.3.1.1.d). L'alignement des exons dérivés de *GUSB* met en évidence des remaniements de type insertions/délétions.

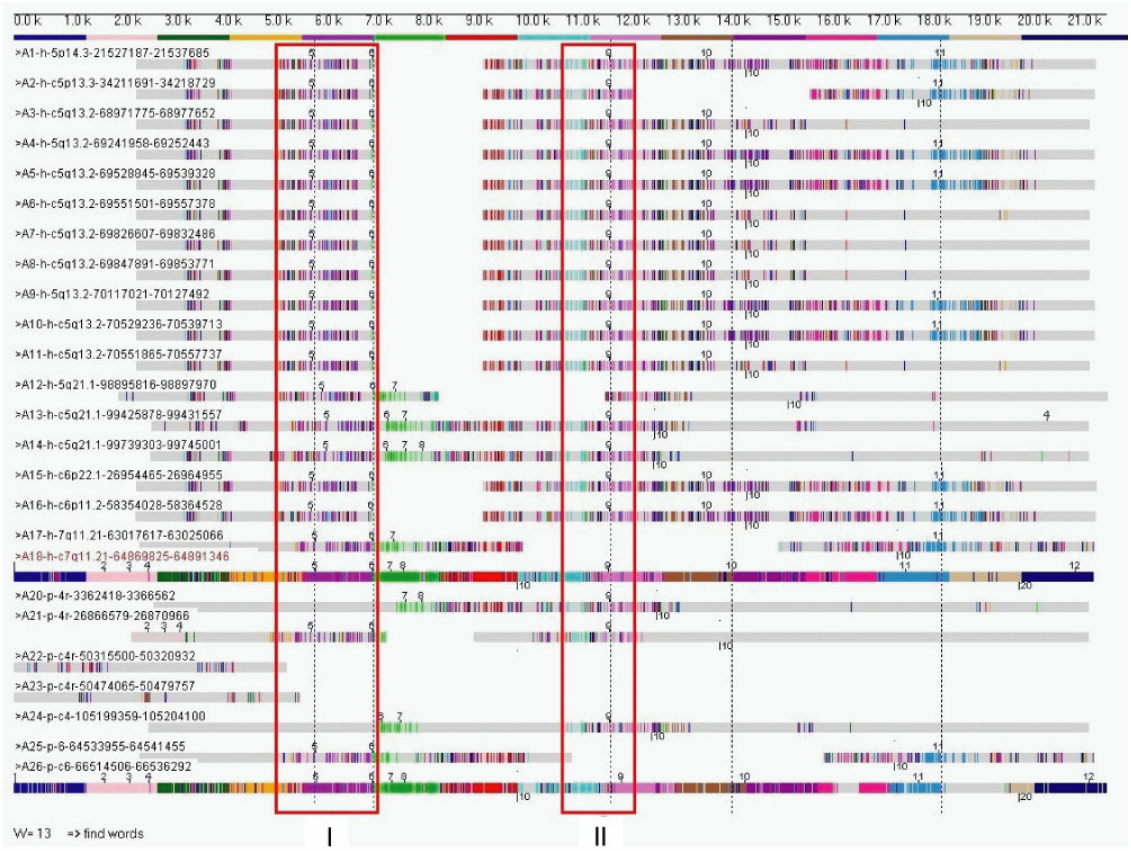


Figure 3.2.3.1.1.d **Analyse visuelle des séquences du groupe A**

L'ordonnée indique une échelle de longueur de séquences (en kb), l'abscisse présente les différentes séquences analysées (A1 à A18 pour l'homme, les suivantes appartenant au génome de Chimpanzé). Leur localisation chromosomique est indiquée. L'une des séquences, ici A18 (qui correspond à *GUSB*), est choisie comme référence. Les autres lui sont comparées. La séquence de référence est d'abord colorée comme indiqué sur l'échelle figurant tout en haut. Ensuite, pour chaque séquence, un domaine similaire à la séquence de référence est coloré de la même façon. Les exons dérivés de *GUSB* sont ensuite alignés pour toutes les séquences, les délétions apparaissent alors en blanc.
Les extrémités des séquences apparaissent en gris, ce qui indique que les bornes des duplications ont été atteintes. Les régions I et II sont présentes dans la majorité des séquences et seront utilisées pour les analyses phylogénétiques.
Figure réalisée par Nicolas Pantel au cours de son M1 au laboratoire de biologie virtuelle à Nice.

Certaines séquences dupliquées sont manifestement très similaires (A1, A4, A5, A9, A10, A15 et A16, par exemple). Les exons 7, 8 et 12 sont absents de la majorité des séquences, à l'inverse des exons 5, 6 et 9.

- **Duplications du groupe B**

Le même travail est effectué pour le second groupe de duplications (données non présentées).

- **Analyse des séquences présentes en 5'**

Les séquences peuvent être davantage étendues, dans le but de retrouver l'intégralité des éléments identifiés lors de l'analyse manuelle. Une extension de 200 kb des séquences analysées par Wlast (et par Nicolas Pantel) indique que les séquences A1 à A11 et A15 possèdent une grande région commune en 5' (données non présentées). Cette région comporte les exons Rp1, Rp4, X1 et X1Δ, identifiés précédemment (Courseaux et al., 2003).

Les premières analyses manuelles et les alignements de séquences obtenus permettent en outre de préciser les structures d'exons et introns, et d'identifier quelques éléments d'intérêt au sein des gènes. On repère ainsi des séries d'adénines (polyA) entre l'exon X1Δ et l'exon 5, entre l'exon 10 et l'exon 5 lorsque l'élément Glu 5-10 est dupliqué au sein du gène et enfin après l'exon 11.

3.2.3.1.2. L'histoire de la mise en place des gènes *GUSL*

Des analyses phylogénétiques ont ensuite été menées par Nicolas Pantel et Richard Christen à partir des séquences ainsi extraites et alignées. Elles portent sur des blocs de séquences partagés par un nombre significatif de gènes *GUSL*.

- **Phylogénie du groupe A**

L'observation des alignements met en evidence deux regions communes à la majorité des sequences (cf Figure 3.2.3.1.1.d) : le domaine I (correspondant aux exons 5 et 6) est partagé par 18 séquences humaines, et le domaine II (correspondant à l'exon 9) est présent dans 16 séquences, comprenant toutes également le domaine I.

L'analyse phylogénétique des séquences du groupe AII conduit à la proposition d'une topologie d'arbre présentée dans la Figure 3.2.3.1.2.a. Cette analyse, réalisée par Nicolas Pantel, suggère que :

- Le gène ancestral α, à l'origine de l'ensemble des séquences, a été initiallement dupliqué du chromosome 7 au chromosome 5 ou 6 (gène ancestral ζ).

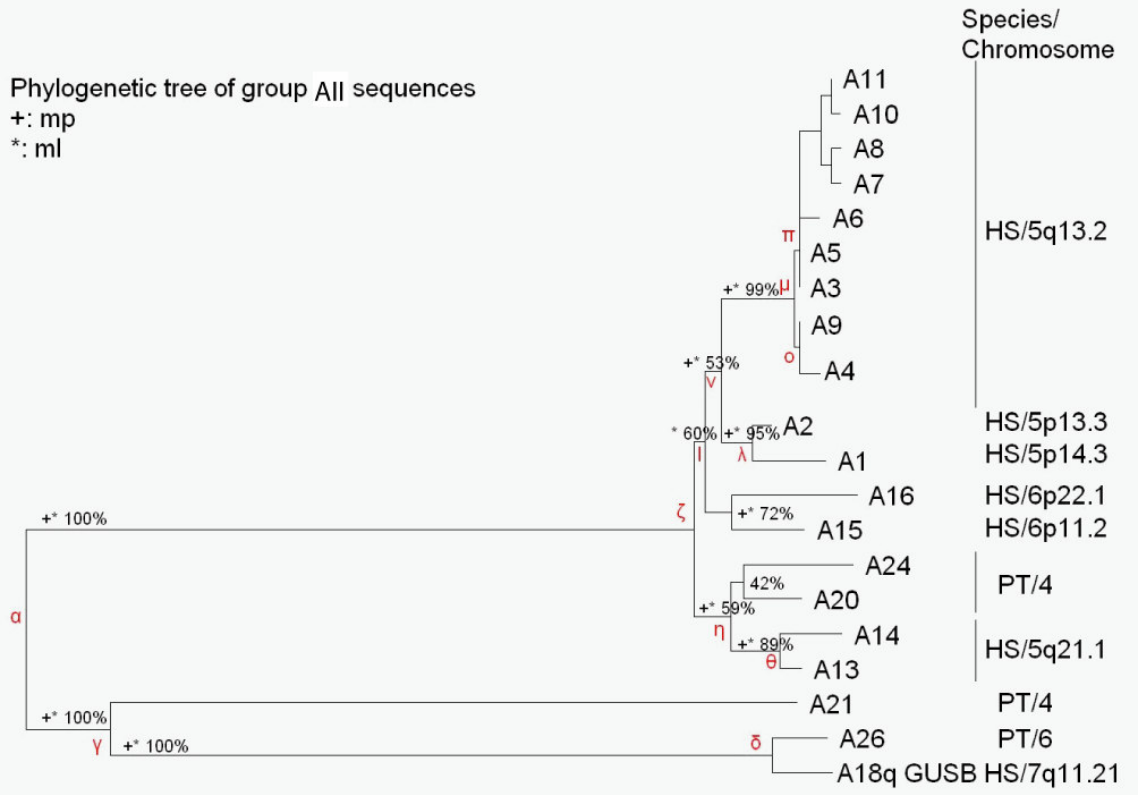


Figure 3.2.3.1.2.a **Analyse phylogénétique des séquences du groupe A**

Les symboles * et + indiquent respectivement que l'embranchement est obtenu par les méthodes de maximum de vraisemblance et de parcimonie. Les pourcentages sont obtenus par la méthode de bootstrap et donnent une indication de la robustesse des branches obtenues par la méthode de Neighbour-Joining. A droite figurent les localisations chromosomiques de chaque séquence, ainsi que l'espèce à laquelle elles appartiennent (HS *Homo sapiens*, PT *Pan troglodytes*). En rouge et en lettres grecques apparaissent les noms des séquences ancestrales présumées. Figure réalisée par Nicolas Pantel au cours de son M1 au laboratoire de biologie virtuelle à Nice.

- Les séquences A20 et A24 (du groupe η) se trouvent sur le chromosome 4 du Chimpanzé (équivalent du chromosome 5 humain), ce qui suggère que la duplication de ζ est intervenue avant la spéciation. Après la spéciation, deux duplications indépendantes ont donné naissance à A20 et A24 chez le Chimpanzé, et à A13 et A14 chez l'Homme.
- Le gène ancestral η existe chez l'Homme et chez le Chimpanzé, contrairement au gène ι, retrouvé uniquement chez l'Homme, ce qui permet de préciser l'histoire du gène ancestral ζ. Ce gène, résultant d'une duplication du gène α, a été intégré sur le chromosome 5 pour devenir le gène η. Le gène η a été ensuite dupliqué pour donner le gène ι. Ce gène, positionné initialement sur le chromosome 5 ou sur le chromosome 6, a été ensuite dupliqué sur l'autre de ces deux chromosomes après la spéciation Homme/Chimpanzé. Une copie est alors présente sur chacun de ces chromosomes humains. Une duplication ultérieure sur le chromosome 6 donne naissance aux séquences A15 et A16, tandis que la séquence présente sur le chromosome 5 a subi toute une série de duplications rapides.

Une analyse phylogénétique des séquences du groupe AI a également été menée par Nicolas Pantel et Richard Christen. La topologie obtenue est identique et confirme les résultats précédents. L'analyse des séquences A12, A17 et A25, absentes de l'analyse précédente, apporte de nouvelles informations (données non présentées).

- Le gène α a été dupliqué de façon précoce pour donner la séquence ε qui a ensuite évolué de façon indépendante jusqu'à la spéciation Homme/Chimpanzé, pour devenir finalement A17 (chez l'Homme) et A25 (chez le Chimpanzé).
- Les séquences A13 et A14 se trouvent associées avec la séquence A12 en 5q21.1, sans que les modalités de l'apparition de cette séquence ne puissent être précisées par notre analyse.

- **Phylogénie du groupe B**

Les séquences du groupe B ont été analysées par Nicolas Pantel de la même façon que les séquences du groupe A (cf Figure 3.2.3.1.2.b). Les séquences B1 et B7 (non identifiées par l'analyse manuelle en 2005, et localisées sur les chromosomes 4 de l'Homme et 3 du Chimpanzé, respectivement) ne présentent pas d'introns et ont donc vraisemblablement été créées par un évènement de rétrotransposition. Elles ne peuvent être directement comparées aux autres séquences du fait de trop courtes régions homologues. Dans cette analyse, l'horloge moléculaire semble pouvoir être appliquée, comme en attestent les

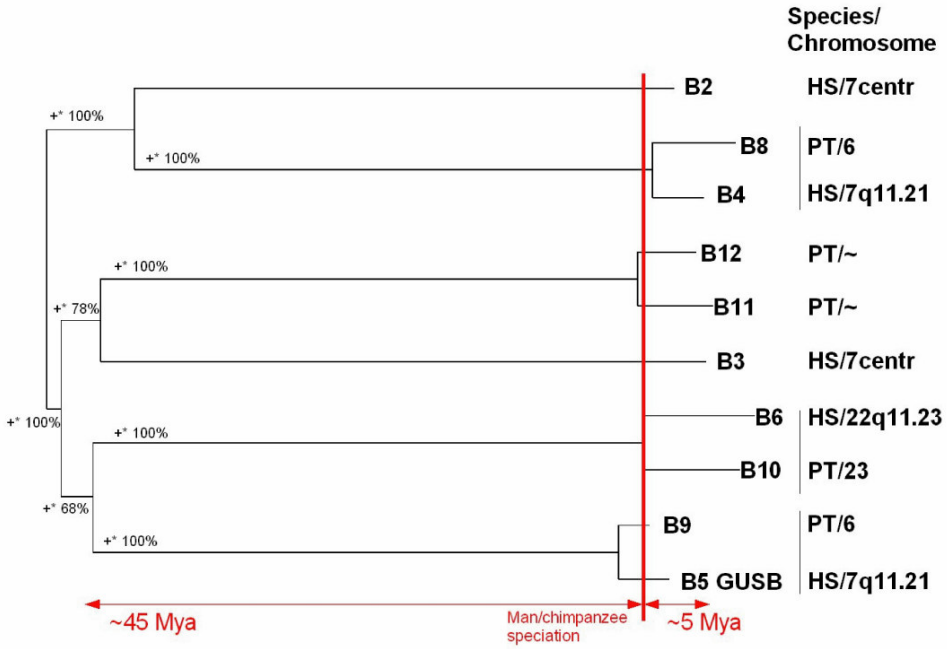


Figure 3.2.3.1.2.b **Analyse phylogénétique des séquences du groupe B**

Les symboles * et + indiquent respectivement que l'embranchement est obtenu par les méthodes de maximum de vraisemblance et de parcimonie. Les pourcentages sont obtenus par la méthode de bootstrap et donnent une indication de la robustesse des branches obtenues par la méthode de Neighbour-Joining. A droite figurent les localisations chromosomiques de chaque séquence, ainsi que l'espèce à laquelle elles appartiennent (HS *Homo sapiens*, PT *Pan troglodytes*).
L'horloge moléculaire est bien respectée. En faisant l'hypothèse que la spéciation Homme/Chimpanzé a eu lieu il y a 5 millions d'années, on peut estimer le pic de duplication à il y a près de 50 millions d'années (Mya). Figure réalisée par Nicolas Pantel au cours de son M1 au laboratoire de biologie virtuelle à Nice.

longueurs des branches correspondant aux évènements de spéciation Homme/Chimpanzé, de sorte qu'il est possible d'estimer les moments des évènements de duplication. Quatre couples de séquences (B4/B8, B11/B12, B6/B10 et B5/B9) ont été mis en place lors de la spéciation Homme/Chimpanzé. Les longueurs de leurs branches correspondent donc à une période d'environ 5 millions d'années. Il est ainsi possible de dater grossièrement les autres duplications du groupe B dans une fourchette allant de 40 à 55 millions d'années. On peut également essayer de dater l'apparition des séquences B1 et B7 en comparant les séquences des exons 1, 2 et 3. L'arbre résultant (non présenté) suggère que l'évènement de rétrotransposition à l'origine de ces gènes a eu lieu très rapidement après le premier évènement de duplication du gène *GUSB* ancestral.

3.2.3.1.3. Gènes *GUSL* au locus SMA

Quatorze des vingt-deux gènes dérivés de *GUSB* identifiés chez l'Homme se situent sur le chromosome 5 (cf Figure 3.2.3.1.3 **(A)**). Les séquences A1 et A2 sont isolées, les séquences A12, A13 et A14 sont co-localisées, et les séquences A3 à A11 sont groupées dans un intervalle de 1,7 Mb. L'outil employé par Nicolas Pantel et Richard Christen dans le cadre de notre collaboration permet d'analyser plus finement ces neuf séquences. On observe (Figure 3.2.3.1.3 **(B)**) une forte similarité de structure entre deux régions, nommées en conséquence duplicon 1 et duplicon 2. Ceci semble indiquer une duplication récente de grande taille. L'extraction de ces régions et leur comparaison par Wlast (données non présentées et analysées par Nicolas Pantel) montrent que ces domaines peuvent être facilement alignés, à l'exception des séquences correspondant aux gènes *PMCHL2* et *NAIP*, ce qui confirme l'hypothèse de la duplication en tandem, dont la date d'apparition sera discutée.

3.2.3.2. Données d'expression des gènes GUSL 5-9-10

3.2.3.2.1. Chez le macaque

Le macaque présente un signal d'hybridation en FISH réalisé avec une sonde Glu 5-10 (Courseaux et al., 2003) et possède donc des gènes dits *GUSL*. Des PCR ont été réalisées sur des ADN complémentaires de tissus centraux de macaques afin d'identifier d'éventuels messagers de gènes *GUSL*. Aucun messager GUSL n'a pu être mis en évidence par PCR et séquençage. Seuls des messagers du gène *GUSB* authentique apparaissent lors de PCR entre les exons 5 et 10 ou 5 et 11 (données non présentées, obtenues par RT-PCR et séquençage).

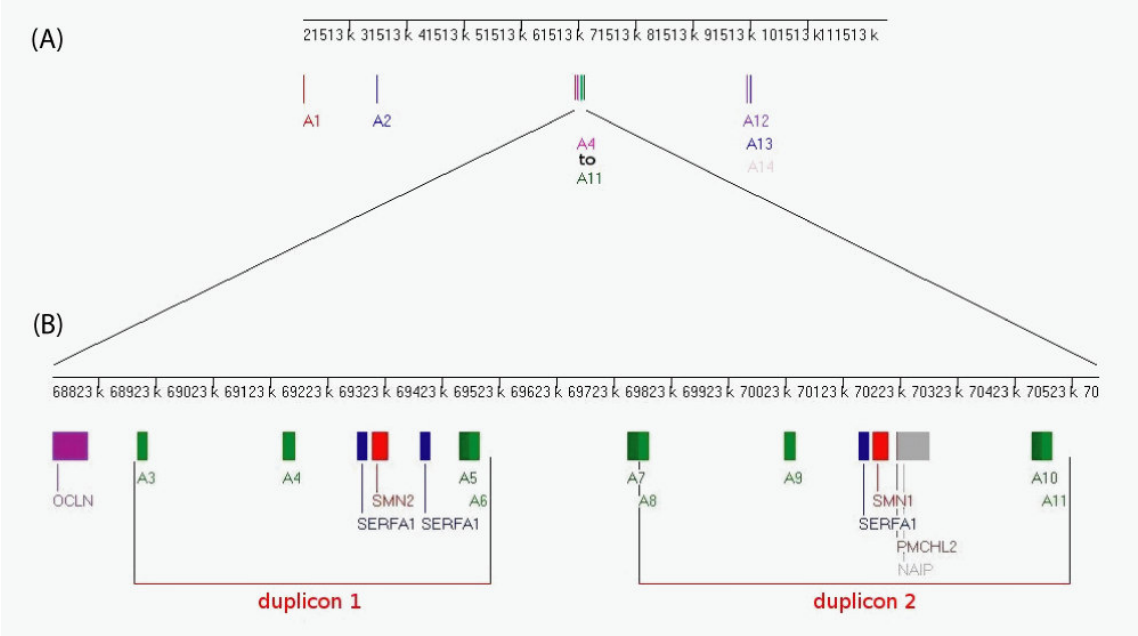
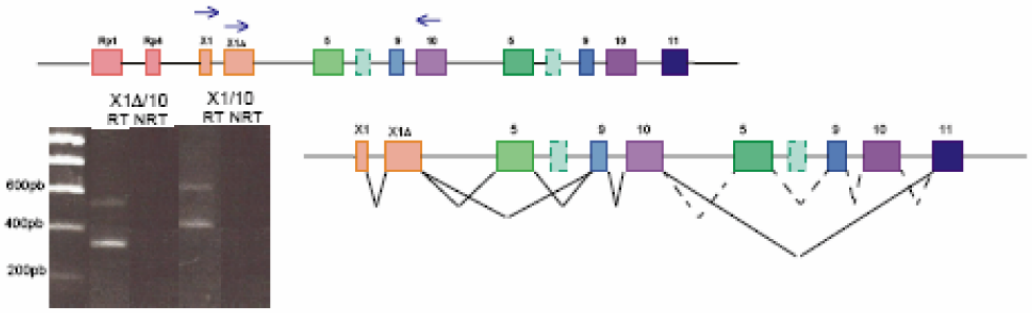


Figure 3.2.3.1.3 **Etude des duplications *GUSL* du chromosome 5 humain**

(A), Positionnement des séquences le long du chromosome, entre les positions 21,5 Mb et 11,5 Mb. **(B)**, Zoom sur la région de 68,9 Mb à 70,9 Mb, correspondant à une région étendue autour du locus SMA. Les positions des séquences et leur étendue sont indiquées. D'autres gènes d'intérêt de cette région sont également figurés. L'alignement des séquences des duplicons 1 et 2 met en évidence une large duplication en tandem. Figure réalisée par Nicolas Pantel au cours de son M1 au laboratoire de biologie virtuelle à Nice.

A



B

		X1-10	X1Δ-10	5-10	9-10
Adult	CBL	+ (BET 2)	+ (BET 2)	+ (BET)	+ (BET 2)
	HPC	+ (BET 2)	+ (BET 2)	+ (BET)	+ (BET 2)
	SPL	-	-	+ (South.)	+ (South. 2)
Newborn	CBL	+ (BET 2)	+ (BET 2)	+ (BET)	+ (BET 2)
	CX	+ (BET 2)	+ (BET 2)	+ (BET)	+ (BET 2)
	HPC	+ (BET 2)	+ (South. 2)	+ (South.)	+ (BET 2)
	HPT	+ (BET 2)	+ (BET 2)	+ (South.)	+ (BET 2)
	COL	+ (South. 2)	+ (South. 2)	-	+ (South. 2)
	LIV	-	-	-	+ (South. 2)
	SPL	-	-	-	-
Fœtus	CBL	+ (BET 2)	+ (BET 2)	+ (BET)	+ (BET 2)
	CX	+ (BET 2)	+ (BET 2)	+ (South.)	+ (BET 2)
	HPC	+ (BET 2)	+ (BET 2)	+ (BET)	+ (BET 2)
	HPT	+ (BET 2)	+ (BET 2)	+ (BET)	+ (BET 2)
	DUO	+ (South. 2)	+ (South. 2)	+ (BET)	+ (BET 2)
	OV	-	-	+ (BET)	+ (South. 2)
	SPL	+ (South. 2)	+ (South. 2)	+ (BET)	+ (BET 2)
	COL	-	+ (South. 2)	+ (South.)	+ (South. 2)
	LIV	-	+ (South. 2)	+ (South.)	+ (South. 2)
	TES	-	+ (South. 2)	+ (South.)	+ (BET 2)

Figure 3.2.3.2.2.a **Expression transcriptionnelle des gènes *GUSL* chez l'Homme**

(A), Structure maximale des gènes GUSL comportant l'élément Glu 5-10. Les exons 5, 9 et 10 sont représentés deux fois, dans la mesure où certains gènes (en l'occurrence A5/6, A7/8 et A10/11) présentent une duplication interne. Des expériences de RT-PCR avec des amorces dans les exons X1 ou X1Δ (amorce sens) et l'exon 10 (amorce antisens) sont présentées. Un épissage alternatif excluant l'exon 5 est ainsi mis en évidence.
(B), Résumé de l'ensemble des expériences de RT-PCR réalisées à partir d'ARN de fœtus, de nouveau-né et d'adulte. CBL, HPC, CX et HPT : cervelet, hippocampe, cortex et hypothalamus, respectivement. SPL, COL, LIV, DUO, OV, TES : rate, colon, foie, duodénum, ovaire et testicule, respectivement. Les résultats sont indiqués par les + et - (détecté ou pas), suivis d'une indication du mode de détection (en BEt ou par Southern Blot). Le – indique qu'aucun messager n'a pu être détecté dans aucune de nos conditions. Le chiffre 2 indique que deux bandes sont identifiées, l'une d'elle correspondant au transcrit alternatif.

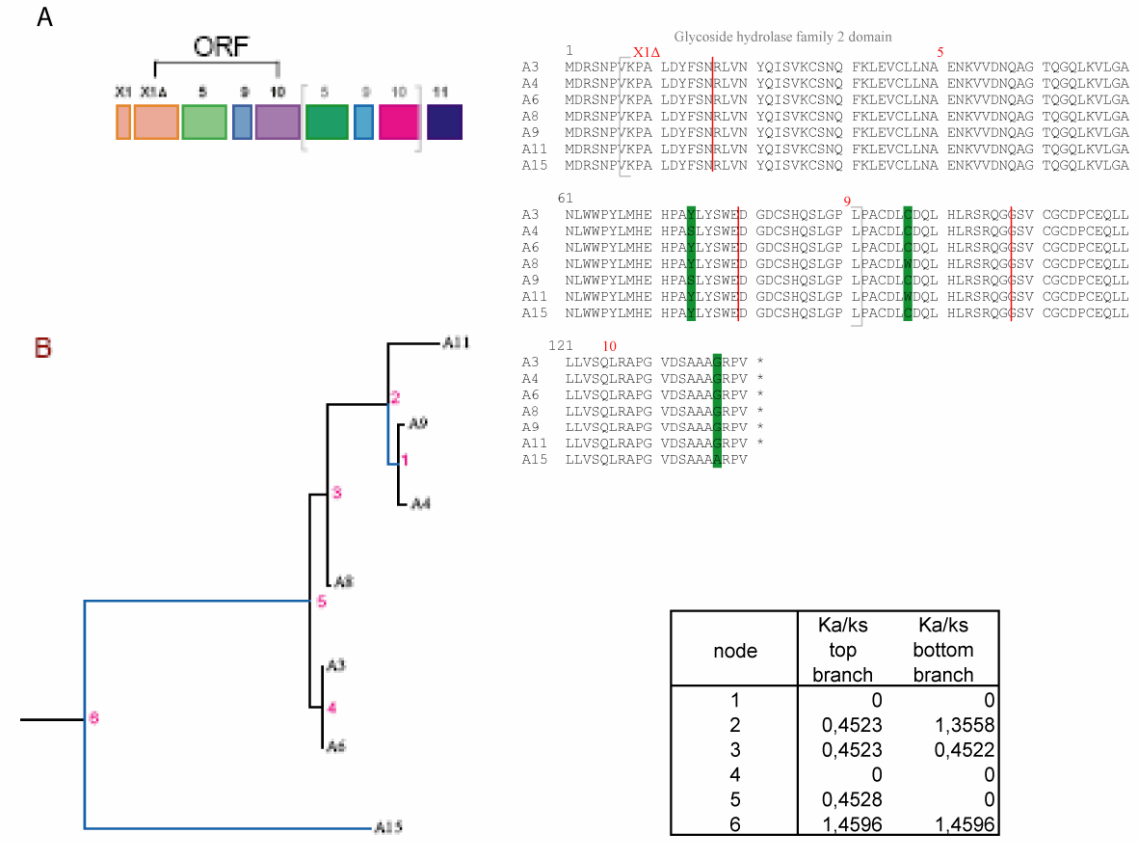
3.2.3.2.2. Chez l'homme

Des analyses de RT-PCR ont été menées, par Christine Payré et moi-même, sur différents tissus centraux et périphériques humains, afin d'amplifier d'éventuels messagers comportant les exons Rp1 et 10, X1Δ et 10, 5 et 10 et enfin 9 et 10. L'expression des gènes *GUSL* est mise en évidence dans tous les tissus analysés (rate, foie, duodénum, testicule, ovaire, colon cervelet, hippocampe, hypothalamus et cortex) et aux différents stades développementaux (fœtus, nouveau-né et adulte), mais de façon plus systématique et plus prononcée dans le système nerveux central. L'analyse de la séquence et la structure (en terme d'enchaînement d'exons) des messagers amplifiés indique qu'ils proviennent essentiellement des gènes des chromosomes 5 et 6. L'origine de chacun des messagers reste cependant difficile à préciser, compte tenu du très faible nombre de mutations distinguant les différents paralogues de la famille *GUSL*. La capacité transcriptionnelle générale peut néanmoins être déduite et il apparaît que l'exaptation des exons X1 et X1Δ est nécessaire à l'établissement de la transcription. On identifie également un épissage alternatif excluant l'exon 5 aux trois stades développementaux (cf Figure 3.2.3.2.2.a).

Il apparaît également que ces messagers comportent une phase ouverte de lecture significative, de 140 acides aminés, correspondant aux exons X1Δ, 5, 9 et 10. Les protéines putatives ne partagent avec la β-Glucuronidase authentique que la séquence codée par l'exon 5, elle-même ponctuée de quelques mutations. Elles comportent un domaine glycoside hydroxylase. Les phases ouvertes de lecture des sept gènes susceptibles de produire la protéine de 140 acides aminés ont été alignées ainsi que celles des peptides ainsi générés. Un calcul de Ka/Ks a alors été opéré sur ces séquences afin d'identifier d'éventuelles pressions de sélection différentielles sur l'un ou l'autre des paralogues. Des disparités apparaissent alors, les valeurs de Ka/Ks étant comprises entre 0 et près de 1,5 selon les branches (cf Figure 3.2.3.2.2.b).

3.2.4. Discussion

Les recherches menées sur le génome entier humain, de façon manuelle puis à l'aide d'un outil automatique, permettent ainsi de caractériser une nouvelle famille de gènes chimériques spécifiques des Primates : plus d'une vingtaine de paralogues *GUSL* sont identifiés. Leurs loci et organisations structurales sont précisés. L'analyse automatique met en évidence l'ensemble des gènes identifiés par l'approche manuelle par Blast, à l'exception de



node	Ka/ks top branch	Ka/ks bottom branch
1	0	0
2	0,4523	1,3558
3	0,4523	0,4522
4	0	0
5	0,4528	0
6	1,4596	1,4596

Figure 3.2.3.2.2.b **Expression traductionnelle potentielle des gènes *GUSL* chez l'Homme**

(A), La phase ouverte de lecture principale comprend des séquences des exons X1Δ, 5, 9 et 10 et est portée par les gènes comportant l'exon X1Δ (soit A3, A4, A5/6, A7/8, A9 , A10/11 et A15). La protéine putative compte 140 acides aminés et présente un domaine glycoside hydroxylase codé par la fin de l'exon X1Δ, l'exon 5 et le début de l'exon 9.
(B), La construction d'un arbre phylogénétique comprenant les paralogues précédemment identifiés (cf (A)) et tenant compte des données de topologie obtenues précédemment (cf Figure 3.2.3.1.2.a) permet le calcul de ratios Ka/Ks. Les valeurs obtenues sont indiquées dans le tableau et les branches de couleur bleue présentent des rapports Ka/Ks supérieurs à 1.

celui du locus 22q11.23, dont une insertion de 42 kb distance trop les exons dérivés de *GUSB* pour que WLAST parvienne à l'identifier. Des gènes additionnels sont également trouvés, probablement du fait du nouvel assemblage du génome humain. Tous ces paralogues correspondent à l'ensemble des sites identifiés et caractérisés préalablement par FISH et analyse de BAC (Courseaux et al., 2003).

L'outil WLAST est indéniablement un outil novateur et très utile. Il permet de visualiser les alignements de séquences d'intérêt identifiées dans les génomes entiers, de sorte à repérer immédiatement les duplications, insertions et délétions. L'exploitation des données obtenues par Wlast, particulièrement après extension des sequences qui flanquent les exons dérivés de *GUSB*, apporte immédiatement deux éléments majeurs pour la compréhension de l'histoire des duplications des gènes *GUSB* et *GUSL*. D'une part, certaines de ces séquences dupliquées sont présentes à la fois dans les génomes de l'Homme et du Chimpanzé et sont très similaires entre les deux espèces (par exemple A17 et A25), ce qui suggère un évènement de duplication précédant de peu la spéciation Homme/Chimpanzé. D'autre part, d'autres séquences dupliquées peuvent être rassemblées en familles. Par exemple les éléments A1, A4, A5, A9, A10, A15 et A16 présentent les mêmes régions 5' et 3' flanquantes, ce qui indique que les évènements de duplication par lesquels ils ont été créés sont très récents. Leur absence du gènome de Chimpanzé confirme cette hypothèse.

Les analyses phylogénétiques qui ont ensuite pu être menées sur ces séquences enrichissent substantiellement notre connaissance de l'histoire évolutive de ces gènes. Les données de localisation et de structure, associées aux résultats de ces analyses phylogénétiques permettent alors de proposer un modèle hypothétique complet et complexe de l'évolution de la sous-famille de gènes *GUSL* du groupe A (possédant l'élément Glu 5-10). Ce modèle, qui fait l'objet de la figure 3.2.4, tient également compte des connaissances actuelles sur les sites donneurs et accepteurs de séquences (Nahon, 2003), ainsi que des données obtenues par FISH dans d'autres espèces de Primates (Courseaux et al., 2003). Il implique des évènements de recombinaison homologue, de duplications successives et d'exaptation de séquences environnantes. Il réconcilie en outre les données obtenues par les différentes méthodes, le positionnement du gène A1 restant seul problématique dans cette histoire évolutive. Cette incertitude reste naturellement à résoudre. L'achèvement du séquençage et de l'assemblage des génomes entiers de *Pan troglodytes* et *Macaca fascicularis* et l'identification, par WLAST, des gènes *GUSL* dans ces génomes permettront certainement

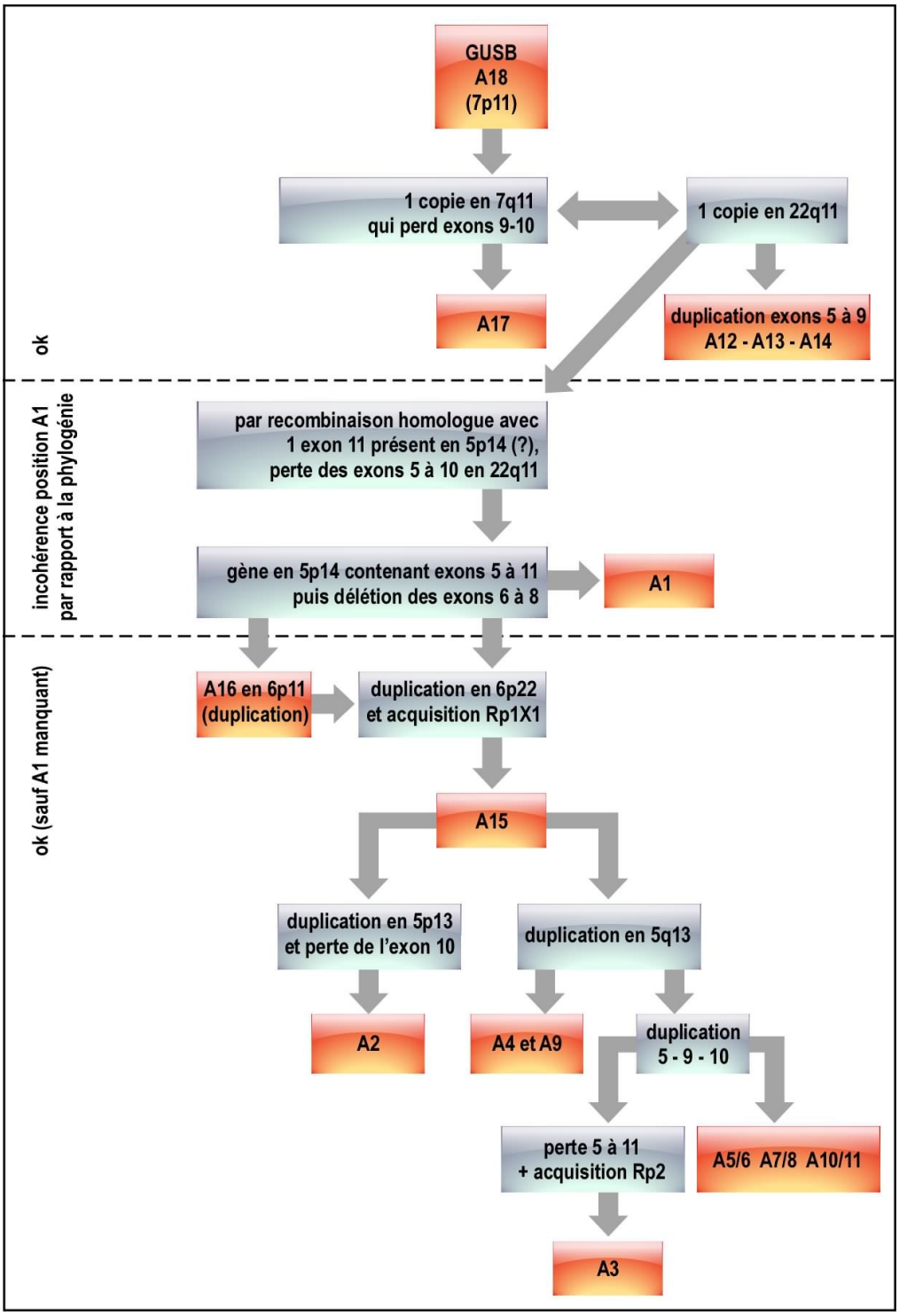


Figure 3.2.4 **Modèle de mise en place des gènes *GUSL* du groupe A chez l'Homme**

Le modèle proposé intègre les données structurales obtenues manuellement et les données de FISH. La concordance avec les données phylogénétiques, quasi-parfaite, est indiquée à gauche. Les gènes actuels apparaissent dans les compartiments rouges, les séquences ancestrales et les processus impliqués figurent dans les compartiments bleus.

de progresser dans la connaissance de l'histoire évolutive de cette grande et complexe famille de gènes.

En outre, l'analyse phylogénétique appliquée aux séquences des gènes *GUSL* du groupe B, ne comportant donc pas l'élément Glu 5-10, indique que la mise en place de ces séquences par duplication date, pour les premières d'entre elles, d'il y a près de 50 millions d'années, soit avant la divergence des Catarhiniens et des Platyrhiniens. Ainsi, ces évènements de duplication auraient précédé les évènements de mise en place des séquences du groupe A et seraient ainsi définitivement indépendants.

Nous montrons ainsi que l'étude de ces paralogues (et non plus d'orthologues comme dans l'étude de la famille de gènes *PMCHL* [*]) permet d'appréhender l'histoire récente des gènes et les modalités de leur création.

L'analyse transcriptionnelle des gènes *GUSL* met en évidence d'intéressantes propriétés. Il apparaît d'abord que les paralogues, non encore caractérisés, hormis en ce qui concerne leur localisation, présents chez le macaque ne possèdent pas d'activité transcriptionnelle dans les tissus analysés. Chez l'Homme, ces études confirment d'abord l'existence de gènes comportant une duplication interne des exons 5-9-10, duplication qui ne résulte donc pas d'une erreur d'assemblage, comme cela pouvait légitimement être redouté (Schmutz and Grimwood, 2004; Schmutz et al., 2004). Elles indiquent également que la capacité transcriptionnelle serait conférée par des séquences recrutées en 5' sur les chromosomes 5 et 6. Ainsi, seuls les gènes A15, A2, A3, A4, A5/6, A7/8, A9, A10/11 semblent transcrits. Ces séquences recrutées en 5', comprenant au minimum les exons X1 et X1Δ comportent donc probablement des éléments promoteurs de reconnaissance pour les polymérases. En outre, bien que ces messagers soient retrouvés de façon relativement ubiquitaire dans les tissus analysés (centraux et périphériques, à différents stades développementaux), l'existence d'un épissage alternatif et de niveaux d'expressions manifestement supérieurs dans les tissus centraux (cervelet, hippocampe, hypothalamus et cortex) relativement aux tissus périphériques (rate, foie, duodénum, testicule, ovaire, colon) indique que les séquences exaptées en 5' possèdent également des régions régulatrices. Les données d'expression chez l'Homme et le Macaque montrent enfin que, d'un point de vue

[*] « Evolutionnary variations in structure and brain expression of the *PMCHL1* gene in macaque and human », Fleur Darré-Toulemonde, Alain Corinus, Audrey Delerue-Audegond, Richard Christen and Jean-Louis Nahon, soumis à Molecular Biology and Evolution.

transcriptionnel, les gènes *GUSL* sont spécifiques des Hominoïdes. Des analyses transcriptionnelles chez les Chimpanzés et Gibbons préciseraient utilement ce dernier point.

Les expériences de RT-PCR et de séquençage indiquent en outre que ces messagers sont susceptibles de coder pour une protéine putative présentant un domaine glycoside hydroxylase. De façon très intéressante dans le contexte de l'étude de gènes spécifiques des Primates ou des Hominoïdes et de l'éventuelle pression de sélection darwinienne positive à laquelle ils pourraient être soumis, des calculs de Ka/Ks ont pu être effectués à partir des sequences codantes de ces paralogues (Dorus et al., 2004; Eichler et al., 2001; Luchetta et al., 2005). Des valeurs très différentes apparaissent, suggérant une accélération de l'évolution des séquences de la copie A15 et de l'ancêtre commun des copies A4 et A9, toutes semblant présentes uniquement chez l'Homme, ce qui pourrait indiquer l'existence d'une pression de sélection positive sur ces gènes. Il semble, de manière plus générale, que la sélection positive darwinienne affecte communément les gènes issus de duplications segmentaires (Bailey and Eichler, 2006; Johnson et al., 2001).

De manière plus générale, la réalisation de ces études a également été l'occasion de réexaminer (Darre-Toulemonde and Nahon, 2004) l'état d'avancement du séquençage des génomes de l'Homme et du Chimpanzé *Pan troglodytes*. De nombreuses erreurs d'assemblage ont été mises en évidence et signifiées à J. Martin, responsable du projet de séquençage du chromosome 5 humain (Schmutz et al., 2004). Elles résultent de deux phénomènes principaux : la présence de duplications segmentaires, notamment au locus SMA en 5q13 en ce qui nous concerne, qui rendent l'assemblage de séquences périlleux (She et al., 2004), et le séquençage de trois haplotypes dont l'homologie de séquence est parfois ignorée au cours de l'assemblage tant le polymorphisme est élevé en ces loci (Fredman et al., 2004). Ce dernier point est d'ailleurs désormais résolu par une nouvelle gestion des séquences, qui distingue les différents haplotypes en ces loci (cf http://www.ensembl.org). La plus grande attention reste donc nécessaire lors de l'exploitation de ces données de séquençage de génome entier.

Nos travaux mettent d'ailleurs également en évidence de nouvelles duplications segmentaires. Si l'essentiel de nos études a porté sur les duplications des seuls gènes *GUSL*, qui constituent déjà des duplications segmentaires, elles ont également permis de positionner certains de ces gènes dans une duplication plus large au locus SMA et d'en préciser la mise en place. Deux copies d'un duplicon comprenant des gènes *GUSL*, mais aussi les gènes *SMN*, sont ainsi présentes en tandem en ce locus. Cette duplication en tandem n'est pas identifiée

dans le génome de Chimpanzé, comme en témoigne l'absence de séquences *GUSL* sur l'équivalent du chromosome 5q13 humain. Or le gène *PMCHL2* existe chez le Chimpanzé et est donc apparu avant cette large duplication. L'ensemble de ces données suggère donc une duplication de A8-A9-SERFA1-SMN1-A10-A11 pour créer A3-A4-SERFA1-SMN2-A5-A6, suivie par une délétion des gènes *PMCHL2* et *NAIP* du duplicon 1 (cf Figure 3.2.3.1.3). Des phénomènes de conversion génique (Ezawa et al., 2006) peuvent enfin expliquer les forts pourcentages d'identités de séquences entre A3, A5 et A6, d'une part, A8, A10 et A11, d'autre part plutôt qu'entre A3 et A8, A5 et A10, A6 et A11, comme attendu en vertu de la duplication précédemment décrite.

Compte tenu de leur localisation, dans ou à proximité du locus SMA, et de leur capacité transcritpionnelle et éventuellement traductionnelle, les gènes *A3*, *A4*, *A5/6*, *A7/8*, *A9*, *A10/11* retiendront toute l'attention du laboratoire à l'avenir.

DISCUSSION

A l'issue de ce temps et de ces travaux de thèse, et après les avoir présentés et discutés un à un, il me semble nécessaire de les replacer dans leur contexte global, et de mettre en perspective quelques uns de leurs principaux traits. Ainsi, les caractéristiques des familles de gènes étudiées seront comparées afin de montrer comme leurs analyses sont complémentaires. Deux aspects de ces familles de gènes seront ensuite approfondis : leur appartenance aux duplications segmentaires et leurs caractéristiques transcriptionnelles. Finalement, je reviendrai sur quelques interrogations relatives aux gènes spécifiques des Primates.

1. Caractéristiques générales des gènes « chimériques » étudiés : des modèles complémentaires

Les expériences réalisées et les réfexions menées, principalement celles relatives aux gènes chimériques des familles *PMCHL* et *GUSL*, permettent d'identifier quelques aspects généraux, conservés ou non, des gènes « chimériques ». Il me semble donc intéressant de présenter tout à la fois les similitudes et les dissemblances des familles de gènes étudiées.

1.1. Aspects méthodologiques

Les modèles de familles de gènes choisis l'ont été sur différents critères : la présence de copies sur le chromosome 5 humain, l'aspect chimérique et la mise en place au cours de l'évolution des génomes de Primates (Courseaux et al., 2003; Viale et al., 1998a; Viale et al., 1998b).

Ils ont cependant ensuite été caractérisés et analysés de manières diverses.

1.1.1. Séquences analysées

L'étude des gènes *PMCHL* a porté sur les orthologues, tandis que les paralogues *GUSL* ont servi à l'analyse de cette deuxième famille de gènes. Cette distinction est liée à des caractéristiques et disponibilités différentielles de séquences.

Seule(s) une ou deux copie(s) compose(nt) la famille *PMCHL*. Par conséquent, les analyses portent nécessairement sur les orthologues, et ce, afin de disposer de données en quantité suffisante. Cette approche a été facilitée par l'accessibilité d'ADN génomiques d'un nombre conséquent d'espèces de Primates et par la relativement petite taille des gènes *PMCHL* (12 kb environ), qui rend possible leur amplification et leur séquençage au laboratoire. Nous disposons dès lors de l'alignement des gènes *PMCHL* de sept espèces de Primates (incluant l'Homme), ce qui permet de les comparer selon différentes approches, qui seront rappelées ultérieurement.

Pour les gènes *GUSL*, à l'inverse, la longueur des gènes exclut d'envisager de les amplifier et les séquencer. En outre, leur très grande similarité nucléotidique au sein d'une espèce empêche de les distinguer au cours du séquençage, à moins de ne travailler qu'à partir de chromosomes bactériens artificiels de sorte à n'amplifier qu'un paralogue à la fois. Par conséquent, nous avons choisi d'étudier les paralogues humains, dont les séquences sont disponibles du fait du séquençage du génome entier. Cette acquisition de données génomiques n'est en revanche pas possible de façon exhaustive pour le Chimpanzé *Pan troglodytes* et le Macaque *Macaca mulatta*, dont les génomes dits entiers sont encore très incomplets, et présentent notamment de grosses lacunes au niveau des régions hétérochromatiques… qu'occupent un certain nombre de nos paralogues ! Les très grandes similarités de séquences entre les paralogues limitent également la possibilité de distinguer leurs messagers. Seules les données structurales seront donc réellement informatives de ce point de vue là.

1.1.2. Evolution : structure et relations phylogénétiques

En raison, partiellement au moins, de la nature des relations entre les séquences comparées (orthologues ou paralogues) et de la longueur de ces séquences pour chaque famille de gènes chimériques étudiée, des aspects différents ont été examinés afin de déterminer les caractéristiques évolutives des gènes. Les séquences *PMCHL*, compte tenu de leur relative petite taille, ont fait l'objet d'une analyse structurale précise, visant à identifier de courtes structures (séquences répétées, insertions, délétions à partir de quelques paires de bases…) et leurs différences entre les orthologues. Pour les gènes de la famille *GUSL*, une analyse structurale a également pu être menée, mais à plus grande échelle : la présence (ou l'absence) d'exons d'intérêt, de grandes duplications, insertions ou délétions a été recherchée. Du fait du grand nombre de paralogues étudiés (relativement au cas de *PMCHL*), une analyse phylogénétique assez fine a pu être effectuée ; elle se révèle très informative.

1.1.3. Expression : caractérisation qualitative et quantitative

Les deux familles de gènes chimériques étudiées ont fait l'objet d'analyses transcriptionnelles similaires. Les caractéristiques qualitatives et quantitatives des messagers PMCHL et GUSL ont été étudiées chez l'Homme et le Macaque, sans que les sites de capping et de polyadénylation n'aient été véritablement recherchés au cours de ma thèse, certains d'entre eux ayant néanmoins été défini préablablement (Courseaux and Nahon, 2001; Viale et al., 2000). Dans les deux familles de gènes étudiées, la distinction entre les paralogues, et par conséquent les messagers qui en sont issus, n'est pas aisée, tant les mutations qui les séparent sont peu nombreuses.

Enfin, les familles de gènes *GUSL* et *PMCHL* se distinguent également par la présence, ou non, de phases ouvertes de lectures significatives, et donc par la capacité probable, ou non, d'être codantes pour une protéine. Par conséquent, l'étude portant sur l'évolution de la séquence protéique n'a pu être menée que pour les séquences GUSL.

1.2. Histoires évolutives

L'un des principaux objectifs de mon travail de thèse a été de définir les modalités de mise en place et d'évolution des familles de gènes « chimériques » spécifiques des Primates *PMCHL* et *GUSL*.

Il apparaît, au terme de nos études, que des mécanismes tout à fait différents sont à l'origine de ces deux familles de gènes. La rétrotransposition est l'évènement déterminant de la mise en place du gène *PMCHL1* (Courseaux and Nahon, 2001), mais reste en revanche presque anecdotique dans l'histoire de la famille des gènes *GUSL*, puisqu'elle serait impliquée uniquement dans l'apparition des séquences B1 et B7. Les gènes *GUSL* ont en effet essentiellement été mis en place par des processus de recombinaison et de duplication segmentaire. En outre, la mise en place de *PMCHL1*, comme celle des gènes *GUSL*, implique l'exaptation de séquences environnantes. Cependant, les exons ainsi créés pour finaliser la mise en place de *PMCHL1* ne sont pas présents en d'autres loci du génome et constituent donc une innovation propre à *PMCHL*, tandis que les exons de *GUSL* qui ne dérivent pas du gène de la *β-Glucuronidase* sont également identifiés en d'autres loci du génome. A ce titre, les gènes *GUSL* méritent davantage l'appellation de « gènes chimériques » puisqu'ils se composent effectivement d'un assemblage d'exons issus de différents gènes ou séquences du génome.

Enfin, une différence notable dans l'évolution de ces deux familles de gènes réside dans la phase d'expansion par duplication. Nous confirmons en effet l'étendue de cette expansion dans la famille *GUSL* (Courseaux et al., 2003), qui compte au moins 15 paralogues dans l'espèce humaine. Les gènes *PMCHL* n'ont, eux, pas subi une telle expansion, puisque les espèces d'Hominoïdes ne comptent chacune que deux gènes de cette famille.

1.3. Questions de spécificité

De manière très intéressante, des précisions ont pu être apportées au cours de mon travail de thèse quant à la réelle spécificité des gènes « chimériques » spécifiques des Primates étudiés. Si les gènes *PMCHL* et *GUSL* du groupe A sont en effet apparus au même moment (Courseaux et al., 2003), après la divergence des Catarhiniens et des Platyrhiniens, l'acquisition d'une capacité transcriptionnelle opère, elle, à des temps différents.

En effet, bien que les sites d'épissage ne soient manifestement pas encore constitués et actifs dans cette espèce, le gène *PMCHL1* est transcrit chez le Macaque *Macaca fascicularis*. Ainsi, la capacité transcriptionnelle semble être acquise dès la création de cette famille de gènes, à moins qu'elle ne le soit qu'ultérieurement et de façon convergente (en terme de profil d'expression) chez l'Homme et le Macaque, ce qui est fort peu probable. A tous points de vue, les gènes *PMCHL* sont donc spécifiques des Catarhiniens.

Il apparaît en revanche dans nos études que les gènes *GUSL* du groupe A ne sont pas transcrits chez le Macaque. Ils le sont chez l'Homme. Cette spécificité nécessitera naturellement d'être précisée par une analyse transcriptionnelle chez les autres Hominoïdes. Il est néanmoins d'ores et déjà possible d'indiquer que la capacité transcriptionnelle n'a, ici, été acquise que quelques temps après le début de la mise en place de cette famille de gènes. Ainsi, les gènes *GUSL* du groupe A sont donc spécifiques des Catarhiniens, mais leur expression semble, elle, spécifique des Hominoïdes.

L'étude et la comparaison des modèles de familles de gènes « chimériques » spécifiques des Primates *PMCHL* et *GUSL* s'avèrent donc particulièrement riches et représentatives des différents modes d'évolution des génomes de Primates par création de gènes. Recombinaison, duplications, rétrotranspositions, brassages d'exons, exaptation de séquences constituent en effet le socle des histoires évolutives de ces familles de gènes. Les capacités transcriptionnelles, qui confèrent à ces séquences leurs caractéristiques de gènes plutôt que de pseudogènes, ont également été analysées et comparées.

2. *PMCHL* et *GUSL* : éléments de duplications segmentaires

Du fait de la duplication de *PMCHL1* en *PMCHL2* et des multiples évènements de duplication à l'origine de la famille *GUSL*, et compte tenu de la taille de ces gènes (12 à 150 kb, environ), ces deux familles de gènes spécifiques des Primates participent aux duplications segmentaires.

Les duplications segmentaires sont, rappelons-le, des segments d'ADN génomique, couvrant 1 à 400 kb, présents en plusieurs loci dans le génome, et partageant typiquement de hauts niveaux d'identité nucléotidique (> 90% d'identité entre les paralogues), ce qui implique une origine relativement récente (moins de 35 millions d'années) (Bailey et al., 2002b; Eichler, 2001). Les modalités de leur mise en place ont été examinées en introduction, nous discuterons ici des conséquences qui peuvent découler de la présence de ces séquences dupliquées dans les génomes de Primates.

2.1. Duplications segmentaires simples et composées

Les duplications segmentaires, à l'origine de réarrangements chromosomiques précédemment mentionnés et sur lesquels je reviendrai ultérieurement, présentent des configurations organisationnelles variées. Elles peuvent être de structure simple, comme dans le cas des gènes *PMCHL*, ou contenir de complexes arrangements de modules dupliqués (Emanuel and Shaikh, 2001). La structure de ces régions est alors très complexe, du fait d'un grand nombre d'évènements de transpositions, juxtapositions et réarrangements segmentaires successifs. Il est possible de définir des blocs unitaires de séquences ayant vraisemblablement connu la même histoire évolutive, de sorte que les plus grandes duplications se décomposent en de multiples duplications de plus petites tailles et d'origines évolutives variées (Bailey et al., 2002b; Samonte and Eichler, 2002). En outre, les duplications segmentaires, aussi bien intrachromosomiques qu'interchromosomiques, sont fréquemment identifiées à proximité d'autres duplications segmentaires indépendantes (Samonte and Eichler, 2002). Les gènes *GUSL* participent, comme cela était suggéré (Courseaux et al., 2003) et comme nous le montrons notamment au niveau des duplicons 1 et 2, de ce deuxième type de duplications segmentaires. Les duplications complexes organisées en modules sont impliquées dans le plus

grand nombre des remaniements chromosomiques à l'origine de désordres génomiques associés à des pathologies (Emanuel and Shaikh, 2001).

2.2. Duplications segmentaires des Primates

Les duplications segmentaires sont identifiées dans de nombreuses espèces métazoaires. Certaines particularités de ces séquences apparaissent cependant nettement dans les génomes de Primates.

2.2.1. Abondance des duplications récentes

En effet, en comparaison des autres mammifères, les génomes de l'Homme et des autres Primates présentent un enrichissement en grandes duplications segmentaires possédant de forts taux d'identité de séquences (Bailey and Eichler, 2006; She et al., 2006). Des données récentes suggèrent une origine complexe des duplications segmentaires de Primates, impliquant deux vagues de duplications au cours de l'évolution de ces espèces (Bailey and Eichler, 2006). En particulier, les estimations globales de divergence des séquences semblent indiquer une période d'expansion de ces séquences au cours de l'histoire récente des Hominoïdes, comme nous l'observons dans la famille de gènes *GUSL* (She et al., 2006). Ainsi, ces duplications manifestement abondantes, pour des raisons non encore précisées, ont participé à la création de nouvelles familles de gènes spécifiques des primates et ont également influencé les modifications phénotypiques et géniques de l'Homme de façon plus fréquente que présupposé (Bailey and Eichler, 2006; She et al., 2006).

2.2.2. Localisation des duplications

En dehors de ces différences d'abondance, un autre différence majeure est mise en évidence lors de la comparaison des duplications segmentaires de Rongeurs, d'une part, et de primates, d'autre part. Il apparaît que si les duplications segmentaires de Rongeurs sont, comme attendu, essentiellement confinées aux régions hétérochromatiques des péricentromères et subtélomères, celles des Primates ont en revanche envahi l'euchromatine (Courseaux et al., 2003). C'est d'ailleurs pour cette raison qu'une fraction euchromatique du génome humain résiste encore aux épreuves du séquençage et de l'assemblage des séquences : elle est constituée de duplications segmentaires (2004).

2.3. Duplications segmentaires et polymorphisme

2.3.1. Polymorphisme du nombre de copies

Par définition même, les duplications segmentaires impliquent de multiples copies d'une même séquence. Les difficultés de séquençage et d'assemblage des régions du génome humain riches en répétitions de ce type ont mis en évidence des variations polymorphiques du nombre de copies (Courseaux et al., 2003; Sharp et al., 2005). L'enrichissement des régions particulièrement sujettes aux remaniements génomiques en sites présentant un polymorphisme du nombre des copies suggère que les duplications segmentaires constituent un catalyseur des variations à grande échelle du génome humain (Perry et al., 2006; Sharp et al., 2005). On observe d'ailleurs réciproquement un enrichissement en duplications segmentaires des sites présentant un polymorphisme du nombre des copies (Sharp et al., 2005). Enfin, les séquences dupliquées pourraient prédisposer à des polymorphismes mitotiques à grande échelle, liés à ces duplications interstitielles et associés avec des caractères génétiques complexes, tels que des susceptibilités à des désordres phobiques (Gratacos et al., 2001). Ces dernières données sont cependant controversées.

2.3.2. Polymorphisme nucléotidique

En outre, un polymorphisme nucléotidique particulièrement élevé (deux fois plus que dans les gènes uniques) a été détecté au sein des duplications segmentaires; il a d'abord été imputé à une contamination par des séquences paralogues (Bailey et al., 2002a; Cheung et al., 2003). Une génotypage quantitatif de sites de polymorphismes nucléotidiques (SNP Single Nucleotid Polymorphism) au sein des duplications segmentaires a néanmoins permis de confirmer une véritable richesse en SNP de ces duplications (Bailey and Eichler, 2006). De même, des taux de mutations élevés sont observés entre éléments paralogues de ces régions, en dépit des possibles conversions géniques (Bailey and Eichler, 2006; She et al., 2006). Ces données sont naturellement à rapprocher de la grande variabilité également décrite pour les séquences protéiques synthétisées à partir d'éléments de duplications segmentaires (Bailey and Eichler, 2006; Johnson et al., 2001).

2.4. Duplications segmentaires et pathologies

Les relations entre ces duplications segmentaires et un grand nombre de pathologies font l'objet de quantité d'articles de la littérature de ces cinq dernières années. Les duplications segmentaires pourraient en effet sous-tendre un grand nombre de variations phénotypiques, comme nous l'avons déjà indiqué, mais également de maladies génétiques (Bailey et al., 2002a; Bailey et al., 2002b).

2.4.1. Mécanistique

Les duplications segmentaires favorisent, de par la grande conservation de séquences entre les paralogues, et comme cela a été suggéré en introduction, des duplications, délétions, et inversions de segments génomiques par recombinaison homologue non allélique (Bailey and Eichler, 2006). Une exceptionnelle richesse en duplications segmentaires a de même été décrite aux bornes des blocs de synténie, sans que les duplications ne soient nécessairement la cause de ces réarrangements majeurs, comme l'indique la corrélation inattendue entre les sites des duplications segmentaires spécifiques des Primates et des réarrangements spécifiques des Rongeurs (Bailey et al., 2004). Les duplications segmentaires sont néanmoins susceptibles de créer de l'instabilité génomique spécifique en prédisposant certains chormosomes ou certaines régions chromosomiques à subir des réarrangements médiés par des crossing-over inégaux. Ces remaniements affectent ensuite naturellement la structure et l'expression des gènes et pourraient donc conduire au développement de pathologies (Bailey et al., 2004; Bailey et al., 2002a; Lupski, 1998a; Lupski, 1998b; Samonte and Eichler, 2002). Un certain nombre de désordres génétiques ont été classifiés dans les syndromes d'aneusomie segmentaire, et résultent de dosages inappropriés de gènes en un locus particulier, du fait des dites duplications segmentaires et selon des modalités largement décrites dans la littérature (Bailey and Eichler, 2006; Emanuel and Shaikh, 2001).

2.4.2. Exemples de pathologies associées aux duplications segmentaires

Ainsi, des données récentes impliquent de façon nette les duplications segmentaires, intrachromosomiques en particulier, dans l'étiologie des réarrangements chromosomiques associés à des désordres génomiques (Emanuel and Shaikh, 2001; Ji et al., 2000; Lupski, 1998b). Ces troubles affectent en outre uniquement les Primates, et les duplications qui les fondent sont apparues au cours de l'évolution de ces espèces, particulièrement dans les 30 derniers millions d'années.

Parmi les pathologies les mieux décrites du point de vue de leur association avec des duplications segmentaires, citons la maladie de Charcot-Marie-Tooth de type 1A [MIM 118220], localisée en 17p11.2 (Lupski, 1998a), qui est associée à une duplication simple interstitielle. Cette pathologie est décrite chez l'Homme et le Chimpanzé, mais pas chez le Gorille qui ne présente, de façon significative, qu'une copie unique de la séquence CMT1A, ce qui indique que la duplication est postérieure à la divergence des Hominoïdes (Samonte and Eichler, 2002). Les syndromes d'Angelman [MIM 105830] et de Prader-Willi [MIM 176270] seraient tous deux dus à une grande délétion de 4 Mb, médiée par des duplications segmentaires complexes, composées essentiellement de duplications du gène *HERC2*, en 15q11-15q13 (Amos-Landgraf et al., 1999). De façon similaire, le développement de la neurofibromatose de type 1 [MIM 162200] repose dans une minorité de cas (2 à 13% des patients) sur une délétion d'1,5 Mb en 17q11.2, associée à des duplications segmentaires composées de multiples répétitions de l'élément NF1-REPs (de 15 à 100 kb) et contenant quelques séquences transcrites et au moins un pseudogène (Dorschner et al., 2000). Le syndrome de Di George [MIM 188400] et le syndrome vélocardifacial [MIM 192430] tiennent également à une large déletion, située en 22q11.2. Cette délétion n'est cependant que l'un des nombreux réarrangements chromosomiques liés aux duplications segmentaires et identifiés en ce locus (Edelmann et al., 1999; Shaikh et al., 2000).

2.4.3. Qu'en est-il de *PMCHL* et *GUSL* ?

Il apparaît ainsi nettement qu'un certain nombre de duplications segmentaires, et par conséquent d'éléments qui les composent, sont associés à la mise en place de pathologies. Les gènes *PMCHL* et *GUSL* appartiennent clairement à des duplications segmentaires (de structures diverses, comme nous l'avons souligné précédemment) et certains des gènes paralogues de ces deux familles se situent en outre au locus de susceptibilité à l'amyotrophie spinale infantile (SMA) en 5q13.3 ou à proximité immédiate (Schmutz et al., 2004). On peut donc légitimement s'interroger sur la participation des gènes *PMCHL2* et *A4* à *A11* (de la famille *GUSL*) aux remaniements associés aux duplications segmentaires et à la pathologie SMA. Les gènes directement impliqués dans le développement de cette maladie, en l'occurrence les gènes *SMN1* et *SMN2*, et les variations alléliques qui les affectent sont précisément décrits (Melki et al., 1994; Monani et al., 1999; Ogino et al., 2004). Cela n'exclut pas, cependant, d'envisager une intervention, entre autres séquences de cette large duplication segmentaire, des éléments des familles *PMCHL* et *GUSL* présents en ce locus dans les remaniements susceptibles d'avoir conduit chez certains individus atteints par la maladie à la

perte de l'une ou l'autre des copies des gènes *SMN*, notamment pour les formes de type I de la pathologie (Ogino et al., 2004). Enfin, on peut imaginer que l'expression même, et donc potentiellement la fonction, des gènes *PMCHL2* et *A4* à *A11*, sont conditionnées par leur présence au sein de ces duplications segmentaires et par les remaniements conséquents.

3. Nouveaux gènes et transcription : les épissages alternatifs

L'analyse de l'expression des gènes *PMCHL* et *GUSL*, et les données bibliographiques concernant d'autres familles de gènes apparus récemment, mettent en évidence quelques caractères récurrents de l'expression de ces gènes. Il apparaît que leur profil transcriptionnel est souvent complexe, tant du point de vue de la qualité des ARNm que de leur répartition tissulaire et développementale.

Il est apparu au cours de nos études que les gènes *PMCHL* et *GUSL* présentent, pour tous ceux d'entre eux qui ont une capacité transcriptionnelle, des messagers alternativement épissés. Si seul un exon cassette est identifié dans la famille de gènes *GUSL*, de nombreux épissages alternatifs de différents types sont en revanche mis en évidence dans la famille des gènes *PMCHL*. Il me paraît donc judicieux de replacer ces données dans le contexte global des mécanismes et conséquences des épissages alternatifs.

3.1. Diversifier le protéome…

La division des gènes eucaryotes en introns et exons est une source potentielle de diversité. En effet, dès lors qu'un gène comporte plus de deux exons, des processus d'épissages alternatifs peuvent être mis en place. Ils constituent le mécanisme le plus utilisé pour augmenter la diversité protéique chez les eucaryotes supérieurs (Graveley, 2001; Smith and Valcarcel, 2000). La proportion de gènes humains subissant un épissage alternatif est estimée désormais à plus de 40% par des analyses d'EST (Modrek et al., 2001). Mais ces analyses sous-estiment vraisemblablement la fréquence de l'épissage alternatif. En outre, de nombreux évènements d'épissage alternatif sont très rares et n'adviennent que dans un type cellulaire ou à un stade développemental précis, ce qui indique une régulation fine et précise de l'épissage alternatif (Modrek et al., 2001; Smith and Valcarcel, 2000). Tous ces biais suggèrent qu'au moins la moitié des gènes humains subit en réalité un épissage alternatif. Il faut également considérer le nombre d'isoformes d'ARNm pouvant être produits à partir d'un unique gène par épissage alternatif. A titre d'exemple, extrême certes, 38 016 ARN messagers distincts peuvent être synthétisés à partir du gène *Dscam* (*Drosophila down syndrome cell adhesion molecule*) de la Drosophile (Schmucker et al., 2000). L'épissage alternatif de ce gène est en outre régulé tout au long du développement, et de façon spécifique du tissu

(Celotto and Graveley, 2001). Les transcrits sont ensuite maturés et traduits. Cet exemple témoigne ainsi de l'importance de l'épissage alternatif, et dans certains cas de l'édition de l'ARNm, dans la diversification et l'accroissement de la taille du protéome (Graveley, 2001; Modrek et al., 2001; Smith and Valcarcel, 2000). Ainsi, une fraction significative des gènes des eucaryotes supérieurs pourrait donner naissance à plus d'une protéine, ce qui pourrait expliquer le paradoxe de la relation entre nombre de gènes et complexité d'un organisme (Graveley, 2001).

En outre, et de manière générale, il apparaît que la très grande majorité des messagers alternatifs identifiés conduisent à une fonctionnalité pertinente d'une point de vue biologique et hautement spécifique, et ce, aussi bien pour les messagers issus de gènes conservés que de nouveaux gènes spécifiques des Primates (Modrek et al., 2001).

3.2. … mais pas seulement !

Nous avons montré comment l'épissage alternatif contribue à la diversification et à l'expansion du protéome des eucaryotes supérieurs. Mais ce mécanisme versatile de régulation de l'expression des gènes peut également altérer l'expression desdits gènes par introduction d'un codon stop prématuré (Smith and Valcarcel, 2000). Ceci se produirait dans 19% des cas d'épissage alternatif, du fait d'un décalage de phase de lecture consécutif à l'épissage alternatif (Modrek et al., 2001).

Il apparaît en outre que 15% des pathologies génétiques humaines sont causées par des mutations affectant des sites d'épissages fonctionnels ou en créant de nouveaux (Smith and Valcarcel, 2000).

Enfin, il a été proposé que l'épissage alternatif pourrait également toucher des régions ou des messagers non traduits et affecter ainsi la localisation cellulaire de l'ARN messager ou sa stabilité. Plus d'un quart des épissages alternatifs concernent en effet des sites localisés dans les régions non traduites (22% en 5'UTR et 4% en 3' UTR) (Modrek et al., 2001).

3.3. Qu'en est-il pour les transcrits PMCHL et GUSL ?

Nous avons mis en évidence un épissage alternatif des messagers GUSL, excluant l'exon 5. Cet épissage est identifié à tous les stades développementaux et dans tous les tissus analysés. En outre, les formes épissées sont relativement abondantes, notamment dans les tissus centraux. Nous montrons par ailleurs que les messagers présumés constitutifs des gènes

GUSL codent potentiellement pour une protéine. L'exclusion de l'exon 5 par épissage alternatif conduirait alors à une protéine de plus petite taille, mais conservant les extrémités N et C-terminales de la protéine supposée constitutive, puisqu'elle maintient le cadre de lecture. Un grand nombre d'expériences restent à effectuer pour déterminer l'exacte fonction de ces protéines putatives, mais il est probable que cet épissage alternatif conduirait à une invalidation de la protéine constitutive, plus qu'à une diversification du protéome. La protéine issue de la traduction du messager épissé est en effet de petite taille (moins de 70 acides aminés) et le seul domaine peptidique identifié dans la protéine constitutive en est absent. Enfin, l'hypothèse d'un épissage alternatif altérant les messagers eux-mêmes ne peut, à ce stade, être exclue.

Les épissages alternatifs décrits pour les messagers PMCHL sont plus nombreux et moins conventionnels, utilisant des sites d'épissages non consensuels, parfois même au sein d'exons constitutifs. Ils n'aboutissent en outre en aucun cas à une éventuelle modification protéomique. Il s'agirait donc vraisemblablement de la génération d'une grande diversité de messagers, susceptibles d'être adressés et stabilisés de manière différentielle. De façon tout à fait significative, cette grande diversité des isoformes transcriptionnelles de *PMCHL* chez l'Homme est essentiellement concentrée dans les testicules. De manière générale, une grande diversité de transcrits, de différents gènes et issus d'épissages alternatifs, est observée dans ce tissu. Ainsi, de ce point de vue, les gènes PMCHL semblent davantage en recherche de fonction(s), fussent-elle(s) directement médiée(s) par les ARN messagers.

CONCLUSION GENERALE

Les processus évolutifs des génomes retiennent l'attention de la communauté scientifique du fait de leur implication dans l'évolution des espèces. En particulier, de façon tout à fait anthropocentrique, l'évolution des Primates fait l'objet de nombreuses études. Ainsi, les gènes spécifiques des Primates pourraient constituer l'un des éléments déterminants des particularités de ces espèces.

Mes travaux de thèse ont principalement porté sur l'étude de deux familles de gènes « chimériques » spécifiques des Primates. Leurs structures, leurs histoires évolutives et leurs expressions ont été examinées, et de nouveaux outils ont été développés à ces fins. Nous pouvons ainsi proposer deux modèles évolutifs tout à fait originaux, reposant sur des mécanismes distincts et présentant des caractéristiques structurales et transcriptionnelles différentes. Dans le contexte actuel des séquençages de génomes entiers et d'analyses des gènes chimériques à grande échelle, les études portant sur ces gènes constituent également deux des quelques exemples approfondis et complets de gènes spécifiques des Primates. Ces travaux contribuent ainsi à une meilleure compréhension de l'évolution des Primates.

Différentes perspectives s'ouvrent alors.

L'étude des familles de génes « chimériques » spécifiques des Primates *PMCHL* et *GUSL* est et sera poursuivie afin d'examiner plus précisément leurs capacités transcriptionnelles et traductionnelles. L'expression de ces gènes sera également analysée dans le contexte de différentes pathologies spécifiques de l'Homme, telles que les maladies neurodégénératives d'Alzheimer et de Parkinson.

D'autres familles de gènes de ce type seront étudiées et finement détaillées. Le choix de ces modèles de gènes spécifiques des Primates s'effectuera en collaboration avec différentes équipes de recherche européennes dans le cadre du projet de recherche APES.

Finalement et globalement, les caractéristiques des gènes spécifiques des Primates (dont le nombre est estimé à moins d'un millier) seront examinées d'un point de vue structural, comme cela a été initié par l'étude des gènes chimériques, mais également du point de vue de leur expression. Ce dernier type de donnée fait en effet cruellement défaut dans la littérature actuelle.

BIBLIOGRAPHIE

(2004) Finishing the euchromatic sequence of the human genome. *Nature*, **431**, 931-945.

Abrao, M.S., Castrucci, A.M., Hadley, M.E. and Hruby, V.J. (1991) Protein-kinase C mediates MCH signal transduction in teleost, synbranchus marmoratus, melanocytes. *Pigment Cell Res*, **4**, 66-70.

Adams, K.L., Cronn, R., Percifield, R. and Wendel, J.F. (2003) Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proc Natl Acad Sci U S A*, **100**, 4649-4654.

Adkins, R.M., Walton, A.H. and Honeycutt, R.L. (2003) Higher-level systematics of rodents and divergence time estimates based on two congruent nuclear genes. *Mol Phylogenet Evol*, **26**, 409-420.

Alba, M.M. and Castresana, J. (2005) Inverse relationship between evolutionary rate and age of mammalian genes. *Mol Biol Evol*, **22**, 598-606.

Amor, D.J. and Choo, K.H. (2002) Neocentromeres: role in human disease, evolution, and centromere study. *Am J Hum Genet*, **71**, 695-714.

Amos-Landgraf, J.M., Ji, Y., Gottlieb, W., Depinet, T., Wandstrat, A.E., Cassidy, S.B., Driscoll, D.J., Rogan, P.K., Schwartz, S. and Nicholls, R.D. (1999) Chromosome breakage in the Prader-Willi and Angelman syndromes involves recombination between large, transcribed repeats at proximal and distal breakpoints. *Am J Hum Genet*, **65**, 370-386.

Andersson, J.O. (2005) Lateral gene transfer in eukaryotes. *Cell Mol Life Sci*, **62**, 1182-1197.

Andersson, T.P., Skold, H.N. and Svensson, S.P. (2003) Phosphoinositide 3-kinase is involved in Xenopus and Labrus melanophore aggregation. *Cell Signal*, **15**, 1119-1127.

Andrews, P. (1992) Evolution and environment in the Hominoidea. *Nature*, **360**, 641-646.

Anxolabéhère, D., Nouaud, D. and Miller, W.J. (2000) Eléments transposables et nouveautés génétiques chez les eucaryotes. *Médecine/Sciences*.

Anzai, T., Shiina, T., Kimura, N., Yanagiya, K., Kohara, S., Shigenari, A., Yamagata, T., Kulski, J.K., Naruse, T.K., Fujimori, Y., Fukuzumi, Y., Yamazaki, M., Tashiro, H., Iwamoto, C., Umehara, Y., Imanishi, T., Meyer, A., Ikeo, K., Gojobori, T., Bahram, S. and Inoko, H. (2003) Comparative sequencing of human and chimpanzee MHC class I regions unveils insertions/deletions as the major path to genomic divergence. *Proc Natl Acad Sci U S A*, **100**, 7708-7713.

Apic, G., Gough, J. and Teichmann, S.A. (2001) Domain combinations in archaeal, eubacterial and eukaryotic proteomes. *J Mol Biol*, **310**, 311-325.

Arney, K.L. (2003) H19 and Igf2--enhancing the confusion? *Trends Genet*, **19**, 17-23.

Arratia, G. (2000) Phylogenetic relationships of Teleostei. Past and present. *Estud Oceanol*, 19-51.

Audinot, V., Lahaye, C., Suply, T., Rovere-Jovene, C., Rodriguez, M., Nicolas, J.P., Beauverger, P., Cardinaud, B., Galizzi, J.P., Fauchere, J.L., Nahon, J.L. and Boutin, J.A. (2002) SVK14 cells express an MCH binding site different from the MCH1 or MCH2 receptor. *Biochem Biophys Res Commun*, **295**, 841-848.

Bailey, J.A., Baertsch, R., Kent, W.J., Haussler, D. and Eichler, E.E. (2004) Hotspots of mammalian chromosomal evolution. *Genome Biol*, **5**, R23.

Bailey, J.A. and Eichler, E.E. (2006) Primate segmental duplications: crucibles of evolution, diversity and disease. *Nat Rev Genet*, **7**, 552-564.

Bailey, J.A., Gu, Z., Clark, R.A., Reinert, K., Samonte, R.V., Schwartz, S., Adams, M.D., Myers, E.W., Li, P.W. and Eichler, E.E. (2002a) Recent segmental duplications in the human genome. *Science*, **297**, 1003-1007.

Bailey, J.A., Yavor, A.M., Viggiano, L., Misceo, D., Horvath, J.E., Archidiacono, N., Schwartz, S., Rocchi, M. and Eichler, E.E. (2002b) Human-specific duplication and mosaic transcripts: the recent paralogous structure of chromosome 22. *Am J Hum Genet*, **70**, 83-100.

Bailey, W.J. (1993) Hominoid trichotomy: a molecular overview. *Evolutionnary Anthropology*, **2**, 100-108.

Baker, B. and Kawauchi, H. (1997) *MCH in non-mammalian vertebrates: neuronal topography and functions. in: Knigge K, Prased A, Preteland S, Wagner JE, editors. MCH and seizures: neuromolecular and neuroendocrine aspects.*

Balakirev, E.S. and Ayala, F.J. (2003) Pseudogenes: are they "junk" or functional DNA? *Annu Rev Genet*, **37**, 123-151.

Barriel, V. (2004) Ces 1,4 % qui nous séparent des chimpanzés! *Médecine/Sciences*.

Batzer, M.A. and Deininger, P.L. (2002) Alu repeats and human genomic diversity. *Nat Rev Genet*, **3**, 370-379.

Beard, K.C. (1993) Phylogenetic systematics of the Primatomorpha, with special reference to Dermoptera. *mammals Phylogeny*.

Bedell, J., Korf, I. and Yandell, M. (2003) *BLAST*. O'Reilly & Associates.

Bednarek, M.A., Feighner, S.D., Hreniuk, D.L., Palyha, O.C., Morin, N.R., Sadowski, S.J., MacNeil, D.J., Howard, A.D. and Van der Ploeg, L.H. (2001) Short segment of human melanin-concentrating hormone that is sufficient for full activation of human melanin-concentrating hormone receptors 1 and 2. *Biochemistry*, **40**, 9379-9386.

Benton, M.J. and Ayala, F.J. (2003) Dating the tree of life. *Science*, **300**, 1698-1700.

Bezard, E., Imbert, C., Deloire, X., Bioulac, B. and Gross, C.E. (1997) A chronic MPTP model reproducing the slow evolution of Parkinson's disease: evolution of motor symptoms in the monkey. *Brain Res*, **766**, 107-112.

Bird, D., Baker, B., Eberle, A. and Swann, R. (1990) The biosynthesis of melanin-concentrating hormone in a fish. *J Neuroendocrinol*, 309-315.

Bittencourt, J.C., Presse, F., Arias, C., Peto, C., Vaughan, J., Nahon, J.L., Vale, W. and Sawchenko, P.E. (1992) The melanin-concentrating hormone system of the rat brain: an immuno- and hybridization histochemical characterization. *J Comp Neurol*, **319**, 218-245.

Blin, N. and Stafford, D.W. (1976) A general method for isolation of high molecular weight DNA from eukaryotes. *Nucleic Acids Res*, **3**, 2303-2308.

Bluet-Pajot, M.T., Presse, F., Voko, Z., Hoeger, C., Mounier, F., Epelbaum, J. and Nahon, J.L. (1995) Neuropeptide-E-I antagonizes the action of melanin-concentrating hormone on stress-induced release of adrenocorticotropin in the rat. *J Neuroendocrinol*, **7**, 297-303.

Bordner, A.J. and Abagyan, R. (2005) REVCOM: a robust Bayesian method for evolutionary rate estimation. *Bioinformatics*, **21**, 2315-2321.

Borowsky, B., Durkin, M.M., Ogozalek, K., Marzabadi, M.R., DeLeon, J., Lagu, B., Heurich, R., Lichtblau, H., Shaposhnik, Z., Daniewska, I., Blackburn, T.P., Branchek, T.A., Gerald, C., Vaysse, P.J. and Forray, C. (2002) Antidepressant, anxiolytic and anorectic effects of a melanin-concentrating hormone-1 receptor antagonist. *Nat Med*, **8**, 825-830.

Borsu, L., Presse, F. and Nahon, J.L. (2000) The AROM gene, spliced mRNAs encoding new DNA/RNA-binding proteins are transcribed from the opposite strand of the melanin-concentrating hormone gene in mammals. *J Biol Chem*, **275**, 40576-40587.

Bradley, R.L., Mansfield, J.P., Maratos-Flier, E. and Cheatham, B. (2002) Melanin-concentrating hormone activates signaling pathways in 3T3-L1 adipocytes. *Am J Physiol Endocrinol Metab*, **283**, E584-592.

Breton, C., Schorpp, M. and Nahon, J.L. (1993) Isolation and characterization of the human melanin-concentrating hormone gene and a variant gene. *Brain Res Mol Brain Res*, **18**, 297-310.

Britten, R.J. (2002) Divergence between samples of chimpanzee and human DNA sequences is 5%, counting indels. *Proc Natl Acad Sci U S A*, **99**, 13633-13635.

Britten, R.J., Rowen, L., Williams, J. and Cameron, R.A. (2003) Majority of divergence between closely related DNA samples is due to indels. *Proc Natl Acad Sci U S A*, **100**, 4661-4665.

Brosius, J. (1991) Retroposons--seeds of evolution. *Science*, **251**, 753.

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

Brown, J.R. (2003) Ancient horizontal gene transfer. *Nat Rev Genet*, **4**, 121-132.

Burgaud, J.L., Poosti, R., Fehrentz, J.A., Martinez, J. and Nahon, J.L. (1997) Melanin-concentrating hormone binding sites in human SVK14 keratinocytes. *Biochem Biophys Res Commun*, **241**, 622-629.

Burki, F. and Kaessmann, H. (2004) Birth and adaptive evolution of a hominoid gene that supports high neurotransmitter flux. *Nat Genet*, **36**, 1061-1063.

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. (1999) The dynamics of chromosome evolution in birds and mammals. *Nature*, **402**, 411-413.

Caceres, M., Lachuer, J., Zapala, M.A., Redmond, J.C., Kudo, L., Geschwind, D.H., Lockhart, D.J., Preuss, T.M. and Barlow, C. (2003) Elevated gene expression levels distinguish human from non-human primate brains. *Proc Natl Acad Sci U S A*, **100**, 13030-13035.

Capy, P., Bazin, C., Higuet, D., Langin, T. and. (1997) *Dynamics and Evolution of Transposable Elements*. Landes Biosciences, Austin, TX.

Cardinaud, B., Darre-Toulemonde, F., Duhault, J., Boutin, J.A. and Nahon, J.L. (2004) Comparative analysis of melanin-concentrating hormone structure and activity in fishes and mammals. *Peptides*, **25**, 1623-1632.

Celotto, A.M. and Graveley, B.R. (2001) Alternative splicing of the Drosophila Dscam pre-mRNA is both temporally and spatially regulated. *Genetics*, **159**, 599-608.

Chambers, J., Ames, R.S., Bergsma, D., Muir, A., Fitzgerald, L.R., Hervieu, G., Dytko, G.M., Foley, J.J., Martin, J., Liu, W.S., Park, J., Ellis, C., Ganguly, S., Konchar, S., Cluderay, J., Leslie, R., Wilson, S. and Sarau, H.M. (1999) Melanin-concentrating

hormone is the cognate ligand for the orphan G-protein-coupled receptor SLC-1. *Nature*, **400**, 261-265.

Chen, F.C. and Li, W.H. (2001) Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. *Am J Hum Genet*, **68**, 444-456.

Chen, Y., Hu, C., Hsu, C.K., Zhang, Q., Bi, C., Asnicar, M., Hsiung, H.M., Fox, N., Slieker, L.J., Yang, D.D., Heiman, M.L. and Shi, Y. (2002) Targeted disruption of the melanin-concentrating hormone receptor-1 results in hyperphagia and resistance to diet-induced obesity. *Endocrinology*, **143**, 2469-2477.

Cheung, B., Holmes, R.S., Easteal, S. and Beacham, I.R. (1999) Evolution of class I alcohol dehydrogenase genes in catarrhine primates: gene conversion, substitution rates, and gene regulation. *Mol Biol Evol*, **16**, 23-36.

Cheung, J., Estivill, X., Khaja, R., MacDonald, J.R., Lau, K., Tsui, L.C. and Scherer, S.W. (2003) Genome-wide detection of segmental duplications and potential assembly errors in the human genome sequence. *Genome Biol*, **4**, R25.

Chou, H.H., Hayakawa, T., Diaz, S., Krings, M., Indriati, E., Leakey, M., Paabo, S., Satta, Y., Takahata, N. and Varki, A. (2002) Inactivation of CMP-N-acetylneuraminic acid hydroxylase occurred prior to brain expansion during human evolution. *Proc Natl Acad Sci U S A*, **99**, 11736-11741.

Chou, H.H., Takematsu, H., Diaz, S., Iber, J., Nickerson, E., Wright, K.L., Muchmore, E.A., Nelson, D.L., Warren, S.T. and Varki, A. (1998) A mutation in human CMP-sialic acid hydroxylase occurred after the Homo-Pan divergence. *Proc Natl Acad Sci U S A*, **95**, 11751-11756.

Clark, A.G., Glanowski, S., Nielsen, R., Thomas, P.D., Kejariwal, A., Todd, M.A., Tanenbaum, D.M., Civello, D., Lu, F., Murphy, B., Ferriera, S., Wang, G., Zheng, X., White, T.J., Sninsky, J.J., Adams, M.D. and Cargill, M. (2003) Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios. *Science*, **302**, 1960-1963.

Claverie, J.M. (2001) Gene number. What if there are only 30,000 human genes? *Science*, **291**, 1255-1257.

Clegg, D.J., Air, E.L., Benoit, S.C., Sakai, R.S., Seeley, R.J. and Woods, S.C. (2003) Intraventricular melanin-concentrating hormone stimulates water intake independent of food intake. *Am J Physiol Regul Integr Comp Physiol*, **284**, R494-499.

Comeron, J.M. (2001) What controls the length of noncoding DNA? *Curr Opin Genet Dev*, **11**, 652-659.

Copeland, N.G., Jenkins, N.A. and Court, D.L. (2001) Recombineering: a powerful new tool for mouse functional genomics. *Nat Rev Genet*, **2**, 769-779.

Corbet, G.B. and Hill, J.E. (1991) *A world list of Mammalian Species*. Oxford University Press, Natural History Museum Publications.

Courseaux, A. and Nahon, J.L. (2001) Birth of two chimeric genes in the Hominidae lineage. *Science*, **291**, 1293-1297.

Courseaux, A., Richard, F., Grosgeorge, J., Ortola, C., Viale, A., Turc-Carel, C., Dutrillaux, B., Gaudray, P. and Nahon, J.L. (2003) Segmental duplications in euchromatic regions of human chromosome 5: a source of evolutionary instability and transcriptional innovation. *Genome Res*, **13**, 369-381.

Crosier, M., Viggiano, L., Guy, J., Misceo, D., Stones, R., Wei, W., Hearn, T., Ventura, M., Archidiacono, N., Rocchi, M. and Jackson, M.S. (2002) Human paralogs of KIAA0187 were created through independent pericentromeric-directed and chromosome-specific duplication mechanisms. *Genome Res*, **12**, 67-80.

Darre-Toulemonde, F. and Nahon, J.L. (2004) Ce que nous apprend le génome du Chimpanzé et des autres Primates. *Le Vivant*.

Day, R. and Gorr, S.U. (2003) Secretory granule biogenesis and chromogranin A: master gene, on/off switch or assembly factor? *Trends Endocrinol Metab*, **14**, 10-13.

Della-Zuana, O., Presse, F., Ortola, C., Duhault, J., Nahon, J.L. and Levens, N. (2002) Acute and chronic administration of melanin-concentrating hormone enhances food intake and body weight in Wistar and Sprague-Dawley rats. *Int J Obes Relat Metab Disord*, **26**, 1289-1295.

Deloukas, P., Earthrowl, M.E., Grafham, D.V., Rubenfield, M., French, L., Steward, C.A., Sims, S.K., Jones, M.C., Searle, S., Scott, C., Howe, K., Hunt, S.E., Andrews, T.D., Gilbert, J.G., Swarbreck, D., Ashurst, J.L., Taylor, A., Battles, J., Bird, C.P., Ainscough, R., Almeida, J.P., Ashwell, R.I., Ambrose, K.D., Babbage, A.K., Bagguley, C.L., Bailey, J., Banerjee, R., Bates, K., Beasley, H., Bray-Allen, S., Brown, A.J., Brown, J.Y., Burford, D.C., Burrill, W., Burton, J., Cahill, P., Camire, D., Carter, N.P., Chapman, J.C., Clark, S.Y., Clarke, G., Clee, C.M., Clegg, S., Corby, N., Coulson, A., Dhami, P., Dutta, I., Dunn, M., Faulkner, L., Frankish, A., Frankland, J.A., Garner, P., Garnett, J., Gribble, S., Griffiths, C., Grocock, R., Gustafson, E., Hammond, S., Harley, J.L., Hart, E., Heath, P.D., Ho, T.P., Hopkins, B., Horne, J., Howden, P.J., Huckle, E., Hynds, C., Johnson, C., Johnson, D., Kana, A., Kay, M., Kimberley, A.M., Kershaw, J.K., Kokkinaki, M., Laird, G.K., Lawlor, S., Lee, H.M., Leongamornlert, D.A., Laird, G., Lloyd, C., Lloyd, D.M., Loveland, J., Lovell, J., McLaren, S., McLay, K.E., McMurray, A., Mashreghi-Mohammadi, M., Matthews, L., Milne, S., Nickerson, T., Nguyen, M., Overton-Larty, E., Palmer, S.A., Pearce, A.V., Peck, A.I., Pelan, S., Phillimore, B., Porter, K., Rice, C.M., Rogosin, A., Ross, M.T., Sarafidou, T., Sehra, H.K., Shownkeen, R., Skuce, C.D., Smith, M., Standring, L., Sycamore, N., Tester, J., Thorpe, A., Torcasso, W., Tracey, A., Tromans, A., Tsolas, J., Wall, M., Walsh, J., Wang, H., Weinstock, K., West, A.P., Willey, D.L., Whitehead, S.L., Wilming, L., Wray, P.W., Young, L., Chen, Y., Lovering, R.C., Moschonas, N.K., Siebert, R., Fechtel, K., Bentley, D., Durbin, R., Hubbard, T., Doucette-Stamm, L., Beck, S., Smith, D.R. and Rogers, J. (2004) The DNA sequence and comparative analysis of human chromosome 10. *Nature*, **429**, 375-381.

Dewannieux, M., Esnault, C. and Heidmann, T. (2003) LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet*, **35**, 41-48.

Doolittle, W.F. and Sapienza, C. (1980) Selfish genes, the phenotype paradigm and genome evolution. *Nature*, **284**, 601-603.

Dorschner, M.O., Sybert, V.P., Weaver, M., Pletcher, B.A. and Stephens, K. (2000) NF1 microdeletion breakpoints are clustered at flanking repetitive sequences. *Hum Mol Genet*, **9**, 35-46.

Dorus, S., Vallender, E.J., Evans, P.D., Anderson, J.R., Gilbert, S.L., Mahowald, M., Wyckoff, G.J., Malcom, C.M. and Lahn, B.T. (2004) Accelerated evolution of nervous system genes in the origin of Homo sapiens. *Cell*, **119**, 1027-1040.

Douzery, E.J., Snell, E.A., Bapteste, E., Delsuc, F. and Philippe, H. (2004) The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils? *Proc Natl Acad Sci U S A*, **101**, 15386-15391.

Drozdz, R. and Eberle, A.N. (1995) Synthesis and iodination of human (phenylalanine 13, tyrosine 19) melanin-concentrating hormone for radioreceptor assay. *J Pept Sci*, **1**, 58-65.

Dutrillaux, B. (1980) Chromosomal evolution of the great apes and man. *J Reprod Fertil Suppl*, **Suppl 28**, 105-111.

Eberle, A. (1988) *Structure-activity relationships of the melanotropins. In: The Melanotropins.* Karger, Basel.

Ebersberger, I., Metzler, D., Schwarz, C. and Paabo, S. (2002) Genomewide comparison of DNA sequences between humans and chimpanzees. *Am J Hum Genet*, **70**, 1490-1497.

Edelmann, L., Pandita, R.K., Spiteri, E., Funke, B., Goldberg, R., Palanisamy, N., Chaganti, R.S., Magenis, E., Shprintzen, R.J. and Morrow, B.E. (1999) A common molecular basis for rearrangement disorders on chromosome 22q11. *Hum Mol Genet*, **8**, 1157-1167.

Eichler, E.E. (2001) Recent duplication, domain accretion and the dynamic mutation of the human genome. *Trends Genet*, **17**, 661-669.

Eichler, E.E., Johnson, M.E., Alkan, C., Tuzun, E., Sahinalp, C., Misceo, D., Archidiacono, N. and Rocchi, M. (2001) Divergent origins and concerted expansion of two segmental duplications on chromosome 16. *J Hered*, **92**, 462-468.

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

Elemento, O., Gascuel, O. and Lefranc, M.P. (2002) Reconstructing the duplication history of tandemly repeated genes. *Mol Biol Evol*, **19**, 278-288.

Emanuel, B.S. and Shaikh, T.H. (2001) Segmental duplications: an 'expanding' role in genomic instability and disease. *Nat Rev Genet*, **2**, 791-800.

Enard, W., Khaitovich, P., Klose, J., Zollner, S., Heissig, F., Giavalisco, P., Nieselt-Struwe, K., Muchmore, E., Varki, A., Ravid, R., Doxiadis, G.M., Bontrop, R.E. and Paabo, S.

(2002a) Intra- and interspecific variation in primate gene expression patterns. *Science*, **296**, 340-343.

Enard, W. and Paabo, S. (2004) Comparative primate genomics. *Annu Rev Genomics Hum Genet*, **5**, 351-378.

Enard, W., Przeworski, M., Fisher, S.E., Lai, C.S., Wiebe, V., Kitano, T., Monaco, A.P. and Paabo, S. (2002b) Molecular evolution of FOXP2, a gene involved in speech and language. *Nature*, **418**, 869-872.

Evans, P.D., Anderson, J.R., Vallender, E.J., Gilbert, S.L., Malcom, C.M., Dorus, S. and Lahn, B.T. (2004) Adaptive evolution of ASPM, a major determinant of cerebral cortical size in humans. *Hum Mol Genet*, **13**, 489-494.

Ezawa, K., S, O.O. and Saitou, N. (2006) Proceedings of the SMBE Tri-National Young Investigators' Workshop 2005. Genome-wide search of gene conversions in duplicated genes of mouse and rat. *Mol Biol Evol*, **23**, 927-940.

Fairbrother, W.G., Holste, D., Burge, C.B. and Sharp, P.A. (2004) Single nucleotide polymorphism-based validation of exonic splicing enhancers. *PLoS Biol*, **2**, E268.

Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L. and Postlethwait, J. (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*, **151**, 1531-1545.

Forray, C. (2003) The MCH receptor family: feeding brain disorders? *Curr Opin Pharmacol*, **3**, 85-89.

Frazer, K.A., Chen, X., Hinds, D.A., Pant, P.V., Patil, N. and Cox, D.R. (2003) Genomic DNA insertions and deletions occur frequently between humans and nonhuman primates. *Genome Res*, **13**, 341-346.

Fredman, D., White, S.J., Potter, S., Eichler, E.E., Den Dunnen, J.T. and Brookes, A.J. (2004) Complex SNP-related sequence variation in segmental genome duplications. *Nat Genet*, **36**, 861-866.

Fujiyama, A., Watanabe, H., Toyoda, A., Taylor, T.D., Itoh, T., Tsai, S.F., Park, H.S., Yaspo, M.L., Lehrach, H., Chen, Z., Fu, G., Saitou, N., Osoegawa, K., de Jong, P.J., Suto, Y., Hattori, M. and Sakaki, Y. (2002) Construction and analysis of a human-chimpanzee comparative clone map. *Science*, **295**, 131-134.

Gagneux, P., Amess, B., Diaz, S., Moore, S., Patel, T., Dillmann, W., Parekh, R. and Varki, A. (2001) Proteomic comparison of human and great ape blood plasma reveals conserved glycosylation and differences in thyroid hormone metabolism. *Am J Phys Anthropol*, **115**, 99-109.

Gagneux, P., Cheriyan, M., Hurtado-Ziola, N., van der Linden, E.C., Anderson, D., McClure, H., Varki, A. and Varki, N.M. (2003) Human-specific regulation of alpha 2-6-linked sialic acids. *J Biol Chem*, **278**, 48245-48250.

Gagneux, P. and Varki, A. (2001) Genetic differences between humans and great apes. *Mol Phylogenet Evol*, **18**, 2-13.

Galtier, N., Gouy, M. and Gautier, C. (1996) SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. *Comput Appl Biosci*, **12**, 543-548.

Gascuel, O. (1997) BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol Biol Evol*, **14**, 685-695.

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

Gianfrancesco, F., Esposito, T., Casu, G., Maninchedda, G., Roberto, R. and Pirastu, M. (2004) Emergence of Talanin protein associated with human uric acid nephrolithiasis in the Hominidae lineage. *Gene*, **339**, 131-138.

Gibbons, A. (2002) Becoming human. In search of the first hominids. *Science*, **295**, 1214-1219.

Glazko, G.V. and Nei, M. (2003) Estimation of divergence times for major lineages of primate species. *Mol Biol Evol*, **20**, 424-434.

Gonzalez, E., Kulkarni, H., Bolivar, H., Mangano, A., Sanchez, R., Catano, G., Nibbs, R.J., Freedman, B.I., Quinones, M.P., Bamshad, M.J., Murthy, K.K., Rovin, B.H., Bradley, W., Clark, R.A., Anderson, S.A., O'Connell R, J., Agan, B.K., Ahuja, S.S., Bologna, R., Sen, L., Dolan, M.J. and Ahuja, S.K. (2005) The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science*, **307**, 1434-1440.

Gonzalez, M.I., Baker, B.I., Hole, D.R. and Wilson, C.A. (1998) Behavioral effects of neuropeptide E-I (NEI) in the female rat: interactions with alpha-MSH, MCH and dopamine. *Peptides*, **19**, 1007-1016.

Gonzalez, M.I., Baker, B.I. and Wilson, C.A. (1997) Stimulatory effect of melanin-concentrating hormone on luteinising hormone release. *Neuroendocrinology*, **66**, 254-262.

Goodman, M. (1999) The genomic record of Humankind's evolutionary roots. *Am J Hum Genet*, **64**, 31-39.

Goodman, M., Bailey, W.J., Hayasaka, K., Stanhope, M.J., Slightom, J. and Czelusniak, J. (1994) Molecular evidence on primate phylogeny from DNA sequences. *Am J Phys Anthropol*, **94**, 3-24.

Gorr, S.U., Huang, X.F., Cowley, D.J., Kuliawat, R. and Arvan, P. (1999) Disruption of disulfide bonds exhibits differential effects on trafficking of regulated secretory proteins. *Am J Physiol*, **277**, C121-131.

Gorr, S.U., Jain, R.K., Kuehn, U., Joyce, P.B. and Cowley, D.J. (2001) Comparative sorting of neuroendocrine secretory proteins: a search for common ground in a mosaic of sorting models and mechanisms. *Mol Cell Endocrinol*, **172**, 1-6.

Gould, S.J. (1970) Dollo on Dollo's law: irreversibility and the status of evolutionary laws. *J Hist Biol*, **3**, 189-212.

Govin, J. (2006) Réorganisation de l'épigénome au cours de la spermiogénèse. *Biologie*. Université Joseph Fourier, Grenoble, France, p. 326.

Gratacos, M., Nadal, M., Martin-Santos, R., Pujana, M.A., Gago, J., Peral, B., Armengol, L., Ponsa, I., Miro, R., Bulbena, A. and Estivill, X. (2001) A polymorphic genomic duplication on human chromosome 15 is a susceptibility factor for panic and phobic disorders. *Cell*, **106**, 367-379.

Graveley, B.R. (2001) L'épissage alternatif: diversifier le protéome. *Biofutur*, 39-43.

Gray, I.C. and Jeffreys, A.J. (1991) Evolutionary transience of hypervariable minisatellites in man and the primates. *Proc Biol Sci*, **243**, 241-253.

Gregory, T.R. (2001) Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. *Biol Rev Camb Philos Soc*, **76**, 65-101.

Griffond, B. and Baker, B.I. (2002) Cell and molecular cell biology of melanin-concentrating hormone. *Int Rev Cytol*, **213**, 233-277.

Groneveld, D., Balm, P.H. and Wendelaar Bonga, S.E. (1995) Identification, cellular localization and in vitro release of a novel teleost melanin-concentrating hormone gene-related peptide. *Neuroendocrinology*, **62**, 498-505.

Grossman, L.I., Wildman, D.E., Schmidt, T.R. and Goodman, M. (2004) Accelerated evolution of the electron transport chain in anthropoid primates. *Trends Genet*, **20**, 578-585.

Groves, C.P. (1993) Order Primates. *Mammal Species of the World, a Taxonomic and Geographis Reference*.

Gu, J. and Gu, X. (2003) Induced gene expression in human brain after the split from chimpanzee. *Trends Genet*, **19**, 63-65.

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

Hacia, J.G. (2001) Genome of the apes. *Trends Genet*, **17**, 637-645.

Harpending, H.C., Batzer, M.A., Gurven, M., Jorde, L.B., Rogers, A.R. and Sherry, S.T. (1998) Genetic traces of ancient demography. *Proc Natl Acad Sci U S A*, **95**, 1961-1967.

Harrison, P., Kumar, A., Lan, N., Echols, N., Snyder, M. and Gerstein, M. (2002) A small reservoir of disabled ORFs in the yeast genome and its implications for the dynamics of proteome evolution. *J Mol Biol*, **316**, 409-419.

Hauser, M.D., Chomsky, N. and Fitch, W.T. (2002) The faculty of language: what is it, who has it, and how did it evolve? *Science*, **298**, 1569-1579.

Hawes, B.E., Kil, E., Green, B., O'Neill, K., Fried, S. and Graziano, M.P. (2000) The melanin-concentrating hormone receptor couples to multiple G proteins to activate diverse intracellular signaling pathways. *Endocrinology*, **141**, 4524-4532.

Hellmann, I., Zollner, S., Enard, W., Ebersberger, I., Nickel, B. and Paabo, S. (2003) Selection on human genes as revealed by comparisons to chimpanzee cDNA. *Genome Res*, **13**, 831-837.

Hervieu, G. and Nahon, J.L. (1995) Pro-melanin concentrating hormone messenger ribonucleic acid and peptides expression in peripheral tissues of the rat. *Neuroendocrinology*, **61**, 348-364.

Hervieu, G.J., Cluderay, J.E., Harrison, D., Meakin, J., Maycox, P., Nasir, S. and Leslie, R.A. (2000) The distribution of the mRNA and protein products of the melanin-concentrating hormone (MCH) receptor gene, slc-1, in the central nervous system of the rat. *Eur J Neurosci*, **12**, 1194-1216.

Hirotsune, S., Yoshida, N., Chen, A., Garrett, L., Sugiyama, F., Takahashi, S., Yagami, K., Wynshaw-Boris, A. and Yoshiki, A. (2003) An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene. *Nature*, **423**, 91-96.

Ho, M.K. and Wong, Y.H. (2001) G(z) signaling: emerging divergence from G(i) signaling. *Oncogene*, **20**, 1615-1625.

Holland, P.W. (1999) Gene duplication: past, present and future. *Semin Cell Dev Biol*, **10**, 541-547.

Hoskins, R.A., Smith, C.D., Carlson, J.W., Carvalho, A.B., Halpern, A., Kaminker, J.S., Kennedy, C., Mungall, C.J., Sullivan, B.A., Sutton, G.G., Yasuhara, J.C., Wakimoto, B.T., Myers, E.W., Celniker, S.E., Rubin, G.M. and Karpen, G.H. (2002) Heterochromatic sequences in a Drosophila whole-genome shotgun assembly. *Genome Biol*, **3**, RESEARCH0085.

Hsieh, W.P., Chu, T.M., Wolfinger, R.D. and Gibson, G. (2003) Mixed-model reanalysis of primate data suggests tissue and species biases in oligonucleotide-based gene expression profiles. *Genetics*, **165**, 747-757.

Huby, T., Dachet, C., Lawn, R.M., Wickings, J., Chapman, M.J. and Thillet, J. (2001) Functional analysis of the chimpanzee and human apo(a) promoter sequences: identification of sequence variations responsible for elevated transcriptional activity in chimpanzee. *J Biol Chem*, **276**, 22209-22214.

Humphray, S.J., Oliver, K., Hunt, A.R., Plumb, R.W., Loveland, J.E., Howe, K.L., Andrews, T.D., Searle, S., Hunt, S.E., Scott, C.E., Jones, M.C., Ainscough, R., Almeida, J.P., Ambrose, K.D., Ashwell, R.I., Babbage, A.K., Babbage, S., Bagguley, C.L., Bailey, J., Banerjee, R., Barker, D.J., Barlow, K.F., Bates, K., Beasley, H., Beasley, O., Bird, C.P., Bray-Allen, S., Brown, A.J., Brown, J.Y., Burford, D., Burrill, W., Burton, J., Carder, C., Carter, N.P., Chapman, J.C., Chen, Y., Clarke, G., Clark, S.Y., Clee, C.M., Clegg, S., Collier, R.E., Corby, N., Crosier, M., Cummings, A.T., Davies, J., Dhami, P., Dunn, M., Dutta, I., Dyer, L.W., Earthrowl, M.E., Faulkner, L., Fleming, C.J., Frankish, A., Frankland, J.A., French, L., Fricker, D.G., Garner, P., Garnett, J., Ghori, J., Gilbert, J.G., Glison, C., Grafham, D.V., Gribble, S., Griffiths, C., Griffiths-Jones,

S., Grocock, R., Guy, J., Hall, R.E., Hammond, S., Harley, J.L., Harrison, E.S., Hart, E.A., Heath, P.D., Henderson, C.D., Hopkins, B.L., Howard, P.J., Howden, P.J., Huckle, E., Johnson, C., Johnson, D., Joy, A.A., Kay, M., Keenan, S., Kershaw, J.K., Kimberley, A.M., King, A., Knights, A., Laird, G.K., Langford, C., Lawlor, S., Leongamornlert, D.A., Leversha, M., Lloyd, C., Lloyd, D.M., Lovell, J., Martin, S., Mashreghi-Mohammadi, M., Matthews, L., McLaren, S., McLay, K.E., McMurray, A., Milne, S., Nickerson, T., Nisbett, J., Nordsiek, G., Pearce, A.V., Peck, A.I., Porter, K.M., Pandian, R., Pelan, S., Phillimore, B., Povey, S., Ramsey, Y., Rand, V., Scharfe, M., Sehra, H.K., Shownkeen, R., Sims, S.K., Skuce, C.D., Smith, M., Steward, C.A., Swarbreck, D., Sycamore, N., Tester, J., Thorpe, A., Tracey, A., Tromans, A., Thomas, D.W., Wall, M., Wallis, J.M., West, A.P., Whitehead, S.L., Willey, D.L., Williams, S.A., Wilming, L., Wray, P.W., Young, L., Ashurst, J.L., Coulson, A., Blocker, H., Durbin, R., Sulston, J.E., Hubbard, T., Jackson, M.J., Bentley, D.R., Beck, S., Rogers, J. and Dunham, I. (2004) DNA sequence and analysis of human chromosome 9. *Nature*, **429**, 369-374.

Jackson, M. (2003) Duplicate, decouple, disperse: the evolutionary transience of human centromeric regions. *Curr Opin Genet Dev*, **13**, 629-635.

Ji, Y., Eichler, E.E., Schwartz, S. and Nicholls, R.D. (2000) Structure of chromosomal duplicons and their role in mediating human genomic disorders. *Genome Res*, **10**, 597-610.

Johnson, M.E., Viggiano, L., Bailey, J.A., Abdul-Rauf, M., Goodwin, G., Rocchi, M. and Eichler, E.E. (2001) Positive selection of a gene family during the emergence of humans and African apes. *Nature*, **413**, 514-519.

Johnston, C.M., Nesterova, T.B., Formstone, E.J., Newall, A.E., Duthie, S.M., Sheardown, S.A. and Brockdorff, N. (1998) Developmentally regulated Xist promoter switch mediates initiation of X inactivation. *Cell*, **94**, 809-817.

Jolly, C.J. (2001) A proper study for mankind: Analogies from the Papionin monkeys and their implications for human evolution. *Am J Phys Anthropol*, **Suppl 33**, 177-204.

Jones, J.M. and Gellert, M. (2004) The taming of a transposon: V(D)J recombination and the immune system. *Immunol Rev*, **200**, 233-248.

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

Jorde, L.B., Rogers, A.R., Bamshad, M., Watkins, W.S., Krakowiak, P., Sung, S., Kere, J. and Harpending, H.C. (1997) Microsatellite diversity and the demographic history of modern humans. *Proc Natl Acad Sci U S A*, **94**, 3100-3103.

Jurka, J., Kohany, O., Pavlicek, A., Kapitonov, V.V. and Jurka, M.V. (2004) Duplication, coclustering, and selection of human Alu retrotransposons. *Proc Natl Acad Sci U S A*, **101**, 1268-1272.

Kaessmann, H. and Paabo, S. (2002) The genetical history of humans and the great apes. *J Intern Med*, **251**, 1-18.

Kaessmann, H., Zollner, S., Nekrutenko, A. and Li, W.H. (2002) Signatures of domain shuffling in the human genome. *Genome Res*, **12**, 1642-1650.

Kawauchi, H., Kawazoe, I., Tsubokawa, M., Kishida, M. and Baker, B.I. (1983) Characterization of melanin-concentrating hormone in chum salmon pituitaries. *Nature*, **305**, 321-323.

Kawazoe, I., Kawauchi, H., Hirano, T. and Naito, N. (1987) Structure-activity relationships of melanin-concentrating hormone. *Int J Pept Protein Res*, **29**, 714-721.

Kay, R.F., Ross, C. and Williams, B.A. (1997) Anthropoid origins. *Science*, **275**, 797-804.

Khaitovich, P., Enard, W., Lachmann, M. and Paabo, S. (2006) Evolution of primate gene expression. *Nat Rev Genet*, **7**, 693-702.

Khaitovich, P., Muetzel, B., She, X., Lachmann, M., Hellmann, I., Dietzsch, J., Steigele, S., Do, H.H., Weiss, G., Enard, W., Heissig, F., Arendt, T., Nieselt-Struwe, K., Eichler, E.E. and Paabo, S. (2004a) Regional patterns of gene expression in human and chimpanzee brains. *Genome Res*, **14**, 1462-1473.

Khaitovich, P., Weiss, G., Lachmann, M., Hellmann, I., Enard, W., Muetzel, B., Wirkner, U., Ansorge, W. and Paabo, S. (2004b) A neutral model of transcriptome evolution. *PLoS Biol*, **2**, E132.

Kilduff, T.S. and de Lecea, L. (2001) Mapping of the mRNAs for the hypocretin/orexin and melanin-concentrating hormone receptors: networks of overlapping peptide systems. *J Comp Neurol*, **435**, 1-5.

King, M.C. and Wilson, A.C. (1975) Evolution at two levels in humans and chimpanzees. *Science*, **188**, 107-116.

Knigge, K.M., Baxter-Grillo, D., Speciale, J. and Wagner, J. (1996) Melanotropic peptides in the mammalian brain: the melanin-concentrating hormone. *Peptides*, **17**, 1063-1073.

Knigge, K.M. and Wagner, J.E. (1997) Melanin-concentrating hormone (MCH) involvement in pentylenetetrazole (PTZ)-induced seizure in rat and guinea pig. *Peptides*, **18**, 1095-1097.

Kokkotou, E., Mastaitis, J.W., Qu, D., Hoersch, D., Slieker, L., Bonter, K., Tritos, N.A. and Maratos-Flier, E. (2000) Characterization of [Phe(13), Tyr(19)]-MCH analog binding activity to the MCH receptor. *Neuropeptides*, **34**, 240-247.

Kuryshev, V.Y., Skryabin, B.V., Kremerskothen, J., Jurka, J. and Brosius, J. (2001) Birth of a gene: locus of neuronal BC200 snmRNA in three prosimians and human BC200 pseudogenes as archives of change in the Anthropoidea lineage. *J Mol Biol*, **309**, 1049-1066.

Lai, C.S., Fisher, S.E., Hurst, J.A., Vargha-Khadem, F. and Monaco, A.P. (2001) A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature*, **413**, 519-523.

Lakaye, B., Minet, A., Zorzi, W. and Grisar, T. (1998) Cloning of the rat brain cDNA encoding for the SLC-1 G protein-coupled receptor reveals the presence of an intron in the gene. *Biochim Biophys Acta*, **1401**, 216-220.

Lebl, M., Hruby, V.J., Castrucci, A.M. and Hadley, M.E. (1989) Melanin concentrating hormone analogues: contraction of the cyclic structure. II. Antagonist activity. *Life Sci*, **44**, 451-457.

Lebl, M., Hruby, V.J., Castrucci, A.M., Visconti, M.A. and Hadley, M.E. (1988) Melanin concentrating hormone analogues: contraction of the cyclic structure. 1. Agonist activity. *J Med Chem*, **31**, 949-954.

Lee, J.T. (2003) Molecular biology: Complicity of gene and pseudogene. *Nature*, **423**, 26-28.

Lehner, B., Williams, G., Campbell, R.D. and Sanderson, C.M. (2002) Antisense transcripts in the human genome. *Trends Genet*, **18**, 63-65.

Liao, D. (1999) Concerted evolution: molecular mechanism and biological implications. *Am J Hum Genet*, **64**, 24-30.

Liu, G., Zhao, S., Bailey, J.A., Sahinalp, S.C., Alkan, C., Tuzun, E., Green, E.D. and Eichler, E.E. (2003) Analysis of primate genomic variation reveals a repeat-driven expansion of the human genome. *Genome Res*, **13**, 358-368.

Logan, D.W., Bryson-Richardson, R.J., Pagan, K.E., Taylor, M.S., Currie, P.D. and Jackson, I.J. (2003) The structure and evolution of the melanocortin and MCH receptors in fish and mammals. *Genomics*, **81**, 184-191.

Long, M. (2000) A new function evolved from gene fusion. *Genome Res*, **10**, 1655-1657.

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

Long, M., Betran, E., Thornton, K. and Wang, W. (2003a) The origin of new genes: glimpses from the young and old. *Nat Rev Genet*, **4**, 865-875.

Long, M., Deutsch, M., Wang, W., Betran, E., Brunet, F.G. and Zhang, J. (2003b) Origin of new genes: evidence from experimental and computational analyses. *Genetica*, **118**, 171-182.

Lopez, P., Casane, D. and Philippe, H. (2002) Bio-informatique (5): Phylogénie et évolution moléculaires. *Médecine/Sciences*.

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

Ludwig, D.S., Tritos, N.A., Mastaitis, J.W., Kulkarni, R., Kokkotou, E., Elmquist, J., Lowell, B., Flier, J.S. and Maratos-Flier, E. (2001) Melanin-concentrating hormone overexpression in transgenic mice leads to obesity and insulin resistance. *J Clin Invest*, **107**, 379-386.

Lupski, J.R. (1998a) Charcot-Marie-Tooth disease: lessons in genetic mechanisms. *Mol Med*, **4**, 3-11.

Lupski, J.R. (1998b) Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends Genet*, **14**, 417-422.

Makalowski, W. (2000) Genomic scrap yard: how genomes utilize all that junk. *Gene*, **259**, 61-67.

Marcus, G.F. and Fisher, S.E. (2003) FOXP2 in focus: what can genes tell us about speech and language? *Trends Cogn Sci*, **7**, 257-262.

Marques, A.C., Dupanloup, I., Vinckenbosch, N., Reymond, A. and Kaessmann, H. (2005) Emergence of young human genes after a burst of retroposition in primates. *PLoS Biol*, **3**, e357.

Marsh, D.J., Weingarth, D.T., Novi, D.E., Chen, H.Y., Trumbauer, M.E., Chen, A.S., Guan, X.M., Jiang, M.M., Feng, Y., Camacho, R.E., Shen, Z., Frazier, E.G., Yu, H., Metzger, J.M., Kuca, S.J., Shearman, L.P., Gopal-Truter, S., MacNeil, D.J., Strack, A.M., MacIntyre, D.E., Van der Ploeg, L.H. and Qian, S. (2002) Melanin-concentrating hormone 1 receptor-deficient mice are lean, hyperactive, and hyperphagic and have altered metabolism. *Proc Natl Acad Sci U S A*, **99**, 3240-3245.

Marshall, C.R., Raff, E.C. and Raff, R.A. (1994) Dollo's law and the death and resurrection of genes. *Proc Natl Acad Sci U S A*, **91**, 12283-12287.

Martensson, L.G. and Andersson, R.G. (2000) Is Ca2+ the second messenger in the response to melatonin in cuckoo wrasse melanophores? *Life Sci*, **66**, 1003-1010.

Martignetti, J.A. and Brosius, J. (1993) BC200 RNA: a neural RNA polymerase III product encoded by a monomeric Alu element. *Proc Natl Acad Sci U S A*, **90**, 11563-11567.

Maston, G.A. and Ruvolo, M. (2002) Chorionic gonadotropin has a recent origin within primates and an evolutionary history of selection. *Mol Biol Evol*, **19**, 320-335.

Matsunaga, T.O., Hruby, V.J., Lebl, M., Castrucci, A.M. and Hadley, M.E. (1989) Melanin concentrating hormone (MCH): structure-function aspects of its melanocyte stimulating hormone-like (MSH-like) activity. *Peptides*, **10**, 773-778.

Mattick, J.S. and Gagen, M.J. (2001) The evolution of controlled multitasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms. *Mol Biol Evol*, **18**, 1611-1630.

Mattick, J.S. and Makunin, I.V. (2005) Small regulatory RNAs in mammals. *Hum Mol Genet*, **14 Spec No 1**, R121-132.

Mayr, E. (1942) *Systematics and the Origin of Species*. Harvard edition.

Mayr, E. (1963) *Animal species and evolution*. Belknap press.

McClintock, B. (1956) *Mutation in maize.* Carnegie Institute of Washington.

McClintock, B. (1984) The significance of responses of the genome to challenge. *Science*, **226**, 792-801.

McLysaght, A., Hokamp, K. and Wolfe, K.H. (2002) Extensive genomic duplication during early chordate evolution. *Nat Genet*, **31**, 200-204.

Mekel-Bobrov, N., Gilbert, S.L., Evans, P.D., Vallender, E.J., Anderson, J.R., Hudson, R.R., Tishkoff, S.A. and Lahn, B.T. (2005) Ongoing adaptive evolution of ASPM, a brain size determinant in Homo sapiens. *Science*, **309**, 1720-1722.

Melki, J., Lefebvre, S., Burglen, L., Burlet, P., Clermont, O., Millasseau, P., Reboullet, S., Benichou, B., Zeviani, M., Le Paslier, D. and et al. (1994) De novo and inherited deletions of the 5q13 region in spinal muscular atrophies. *Science*, **264**, 1474-1477.

Mignone, F., Horner, D.S. and Pesole, G. (2004) WebVar: A resource for the rapid estimation of relative site variability from multiple sequence alignments. *Bioinformatics*, **20**, 1331-1333.

Miller, C.L., Hruby, V.J., Matsunaga, T.O. and Bickford, P.C. (1993) Alpha-MSH and MCH are functional antagonists in a CNS auditory gating paradigm. *Peptides*, **14**, 431-440.

Minth, C.D., Qiu, H., Akil, H., Watson, S.J. and Dixon, J.E. (1989) Two precursors of melanin-concentrating hormone: DNA sequence analysis and in situ immunochemical localization. *Proc Natl Acad Sci U S A*, **86**, 4292-4296.

Mitra, J., Tang, X., Almo, S.C. and Shields, D. (1998) Temperature-induced conformational changes in prosomatostatin-II: implications for processing. *Biochem J*, **334 (Pt 1)**, 275-282.

Modrek, B., Resch, A., Grasso, C. and Lee, C. (2001) Genome-wide detection of alternative splicing in expressed sequences of human genes. *Nucleic Acids Res*, **29**, 2850-2859.

Monani, U.R., Lorson, C.L., Parsons, D.W., Prior, T.W., Androphy, E.J., Burghes, A.H. and McPherson, J.D. (1999) A single nucleotide difference that alters splicing patterns distinguishes the SMA gene SMN1 from the copy gene SMN2. *Hum Mol Genet*, **8**, 1177-1183.

Monzon, M.E., de Souza, M.M., Izquierdo, L.A., Izquierdo, I., Barros, D.M. and de Barioglio, S.R. (1999) Melanin-concentrating hormone (MCH) modifies memory retention in rats. *Peptides*, **20**, 1517-1519.

Moran, J.V., DeBerardinis, R.J. and Kazazian, H.H., Jr. (1999) Exon shuffling by L1 retrotransposition. *Science*, **283**, 1530-1534.

Mori, M., Harada, M., Terao, Y., Sugo, T., Watanabe, T., Shimomura, Y., Abe, M., Shintani, Y., Onda, H., Nishimura, O. and Fujino, M. (2001) Cloning of a novel G protein-coupled receptor, SLT, a subtype of the melanin-concentrating hormone receptor. *Biochem Biophys Res Commun*, **283**, 1013-1018.

Mouchantaf, R., Kumar, U., Sulea, T. and Patel, Y.C. (2001) A conserved alpha-helix at the amino terminus of prosomatostatin serves as a sorting signal for the regulated secretory pathway. *J Biol Chem*, **276**, 26308-26316.

Murray, J.F., Adan, R.A., Walker, R., Baker, B.I., Thody, A.J., Nijenhuis, W.A., Yukitake, J. and Wilson, C.A. (2000) Melanin-concentrating hormone, melanocortin receptors and regulation of luteinizing hormone release. *J Neuroendocrinol*, **12**, 217-223.

Nadeau, J.H. and Taylor, B.A. (1984) Lengths of chromosomal segments conserved since divergence of man and mouse. *Proc Natl Acad Sci U S A*, **81**, 814-818.

Nahon, J. (2001) Des gènes "chimères" sont apparus dans la lignée des Hominidés: l'indice d'une spécificité génomique humaine? *Médecine/Sciences*, 411-413.

Nahon, J.L. (1994) The melanin-concentrating hormone: from the peptide to the gene. *Crit Rev Neurobiol*, **8**, 221-262.

Nahon, J.L. (2003) Birth of 'human-specific' genes during primate evolution. *Genetica*, **118**, 193-208.

Nahon, J.L. (2006) The melanocortins and melanin-concentrating hormone in the central regulation of feeding behavior and energy homeostasis. *C R Biol*, **329**, 623-638; discussion 653-625.

Nahon, J.L., Presse, F., Bittencourt, J.C., Sawchenko, P.E. and Vale, W. (1989) The rat melanin-concentrating hormone messenger ribonucleic acid encodes multiple putative neuropeptides coexpressed in the dorsolateral hypothalamus. *Endocrinology*, **125**, 2056-2065.

Navarro, A. and Barton, N.H. (2003) Chromosomal speciation and molecular divergence--accelerated evolution in rearranged chromosomes. *Science*, **300**, 321-324.

Nekrutenko, A. and Li, W.H. (2001) Transposable elements are found in a large number of human protein-coding genes. *Trends Genet*, **17**, 619-621.

Ogino, S., Wilson, R.B. and Gold, B. (2004) New insights on the evolution of the SMN1 and SMN2 region: simulation and meta-analysis for allele and haplotype frequency calculations. *Eur J Hum Genet*, **12**, 1015-1023.

Ohno, S. (1970) *Evolution by gene duplication*. Springer-Verlag, Berlin.

Ohta, T. (1987) Very slightly deleterious mutations and the molecular clock. *J Mol Evol*, **26**, 1-6.

Ohta, T. (2000) Evolution of gene families. *Gene*, **259**, 45-52.

Ono, M., Wada, C., Oikawa, I., Kawazoe, I. and Kawauchi, H. (1988) Structures of two kinds of mRNA encoding the chum salmon melanin-concentrating hormone. *Gene*, **71**, 433-438.

Orth, M. and Tabrizi, S.J. (2003) Models of Parkinson's disease. *Mov Disord*, **18**, 729-737.

Oshima, N., Nakamaru, N., Araki, S. and Sugimoto, M. (2001) Comparative analyses of the pigment-aggregating and -dispersing actions of MCH on fish chromatophores. *Comp Biochem Physiol C Toxicol Pharmacol*, **129**, 75-84.

Otto, S.P. and Whitton, J. (2000) Polyploid incidence and evolution. *Annu Rev Genet*, **34**, 401-437.

Panning, B., Dausman, J. and Jaenisch, R. (1997) X chromosome inactivation is mediated by Xist RNA stabilization. *Cell*, **90**, 907-916.

Parkes, D. and Vale, W. (1992) Secretion of melanin-concentrating hormone and neuropeptide-EI from cultured rat hypothalamic cells. *Endocrinology*, **131**, 1826-1831.

Patterson, N., Richter, D.J., Gnerre, S., Lander, E.S. and Reich, D. (2006) Genetic evidence for complex speciation of humans and chimpanzees. *Nature*, **441**, 1103-1108.

Patthy, L. (2003) Modular assembly of genes and the evolution of new functions. *Genetica*, **118**, 217-231.

Pedeutour, F., Szpirer, C. and Nahon, J.L. (1994) Assignment of the human pro-melanin-concentrating hormone gene (PMCH) to chromosome 12q23-q24 and two variant genes (PMCH1 and PMCHL2) to chromosome 5p14 and 5q12-q13. *Genomics*, **19**, 31-37.

Perry, G.H., Tchinda, J., McGrath, S.D., Zhang, J., Picker, S.R., Caceres, A.M., Iafrate, A.J., Tyler-Smith, C., Scherer, S.W., Eichler, E.E., Stone, A.C. and Lee, C. (2006) Hotspots for copy number variation in chimpanzees and humans. *Proc Natl Acad Sci U S A*, **103**, 8006-8011.

Pissios, P. and Maratos-Flier, E. (2003) Melanin-concentrating hormone: from fish skin to skinny mammals. *Trends Endocrinol Metab*, **14**, 243-248.

Presgraves, D.C., Balagopalan, L., Abmayr, S.M. and Orr, H.A. (2003) Adaptive evolution drives divergence of a hybrid inviability gene between two species of Drosophila. *Nature*, **423**, 715-719.

Presse, F., Nahon, J.L., Fischer, W.H. and Vale, W. (1990) Structure of the human melanin concentrating hormone mRNA. *Mol Endocrinol*, **4**, 632-637.

Prince, V.E. and Pickett, F.B. (2002) Splitting pairs: the diverging fates of duplicated genes. *Nat Rev Genet*, **3**, 827-837.

Ptak, S.E. and Przeworski, M. (2002) Evidence for population growth in humans is confounded by fine-scale population structure. *Trends Genet*, **18**, 559-563.

Qu, D., Ludwig, D.S., Gammeltoft, S., Piper, M., Pelleymounter, M.A., Cullen, M.J., Mathes, W.F., Przypek, R., Kanarek, R. and Maratos-Flier, E. (1996) A role for melanin-concentrating hormone in the central regulation of feeding behaviour. *Nature*, **380**, 243-247.

Roberts, G.C. and Smith, C.W. (2002) Alternative splicing: combinatorial output from the genome. *Curr Opin Chem Biol*, **6**, 375-383.

Rodionov, V., Yi, J., Kashina, A., Oladipo, A. and Gross, S.P. (2003) Switching between microtubule- and actin-based transport systems in melanophores is controlled by cAMP levels. *Curr Biol*, **13**, 1837-1847.

Rodriguez, M., Beauverger, P., Naime, I., Rique, H., Ouvry, C., Souchaud, S., Dromaint, S., Nagel, N., Suply, T., Audinot, V., Boutin, J.A. and Galizzi, J.P. (2001) Cloning and molecular characterization of the novel human melanin-concentrating hormone receptor MCH2. *Mol Pharmacol*, **60**, 632-639.

Rogers, J.H. (1985) The origin and evolution of retroposons. *Int Rev Cytol*, **93**, 187-279.

Rosenberg, M.S. and Kumar, S. (2003) Heterogeneity of nucleotide frequencies among evolutionary lineages and phylogenetic inference. *Mol Biol Evol*, **20**, 610-621.

Rowen, L., Koop, B.F. and Hood, L. (1996) The complete 685-kilobase DNA sequence of the human beta T cell receptor locus. *Science*, **272**, 1755-1762.

Rubinsztein, D.C., Amos, W., Leggo, J., Goodburn, S., Jain, S., Li, S.H., Margolis, R.L., Ross, C.A. and Ferguson-Smith, M.A. (1995a) Microsatellite evolution--evidence for directionality and variation in rate between species. *Nat Genet*, **10**, 337-343.

Rubinsztein, D.C., Leggo, J. and Amos, W. (1995b) Microsatellites evolve more rapidly in humans than in chimpanzees. *Genomics*, **30**, 610-612.

Ruvolo, M. (2004) Comparative primate genomics: the year of the chimpanzee. *Curr Opin Genet Dev*, **14**, 650-656.

Saito, Y., Nothacker, H.P., Wang, Z., Lin, S.H., Leslie, F. and Civelli, O. (1999) Molecular characterization of the melanin-concentrating-hormone receptor. *Nature*, **400**, 265-269.

Saito, Y., Wang, Z., Hagino-Yamagishi, K., Civelli, O., Kawashima, S. and Maruyama, K. (2001) Endogenous melanin-concentrating hormone receptor SLC-1 in human melanoma SK-MEL-37 cells. *Biochem Biophys Res Commun*, **289**, 44-50.

Sammak, P.J., Adams, S.R., Harootunian, A.T., Schliwa, M. and Tsien, R.Y. (1992) Intracellular cyclic AMP not calcium, determines the direction of vesicle movement in melanophores: direct measurement by fluorescence ratio imaging. *J Cell Biol*, **117**, 57-72.

Samonte, R.V. and Eichler, E.E. (2002) Segmental duplications and the evolution of the primate genome. *Nat Rev Genet*, **3**, 65-72.

Sanchez, M., Baker, B.I. and Celis, M. (1997) Melanin-concentrating hormone (MCH) antagonizes the effects of alpha-MSH and neuropeptide E-I on grooming and locomotor activities in the rat. *Peptides*, **18**, 393-396.

Sargent, C.A., Chalmers, I.J., Leversha, M. and Affara, N.A. (1994) A rearrangement on chromosome 5 of an expressed human beta-glucuronidase pseudogene. *Mamm Genome*, **5**, 791-796.

Schlumberger, S.E., Jaggin, V., Tanner, H. and Eberle, A.N. (2002a) Endogenous receptor for melanin-concentrating hormone in human neuroblastoma Kelly cells. *Biochem Biophys Res Commun*, **298**, 54-59.

Schlumberger, S.E., Talke-Messerer, C., Zumsteg, U. and Eberle, A.N. (2002b) Expression of receptors for melanin-concentrating hormone (MCH) in different tissues and cell lines. *J Recept Signal Transduct Res*, **22**, 509-531.

Schmucker, D., Clemens, J.C., Shu, H., Worby, C.A., Xiao, J., Muda, M., Dixon, J.E. and Zipursky, S.L. (2000) Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell*, **101**, 671-684.

Schmutz, J. and Grimwood, J. (2004) Genomes: fowl sequence. *Nature*, **432**, 679-680.

Schmutz, J., Martin, J., Terry, A., Couronne, O., Grimwood, J., Lowry, S., Gordon, L.A., Scott, D., Xie, G., Huang, W., Hellsten, U., Tran-Gyamfi, M., She, X., Prabhakar, S., Aerts, A., Altherr, M., Bajorek, E., Black, S., Branscomb, E., Caoile, C., Challacombe, J.F., Chan, Y.M., Denys, M., Detter, J.C., Escobar, J., Flowers, D., Fotopulos, D., Glavina, T., Gomez, M., Gonzales, E., Goodstein, D., Grigoriev, I., Groza, M., Hammon, N., Hawkins, T., Haydu, L., Israni, S., Jett, J., Kadner, K., Kimball, H., Kobayashi, A., Lopez, F., Lou, Y., Martinez, D., Medina, C., Morgan, J., Nandkeshwar, R., Noonan, J.P., Pitluck, S., Pollard, M., Predki, P., Priest, J., Ramirez, L., Retterer, J., Rodriguez, A., Rogers, S., Salamov, A., Salazar, A., Thayer, N., Tice, H., Tsai, M., Ustaszewska, A., Vo, N., Wheeler, J., Wu, K., Yang, J., Dickson, M., Cheng, J.F., Eichler, E.E., Olsen, A., Pennacchio, L.A., Rokhsar, D.S., Richardson, P., Lucas, S.M., Myers, R.M. and Rubin, E.M. (2004) The DNA sequence and comparative analysis of human chromosome 5. *Nature*, **431**, 268-274.

Segal-Lieberman, G., Bradley, R.L., Kokkotou, E., Carlson, M., Trombly, D.J., Wang, X., Bates, S., Myers, M.G., Jr., Flier, J.S. and Maratos-Flier, E. (2003) Melanin-concentrating hormone is a critical mediator of the leptin-deficient phenotype. *Proc Natl Acad Sci U S A*, **100**, 10085-10090.

Senut, B. (1999) *Les Primates, ancêtres de l'Homme: Des Grands Singes à l'Homme*. Editions ArtCom'.

Seoighe, C. (2003) Turning the clock back on ancient genome duplication. *Curr Opin Genet Dev*, **13**, 636-643.

Shaikh, T.H., Kurahashi, H., Saitta, S.C., O'Hare, A.M., Hu, P., Roe, B.A., Driscoll, D.A., McDonald-McGinn, D.M., Zackai, E.H., Budarf, M.L. and Emanuel, B.S. (2000) Chromosome 22-specific low copy repeats and the 22q11.2 deletion syndrome: genomic organization and deletion endpoint analysis. *Hum Mol Genet*, **9**, 489-501.

Sharp, A.J., Locke, D.P., McGrath, S.D., Cheng, Z., Bailey, J.A., Vallente, R.U., Pertz, L.M., Clark, R.A., Schwartz, S., Segraves, R., Oseroff, V.V., Albertson, D.G., Pinkel, D. and Eichler, E.E. (2005) Segmental duplications and copy-number variation in the human genome. *Am J Hum Genet*, **77**, 78-88.

She, X., Jiang, Z., Clark, R.A., Liu, G., Cheng, Z., Tuzun, E., Church, D.M., Sutton, G., Halpern, A.L. and Eichler, E.E. (2004) Shotgun sequence assembly and recent segmental duplications within the human genome. *Nature*, **431**, 927-930.

She, X., Liu, G., Ventura, M., Zhao, S., Misceo, D., Roberto, R., Cardone, M.F., Rocchi, M., Green, E.D., Archidiacano, N. and Eichler, E.E. (2006) A preliminary comparative analysis of primate segmental duplications shows elevated substitution rates and a great-ape expansion of intrachromosomal duplications. *Genome Res*, **16**, 576-583.

Sherbrooke, W.C. and Hadley, M.E. (1988) Exploring the evolutionary history of melanin-concentrating and melanin-stimulating hormone receptors on melanophores: neopterygian (holostean) and chondrostean fishes. *Pigment Cell Res*, **1**, 344-349.

Shimada, M., Tritos, N.A., Lowell, B.B., Flier, J.S. and Maratos-Flier, E. (1998) Mice lacking melanin-concentrating hormone are hypophagic and lean. *Nature*, **396**, 670-674.

Sikela, J.M. (2006) The jewels of our genome: the search for the genomic changes underlying the evolutionarily unique capacities of the human brain. *PLoS Genet*, **2**, e80.

Singer, S.S., Mannel, D.N., Hehlgans, T., Brosius, J. and Schmitz, J. (2004) From "junk" to gene: curriculum vitae of a primate receptor isoform gene. *J Mol Biol*, **341**, 883-886.

Skofitsch, G., Jacobowitz, D.M. and Zamir, N. (1985) Immunohistochemical localization of a melanin concentrating hormone-like peptide in the rat brain. *Brain Res Bull*, **15**, 635-649.

Skryabin, B.V., Kremerskothen, J., Vassilacopoulou, D., Disotell, T.R., Kapitonov, V.V., Jurka, J. and Brosius, J. (1998) The BC200 RNA gene and its neural expression are conserved in Anthropoidea (Primates). *J Mol Evol*, **47**, 677-685.

Sleutels, F., Zwart, R. and Barlow, D.P. (2002) The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature*, **415**, 810-813.

Smith, C.W. and Valcarcel, J. (2000) Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem Sci*, **25**, 381-388.

Smith, G.P. (1974) Unequal crossover and the evolution of multigene families. *Cold Spring Harb Symp Quant Biol*, **38**, 507-513.

Soltis, D.E. and Soltis, P.S. (1999) Polyploidy: recurrent formation and genome evolution. *Trends In Ecology And Evolution*, **14**, 348-352.

Stankiewicz, P., Shaw, C.J., Withers, M., Inoue, K. and Lupski, J.R. (2004) Serial segmental duplications during primate evolution result in complex human genome architecture. *Genome Res*, **14**, 2209-2220.

Stedman, H.H., Kozyak, B.W., Nelson, A., Thesier, D.M., Su, L.T., Low, D.W., Bridges, C.R., Shrager, J.B., Minugh-Purvis, N. and Mitchell, M.A. (2004) Myosin gene mutation correlates with anatomical changes in the human lineage. *Nature*, **428**, 415-418.

Storz, G. (2002) An expanding universe of noncoding RNAs. *Science*, **296**, 1260-1263.

Strachan, T. and Read, A. (1996) *Human Molecular Genetics*. BIOS scientific publishers, Oxford, UK.

Suply, T., Della Zuana, O., Audinot, V., Rodriguez, M., Beauverger, P., Duhault, J., Canet, E., Galizzi, J.P., Nahon, J.L., Levens, N. and Boutin, J.A. (2001) SLC-1 receptor mediates effect of melanin-concentrating hormone on feeding behavior in rat: a structure-activity study. *J Pharmacol Exp Ther*, **299**, 137-146.

Svensson, S.P., Norberg, T., Andersson, R.G., Grundstrom, N. and Karlsson, J.O. (1991) MCH-induced pigment aggregation in teleost melanophores is associated with a cAMP reduction. *Life Sci*, **48**, 2043-2046.

Tadayyon, M., Welters, H.J., Haynes, A.C., Cluderay, J.E. and Hervieu, G. (2000) Expression of melanin-concentrating hormone receptors in insulin-producing cells: MCH stimulates insulin release in RINm5F and CRI-G1 cell-lines. *Biochem Biophys Res Commun*, **275**, 709-712.

Takekawa, S., Asami, A., Ishihara, Y., Terauchi, J., Kato, K., Shimomura, Y., Mori, M., Murakoshi, H., Kato, K., Suzuki, N., Nishimura, O. and Fujino, M. (2002) T-226296: a novel, orally active and selective melanin-concentrating hormone receptor antagonist. *Eur J Pharmacol*, **438**, 129-135.

Tan, C.P., Sano, H., Iwaasa, H., Pan, J., Sailer, A.W., Hreniuk, D.L., Feighner, S.D., Palyha, O.C., Pong, S.S., Figueroa, D.J., Austin, C.P., Jiang, M.M., Yu, H., Ito, J., Ito, M., Ito, M., Guan, X.M., MacNeil, D.J., Kanatani, A., Van der Ploeg, L.H. and Howard, A.D. (2002) Melanin-concentrating hormone receptor subtypes 1 and 2: species-specific gene expression. *Genomics*, **79**, 785-792.

Taylor, J.S., Van de Peer, Y. and Meyer, A. (2001) Genome duplication, divergent resolution and speciation. *Trends Genet*, **17**, 299-301.

Taylor, M.S., Kai, C., Kawai, J., Carninci, P., Hayashizaki, Y. and Semple, C.A. (2006) Heterotachy in mammalian promoter evolution. *PLoS Genet*, **2**, e30.

Teaford, M.F. and Ungar, P.S. (2000) Diet and the evolution of the earliest human ancestors. *Proc Natl Acad Sci U S A*, **97**, 13506-13511.

Thanaraj, T.A. and Clark, F. (2001) Human GC-AG alternative intron isoforms with weak donor sites show enhanced consensus at acceptor exon positions. *Nucleic Acids Res*, **29**, 2581-2593.

Theodosiou, A.M., Morrison, K.E., Nesbit, A.M., Daniels, R.J., Campbell, L., Francis, M.J., Christodoulou, Z. and Davies, K.E. (1994) Complex repetitive arrangements of gene sequence in the candidate region of the spinal muscular atrophy gene in 5q13. *Am J Hum Genet*, **55**, 1209-1217.

Thomas, H. (1999) *Les Primates, ancêtres de l'Homme: Les premiers Primates*. Editions ArtCom'.

Thomson, T.M., Lozano, J.J., Loukili, N., Carrio, R., Serras, F., Cormand, B., Valeri, M., Diaz, V.M., Abril, J., Burset, M., Merino, J., Macaya, A., Corominas, M. and Guigo, R. (2000) Fusion of the human gene for the polyubiquitination coeffector UEV1 with Kua, a newly identified gene. *Genome Res*, **10**, 1743-1756.

TomHon, C., Zhu, W., Millinoff, D., Hayasaka, K., Slightom, J.L., Goodman, M. and Gumucio, D.L. (1997) Evolution of a fetal expression pattern via cis changes near the gamma globin gene. *J Biol Chem*, **272**, 14062-14066.

Tsukamura, H., Thompson, R.C., Tsukahara, S., Ohkura, S., Maekawa, F., Moriyama, R., Niwa, Y., Foster, D.L. and Maeda, K. (2000) Intracerebroventricular administration of melanin-concentrating hormone suppresses pulsatile luteinizing hormone release in the female rat. *J Neuroendocrinol*, **12**, 529-534.

Tuma, M.C. and Gelfand, V.I. (1999) Molecular mechanisms of pigment transport in melanophores. *Pigment Cell Res*, **12**, 283-294.

Uddin, M., Wildman, D.E., Liu, G., Xu, W., Johnson, R.M., Hof, P.R., Kapatos, G., Grossman, L.I. and Goodman, M. (2004) Sister grouping of chimpanzees and humans as revealed by genome-wide phylogenetic analysis of brain gene expression profiles. *Proc Natl Acad Sci U S A*, **101**, 2957-2962.

Varki, A. (2004) How to make an ape brain. *Nat Genet*, **36**, 1034-1036.

Varki, A. and Altheide, T.K. (2005) Comparing the human and chimpanzee genomes: searching for needles in a haystack. *Genome Res*, **15**, 1746-1758.

Vaughan, J.M., Fischer, W.H., Hoeger, C., Rivier, J. and Vale, W. (1989) Characterization of melanin-concentrating hormone from rat hypothalamus. *Endocrinology*, **125**, 1660-1665.

Verret, L., Goutagny, R., Fort, P., Cagnon, L., Salvert, D., Leger, L., Boissard, R., Salin, P., Peyron, C. and Luppi, P.H. (2003) A role of melanin-concentrating hormone producing neurons in the central regulation of paradoxical sleep. *BMC Neurosci*, **4**, 19.

Viale, A., Courseaux, A., Presse, F., Ortola, C., Breton, C., Jordan, D. and Nahon, J.L. (2000) Structure and expression of the variant melanin-concentrating hormone genes: only PMCHL1 is transcribed in the developing human brain and encodes a putative protein. *Mol Biol Evol*, **17**, 1626-1640.

Viale, A., Ortola, C., Hervieu, G., Furuta, M., Barbero, P., Steiner, D.F., Seidah, N.G. and Nahon, J.L. (1999) Cellular localization and role of prohormone convertases in the processing of pro-melanin concentrating hormone in mammals. *J Biol Chem*, **274**, 6536-6545.

Viale, A., Ortola, C., Richard, F., Vernier, P., Presse, F., Schilling, S., Dutrillaux, B. and Nahon, J.L. (1998a) Emergence of a brain-expressed variant melanin-concentrating hormone gene during higher primate evolution: a gene "in search of a function". *Mol Biol Evol*, **15**, 196-214.

Viale, A., Ortola, C., Vernier, P., Breton, C., Presse, F. and Nahon, J.L. (1998b) Structure, expression, and evolution of the variant MCH gene in primates. *Ann N Y Acad Sci*, **839**, 214-218.

Viale, A., Zhixing, Y., Breton, C., Pedeutour, F., Coquerel, A., Jordan, D. and Nahon, J.L. (1997) The melanin-concentrating hormone gene in human: flanking region analysis,

fine chromosome mapping, and tissue-specific expression. *Brain Res Mol Brain Res*, **46**, 243-255.

Wall, J.D. (2003) Estimating ancestral population sizes and divergence times. *Genetics*, **163**, 395-404.

Wang, Y.Q. and Su, B. (2004) Molecular evolution of microcephalin, a gene determining human brain size. *Hum Mol Genet*, **13**, 1131-1137.

Watanabe, H., Fujiyama, A., Hattori, M., Taylor, T.D., Toyoda, A., Kuroki, Y., Noguchi, H., BenKahla, A., Lehrach, H., Sudbrak, R., Kube, M., Taenzer, S., Galgoczy, P., Platzer, M., Scharfe, M., Nordsiek, G., Blocker, H., Hellmann, I., Khaitovich, P., Paabo, S., Reinhardt, R., Zheng, H.J., Zhang, X.L., Zhu, G.F., Wang, B.F., Fu, G., Ren, S.X., Zhao, G.P., Chen, Z., Lee, Y.S., Cheong, J.E., Choi, S.H., Wu, K.M., Liu, T.T., Hsiao, K.J., Tsai, S.F., Kim, C.G., S, O.O., Kitano, T., Kohara, Y., Saitou, N., Park, H.S., Wang, S.Y., Yaspo, M.L. and Sakaki, Y. (2004) DNA sequence and comparative analysis of chimpanzee chromosome 22. *Nature*, **429**, 382-388.

Webster, A.J., Payne, R.J. and Pagel, M. (2003) Molecular phylogenies link rates of evolution and speciation. *Science*, **301**, 478.

Weissenbach, J. (2004) Genome sequencing: differences with the relatives. *Nature*, **429**, 353-355.

Wildman, D.E., Uddin, M., Liu, G., Grossman, L.I. and Goodman, M. (2003) Implications of natural selection in shaping 99.4% nonsynonymous DNA identity between humans and chimpanzees: enlarging genus Homo. *Proc Natl Acad Sci U S A*, **100**, 7181-7188.

Wilhelm, M. and Wilhelm, F.X. (2001) Reverse transcription of retroviruses and LTR retrotransposons. *Cell Mol Life Sci*, **58**, 1246-1262.

Wolfe, K.H. (2001) Yesterday's polyploids and the mystery of diploidization. *Nat Rev Genet*, **2**, 333-341.

Wu, C. and Morris, J.R. (2001) Genes, genetics, and epigenetics: a correspondence. *Science*, **293**, 1103-1105.

Wu, C.I. and Ting, C.T. (2004) Genes and speciation. *Nat Rev Genet*, **5**, 114-122.

Yang, Z. (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci*, **13**, 555-556.

Yi, P., Zhang, W., Zhai, Z., Miao, L., Wang, Y. and Wu, M. (2003) Bcl-rambo beta, a special splicing variant with an insertion of an Alu-like cassette, promotes etoposide- and Taxol-induced cell death. *FEBS Lett*, **534**, 61-68.

Zdobnov, E.M., von Mering, C., Letunic, I. and Bork, P. (2005) Consistency of genome-based methods in measuring Metazoan evolution. *FEBS Lett*, **579**, 3355-3361.

Zhang, J., Rosenberg, H.F. and Nei, M. (1998) Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc Natl Acad Sci U S A*, **95**, 3708-3713.

Zhang, J., Wang, X. and Podlaha, O. (2004) Testing the chromosomal speciation hypothesis for humans and chimpanzees. *Genome Res*, **14**, 845-851.

Zietkiewicz, E., Richer, C., Makalowski, W., Jurka, J. and Labuda, D. (1994) A young Alu subfamily amplified independently in human and African great apes lineages. *Nucleic Acids Res*, **22**, 5608-5612.

Ref Peptides NE PAS DELETER, jeter après impression (ou passage à pdf)

(Abrao et al., 1991; Adams et al., 2003; Adkins et al., 2003; Alba and Castresana, 2005; Amor and Choo, 2002; Andersson, 2005; Andersson et al., 2003; Andrews, 1992; Anxolabéhère et al., 2000; Apic et al., 2001; Arratia, 2000; Audinot et al., 2002; Bailey and Eichler, 2006; Bailey et al., 2002b; Bailey, 1993; Baker and Kawauchi, 1997; Balakirev and Ayala, 2003; Barriel, 2004; Batzer and Deininger, 2002; Beard, 1993; Bednarek et al., 2001; Benton and Ayala, 2003; Bird et al., 1990; Bittencourt et al., 1992; Blin and Stafford, 1976; Bluet-Pajot et al., 1995; Borowsky et al., 2002; Borsu et al., 2000; Bradley et al., 2002; Breton et al., 1993; Brosius, 1991; Brosius, 2003; Brown, 2003; Burgaud et al., 1997; Burt et al., 1999; Caceres et al., 2003; Capy et al., 1997; Cardinaud et al., 2004; Chambers et al., 1999; Chen and Li, 2001; Chen et al., 2002; Chou et al., 2002; Chou et al., 1998; Clark et al., 2003; Claverie, 2001; Clegg et al., 2003; Comeron, 2001; Corbet and Hill, 1991; Courseaux and Nahon, 2001; Courseaux et al., 2003; Day and Gorr, 2003; Della-Zuana et al., 2002; Dewannieux et al., 2003; Doolittle and Sapienza, 1980; Dorus et al., 2004; Douzery et al., 2004; Drozdz and Eberle, 1995; Eberle, 1988; Ebersberger et al., 2002; Eichler, 2001; Eichler and Sankoff, 2003; Elemento et al., 2002; Enard et al., 2002a; Enard and Paabo, 2004; Enard et al., 2002b; Evans et al., 2004; Forray, 2003; Frazer et al., 2003; Gagneux et al., 2001; Gagneux et al., 2003; Gagneux and Varki, 2001; Galtier et al., 1996; Genereux and Logsdon, 2003; Gianfrancesco et al., 2004; Gibbons, 2002; Glazko and Nei, 2003; Gonzalez et al., 2005; Gonzalez et al., 1998; Gonzalez et al., 1997; Goodman, 1999; Goodman et al., 1994; Gorr et al., 1999; Gorr et al., 2001; Govin, 2006; Gray and Jeffreys, 1991; Gregory, 2001; Griffond and Baker, 2002; Groneveld et al., 1995; Groves, 1993; Gu and Gu, 2003; Gu et al., 2002; Hacia, 2001; Harrison et al., 2002; Hawes et al., 2000; Hellmann et al., 2003; Hervieu and Nahon, 1995; Hervieu et al., 2000; Hirotsune et al., 2003; Ho and Wong, 2001; Holland, 1999; Hoskins et al., 2002; Jackson, 2003; Jones and Gellert, 2004; Jones et al., 1992; Kaessmann et al., 2002; Kawauchi et al., 1983; Kawazoe et al., 1987; Khaitovich et al., 2004a; Kilduff and de Lecea, 2001; Knigge et al., 1996; Knigge and Wagner, 1997; Kokkotou et al., 2000; Kuryshev et al., 2001; Lai et al., 2001; Lakaye et al., 1998; Lebl et al., 1989; Lebl et al., 1988; Lee, 2003; Liao, 1999; Logan et al., 2003; Long, 2000; Long, 2001; Long et al., 2003a; Long et al., 2003b; Lopez et al., 2002; Luchetta et al., 2005; Ludwig et al., 2001; Makalowski, 2000; Marques et al., 2005; Marsh et al., 2002; Martensson and Andersson, 2000; Martignetti and Brosius, 1993; Matsunaga et al., 1989; Mayr, 1942; Mayr, 1963; McClintock, 1956; McClintock, 1984; McLysaght et al., 2002; Melki et al., 1994; Miller et al., 1993; Minth et al., 1989; Mitra et al., 1998; Modrek et al., 2001; Monani et al., 1999; Monzon et al., 1999; Moran et al., 1999; Mori et al., 2001; Mouchantaf et al., 2001; Murray et al., 2000; Nadeau and Taylor, 1984; Nahon, 1994; Nahon, 2003; Nahon, 2006; Nahon et al., 1989; Nekrutenko and Li, 2001; Ohno, 1970; Ohta, 1987; Ohta, 2000; Ono et al., 1988; Orth and Tabrizi, 2003; Oshima et al., 2001; Otto and Whitton, 2000; Parkes and Vale, 1992; Patthy, 2003; Pedeutour et al., 1994; Pissios and Maratos-Flier, 2003; Presgraves et al., 2003; Presse et al., 1990; Prince and Pickett, 2002; Qu et al., 1996; Roberts and Smith, 2002; Rodionov et al., 2003; Rodriguez et al., 2001; Rogers, 1985; Rubinsztein et al., 1995a; Rubinsztein et al., 1995b; Ruvolo, 2004; Saito et al., 1999; Saito et al., 2001; Sammak et al., 1992; Samonte and Eichler, 2002; Sanchez et al., 1997; Sargent et al., 1994; Schlumberger et al., 2002a; Schlumberger et al., 2002b; Schmutz and Grimwood, 2004; Schmutz et al., 2004; Segal-Lieberman et al., 2003; Senut, 1999; Seoighe, 2003; Sherbrooke and Hadley, 1988; Shimada et al., 1998;

Sikela, 2006; Skofitsch et al., 1985; Skryabin et al., 1998; Smith, 1974; Soltis and Soltis, 1999; Stankiewicz et al., 2004; Stedman et al., 2004; Storz, 2002; Strachan and Read, 1996; Suply et al., 2001; Svensson et al., 1991; Tadayyon et al., 2000; Takekawa et al., 2002; Tan et al., 2002; Taylor et al., 2001; Thanaraj and Clark, 2001; Theodosiou et al., 1994; Thomas, 1999; Thomson et al., 2000; Tsukamura et al., 2000; Tuma and Gelfand, 1999; Varki and Altheide, 2005; Vaughan et al., 1989; Verret et al., 2003; Viale et al., 2000; Viale et al., 1998a; Viale et al., 1998b; Viale et al., 1997; Webster et al., 2003; Wilhelm and Wilhelm, 2001; Wolfe, 2001; Wu and Morris, 2001; Wu and Ting, 2004; Zdobnov et al., 2005; Zhang et al., 1998; Zietkiewicz et al., 1994)

LE VIVANT

(Anzai et al., 2003; Britten, 2002; Britten et al., 2003; Deloukas et al., 2004; Fairbrother et al., 2004; Fujiyama et al., 2002; Hauser et al., 2002; Humphray et al., 2004; Jolly, 2001; Jurka et al., 2004; Kaessmann and Paabo, 2002; Khaitovich et al., 2004b; Liu et al., 2003; Marcus and Fisher, 2003; Nahon, 2001; Navarro and Barton, 2003; Ptak and Przeworski, 2002; She et al., 2004; Singer et al., 2004; Teaford and Ungar, 2000; Watanabe et al., 2004; Weissenbach, 2004; Yi et al., 2003; Zhang et al., 2004)

PMCHL

(Arney, 2003; Bordner and Abagyan, 2005; Burki and Kaessmann, 2004; Cheung et al., 1999; Eichler et al., 2001; Grossman et al., 2004; Hsieh et al., 2003; Huby et al., 2001; Johnston et al., 1998; Khaitovich et al., 2006; King and Wilson, 1975; Maston and Ruvolo, 2002; Mattick and Gagen, 2001; Mattick and Makunin, 2005; Mekel-Bobrov et al., 2005; Mignone et al., 2004; Panning et al., 1997; Patterson et al., 2006; Rosenberg and Kumar, 2003; Sleutels et al., 2002; Taylor et al., 2006; TomHon et al., 1997; Uddin et al., 2004; Varki, 2004; Wall, 2003; Wang and Su, 2004; Wildman et al., 2003; Yang, 1997)

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Peptides 25 (2004) 1623–1632
PEPTIDES
www.elsevier.com/locate/peptides

Comparative analysis of melanin-concentrating hormone structure and activity in fishes and mammals

Bruno Cardinaud[a,1], Fleur Darré-Toulemonde[a,1], Jacques Duhault[b], Jean A. Boutin[c], Jean-Louis Nahon[a,*]

[a] *Institut de Pharmacologie Moléculaire et Cellulaire, UMR 6097, 660 route des Lucioles, Sophia Antipolis, Valbonne 06560, France*
[b] *Division Métabolisme, Institut de Recherche Servier, Suresnes, France*
[c] *Division de Pharmacologie Moléculaire et Cellulaire, Institut de Recherche Servier, Suresnes, France*

Received 21 February 2004; accepted 26 May 2004
Available online 11 September 2004

Abstract

A comparative analysis of the structure of the melanin-concentrating hormone (MCH) precursor reveals that this sequence has been subjected to a higher selection pressure in mammals than in teleosts, suggesting that the structural constraints have not been the same throughout the vertebrate lineage. In contrast, the MCH peptide sequence has been very well conserved in all species. A sensitive and reproducible eel skin assay was developed and allowed us to define the structural features needed for a full MCH bioactivity. It was shown that the minimal structure carrying the critical residues was the same in fishes and in mammals. A pharmacological approach confirmed that MCH receptor activation decreased the cAMP levels in the fish skin, but this effect appeared to be independent from a Gαi protein. We propose that one of the intracellular signaling pathways of the MCH receptor in fish skin is the activation of one or several cellular phosphodiesterases.
© 2004 Elsevier Inc. All rights reserved.

Keywords: MCH; MCH receptor; Pro-MCH gene; Pro-peptide gene evolution; MCH signaling

1. Introduction

Melanin-concentrating hormone (MCH) is a cyclic peptide that is found in the brain of many vertebrate taxa, the most studied being teleosts and mammals (recently reviewed in [32,35]). The peptide was originally discovered on the basis of its pigment aggregation property in salmon fish skin and characterized as an heptadecapeptide [38]. The ability of MCH to aggregate melanin is not confined to teleost fishes since holostean fishes also respond to this peptide [77]. However, there is no evidence that MCH plays a major role as a neurohypophysial hormone in color change other than in neopterygians, suggesting therefore that its skin paling function evolved later and uniquely in a group ancestral to teleosts and holosteans (reviewed in [32]). In mammals, MCH is a nonadecapeptide that displays strong sequence conservation in the ring structure with the salmon counterpart [87]. In the rat brain, most MCH perikarya are located in sub-zona incerta (ZI) and lateral hypothalamus area (LHA) and project widely throughout the brain, from the olfactory bulb to the spinal cord and with particular abundance in hypothalamic areas, parts of the extrapyramidal system and the reticular formation of the brainstem [11,79]. This broad distribution of MCH projections suggested early on that the peptide could be involved in many brain functions [11]. The recent studies using intracerebroventricular injections of MCH agonists/antagonists [14,23,67,83] and MCH transgenic mouse models [48,76,78] support a major role for MCH in the central regulation of feeding behavior as well as energy homeostasis. However, MCH could be also implicated in other goal-oriented behaviors, such as drinking [21], sleeping [88], reproduction [29,58,85], sensorimotor-integration [42,52] and general arousal [55].

* Corresponding author. Tel.: +33 4 93 95 7753; fax: +33 4 93 95 7708.
E-mail address: nahonjl@ipmc.cnrs.fr (J.-L. Nahon).
[1] Equal contributors.

0196-9781/$ – see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.peptides.2004.05.020

Detailed reviews on the functions of MCH in fishes and mammals may be consulted [8,32,35,41,59].

MCH is generated after cleavage at a dibasic amino acid site in the C-terminal part of a precursor of 132 amino acids (AA) in salmonids [53,61,62] and 165 AA in mammals [60,66]. Apart from MCH, additional peptides could be liberated from pro-MCH. Indeed, cleavage at a $Arg^{99}Arg^{100}$ site may generate a 13 AA-peptide named Mch-gene-related peptide in trout (Mgrp; [10]) or neuropeptide (N)-glutamic acid(E)-valine(V) in coho salmon [61]. The largest Mgrp, of 22 AA, is produced in tilapia [33]. In mammals a 13 AA amidated peptide named neuropeptide (N)-glutamic acid(E)-isoleucine(I) is produced together with MCH in hypothalamic neurons [60,64]. While the functions of Mgrp/NEV remained elusive in teleost fishes, the peptide counterpart in mammals, i.e. NEI, has been associated with the regulation of the stress response and anxiety [13,28] and behavioral effects such as grooming, locomotor activities and female sexual receptivity [28,73].

Initial efforts to characterize the MCH receptor(s) either in fish or mammalian systems were largely restrained due to the lack of a suitable binding assay associated with the highly hydrophobic nature of MCH and its high sensitivity to degradation (reviewed in [75]). The use of an MCH analog named [Phe^{13}, Tyr^{19}]-MCH permitted the discovery of binding sites on various mammalian cells lines including melanoma cells [24] and SVK-14 keratinocyte cells [18]. However, the low reliability of the binding properties using this analog and of intracellular signaling associated with the [Phe^{13}, Tyr^{19}]-MCH binding, made the overt expression of functional MCH receptors in these cells highly questionable [7,43].

The first MCH receptor (named here MCH-R1) was identified almost simultaneously by several laboratories, using a “reverse-pharmacology” strategy [19,70]; reviewed in [15,26,35,65]. This receptor corresponds to a splice variant of a G-protein-coupled-receptor (GPCR) orphan named SLC1/GPR24 [44]. MCH-R1 mRNA and protein are widely distributed in the rat and mouse brain, in agreement with the major projection sites of MCH neurons [36,40]. In particular, very high expression of MCH-R1 is found in hypothalamic ventromedial and dorsomedial nuclei, both major regions that regulate feeding. The importance of MCH-R1 expressing neuronal pathway in the regulation of feeding behavior and energy balance was recently supported by targeted disruption of the MCH-R1 gene in mice, that results in a lean phenotype associated with hyperactivity [20,49], and the use of non-peptidic antagonists that selectively block MCH-R1 and inhibit MCH-induced food intake in the rat [14,83].

In transfected cells models, MCH-R1 couples to multiple G proteins including Gi, Go and Gq; MCH may therefore potentially inhibit cAMP production induced by forskolin, activate MAP kinase activity, and increase intracellular free Ca^{2+} levels [34,65]. In mammalian cells that express MCH-R1 endogenously, MCH stimulates mainly the MAPK pathway [16,71,74] but it may also inhibit [71] or potentialize [82] the forskolin-induced cAMP production.

A second MCH receptor, called MCH-R2, was initially identified in humans by in silico studies [56] (reviewed in [15,26,35,65]), and more recently in fishes [47] but is apparently lacking in rats and mice [84]. MCH-R2 displays only 38% sequence identity to the MCH-R1 and exhibits a partly distinct expression profile in the human brain and peripheral tissues (reviewed in [35]). Furthermore, analysis of the signaling profile of human MCH-R2 in heterologous cell lines suggests that this receptor couples quite exclusively to Gαq protein to increase intracellular Ca^{2+} levels. The functional importance of MCH-R2 remains elusive due to the lack of an appropriate animal model.

Paradoxically, MCH receptor structure and signaling in fish models remain poorly studied. Both MCH-R1 and MCH-R2 orthologues, named mchr1 and mchr2, were recently identified by data mining in genomes of the zebrafish and the pufferfish *Fugu* [47]. A third MCH-R gene was found in zebrafish and could result from a duplication of MCHR1 gene which probably occurred specifically in the zebrafish lineage. Interestingly, the mchr1a gene is expressed only in embryo while mchr1b and mchr2 mRNA can be detected at all stages of zebrafish development.

Information regarding the intracellular pathways activated by MCH-R in fish is rather sparce. MCH triggers melanosome aggregation through protein kinase C activation in *Synbranchus marmoratus* [1] and inhibition of adenylate cyclase activity in the cuckoo wrasse [81], peppered catfish and Nile-tilapia [63].

In this study, we compared first the sequences of pro-MCH in mammalian and fish species in order to identify conservation in broad structures and motifs. Second, we used an eel skin assay to assess the effect of salmon and rat/human MCH analogues on pigment aggregation. Finally, we refined the downstream effector pathways activated by MCH in the same eel skin model.

2. Materials and methods

2.1. DNA samples

DNA samples from individual primates (*Tarsius syrichta*, *Saguinus oedipus*, *Cercopithecus aethiops*, *Hylobates Lar*, *Pan troglodytes* and *Pan paniscus*) were kindly provided by Philippe Dijan (CEREMOD, Dijon, France). Human DNA samples were prepared and stored as reported [17]. Genomic DNA was isolated from the occipital cortex of a *Macaca mulatta* (provided by Christian Gross (Bordeaux, France)) according to the Blin and Stafford’s method [12].

2.2. PCR amplification

2.2.1. Primers

Oligonucleotides were purchased from Eurogentec (Belgium). The primers used to amplify the MCH gene (∼1.5 kb

from exon 1 to 3) are the following. hMCHF2: 5′-TCAAGACTATAATCTACTCAACAG-3′ and hMCHR2: 5′-TGATACATCTTAACACTACTTTCTTTT-3′.

2.2.2. PCR conditions

One hundred to two hundred nanograms of genomic DNA were added to PCR reaction solution following the supplier protocol (LA TAQ,TaKaRa), in a total volume of 25 μl. Thirty-five cycles of amplification were carried out as follows: 30 s at 96 °C (denaturation), 30 s at 56 °C (annealing), 2 min at 72 °C (extension). A final extension step of 7 min at 72 °C was performed.

2.3. Sequencing

The DNA sequence of PCR-amplified fragments was determined from double-strand DNA using the Ampli Taq Polymerase FS and the Big Dye Terminator sequencing kit (Applera) and a Perkin Elmer ABI PRISM 377 sequencer.

2.4. Sequence alignments

Sequences obtained from the databases (NCBI and EBI) and PCR fragments sequenced were aligned in SEAVIEW [27]. Percentages of identity were calculated between two species of eutherians mammals (rat and human) and two species of *Clupeocephala* teleosts fishes (zebrafish and *Fugu*), respectively, considering a gap insertion as a misalignment. Only the perfectly matched amino acids were considered positive. The percentages given represent the number of matched amino acids compared to the total number of caracters (amino acids and gaps) in the alignment.

2.5. Fish skin preparation

Freshwater eels (*Anguilla anguilla*) were bought at a local dealer and kept for 1–3 weeks in aerated fresh water. Animals from both sexes were used.

The day before experiments, one eel was anesthetized by immersion for about 15 min in ice-cold water, and subsequently killed by decapitation. The whole skin was peeled off, and pieces were stamped out from the dark area of the skin using a punch having a diameter of 6 mm. They were kept overnight at 4 °C in a Hepes-buffered saline, pH 7.4, containing 10 mM Hepes, 120 mM NaCl, 2.7 mM KCl, 1.8 mM $CaCl_2$, 1.8 mM $MgCl_2$ and 5.6 mM D-glucose. αMelanocyte-stimulating hormone (αMSH) at low concentration (10^{-11} M) was systematically added to the saline in order to obtain a fully-dispersed state of melanophores pigments.

2.6. Photometric measurements of pigment aggregation

The day of the experiment, skin pieces were individually placed in a well of a plastic 24-wells cell culture plate which had been previously coated with dimethyl dichlorosilane.

The rate and extent of aggregation were quantitated by video image analysis: skin pieces were viewed with a 80 mm lens and imaged in grey-scale with a DCC video camera (JCV KYF55B) into the NIH Image public software (National Institutes of Health). The state of pigment dispersion/aggregation was quantitated just before, and 1 h after addition of MCH or MCH-derived peptides, using the average pixel value (APV) of the skin pieces. The APV were from 0 (white pixel) to 256 (black pixel). Measured on whole skin pieces, the maximal (fully-dispersed state) and minimal (fully aggregated state) APV measured under our experimental conditions were 170 and 0.4, respectively. The melanophore response to MCH or peptide analogues were calculated as the ratio APV (1 h after)/APV (before) × 100.

2.7. Cyclic AMP measurements

Skin pieces in wells of a 24-well plate were treated with 0.2 mM isobutylmethylxantine (IBMX) for 15 min before addition of MCH and/or forskolin for 5 min. Incubations were stopped by rapid freezing in liquid nitrogen. Skin pieces were then homogenized in 3 ml 10% trichloracetic acid (TCA) using a polytron homogeneizer. The homogenates were centrifuged at 5000 × *g* for 10 min at 4 °C. Supernatants were extracted four times with diethyl ether, and the TCA-free aqueous phase were finally lyophilised. The cAMP content of the dried pellets were measured using the cAMP RIA kit from Immunotech (Marseille, France), following the manufacturer's recommendations. Data are expressed as means ± S.E.M.

2.8. Peptides and chemicals

All the peptides used were from Neosystem SA (Strasbourg, France). Chemicals were from Sigma-Aldrich.

2.9. Statistics

All results are expressed as mean ± S.E.M. of *n* observations. Sets of data in Figs. 4 and 5 were compared with Student's *t*-test.

3. Results and discussion

3.1. Pro-MCH evolution

The MCH gene was amplified and sequenced from eight primate species, including human. As shown in Fig. 1, it appears that the intron/exon structure is fully conserved among all the primate and rodent species (mouse, rat). Moreover, the intronic sequences evolved neutrally whereas the exons are well conserved (not shown).

In mammals, both the peptides MCH and NEI are perfectly conserved, suggesting the existence of strong structural constraints on these two sequences. The prodomain is

also highly conserved (rat/human identity: 76%), even if no known function is associated with it. The teleost sequences were all retrieved in silico, either by data mining (*Fugu* and zebrafish sequences) or directly from the NCBI GenBank accession (other fishes). The MCH peptide, and particularly the loop (which is responsible for the interaction with the receptor; see below) is also very well conserved in all the available teleost sequences (Fig. 2). The only differences concern a few residues in the N-terminal and C-terminal parts (see *Fugu* and *Oreochromis* sequences, respectively). In great contrast, there is virtually no structural conservation in teleosts of the sequence which is homologous to the peptide NEI, raising the possibility that this sequence is not functional in this lineage. A previous study suggested that this sequence, called NEV or Mgrp, was processed and released in Tilapia [33], but its function remained unknown. The rest of the propeptide shows virtually no sequence conservation, and a proper alignment cannot be obtained.

As previously described [59], it is almost impossible to match the teleost and mammals precursors.

All the teleost studied belong to the *Clupeocephala* group, which diverged 145 million years ago [5]. The rodent lineage and the primate lineage diverged around 100 million years ago [2]. Although having approximately the same time of divergence, it is clear that the prodomains sequences evolved much more rapidly in the teleost group than in mammals. In other words, this sequence is under higher structural constraints in mammals than in fishes. This manifestation of a strong evolutionary pressure in the mammals lineage could be generalized to other neuropeptide systems: the precursors of neuropeptide Y (NPY), somatostatin (SRIF) and corticotropin-releasing factor (CRF) are also more conserved in the mammalian lineage than in the teleosts (Table 1).

The interpretation of such a phenomenon is hypothetical: sequence and structural similarities have been identified in

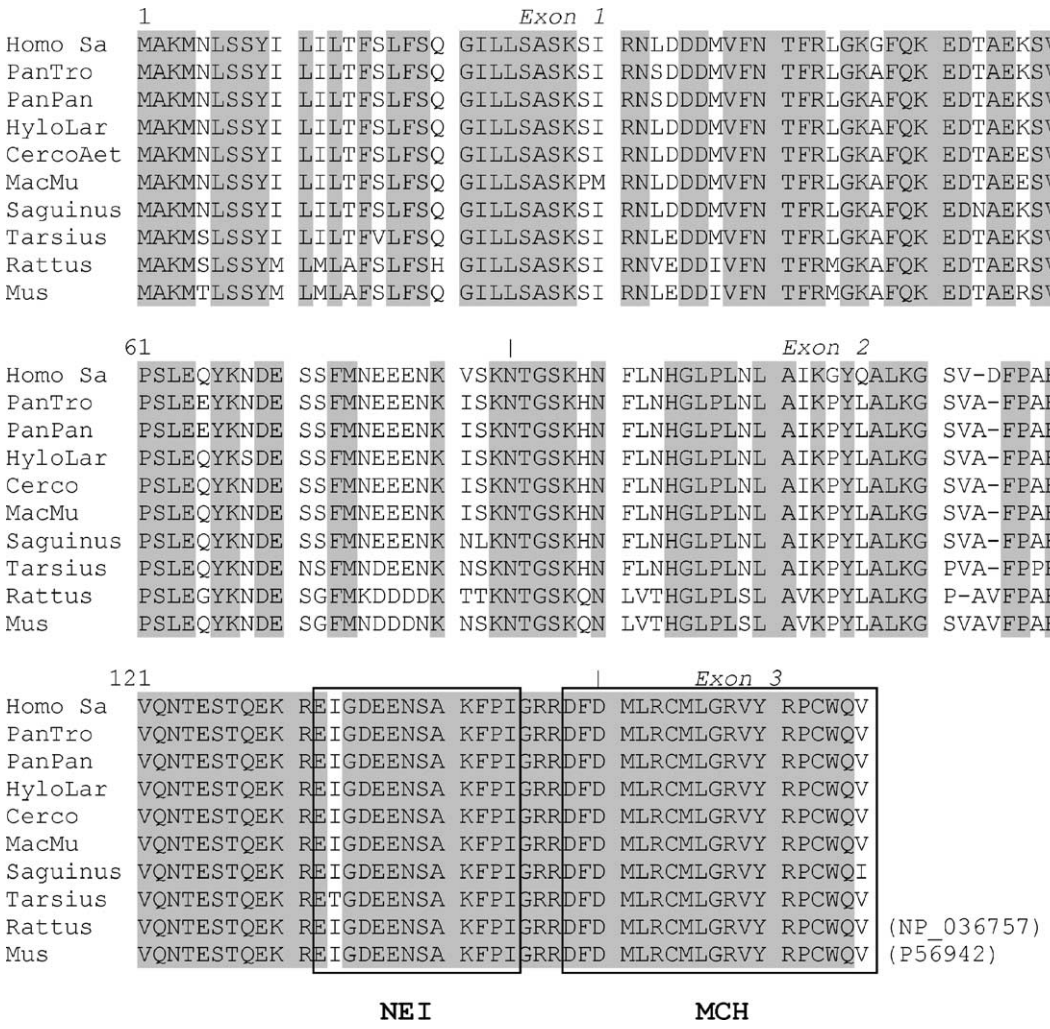


Fig. 1. MCH precursor in mammals. The MCH precursor sequences of 10 eutherians species are aligned. The matched aminoacids (throughout all the sequences) are highlighted in grey. The MCH and NEI peptides sequences are boxed, and the accession numbers, when available, are noted at the end of the sequences. The extent of each exon coding for the amino acid sequence are delineated. Homo Sa: *Homo sapiens*; PanTro: *Pan troglodytes*; PanPan: *Pan paniscus*; HyloLar: *Hylobates lar*; Cerco: *Cercopithecus aethiops*; MacMu: *Macaca mulatta*; Saguinus: *Saguinus oedipus*; Tarsius: *Tarsius syrichta*; Rattus: *Rattus norvegicus*; Mus: *Mus musculus*.

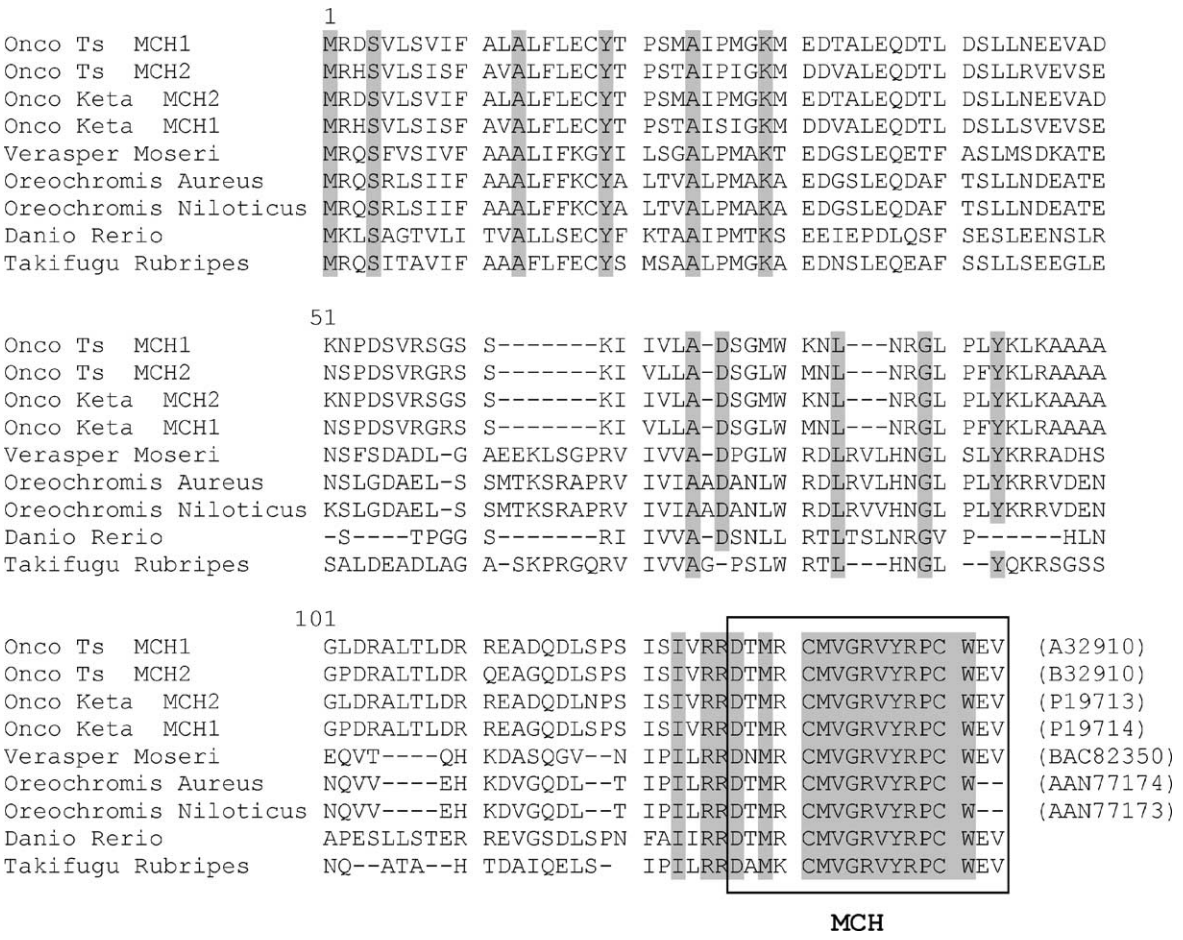


Fig. 2. MCH precursor in teleosts. The MCH precursor sequences of seven *clupeocephala* species are aligned. Note that there are two sequences for each of the *Oncorhynchus*, due to the polyploidization of their genome. The matched aminoacids (throughout all the sequences) are highlighted in grey. The MCH peptide sequence is boxed, and the accession numbers, when available, are noted at the end of the sequences. Onco Ts: *Oncorhynchus Tshawytscha*; Onco Keta: *Oncorhynchus Keta*.

subsets of regulated secretory proteins [22]. Neuroendocrine peptides which are released along the regulated secretory pathway have several structural features in their prodomain which are known to contribute to the correct processing and sorting of the bioactive peptides. Well documented examples include: the importance of the secondary structure [54,57], some post translational modifications (i.e. disulfide loops [30]), and consensus sequences interacting with carboxypeptidase E [31]. Since no fundamental difference between the sorting/processing of fishes and mammals peptides in general have ever been demonstrated, a simple explanation for such a difference in the evolution rates is not apparent. The MCH system could represent an interesting model in this respect since its function is largely different in fishes (where it is principally a neurohormone liberated in the general circulatory system from the neurohypophysis) from that in mammals (which use MCH mainly as a neurotransmitter). This "endocrine to nervous" switch of the peptide could have parallel modifications in the regulation of the sorting/processing mechanisms.

3.2. The eel skin assay

The skin covering the dorsal part of the eel body has a homogenous dark coloring. This allowed us to prepare, from a 40 to 50 cm long animal, up to 200 round skin pieces of 30 mm^2, having an identical coloring, and thus containing the same proportions of chromatophores (essentially melanophores and xanthophores). When they were kept in a physiological saline solution, these pieces underwent pigment aggregation in 1–2 h, resulting in a slight pallor. Thus, the pigment aggregation induced by MCH, MCH-derived peptides or norepinephrine was systematically measured on pieces which had been previously treated overnight with a low dose (10^{-11} M) of αMSH.

The assay we developed, based on the measurement of the average density (average pixel value, APV) of a skin piece just before, and 1 h after the treatment with MCH or other aggregating agents, was highly sensitive and reproducible (Fig. 3A).

Table 1
Percentages of identity of the aminoacids sequences from several neuropeptide encoding precursors

	Pro-MCH	Pro-NPY	Pro-SRIF	Pro-CRF
Rattus/Homo	80	94	96	78
Fugu/Danio	38	63	54	53

The compared species are on the one hand *Rattus norvegicus* and *Homo sapiens* representing eutherians mammals, and on the other hand *Takifugu rubripes* and *Danio rerio* representing clupeocephala teleost fishes.

Typical log-dose–response curves are shown in Fig. 3B. With both salmon and rat/human peptides, the maximum response was observed at 10 nM, and EC_{50} values were around 0.5 nM. In contrast to other teleost species, MCH had only a pigment-aggregating action in *Anguilla*, even at the highest dose tested. This was confirmed by microscopic observations of skin pieces, which revealed that all the chromatophores (melanophores and xanthophores) were fully aggregated at 10 μM (not shown).

This ruled out the existence in *Anguilla* of a low affinity MCH receptor subtype specifically responsible for a dispersing effect of the peptide, as the "β-MCH receptor" suggested in the peppered catfish and the Nile tilapia [63].

Depending on the season and the origin of animals, the skin color can vary from brown–green ("yellow stage") to dark brown–blue ("silver stage"). Thus, the APV of skin pieces can vary from animal to animal. However, the MCH EC_{50} values were all in the nM range and did not depend upon the stage of the animal used.

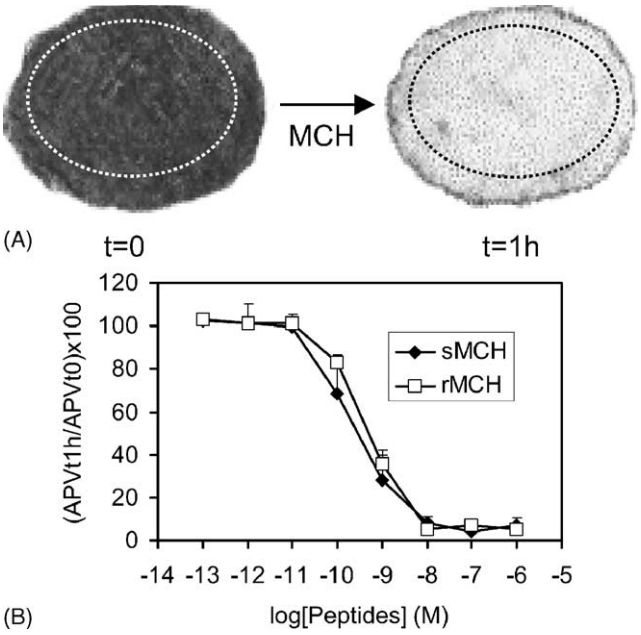


Fig. 3. MCH induced pigment aggregation in *Anguilla* skin. (A) Description of the assay. A representative skin piece (Ø 6 mm) before, and 1 h after treatment with MCH. The average density (measured as the 'average pixel value' using NIH Image 1.6) is measured at $t = 0$ and $t = 1$ h. (B) Typical log dose–response curves for rat/human (□) and salmon (♦) MCH-induced melanosome aggregation. The response is expressed as $[(APVt = 1\,h/APVt = 0) \times 100]$. Each point indicates the mean ± S.E.M. of measurements on four different skin pieces from the same animal.

3.3. $MCH_{6–17}$ is necessary and sufficient for melanosome aggregation

Different assays based on fish melanophores have been used in the past to decipher the structure-function relationships of MCH. Numerous analogues of salmon MCH have been prepared, and used in various fish scales bioassays [39,45,46,51].

Those studies showed that the biological activity did not depend upon the peptide N- and C-termini. The cyclic core was essential, any deletion into it having a deleterious effect, but it was generally not sufficient for a complete bioactivity [25].

The most detailed structure–activity studies have been performed recently on stable cell lines expressing either mammalian MCH-R1 or MCH-R2 [6,9,80]. Experiments of ligand binding, second messengers measurements, and [^{35}S]GTPγS binding, have allowed the structural elements crucial for binding and receptor activation to be identified, and have particularly highlighted the crucial role of Arg^6 (Arg^4 in salmon MCH) in the formation of ligand–receptor complexes. The importance of Trp^{17} in mammalian MCH is more controversial since it was shown to be essential [6] or not essential [9] for a full binding/activity at MCH-R1. The role of the homologous Trp^{15} in salmon MCH seemed to be dependant upon the fish bioassay used [25].

In order to compare the pharmacological profiles of the eel skin MCH receptor to mammalian MCH-R1 and MCH-R2, the abilities of mammalian MCH, salmon MCH, [Phe^{13}, Tyr^{19}]MCH and $MCH_{6–17}$ to induce a melanosome aggregation were assessed. As shown in Table 2, the EC_{50} for mammalian MCH, salmon MCH and $MCH_{6–17}$ were exactly the same in the eel skin assay (1.3 nM). In contrast, [Phe^{13}, Tyr^{19}]MCH was slightly less effective (5.3 nM), confirming the importance of the side group on Tyr^{13} (corresponding to Tyr^{11} in salmon MCH) for peptide–receptor interaction.

When acting on mammalian MCH-R1 and MCH-R2, both salmon MCH and [Phe^{13}, Tyr^{19}]MCH have higher EC_{50} than mammalian MCH and $MCH_{6–17}$.

These comparative results do not provide enough cues to know if the fish skin MCH receptor is orthologous to MCH-R1 or MCH-R2, but it is clear that the critical residues necessary for the agonistic activity of MCH at its receptor(s) are carried by the same minimal structure (the cyclic $MCH_{6–17}$) in fishes and mammals, in spite of several hundred million years of separate evolution.

3.4. MCH receptor activation induces a cyclic AMP reduction which is partially independent of Gαi

Probably because of the highly adaptative function of color changes in lower vertebrates, the pigment movements in chromatophores are tightly controlled by both the nervous and the endocrine systems. Several transduction pathways are involved in the microtubule- and actin-based pigment granules transport systems. The cyclic AMP, IP3/diacylglycerol,

Table 2
Characterization of MCH and MCH-derived peptides response on the eel skin assay (comparison with human MCH-R1 and MCH-R2)

Compound	Sequence	Eel skin assay EC_{50} (nM)	MCH-R1/SLC1 EC_{50} (nM)	MCH-R2/S643b EC_{50} (nM)
MCH	DFDMLRCMLGRVYRPCWQV	1.3 ± 0.3 (*n* = 14)	1.8 ± 0.8	9.5 ± 2.2
Salmon MCH	DT-M-RCMVGRVYRPCWEV	1.3 ± 0.3 (*n* = 12)	3.2 ± 0.7	15.2 ± 0.6
[Phe^{13}, Tyr^{19}]MCH	DFDMLRCMLGRVFRPCWQY	5.3 ± 2.6 (*n* = 3)	3.2 ± 0.7	11.7 ± 1.7
$MCH_{6–17}$	RCMLGRVYRPCW	1.3 ± 0.8 (*n* = 6)	2.1 ± 0.1	2.3 ± 0.8

Potencies are given as means of EC_{50} values ± S.E.M. obtained from separate *n* concentration–response curves. Results for MCH-R1/SLC1 and MCH-R2/S643b are from [69].

MAP kinase and phosphoinositide 3-kinase pathways have been shown to be crucial for the control of pigment movements [3,4,50,68] depending on the model studied (teleost or amphibian chromatophores), and on the agent used to induce pigment movements (MCH, norepinephrine, melatonin, nerve stimulation or light for aggregation; αMSH for dispersion).

Previous articles have proposed that MCH could act through the activation of a Gi protein. [63,81]. This is in agreement with the paling effects of agents known to act through Gαi proteins (melatonin, norepinephrine when acting at α2 receptors), and with the known role of PKA in regulating pigment transport [68,72,86]. We confirm that MCH is indeed able to reduce the cAMP levels in eel skin pieces: MCH (100 nM) significantly inhibited the forskolin-induced intracellular cAMP (Fig. 4). This result was confirmed by three independent experiments showing a statistically significant effect ($P < 0.05$) of MCH on the skin cAMP levels, even in the absence of forskolin (not shown).

These results must take into account the fact that the chromatophores are found only in the dermis layer and represent only a small percentage of the cells in the skin pieces and thus presumably make only a small contribution to the total cAMP. One can assume that at the level of a single chromatophore, the cAMP levels are strongly regulated by MCH. Pertussis toxin (PTX) was used to block (by causing ADP-ribosylation) the Gαi protein expressed in eel skin. Overnight treatment with 0.5 μg/ml PTX completely blocked the melanin aggregation induced by adrenergic α2 receptor activation, achieved with a mixture of norepinephrine (α1/α2 adrenergic receptor agonist) plus prazozin (α1 adrenergic receptor antagonist). In contrast, PTX was ineffective in blocking the MCH-induced aggregation, suggesting that the MCH-induced reduction in cAMP levels might be independent of canonical Gi proteins. The activation of a PTX-insensitive G-protein, like Gαz [37] cannot be ruled out, but one can also hypothesize that MCH activates one or several cyclic nucleotide phosphodiesterase (PDE).

Pre-treatment of skin pieces with increasing concentrations of the general PDE inhibitor IBMX revealed that the effect of 100 nM MCH was impaired, although not completely blocked by the highest concentration of IBMX tested (Fig. 6). This shows that MCH cannot exert its full effect when the cAMP levels are experimentally up-regulated by inhibiting PDE. However the result also suggests MCH may act on eel melanophores by activating additional intracellular transduction pathway(s), which also contribute to skin pallor.

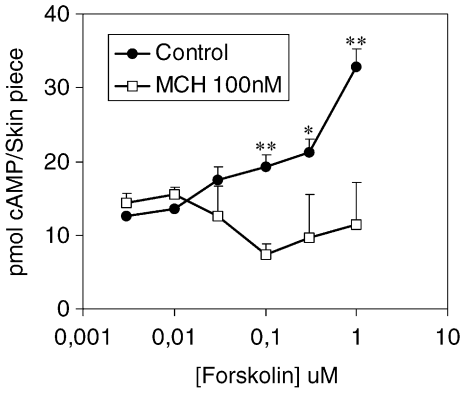


Fig. 4. Effect of MCH on the cAMP content. Dose–response effect of forskolin on cAMP content of eel skin pieces treated (□) or not treated (●) with 100 nM MCH. Forskolin and MCH were applied for 5 min. *Different from the control at $P < 0.05$, **$P < 0.01$, ***$P < 0.001$ (Student's *t*-test).

The confirmation that MCH could activate chromatophores PDE needs further experiments. However, this hypothesis could explain how MCH might reduce cAMP levels by a Gαi-independent pathway. Furthermore, it has been demonstrated recently that melatonin induces pigment aggregation in *Xenopus* melanophores through a melatonin Mel1c receptor able to activate the phosphoinositide 3-kinase

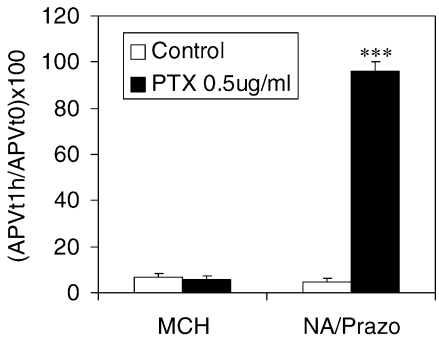


Fig. 5. Effect of pertussis toxin (PTX) treatment on the MCH-induced aggregation. Skin pieces were treated with PTX (0.5 μg/ml) overnight at 4 °C, before challenging with MCH (100 nM) or norepinephrine/prazozin (10 μM each). Pertussis toxin-mediated ADP-ribosylation of Gi proteins had no apparent effect on the MCH response (left), while it completely blocked the adrenergic α2 response (right). ***Different from the control at $P < 0.001$ (Student's *t*-test).

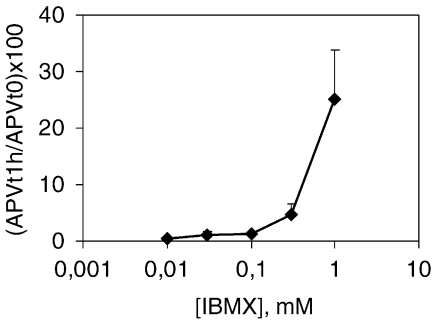


Fig. 6. Effect of IBMX treatment on the effect of MCH. Skin pieces were treated with 0.01–1 mM IBMX for 30 min before measurement of pigment aggregation induced by MCH (100 nM). The paling effect of MCH was significantly less intense (***different from 0.01 mM IBMX at $P < 0.001$) when PDE were inhibited by 1 mM IBMX.

via Gβγ, which directly or indirectly activates PDE [3]. PI-3 kinase is also involved in the effect of melatonin in the melanophores of the fish *Labrus bimaculatus* (cukoo wrasse) [3], and we can hypothesize that such a mechanism is at least in part responsible for the MCH-induced reduction of cAMP levels in the *Anguilla* skin.

4. Conclusion

First, our sequence comparison analysis has demonstrated strong conservation in the MCH precursor apart from the MCH sequence itself, selectively in the mammalian lineage. This suggests that this domain may have important functions in mammals independently from MCH peptide processing. Second, because of the lack of reliable cell models expressing endogenously the mammalian MCH receptors R1 and R2, the informations regarding the signaling pathways and cellular functions of those receptors come essentially from transfected cells. In this respect, the results which can be obtained from a fish skin model are of great interest, as we have tentatively showed here. However, an important issue would be to define which subtype (MCH-R1 or MCH-R2) is responsible for the MCH-induced pigment granule translocation. The two receptors genes have orthologs in the teleost genome [47], and molecular techniques would allow to discover which one is transcribed in the different chromatophores.

Acknowledgments

We have appreciated a lot the helpful comments of both reviewers. We thank very much S. Ricois (I.P.M.C., Valbonne, France) for technical assistance. This work was supported by the Institut de Recherche Servier (Suresnes, France), the Centre National de la Recherche Scientifique (Programme OHLL 2003-2004, CNRS, France) and the Association pour la Recherche sur le Cancer (subvention ARC 3375). F.D.-T. was supported by a fellowship from the French Education Minister (“Allocation couplée” MENRS/ENS).

References

[1] Abrao MS, De L, Castrucci AM, Hadley ME, Hruby VJ. Protein-kinase C mediates MCH signal transduction in teleost, *Synbranchus marmoratus*, melanocytes. Pigment Cell Res 1991;4:66–70.
[2] Adkins RM, Walton AH, Honeycutt RL. Higher-level systematics of rodents and divergence time estimates based on two congruent nuclear genes. Mol Phylogenet Evol 2003;26:409–20.
[3] Andersson TP, Skold HN, Svensson SP. Phosphoinositide 3-kinase is involved in *Xenopus* and *Labrus* melanophore aggregation. Cell Signal 2003;15:1119–27.
[4] Andersson TP, Svensson SP, Karlsson AM. Regulation of melanosome movement by MAP kinase. Pigment Cell Res 2003;16: 215–21.
[5] Arratia G. Phylogenetic relationships of Teleostei. Past and present. Estud Oceanol 2000;19:19–51.
[6] Audinot V, Beauverger P, Lahaye C, Suply T, Rodriguez M, Ouvry C, et al. Structure–activity relationship studies of melanin-concentrating hormone (MCH)-related peptide ligands at SLC-1, the human MCH receptor. J Biol Chem 2001;276:13554–62.
[7] Audinot V, Lahaye C, Suply T, Rovère-Jovène C, Rodriguez M, Nicolas JP, et al. SVK14 cells express an MCH binding site different from the MCH1 or MCH2 receptor. Biochem Biophys Res Commun 2002;295:841–8.
[8] Baker BI, Kawauchi H. MCH in non-mammalian vertebrates: neuronal topography and functions. In: Knigge K, Prased A, Preteland S, Wagner JE, editors. MCH and seizures: neuromolecular and neuroendocrine aspects. India: Research Signpost; 1997. p. 1–30.
[9] Bednarek MA, Feighner SD, Hreniuk DL, Palyha OC, Morin NR, Sadowski SJ, et al. Short segment of human melanin-concentrating hormone that is sufficient for full activation of human melanin-concentrating hormone receptors 1 and 2. Biochemistry 2001;40:9379–86.
[10] Bird DJ, Baker BI, Eberle A, Swann RW. The biosynthesis of melanin-concentrating hormone in a fish. J Neuroendocrinol 1990;2: 309–15.
[11] Bittencourt JC, Presse F, Arias C, Peto C, Vaughan J, Nahon JL, et al. The melanin-concentrating hormone system of the rat brain: an immuno- and hybridization histochemical characterization. J Comp Neurol 1992;319:218–45.
[12] Blin N, Stafford DW. A general method for isolation of high molecular weight DNA from eukaryotes. Nucleic Acids Res 1976;3:2303–8.
[13] Bluet-Pajot MT, Presse F, Vokö Z, Hoeger C, Mounier F, Epelbaum J, et al. Neuropeptide-E-I antagonizes the action of melanin-concentrating hormone on stress-induced release of adrenocorticotropin in the rat. J Neuroendocrinol 1995;7:297–303.
[14] Borowsky B, Durkin MM, Ogozalek K, Marzabadi MR, DeLeon J, Heurich R, et al. Antidepressant, anxiolytic and anorectic effects of a melanin-concentrating hormone-1 receptor antagonist. Nat Med 2002;8:825–30.
[15] Boutin JA, Suply T, Audinot V, Rodriguez M, Beauverger P, Nicolas JP, et al. Melanin-concentrating hormone and its receptors: state of the art. Can J Physiol Pharmacol 2002;80:388–95.
[16] Bradley RL, Mansfield JPR, Maratos-Flier E, Bheatham B. Melanin-concentrating hormone activates signaling pathways in 3T3-L1 adipocytes. Am J Physiol Endocrinol Metab 2002;283:E584–92.
[17] Breton C, Schorpp M, Nahon JL. Isolation and characterization of the human melanin-concentrating hormone gene and a variant gene. Brain Res Mol Brain Res 1993;18:297–310.
[18] Burgaud JL, Poosti R, Fehrentz J-A, Martinez J, Nahon JL. Melanin-concentrating hormone binding sites in human SVK14 keratinocytes. Biochem Biophys Res Commun 1997;241:622–9.
[19] Chambers J, Ames RS, Bergsma D, Muir A, Fitzgerald LR, Hervieu G, et al. Melanin-concentrating hormone is the cognate ligand for the orphan G-protein-coupled receptor SLC-1. Nature 1999;400:261–5.

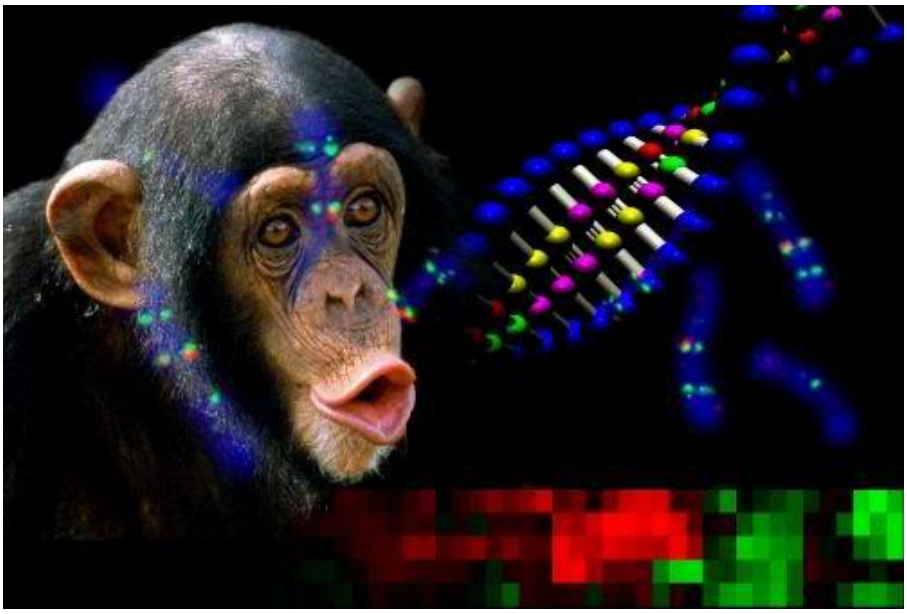
[20] Chen Y, Hu C, Hsu C-K, Zhang Q, Bi C, Asnicar M, et al. Targeted disruption of the melanin-concentrating hormone receptor-1 results in hyperphagia and resistance to diet-induced obesity. Endocrinology 2002;143:2469–77.
[21] Clegg DJ, Air EL, Benoit SC, Sakai RS, Seeley RJ, Woods SC. Intraventricular melanin-concentrating hormone stimulates water intake independent of food intake. Am J Physiol Regul Integr Comp Physiol 2003;284:R494–9.
[22] Day R, Gorr SU. Secretory granule biogenesis and chromogranin A: master gene, on/off switch or assembly factor? Trends Endocrinol Metab 2003;14:10–3.
[23] Della-Zuana O, Presse F, Ortola C, Duhault J, Nahon JL, Levens N. Acute and chronic adminsitration of melanin concentrating hormone enhances food intake and body weight in Wistar and Sprague-Dawley rats. Int J Obesity 2002;26:1289–95.
[24] Drozdz R, Eberle AN. Synthesis and iodiantion of human (phenylalanine13, tyrosine19) melanin-concentrating hormone for radioreceptor assay. J Peptide Sci 1995;1:58–65.
[25] Eberle A. Structure–activity relationships of the the melanotropins. In: The melanotropins. Basel: Karger; 1988.
[26] Forray C. The MCH-receptor family: feeding brain disorders? Curr Opin Pharmacol 2003;3:85–9.
[27] Galtier N, Gouy M, Gautier C. SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. Comput Appl Biosci 1996;12:543–8.
[28] Gonzalez MI, Baker BI, Hole DR, Wilson CA. Behavioral effects of neuropeptide E-I (NEI) in the female rat: interactions with α-MSH, MCH and dopamine. Peptides 1998;19:1007–16.
[29] Gonzalez MI, Baker BI, Wilson CA. Stimulatory effect of melanin concentrating hormone on LH release. Neuroendocrinology 1997;66:254–62.
[30] Gorr SU, Huang XF, Cowley DJ, Kuliawat R, Arvan P. Disruption of disulfide bonds exhibits differential effects on trafficking of regulated secretory proteins. Am J Physiol 1999;277:C121–31.
[31] Gorr SU, Jain RK, Kuehn U, Joyce PB, Cowley DJ. Comparative sorting of neuroendocrine secretory proteins: a search for common ground in a mosaic of sorting models and mechanisms. Mol Cell Endocrinol 2001;172:1–6.
[32] Griffond B, Baker BI. Cell and molecular cell biology of melanin-concentrating hormone. Int Rev Cytol 2002;213:233–77.
[33] Groneveld D, Balm PH, Wendelaar Bonga SE. Identification, cellular localization and in vitro release of a novel teleost melanin-concentrating hormone gene-related peptide. Neuroendocrinology 1995;62:498–505.
[34] Hawes BE, Kil E, Green B, O'Neill K, Fried S, Graziano MP. The melanin-concentrating hormone receptor couples to multiple G proteins to activate diverse intracellular signaling pathways. Endocrinology 2000;141:4524–32.
[35] Hervieu G, Maulon-Feraille L, Chambers JK, Cluderay JE, Wilson S, Presse F, et al. The melanin-concentrating hormone. In: Quirion R, Björklund A, Hökfelt T, editors. Handbook of chemical neuroanatomy. Peptide receptors, Part II, vol. 20. Elsevier Science; 2003. p. 31–101.
[36] Hervieu GJ, Cluderay JE, Harrison D, Meakin J, Maycox P, Nasir S, et al. The distribution of the mRNA and protein products of the melanin-concentrating hormone (MCH) receptor gene, SLC-1, in the central nervous system of the rat. Eur J Neurosci 2000;12:1–23.
[37] Ho MK, Wong YH. G(z) signaling: emerging divergence from G(i) signaling. Oncogene 2001;20:1615–25.
[38] Kawauchi H, Kawazoe I, Tsubokawa M, Kishida M, Baker BI. Characterization of melanin-concentrating hormone in chum salmon pituitaries. Nature 1983;305:321–3.
[39] Kawazoe I, Kawauchi H, Hirano T, Naito N. Structure–activity relationships of melanin-concentrating hormone. Int J Pept Protein Res 1987;29:714–21.
[40] Kilduff TS, de Lecea L. Mapping of the mRNAs for the hypocretin/orexin and melanin-concentrating hormone receptors: networks of overlapping peptide systems. J Comp Neurol 2001;435:1–5.
[41] Knigge KM, Baxter-Grillo D, Speciale J, Wagner J. Melanotropic peptides in the mammalian brain: the melanin-concentrating hormone. Peptides 1996;17:1063–73.
[42] Knigge KM, Wagner JE. Melanin-concentrating hormone (MCH) involvement in pentylenetetrazole (PTZ)-induced seizure in rat and guinea pig. Peptides 1997;18:1095–7.
[43] Kokkotou E, Mastaitis JW, Qu D, Hoersch D, Slieker L, Bonter K, et al. Characterization of [Phe13, Tyr19]-MCH analog binding activity to the MCH receptor. Neuropeptides 2000;34:240–7.
[44] Lakaye B, Minet A, Zorzi W, Grisar T. Cloning of the rat brain cDNA encoding for the SLC-1 G protein-coupled receptor reveals the presence of an intron in the gene. Biochim Biophys Acta 1998;1401:216–20.
[45] Lebl M, Hruby VJ, Castrucci AM, Hadley ME. Melanin concentrating hormone analogues: contraction of the cyclic structure. II. Antagonist activity. Life Sci 1989;44:451–7.
[46] Lebl M, Hruby VJ, Castrucci AM, Visconti MA, Hadley ME. Melanin concentrating hormone analogues: contraction of the cyclic structure. 1. Agonist activity. J Med Chem 1988;31:949–54.
[47] Logan DW, Bryson-Richardson J, Pagan KE, Taylor MS, Currie PD, Jackson IJ. The structure and evolution of the melanocortin and MCH receptors in fish and mammals. Genomics 2003;81:184–91.
[48] Ludwig DS, Tritos NA, Mastaitis JW, Kulkarni R, Kokkotou E, Elmquist J, et al. Melanin-concentrating hormone overexpression in transgenic mice leads to obesity and insulin resistance. J Clin Invest 2001;107:379–86.
[49] Marsh DJ, Weingarth DT, Novi DE, Chen HY, Trumbauer ME, Chen AS, et al. Melanin-concentrating hormone 1 receptor-deficient mice are lean, hyperactive, and hyperphagic and have altered metabolism. Proc Natl Acad Sci USA 2002;99:3240–5.
[50] Martensson LG, Andersson RG. Is Ca^{2+} the second messenger in the response to melatonin in cuckoo wrasse melanophores? Life Sci 2000;66:1003–10.
[51] Matsunaga TO, Castrucci AM, Hadley ME, Hruby VJ. Melanin concentrating hormone (MCH): synthesis and bioactivity studies of MCH fragment analogues. Peptides 1989;10:349–54.
[52] Miller CL, Hruby VJ, Matsunaga TO, Bickford PC. Alpha-MSH and MCH are functional antagonists in a CNS auditory gating paradigm. Peptides 1993;14:431–40.
[53] Minth CD, Qiu H, Akil H, Watson SJ, Dixon JE. Two precursors of melanin-concentrating hormone: DNA sequence analysis, in situ and immunochemical localization. Proc Natl Acad Sci USA 1989;86:4292–6.
[54] Mitra J, Tang X, Almo SC, Shields D. Temperature-induced conformational changes in prosomatostatin-II: implications for processing. Biochem J 1998;334:275–82.
[55] Monzon ME, de Souza MM, Izquierdo LA, Izquierdo I, Barros DM, de Barioglio SR. Melanin-concentrating hormone (MCH) modifies memory retention in rats. Peptides 1999;20:1517–9.
[56] Mori M, Harada M, Terao Y, Sugo T, Watanabe T, Shimomura Y, et al. Cloning of a novel G protein-coupled receptor, SLT, a subtype of the melanin-concentrating hormone receptor. Biochem Biophys Res Commun 2001;283:1013–8.
[57] Mouchantaf R, Kumar U, Sulea T, Patel YC. A conserved alpha-helix at the amino terminus of prosomatostatin serves as a sorting signal for the regulated secretory pathway. J Biol Chem 2001;276:26308–16.
[58] Murray JF, Adan RAH, Baker BI, Thody AJ, Nijenhuis WAJ, Ykitake J, et al. Melanin-concentrating hormone, melanocortin receptors and regulation of luteinizing hormone release. J Neuroendocrinol 2000;12:217–23.
[59] Nahon JL. The melanin-concentrating hormone: from the peptide to the gene. Crit Rev Neurobiol 1994;8:221–62.
[60] Nahon JL, Presse F, Bittencourt JC, Sawchenko P, Vale W. The rat melanin-concentrating hormone mRNA encodes multiple puta-

tive neuropeptides coexpressed in the dorsolateral hypothalamus. Endocrinology 1989;125:2056–65.
[61] Nahon JL, Presse F, Schoepfer R, Vale W. Identification of a single melanin-concentrating hormone mRNA in coho salmon: structural relatedness with 7SL RNA. J Neuroendocrinol 1991;3:173–83.
[62] Ono M, Wada C, Oikawa I, Kawazoe I, Kawauchi H. Structures of two kinds of mRNA encoding the chum salmon melanin-concentrating hormone. Gene 1988;71:433–8.
[63] Oshima N, Nakamaru N, Araki S, Sugimoto M. Comparative analyses of the pigment-aggregating and -dispersing actions of MCH on fish chromatophores. Comp Biochem Physiol C Toxicol Pharmacol 2001;129:75–84.
[64] Parkes DG, Vale W. Secretion of melanin-concentrating hormone and neuropeptide-EI from cultured rat hypothalamic cells. Endocrinology 1992;131:1826–31.
[65] Pissios P, Maratos-Flier E. Melanin-concentrating hormone: from fish skin to skinny mammals. Trends Endocrinol Metab 2003;14:243–8.
[66] Presse F, Nahon JL, Fischer WH, Vale W. Structure of the human melanin-concentrating hormone mRNA. Mol Endocrinol 1990;4:632–7.
[67] Qu D, Ludwig DS, Gammeltoft S, Piper M, Pelleymounter MA, Cullen MJ, et al. A role for melanin-concentrating hormone in the central regulation of feeding behaviour. Nature 1996;380:243–7.
[68] Rodionov V, Yi J, Kashina A, Oladipo A, Gross SP. Switching between microtubule- and actin-based transport systems in melanophores is controlled by cAMP levels. Curr Biol 2003;13: 1837–47.
[69] Rodriguez M, Beauverger P, Naime I, Rique H, Ouvry C, Souchaud S, et al. Cloning and molecular characterization of the novel human melanin-concentrating hormone receptor MCH2. Mol Pharmacol 2001;60:632–9.
[70] Saito Y, Nothacker H-P, Wang Z, Lin SHS, Leslie F, Civelli O. Molecular characterization of the melanin-concentrating-hormone receptor. Nature 1999;400:265–9.
[71] Saito Y, Wang Z, Hagino-Yamagishi K, Civelli O, Kawashima S, Maruyama K. Endogenous melanin-concentrating hormone receptor SLC-1 in human melanoma SK-MEL-37 cells. Biochem Biophys Res Commun 2001;289:44–50.
[72] Sammak PJ, Adams SR, Harootunian AT, Schliwa M, Tsien RY. Intracellular cyclic AMP not calcium, determines the direction of vesicle movement in melanophores: direct measurement by fluorescence ratio imaging. J Cell Biol 1992;117:57–72.
[73] Sanchez M, Baker BI, Celis M. Melanin-concentrating hormone (MCH) antagonizes the effects of α-MSH and neuropeptide E-I on grooming and locomotor activities in the rat. Peptides 1997;18:393–6.
[74] Schlumberger SE, Jäggin V, Tanner H, Eberle AN. Endogenous receptor for melanin-concentrating hormone in human neuroblastoma Kelly cells. Biochem Biophys Res Commun 2002;298:54–9.
[75] Schlumberger SE, Talke-Messerer C, Zumsteg U, Eberle AN. Expression of receptors for melanin-concentrating hormone (MCH) in different tissues and cell lines. J Recept Signal Transduct Res 2002;22:509–31.
[76] Segal-Lieberman G, Bradley RL, Kokkotou E, Carlson M, Trombly DJ, Wang X, et al. Melanin-concentrating hormone is a critical mediator of the leptin-deficient phenotype. Proc Natl Acad Sci USA 2003;100:10085–90.
[77] Sherbrooke WC, Hadley ME. Exploring the evolutionary history of melanin-concentrating and melanin-stimulating hormone receptors on melanophores: neopterygian (holostean) and chondrostean fishes. Pigment Cell Res 1988;1:344–9.
[78] Shimada M, Tritos NA, Lowell BB, Flier JS, Maratos-Flier E. Mice lacking melanin-concentrating hormone are hypophagic and lean. Nature 1998;396:670–4.
[79] Skofitsch G, Jacobowitz DM, Zamir N. Immunohistochemical localization of a melanin concentrating hormone-like peptide in the rat brain. Brain Res Bull 1985;15:635–49.
[80] Suply T, Della Zuana O, Audinot V, Rodriguez M, Beauverger P, Duhault J, et al. SLC-1 receptor mediates effect of melanin-concentrating hormone on feeding behavior in rat: a structure–activity study. J Pharmacol Exp Ther 2001;299:137–46.
[81] Svensson SP, Norberg T, Andersson RG, Grundstrom N, Karlsson JO. MCH-induced pigment aggregation in teleost melanophores is associated with a cAMP reduction. Life Sci 1991;48:2043–6.
[82] Tadayyon M, Welters HJ, Haynes AC, Cluderay JE, Hervieu G. Expression of melanin-concentrating hormone receptors in insulin-producing cells: MCH stimulates insulin release in RINm5F and CRI-G1 cell-lines. Biochem Biophys Res Commun 2000;275:709–12.
[83] Takekawa S, Asami A, Ishihara Y, Terauchi J, Kato K, Shimomura Y, et al. T-226296: a novel, orally active and selective melanin-concentrating hormone receptor antagonist. Eur J Pharmacol 2002;438:129–35.
[84] Tan CP, Sano H, Iwaasa H, Pan J, Sailer AW, Hreniuk DL, et al. Melanin-concentrating hormone receptor subtypes 1 and 2: species-specific gene expression. Genomics 2002;79:785–92.
[85] Tsukamura H, Thompson RC, Tsukahara S, Ohkura S, Maekawa F, Moriyama R, et al. Intracerebroventricular administration of melanin-concentrating hormone suppresses pulsatile luteinizing hormone release in the female rat. J Neuroendocrinol 2000;12:529–34.
[86] Tuma MC, Gelfand VI. Molecular mechanisms of pigment transport in melanophores. Pigment Cell Res 1999;12:283–94.
[87] Vaughan JM, Fischer WH, Hoeger C, Rivier J, Vale W. Characterization of melanin-concentrating hormone from rat hypothalamus. Endocrinology 1989;125:1660–5.
[88] Verret L, Goutagny R, Fort P, Cagnon L, Salvert D, Léger L, et al. A role of melanin-concentrating hormone producing neurons in the central regulation of paradoxical sleep. BMC Neurosci 2003;4:19.

Vivant
L'actualité des sciences et débats sur le vivant

Synthèse

Ce que nous apprend le génome du chimpanzé et des autres primates



La publication de la première séquence complète du génome du chimpanzé était fortement attendue. Comment ne pas penser, en effet, qu'elle pourrait éclairer d'un jour nouveau les changements génétiques qui ont fait « s'évader » l'ancêtre de l'homme moderne de la lignée des grands singes, voilà 5 à 6 millions d'années ? Effectivement, notre cousin singe permet de mettre le doigt sur des remaniements chromosomiques, des « pouponnières à gènes », qui semblent avoir été à l'origine de gènes spécifiques de l'homme.

Depuis les temps anciens et tout le long de son histoire l'homme s'est interrogé sur son origine et sur sa singularité de roseau pensant. Pour le biologiste et tout particulièrement le généticien, la réponse à cette difficile question doit essentiellement être trouvée dans l'analyse structurale et fonctionnelle du support de notre hérédité, l'ADN. Notre vision anthropomorphique du monde vivant nous a donc conduit à lancer dans la dernière décennie une entreprise scientifico-industrielle qui a abouti à la parution simultanée en 2001 de deux « brouillons » de séquences globales de notre génome. Pour les *aficionados*, la séquence consensuelle, quasi achevée, du génome humain est actuellement librement consultable [1].

Toutefois cette séquence ne prend tout son intérêt que si on la compare à celles d'autres espèces. L'analyse des génomes de la levure, du nématode (*Caenorhabditis elegans*), de la mouche drosophile, des poissons fugu (*Fugu rubripes* et *Tetraodon nigroviridis*) ou même de la souris apporte des informations capitales sur les événements et forces qui ont forgé les chromosomes et permis l'émergence de familles de gènes à des étapes clés de l'évolution. Mais elle n'apporte guère de lumières sur les mécanismes qui ont présidé au développement des fonctions propres aux Hominidés et tout particulièrement de l'*Homo sapiens sapiens*, seul représentant actuel des Hominiinés.

La primauté aux primates

Comment, en effet, avec pour seul point de comparaison les génomes de la souris et de l'homme, aborder l'étude des gènes responsables des modifications anatomiques qui ont permis la diversification et l'adaptation du régime alimentaire en fonction des changements climatiques [2], l'apparition de la préhension puis de la bipédie (libérant ainsi la main) [9], le développement spectaculaire du cerveau en concertation avec des modifications des muscles et os faciaux [4]… pour aboutir *in fine* à l'émergence des fonctions cognitives les plus élaborées comme le langage [5]. Il faut, pour cela, s'intéresser aux espèces « cousines » dans l'ordre des primates.

C'est ainsi que différents consortiums se sont attelés dès 1998 au séquençage du génome du chimpanzé commun (*Pan troglodytes*). Une version encore inachevée est désormais en ligne [6,7]. Un certain nombre d'études font déjà référence au génome de chimpanzé et en utilisent les données, notamment pour des comparaisons avec les résultats obtenus pour l'espèce humaine. Deux études ont tenté une comparaison globale des génomes humain et de chimpanzé [8,9]… mais elles ne sont pas exemptes de problèmes méthodologiques ***(voir l'encadré ci-dessous)***.

Question de méthode

Quelle est la qualité des informations disponibles à ce jour sur le génome du chimpanzé ? Deux critiques majeures peuvent être portées.
Tout d'abord, le choix du séquençage de multiples individus, exposé et défendu dans les articles princeps de lancement du séquençage du génome de chimpanzé **(1, 2)**, est fortement discutable, notamment pour l'étude de gènes qui ont été dupliqués. Il est en effet difficile dans ces conditions de distinguer deux gènes paralogues (gènes dont les séquences sont homologues et résultent de la duplication récente d'un même gène ancestral) et deux versions (allèles) du même gène portés par différents individus !

Deuxième critique, le génome de chimpanzé comporte encore des « trous » de séquençage, particulièrement dans les régions sujettes aux remaniements et aux duplications. Cela pourrait conduire à une mauvaise estimation des insertions/délétions (« indels ») entre les génomes de chimpanzé et d'homme. En se basant sur les séquences disponibles en 2002 puis 2003, et en tenant compte de la totalité des séquences (indels inclus), **Roy Britten**, du CalTech (Californie) a proposé une valeur minimale d'environ 5 % de divergence entre les génomes du chimpanzé et de l'homme, valeur due essentiellement aux indels **(3,4)**. Nous sommes donc très éloignés du 1,2 % de divergence proposé initialement en utilisant la première séquence de génome de chimpanzé publiée et en omettant justement les indels **(5)**.

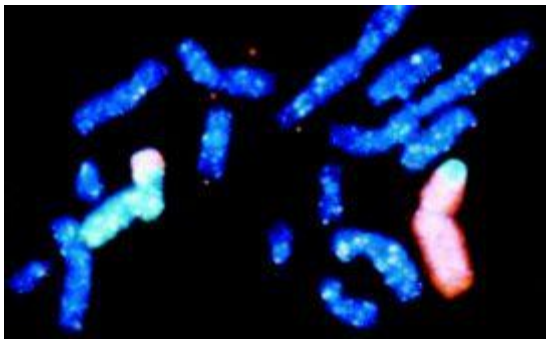
Plus généralement, ces études soulèvent le problème de la fiabilité des séquences réalisées selon la méthode du « *whole genome shotgun* » (WGS), un problème qui se pose d'ailleurs encore pour la séquence du génome humain assemblée selon cette méthode. Comme indiqué dans une étude parue en octobre 2004 par l'équipe d'**Evan E. Eichler** (Université de Washington, Seattle), l'utilisation exclusive de cette méthode conduit à une diminution significative de la taille du génome du fait de la perte des gènes insérés dans les duplications et, par voie de conséquence, à une vision erronée de la structure et de l'évolution des génomes **(6)**. Il est donc primordial de séquencer des chromosomes artificiels (BAC), de manière à déterminer la séquence de grandes régions d'une seule et même molécule d'ADN.

(1) D. Reich et S. Lander (2002) Sequencing the Chimpanzee genome
http://genome.wustl.edu/projects/chimp/ChimpGenome2.pdf.
(2) M. Olson *et al.* (2002) A white paper advocating complete sequencing of the genome of the common chimpanzee, *Pan Troglodytes*
http://genome.wustl.edu/projects/chimp/Chimp_genome1_editted.pdf.
(3) R. J. Britten (2002) Divergence between samples of chimpanzee and human DNA sequences is 5%, counting indels, *PNAS* 99:13633-13635.
http://www.pnas.org
(4) R. J. Britten *et al.* (2003) Majority of divergence between closely related DNA samples is due to indels, *PNAS* 100:4661-4665. http://www.pnas.org
(5) A. Fujiyama *et al.* (2002) Construction and analysis of a Human-Chimpanzee comparative clone map, *Science* 295 (5552):131-4.
(6) X. She *et al.* (2004) Shotgun sequence assembly and recent segmental duplications within the human genome, *Nature* 431 (7011):927-30.

Chromosomes homologues

Pour éviter ces problèmes, d'autres travaux ont analysé un seul chromosome dans l'une des espèces et les séquences « synténiques » (homologues entre deux espèces) dans l'autre. Ainsi, le chromosome 22 de *Pan troglodytes* a été totalement séquencé à partir de clones génomiques strictement ordonnés [10] et comparé avec son homologue chez l'homme, le chromosome 21, lui aussi totalement séquencé.

À l'inverse, la séquence encore imparfaite du génome de chimpanzé a été comparée à celles totalement achevées de certains chromosomes humains, comme le chromosome 5 [11], le chromosome 9 [12] et le chromosome 10 [13]. Par ailleurs, des études comparatives chimpanzé-homme ont été réalisées en utilisant uniquement des séquences entièrement validées [14], ou en se focalisant sur des régions hautement informatives et associées à des différences phénotypiques notables entre l'homme et le chimpanzé, comme la région du Complexe majeur d'histocompatibilité (CMH) [15].



© U.S. Department of Energy, Human Genome Program

Nous avons sélectionné ici quelques faits majeurs qui ressortent de ces travaux, en essayant de compléter les excellents articles déjà parus commentant ces mêmes études [16-18].

Une divergence remontant au plus à 6,2 millions d'années

Tout d'abord, la comparaison (excluant les régions situées autour du centromère et proches des télomères des chromosomes, pauvres en gènes actifs mais riches en séquences dupliquées) des séquences strictement alignées entre le chimpanzé et l'homme indique un taux de divergence moyen dû à des substitutions de bases de 1,4 % (1,2 à 1,75 %). Si l'on admet l'hypothèse de **l'horloge moléculaire**, ce résultat place le temps de divergence de notre espèce et celle du chimpanzé dans une fourchette de 4,6 à 6,2 millions d'années. Une estimation compatible avec les données les plus récentes de l'analyse des fossiles d'Hominidés, incluant notre plus vieil ancêtre hominidé, Toumaï (*Sahelanthropus tchadensis*), daté de 7 millions d'années.

Découvert en 2002, Toumaï, âgé de 6 à 7 millions d'années, serait le plus ancien représentant connu de la lignée humaine (Australopithèques et *Homo*), et semble proche des derniers ancêtres communs au chimpanzé et au genre *Homo*.
© Mission paléoanthropologique franco-tchadienne

Par ailleurs, l'analyse, chez l'homme et le chimpanzé, des sites de polymorphisme ponctuels (SNP, *single nucleotide polymorphisms*), c'est-à-dire de variations portant sur une seule base de l'ADN, a permis de déterminer les allèles ancestraux de groupes de SNP. Comme les études génétiques visant à comprendre les mécanismes moléculaires de maladies complexes comme le diabète requièrent l'analyse de milliers de ces SNP, cette donnée est essentielle pour valider des polymorphismes marqueurs de pathologies chez l'homme.

De plus, une fois un allèle ancestral identifié, il est possible de définir des régions hyperconservées (a priori soumises à une forte pression de sélection) et des régions hypervariables, qui permettent d'accréditer les hypothèses de sélection « positive » (un mouvement pour la diversité et la nouveauté sur lequel nous reviendrons plus loin).

Contraintes fonctionnelles

Ainsi, la comparaison des régions codantes de protéines de chimpanzé et d'homme a révélé un **taux de divergence** particulièrement faible, indiquant que ces régions codantes sont soumises à de fortes contraintes fonctionnelles. De plus, environ 40 % des mutations « synonymes » (c'est-à-dire sans influence sur la séquence protéique) sont éliminées dans les gènes humains codant des protéines par rapport aux régions intergéniques, en comparant aux mêmes régions du génome de chimpanzé [19].

Cette sélection « négative », qui signe le caractère délétère des mutations ainsi éliminées, affecterait principalement des séquences « activatrices » de l'**épissage** (appelées ESE pour « *exonic splicing enhancer* ») des transcrits primaires [20]. Les séquences ESE étant nécessaires pour l'épissage constitutif mais aussi

alternatif, cette contrainte évolutive pourrait indiquer que des altérations dans ces séquences auraient un effet négatif sur le niveau d'expression des ARN messagers concernés, mais aussi sur leur diversité.

Il est tentant de spéculer qu'une telle divergence dans les contraintes évolutives des séquences activatrices de l'épissage serait responsable des variations de l'expression génique entre le chimpanzé et l'homme, qu'elles soient quantitatives ***(voir l'analyse du transcriptome plus loin)*** ou qualitatives (accrétion de domaines protéiques par épissage alternatif). Cependant, cette analyse est encore trop partielle (deux espèces d'Hominidés seulement dans cette étude) et trop partiale (un seul article paru à ce jour) pour permettre d'en tirer des conclusions définitives…

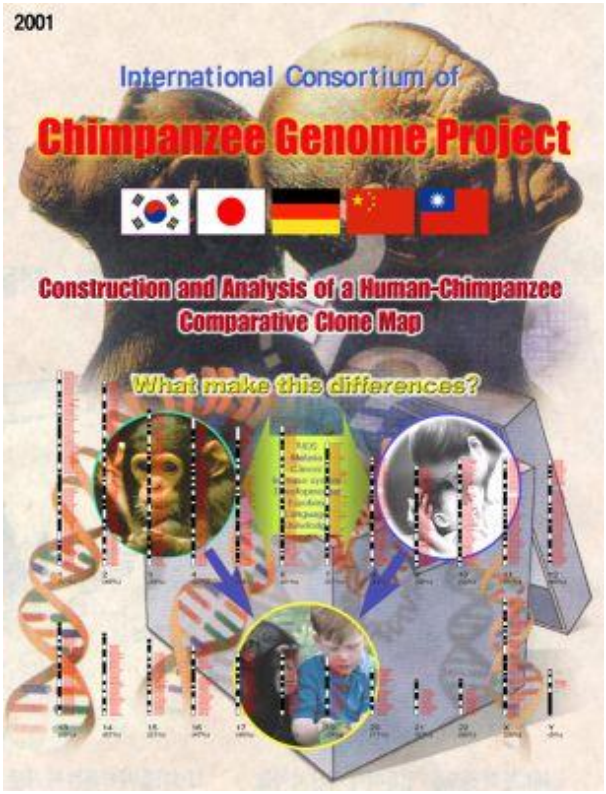
À l'inverse des contraintes évolutives pesant sur les régions codantes, une grande divergence a été trouvé en comparant les régions non codantes d'un même transcrit chez le chimpanzé et l'homme [19]. Ce phénomène, retrouvé au sein des populations humaines et d'autres espèces [10], reflète l'idée que ces régions, porteuses de séquences régulatrices, ont été soumises à une sélection « diversifiante » et suggère que l'évolution de telles séquences a été un élément majeur dans la séparation des deux espèces.

Enfin, la connaissance des allèles ancestraux des SNP permettra de clarifier l'hypothèse d'une expansion récente des populations humaines (il y a 50 0000 ou 150 000 ans selon le type d'analyse) suite à une diminution draconienne de la population ancestrale (théorie du « goulot de bouteille ») [21]. De même, elle devrait fournir des marqueurs adéquats pour retracer les migrations, colonisations et expansions géographiques de nos ancêtres et, en utilisant une stratégie adaptée d'échantillonnage des populations humaines [22], rendre compte de notre histoire démographique, qui reste bien énigmatique et controversée [23].

Des duplications génératrices de gènes propres à l'homme ?

L'analyse du génome de chimpanzé a confirmé l'existence d'une forte proportion de duplications particulières, dites « segmentaires », dans les régions situées autour des centromères et proches des télomères (régions relativement pauvres en gènes et à fort taux de réarrangements chromosomiques), ainsi que dans l'euchromatine (soit les régions transcrites représentant 5 à 10 % du génome humain). Ces duplications, identifiées initialement dans le génome humain, sont apparues dans les derniers 35-40 millions d'années, au cours de l'évolution des primates. Elles sont constituées de multiples morceaux d'**exons** et d'**introns** appartenant à des gènes présents par ailleurs en copie unique chez l'ancêtre commun.

Ces gènes « chimériques » sont compactés ou remaniés dans des structures pouvant atteindre plusieurs millions de bases. Du fait de leur très forte identité de séquence (95-99 %), ces duplications sont à l'origine de multiples remaniements chromosomiques (délétions, inversions, translocations), induites par des recombinaisons homologues inégales, et elles sont potentiellement à l'origine de pathologies chez l'homme, comme l'**adrénoleucodystrophie** ou les syndromes de **Prader-Willi** et d'**Angelman**.



Un consortium de 5 pays (Allemagne, Chine, Japon, Corée, Taïwan) a publié en 2003 la séquence du chromosome 22 du chimpanzé.
© DR

Barrières reproductives

Ces duplications segmentaires ont aussi participé à l'élaboration des chromosomes en prédisposant à de multiples réarrangements différenciant les chromosomes chez les primates. Selon l'hypothèse de **Michael White** (1910-1983) présentée en 1978 dans son livre *Modes of speciation*, de tels réarrangements chromosomiques conduisent à des barrières reproductives et *in fine* à l'apparition d'espèces. Cette hypothèse prédit une divergence nucléotidique accélérée dans les régions fortement réarrangées par rapport aux régions colinéaires pour deux espèces données. La première étude utilisant une sélection de 115 **gènes orthologues** entre le

chimpanzé et l'homme semblait conforter cette hypothèse [24]. Cependant, des résultats très récents, qui reposent sur la comparaison des séquences globales du génome de chimpanzé et d'homme, contredisent cette assertion. Ils suggèrent plutôt l'absence d'hybridation entre les populations ancestrales [25].

Enfin, les duplications segmentaires sont certainement à l'origine de la formation de nouveaux gènes spécifiques de la lignée des hominidés, certains étant plus particulièrement exprimés dans le cerveau. Le cas des gènes *PMCHL1* et *PMCHL2* (*Pro-MCH-like*) est à cet égard illustratif de la complexité des mécanismes moléculaires mis en jeu pour permettre l'émergence d'une nouvelle famille de gènes chez les hominidés ***(voir l'encadré ci-dessous)***.

PMCHL : une famille de gènes propres aux Hominidés

Sans rentrer dans les détails, exposés par ailleurs **(1,2,3)**, l'histoire des gènes *PMCHL1* et *PMCHL2* (*pro-MCH-like*), qui semblent propres à la lignée des **Hominidés** (homme et grands singes), peut être décomposée en trois grandes phases :

1) Un remaniement du gène de **l'hormone de mélano-concentration (MCH)**, présent aujourd'hui sur le chromosome 12, a produit une copie tronquée et inactive de ce gène. Ce remaniement a été provoqué par la production d'une copie inversée du gène *MCH* (transcription inverse), autrement dit d'un ARN antisens « naturel », puis par l'insertion de cet ARN, par **rétrotransposition**, sur le bras court du chromosome 5 ancestral. Cet événement est survenu il y a environ 25-30 millions d'années ;
2) L'accumulation de mutations ou de microinsertions-délétions a entraîné la « création » de multiples exons d'un côté du site d'insertion sur le chromosome 5 et a abouti, il y a environ 18 millions d'années, à l'émergence d'une copie quasi achevée du gène *PMCHL1* ;
3) Enfin, une translocation d'une séquence contenant *PMCHL1* (duplication puis intégration en un autre site) a eu lieu au moment de la divergence des Hominidés, il y a 5-10 millions d'années ; elle a permis l'apparition d'une seconde copie, appelée *PMCHL2*, sur le bras long du chromosome 5.

L'histoire tumultueuse de ces gènes s'accompagne d'une non moins grande complexité quand on analyse leur patron d'expression. En effet, plusieurs dizaines de transcrits sont générés par ces deux gènes chez l'homme. Fait notable, alors que les deux copies *PMCHL1* et *PMCHL2* sont transcrites dans le testicule, seul le gène *PMCHL1* est exprimé dans le cerveau humain mais aussi de macaque. Cela suggère que ce gène acccomplissait une fonction cérébrale dès son origine.

De plus, le gène *PMCHL1* code potentiellement une nouvelle protéine ayant les caractéristiques d'un facteur nucléaire **(4)**. Il reste maintenant à lui trouver une fonction en relation avec son expression tissulaire et son évolution chez les primates. Son histoire contemporaine est donc loin d'être achevée !...

(1) A. Courseaux & J.L. Nahon (2001) Birth of two chimeric genes in the Hominidae lineage, *Science* 291 : 1293-1297.
(2) J. L. Nahon (2001) Des gènes « chimères » sont apparus dans la lignée des Hominidés : l'indice d'une spécificité génomique humaine ?, *médecine/sciences* 17: 411-413.
(3) Voir *CNRS info*
http://www.cnrs.fr/Cnrspresse/n392/html/n392a04.htm 392, avril 2001.
(4) A. Viale *et al.* (2000) Structure and expression of the variant melanin-concentrating hormone genes: only PMCHL1 is transcribed in the developing human brain and encodes a putative protein, *Mol. Biol. Evol.* 17: 1626-1640.

Accélérations de l'évolution

En se focalisant cette fois sur les séquences qui codent les protéines, il est possible de mettre en évidence une évolution accélérée qui résulte d'une sélection positive : les mutations sont conservées car elles créent un avantage sélectif. De multiples études analysant le génome humain et le génome de chimpanzé ont suggéré que les séquences protéiques évoluaient beaucoup plus vite dans les régions chromosomiques à fort taux de réarrangements [24] et dans des familles de gènes particulières, qui participent au contrôle de l'olfaction, de

l'audition, du catabolisme des acides aminés, du développement du squelette et du système nerveux [9].

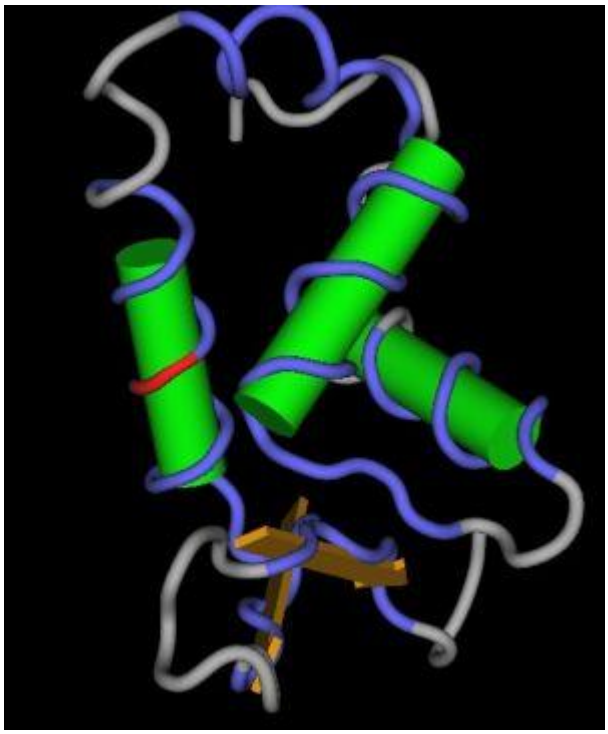
Ces accélérations dans les profils de substitution des acides aminés peuvent être le signe d'une perte de fonction (« pseudogénisation » : la mutation est maintenue car elle entraîne une perte de fonction qui ne donne plus prise à la sélection) ou d'une véritable adaptation évolutive sous la pression de l'environnement. Quant à l'émergence de fonctions propres à l'homme, il est clair que certains des gènes soumis à une forte sélection positive chez l'homme, et pas chez le chimpanzé, sont potentiellement des gènes impliqués dans le développement de l'audition, comme le gène de l'α-tectorine, un composant de la membrane tectoriale de l'oreille interne (gel qui couvre l'épithélium sensoriel de la cochlée) [9].

Dans ce contexte, la découverte en 2001 du gène *FOXP2*, est fascinante. Ce gène, qui code un facteur transcriptionnel, a été abusivement dénommé « gène du langage » sur la base de son implication dans un syndrome congénital comportant des troubles de l'articulation [26,27]. Il offre le paradoxe, pour un « gène du langage », d'être exprimé dans de nombreux organes, outre le cerveau, et d'être extrêmement conservé chez les mammifères. Cependant, il présente un polymorphisme et une évolution rapide dans l'espèce humaine qui serait survenue il y a environ 200 000 ans [28]. On peut donc être fortement tenté de voir en ce gène un prototype des gènes de l'hominisation. Ce serait cependant, à notre sens, un abus… de langage, compte tenu des « paradoxes » évoqués plus haut et de la forte composante environnementale et sociétale de l'évolution humaine.

Des explorations à poursuivre…

Par ailleurs, un certain nombre d'autres études pourraient s'enrichir des données du séquençage du génome de chimpanzé. Certaines de ces études sont d'ores et déjà publiées, mais pourraient être poursuivies en ce sens.

Par exemple, l'analyse de l'implication de séquences Alu (séquences répétées propres aux primates) dans les duplications et l'évolution des génomes, publiée en 2004 [29] serait plus complète si les auteurs s'étaient appuyés sur la totalité des données de séquençage des génomes du chimpanzé et de l'homme. En effet, l'insertion de séquences Alu, en apportant des séquences reconnues comme sites d'épissage, permet un phénomène d'« exonisation » (création de séquences d'introns-exons) propre aux primates. Ce processus, peu exploré chez le chimpanzé, est démontré pour une poignée de gènes chez l'homme : par exemple, le gène *Bcl-rambo bêta*, qui code une protéine variante de la protéine Bcl-rambo, impliquée dans la mort cellulaire [30], ou le gène *p75TNFR*, qui gouverne la synthèse du récepteur de type 2 du facteur de nécrose des tumeurs (TNF) [31].



Structure partielle de la protéine FOXP. Son altération (mutation identifiée en rouge) est responsable de troubles sévères du langage.
© Wellcome Trust Centre for Human Genetics

Des pouponnières à gènes

En outre, cette grande quantité de données génomiques pourrait également servir à l'étude des transcriptomes (ensembles des ARN transcrits à partir des génomes). Quatre études au moins, publiées entre 2002 et 2004, ont fait état de différences d'expression génique entre les différentes espèces de primates. En particulier, le groupe de **Svante Pääbo** (Institut Max-Planck d'anthropologie évolutionniste, Leipzig) a été le premier, en avril 2002, à révéler un niveau d'expression génique plus élevé dans le cerveau humain par rapport à celui du chimpanzé [32].

Une étude plus affinée de la même équipe, parue en août dernier sur la base des données du séquençage, a donné des résultats particulièrement intéressants [33]. Ainsi les transcriptomes de différentes aires cérébrales apparaissent très divers chez un même individu et, pour une aire donnée, au sein de populations d'hommes ou de chimpanzé. De plus, alors que l'expression des gènes codant des protéines de la signalisation et de la différenciation cellulaire diffère selon les aires cérébrales au sein d'une même espèce, elle est très conservée

quand on la compare, pour une aire donnée, entre les deux espèces.

En revanche, environ 10 % des gènes étudiés (soit environ 2 000 gènes) sont exprimés de manière différentielle pour une région donnée du cerveau humain et du cerveau de chimpanzé. Ces gènes se trouvent localisés préférentiellement au sein des duplications segmentaires dont nous avons parlé… ce qui renforce encore l'hypothèse que ces régions chromosomiques seraient des « pouponnières à gènes », porteuses de gènes nouvellement créés et dont l'expression serait spécifique de chaque espèce de primates [34].

Un modèle neutraliste ?

Enfin, le même groupe de l'Institut Max-Planck vient de révéler que la majorité des divergences d'expression entre les transcriptomes de cerveau de différents primates s'accumulent linéairement au cours du temps. Ces divergences seraient donc compatibles avec un modèle d'évolution « neutraliste », c'est-à-dire que les variations observées seraient sans importance fonctionnelle [35] ! En conséquence de quoi, Pääbo et ses collaborateurs proposent de réévaluer, à l'aide de nouveaux modèles mathématiques, tous les travaux démontrant une sélection positive de l'expression globale du génome.

Macaque rhésus (*Macaca mulatta*)
© 1999/2000 Naturschutz-Tierpark Görlitz e.V.

L'ensemble des ces données et les études qui s'appuieront sur les séquences génomiques du chimpanzé et de l'homme concourront probablement à une meilleure compréhension de « ce qui fait l'homme ». En particulier, les fonctions spécifiques de la lignée humaine, ainsi que les pathologies qui y sont associées, pourront être mieux explorées et analysées.

Cependant, il est nécessaire, pour bien analyser les données d'évolution-développement et de physiopathologie, de disposer d'une séquence « externe » relativement proche. Nous ne pouvons donc qu'appeler de nos vœux le séquençage prochain du génome d'un autre primate. Nous avons une préférence marquée pour le macaque rhésus (*Macaca mulatta*) et ce pour de multiples raisons : évidemment pour son implication comme animal modèle de pathologies humaines (infectieuses comme le paludisme, ou chroniques comme les maladies neurodégénératives), mais aussi pour son importance dans les analyses de neuroanatomie fonctionnelle (en PET scan), voire dans l'étude des aspects cognitifs de la sociabilité, une composante à l'origine de l'évolution humaine [36].

Enfin, le macaque rhésus possède un bagage génétique, en particulier au niveau des « pouponnières à gènes », très proche de celui de notre ancêtre anthropoïde qui portait en germe, il y a 25 millions d'années, toutes les modifications anatomiques et comportementales qui ont fait ce que nous sommes.

Fleur DARRE-TOULEMONDE et Jean-Louis NAHON

CNRS, Institut de Pharmacologie Moléculaire et Cellulaire,
Université de Nice-Sophia Antipolis,
660 route des Lucioles,
Sophia Antipolis
06560 VALBONNE

nahonjl@ipmc.cnrs.fr

[1] Séquence en accès libre sur le site Ensembl http://www.ensembl.org/Homo_sapiens.
[2] M.F. Teaford & P.S. Ungar (2000) Diet and the evolution of the earliest human ancestors, *PNAS* 97: 13506-13511. http://www.pnas.org.
[3] O.C. Lovejoy (2005) The natural history of human gait and posture Part 2. Hip and thigh, *Gait Posture* 21: 113-124.
[4] Voir une étude exceptionnellement informative et transdisciplinaire concernant l'évolution du gène *MYH16* chez les hominidés : H.H. Stedman *et al.* (2004) Myosin gene mutation correlates with anatomical changes in the human lineage, *Nature* 428: 415-418.

[5] M.D. Hauser *et al.* (2002) The faculty of language: what is it, who has it, and how did it evolve ?, *Science* 298: 1569-1579.
[6] Voir le consortium ARACHNE
http://genome.ucsc.edu/cgi-bin/hgGateway?org=Chimp, qui utilise un alignement du génome du chimpanzé et du génome humain pour valider l'assemblage des contigs (ensemble des fragments d'ADN contigus et ordonnancés utilisés pour reconstituer la carte physique du génome).
[7] Le site http://www.ensembl.org/pan_Troglodytes propose une série de données reposant sur l'assemblage de séquences issues de différents individus et réalisées par le Chimpanzee Sequencing Consortium dirigé par le GSC (St. Louis) et par The Broad Institute (MIT).
[8] A. Fujiyama *et al.* (2002) Construction and analysis of a Human-Chimpanzee comparative clone map, *Science* 295(5552):131-134.
[9] A.G. Clark *et al.* (2003) Inferring nonneutral evolution from Human-Chimp-Mouse orthologous gene trios, *Science* 302(5652):1960-1963.
[10] H. Watanabe *et al.* (2004) DNA sequence and comparative analysis of chimpanzee chromosome 22, *Nature* 429(6990):382-8. **[11]** J. Schmutz *et al.* (2004) The DNA sequence and comparative analysis of human chromosome 5, *Nature* 431 : 268-274.
[12] S.J. Humphray *et al.* (2004) DNA sequence and analysis of human chromosome 9, *Nature* 429 : 369-374.
[13] P. Deloukas *et al.* (2004) The DNA sequence and comparative analysis of human chromosome 10, *Nature* 429 : 375-381.
[14] G. Liu et al. (2003) Analysis of primate genomic variation reveals a repeat-driven expansion of the uman genome, *Genome Res.* 13:358.
http://www.genome.org/cgi/content/full/13/3/358.
[15] T. Anzai *et al.* (2003) Comparative sequencing of human and chimpanzee MHC class I regions unveils insertions/deletions as the major path to genomic divergence, *PNAS* 100 (13):7708-7713.
http://www.pnas.org/.
[16] J. Weissenbach (2004) Differences with the relatives, *Nature* 429: 353-355.
[17] M. Ruvolo (2004) Comparative primate genomics: the year of the chimpanzee, *Current Opinion in Genetics & Development* 14: 650-656.
[18] V. Barriel (2004) Ces 1,4% qui nous séparent des chimpanzés !, *médecine/sciences* 20: 859-861.
[19] I. Hellmann *et al.* (2003) Selection on human genes as revealed by comparisons to chimpanzee cDNA, *Genome Res.* 13(5):831.
http://www.genome.org.
[20] W. Fairbrother *et al.* (2004) Single nucleotide polymorphism-based validation of exonic splicing enhancers, *Plos Biol.* 2(9):E268.
http://www.plosbiology.org.
[21] H. Kaessmann & S. Pääbo (2002) The genetical history of humans and the great apes, *J. Int. Med.* 251: 1-18.
http://email.eva.mpg.de/~paabo/pdf1/Kaessmann_JIM_2002.pdf.
[22] S. E. Ptak & M. Przeworski (2002) Evidence for population growth in humans is confounded by fine-scale population structure, *TRENDS in Genetics* 18: 559-563.
[23] E. Crubézy & J. Braga (2003) Homo Sapiens prend de l'âge, *La Recherche* 368: 30-35.
[24] A. Navarro & N. Barton (2003) Chromosomal speciation and molecular divergence--accelerated evolution in rearranged chromosomes, *Science* 300(5617): 321.
[25] J. Zhang *et al.* (2004) Testing the chromosomal speciation hypothesis for humans and chimpanzees, *Genome Res.* 14: 845-851.
http://www.genome.org.
[26] C.S.L. Lai *et al.* (2001) A forkhead-domain gene is mutated in a severe speech and language disorder, *Nature* 413: 519-523.
[27] G.F. Marcus & S.E. Fisher (2003) FOXP2 in focus: what can genes tell us about speech and language ?, *TRENDS in Cognitive Sciences* 7 : 257-262.
[28] W. Enard *et al.* (2002) Molecular evolution of FOXP2, a gene involved in speech and language, *Nature* 418: 869-872.
http://email.eva.mpg.de/~paabo/pdf1/Enard_Molecular_Nature_2002.pdf.
[29] J. Jurka *et al.* (2004) Duplication, coclustering, and selection of human Alu retrotransposons, *PNAS* 101(5):1268.
http://www.pnas.org/.
[30] P. Yi *et al.* (2003) Bcl-rambo beta, a special splicing variant with an insertion of an alu-like cassette, promotes etoposide- and taxol-induced cell death, *FEBS Letters* 534: 61-68.
[31] S.S. Singer & D.N. Männel (2004) From « junk » to gene: curriculum vitae of a primate receptor isoform gene, *J. Mol. Biol.* 341 : 883-886.
[32] W. Enard *et al.* (2002) Intra- and interspecific variation in primate gene expression patterns, *Science* 296(5566):340-343.
[33] P. Khaitovich *et al.* (2004) Regional patterns of gene expression in human and chimpanzee brains, *Genome Res.* 14:1462-1473.
http://www.genome.org.
[34] J.L. Nahon (2003) Birth of « human-specific » genes during primate evolution, *Genetica* 118(2-3):193.
[35] P. Khaitovich *et al.* (2004) A neutral model of transcriptome evolution, *PLoS Biology* 2: 0682-0689. http://www.plosbiology.org.
[36] C.J. Jolly (2001) A proper study for mankind: analogies from the papionin monkeys and their implications for human evolution, *Yearbook of physical anthropology* 44: 177-204.

Glossaire

L'Adrénoleucodystrophie

Maladie génétique de transmission liée à l'X (transmise par les femmes porteuses), l'adrénoleucodystrophie (ALD) touche les garçons. Il s'agit d'une atteinte métabolique (dystrophie) de la substance blanche (leuco) du système nerveux. C'est une des plus fréquentes des maladies dégénératives du système nerveux central avec une incidence est de 1 sur 12000 naissances.
voir Association française contre l'adrénoleucodystrophie
http://www.med.univ-rennes1.fr/sisrai/art/adrenoleucodystrophie,_p.187.html

Le syndrome d'Angelman

Il est lié à différentes anomalies d'origine génétique localisées sur le chromosome 15, dans la région 15q11-q13. Ses manifestations constantes sont : une déficience mentale sévère, un retard du développement moteur, une absence de langage (le niveau de compréhension est meilleur que l'expression), une ataxie avec démarche raide et saccadée, des comportements particuliers : sourires et rires très faciles avec épisodes d'hyperexcitabilité se traduisant par un battement des avant-bras et enfin des difficultés d'attention.
L'incidence de ce syndrome est estimée très approximativement de 1/12 000 à 1/20 000.
voir Association francophone du syndrome d'Angelman
http://www.angelman-afsa.org

Epissage

Processus qui transforme les molécules d'ARN (transcrits) issues de la transcription de l'ADN et porteuses de séquences non codantes (introns) et de séquences codantes (exons) en ARN messagers porteurs uniquement d'exons. Ce processus engendre une diversité des protéines issues d'un même gène par le jeu du raboutement d'exons qui varient d'un ARN messager à un autre (épissage alternatif).

Hominidés et Homininés

Selon les classifications, on parle d'homininés ou d'hominines pour désigner le groupe phylogénétique de l'homme.

Du fait de la très faible différence génétique entre *Homo* et les deux espèces de chimpanzé (*Pan troglodytes* et le bonobo *Pan paniscus*), les systématiciens appellent cet ensemble Hominidés.

Certains réservent le terme Homininés au groupe Australopithèques + *Homo*. Pour d'autres, ce dernier groupe représente les Hominines, et les Homininés correspondent au groupe (Hominines + chimpanzés). Dans cette logique, on a :

- **Espèce *Homo sapiens***
- **Genre *Homo* : *Homo sapiens* + les espèces fossiles comme néanderthal ou *erectus***
- **Hominines : les genres *Australopithecus* et *Homo***
- **Homininés : les deux espèces de chimpanzés et les hominines**
- **Famille des Hominidés : gorille et Homininés**
- **Hominoïdés : Hominidés + orang-outang**
- **Hominoïdes : Hominoïdés + gibbons**

L'horloge moléculaire

Le taux d'évolution par mutations des protéines au cours du temps semble à peu près constant chez les vertébrés. Si l'on connaît le taux d'accumulation des mutations en un site chromosomique, il est possible d'estimer le temps de divergence entre deux espèces en comparant les séquences présentes dans ce site (gènes homologues des deux espèces) ou les séquences des protéines correspondantes. Plus les régions sont soumises à une forte pression de sélection (régions codant des gènes et protéines essentiels à la survie de l'organisme), plus le taux d'accumulation de mutations est faible et l'horloge lente. Voir Infobiogen
http://www.infobiogen.fr/doc/documents.php?cours=bioinfo_analseq10

L'hormone de mélano-concentration (MCH)

La *Melanin-Concentrating Hormone* (MCH) est un neuropeptide clef du contrôle de la prise alimentaire et de l'homéostasie énergétique, produit par des neurones de l'hypothalamus.

Gènes orthologues

Gènes d'espèces différentes dont les séquences sont homologues. Ils dérivent d'un même gène ancestral et ont divergé à la suite d'un événement de spéciation tel qu'un réarrangement chromosomique. Ils peuvent avoir ou non la même fonction.

Le syndrome de Prader-Willi

Le syndrome de Prader-Willi est une maladie rare (une naissance sur 10 000 à 15 000) qui a pour cause l'absence ou la perte de fonction de gènes au niveau du chromosome 15. Dans plus de 95 % des cas, cette anomalie génétique n'est pas héritée des parents, il s'agit d'une mutation *de novo*.

Les enfants atteints manifestent à la naissance (et même à l'état foetal) une faiblesse musculaire sévère qui entraîne des difficultés d'alimentation chez le nouveau-né. Par la suite, ils ont besoin d'un apport calorique bien moins important que la normale pour ne pas grossir et, à partir de deux à cinq ans, ils ne semblent pas éprouver de satiété. Si un régime alimentaire hypocalorique strict accompagné d'activités physiques régulières n'est pas mis en place, ces enfants développeront une obésité qui pourrait mettre leur vie en danger.

voir Association Prader-Willi France
http://perso.wanadoo.fr/pwillifr/

Rétrotransposition

Evénement de tranposition d'une séquence génique d'un site chomosomique à un autre, au moyen d'un intermédiaire ARN.

Séquences Alu

Familles de séquences répétées dispersées sur l'ensemble du génome des primates uniquement. Elles sont reconnues par l'enzyme de restriction Alu.

Un taux de divergence protéique très bas

Ce taux, dN/dS, est de 0,22 %.
dN : nombre de substitutions d'acides aminés ;
dS : nombre de substitutions « synonymes », c'est-à-dire sans influence sur la séquence protéique.

Pour aller plus loin

- **Chimpanzee Genome**
  http://genome.wustl.edu/projects/chimp/index.php
- **National Human Genome Research Institute (NHGRI)**
  http://www.genome.gov/11008056
- **Nature Focus**
  http://www.nature.com/nature/focus/chimpgenome/
- **Publications en ligne de Svante Pääbo**
  http://email.eva.mpg.de/~paabo/files/public.html

© Vivant Editions – http://www.vivantinfo.com

EVOLUTIONARY VARIATIONS IN STRUCTURE AND BRAIN EXPRESSION OF THE *PMCHL1* GENE IN MACAQUE AND HUMAN

Fleur DARRE-TOULEMONDE [1], Alain CORINUS [1,2], Audrey DELERUE-AUDEGOND [1,2], Richard CHRISTEN [3] and Jean-Louis NAHON *[1,2]

1- CNRS UMR 6097, UNSA, IPMC, Valbonne, France

2- CNRS «Primate Biolab»,1 and CNRS UMR 5543, Bordeaux, France

3- CNRS "Virtual Biology", Nice, France

* Corresponding Author

ABSTRACT

Brain-expressed genes that were created in primate lineage represented obvious candidates to investigate molecular mechanisms that contributed to neural reorganization and emergence of new behavioural functions in *Homo sapiens*. *PMCHL1* arose from retrotansposition of a melanin-concentrating hormone (MCH) antisense mRNA on the ancestral human chromosome 5p14 when Platyrrhines and Catarrhines diverged. Mutations before divergence of Hylobatidae led to creation of new exons and finally *PMCHL1* duplicated in ancestor of Hominids to generate *PMCHL2* at the human chromosome 5q13. A complex pattern of spliced and unspliced PMCHL mRNAs were found in human brain and testis. In this report, we established that the initial insertion event took place within an intron of the *brain cadherin* (*CDH12*) gene, soon after Platyrrhines divergence 30-35 Mya. The region that duplicated on 5q13 to generate *PMCHL2* covered 92 kb of the intronic sequence and encompassed exon 3 of *CDH12*. Numerous short indels and copies of different classes of repeated sequences were mapped within *PMCHL1* exons-introns in Catarrhines. Localized sequence variations among primates were found in regions encompassing these elements, indicative of species-specific divergences in transcribed regions. Indeed, while the cortex/cerebellum-specific distribution of unspliced PMCHL1 transcripts carrying these regions appeared conserved in macaque and human, both quantitative and qualitative differences were found. In particular, unspliced PMCHL1 transcripts of different lengths and abundance were found in macaque and human cortex. Overall, our results provided a framework for further investigations regarding the relative expression and functions of PMCHL1 transcripts in a suitable non-human primate model.

INTRODUCTION

Primate species are very narrow in terms of genomic nucleotidic identities (values for which the speciation barrier doesn't occur in other species such as rodents), but show rather different phenotypic traits. When genomes are distantly related, research is oriented toward conserved sequences, assuming they must be functionally important. For closely related species, such as hominoid species, the basic premise is « if it is different, it might be important in explaining species differences ». However there is an ancient (King and Wilson 1975) but still active debate in the geneticist community about the relative contribution of structural genomic modifications that led to emergence of novel genes (Bailey and Eichler 2006)or loss of genes and genome-wide gene expression changes, particularly in the brain (Khaitovich, Enard et al. 2006) that could account for the phenotypic differences observed between primate species. A genome-wide comparative study of mammalian promoters suggested an accelerated evolution of primate promoters during the last 25 million years (Taylor, Kai et al. 2006). Strikingly, higher mRNA and protein expression levels were found in the human brain relative to other primates (Enard, Khaitovich et al. 2002; Caceres, Lachuer et al. 2003; Uddin, Wildman et al. 2004). However, overall gene expression in the brain has diverged less than in the heart, kidney, liver or testis among human and chimpanzee species(Hsieh, Chu et al. 2003), in agreement with a neutral model of transcriptome evolution not only in primates but also shared among mammals (Khaitovich, Enard et al. 2006). In parallel, contrasting results were found when determining the Ka/Ks ratio, a tentative indicator of positive Darwinian selection, in the coding region of genes expressed in the mammalian brain (Clark, Glanowski et al. 2003; Dorus, Vallender et al. 2004). However, there are few but well established examples of brain expressed genes being positively selected in the human lineage, most of them being involved in brain growth such as microcephalin or ASPM (Evans,

Anderson et al. 2004; Wang and Su 2004; Mekel-Bobrov, Gilbert et al. 2005) and neurotransmission such as GLUD2 (Burki and Kaessmann 2004). Recently, divergence between human and chimpanzee sequences have been reevaluated to almost 5%, resulting mainly from indels events (deriving from specific retrotranspositions, segmental duplications...) (Britten 2002; Britten, Rowen et al. 2003; Frazer, Chen et al. 2003; Bailey and Eichler 2006) and offering therefore a wide variety of sites at which lineage-specific genetic novelty could happen. Indeed, recent segmental duplications are particularly enriched in genes that displayed expression differences between humans and chimpanzees (Khaitovich, Muetzel et al. 2004). To reconcile apparently conflicting data we previously proposed that genomic rearrangements, comprising creation of novel but rare protein coding and/or non-coding genes, may have been a major driving force during primate evolution, in addition to single nucleotide mutations that confer alterations in the gene expression patterns or amino acids sequences (Courseaux and Nahon 2001; Nahon 2003). These accidental events would have been particularly important in shaping the hominoid brain, as exemplified by the GLUD2 gene model (Burki and Kaessmann 2004). Current primate genome and gene expression comparative studies tend to support this hypothesis (reviewed in (Khaitovich, Enard et al. 2006; Sikela 2006)).

Processes of exon shuffling, retrotransposition, gene promoter fusion and segmental duplications have been suggested to lead to novel genes harbouring completely new structure and expression patterns selectively in the primate lineage (reviewed in (Long 2001; Long, Betran et al. 2003)). Because of the supposed rapid evolution of this kind of genes, the study of the creation and early evolution of a gene requires the discovery of a gene that has retained caracteristical features of its youth (Long, Deutsch et al. 2003).

PMCHL1 and *PMCHL2* genes, located respectively on human chromosome 5p14 and 5q13, fully meet these requirements. They have previously been shown (Courseaux and Nahon 2001; Courseaux, Richard et al. 2003) to have been created in the hominoid lineage through retrotransposition at the ancestral

chromosome 5p14 locus of an antisense melanin-concentrating hormone (MCH) gene transcript, local rearrangement leading to a truncated version of the retrogene, sequence remodelling (indels and mutations accumulation that allowed creation of exons) and later duplication on the 5q arm. Furthermore, processed and unprocessed transcripts were characterized in a human fetal brain library (Courseaux and Nahon 2001) as well as in developing human brain (Viale, Courseaux et al. 2000). One of these mRNAs was found encoding a putative nuclear protein of 8 kDa that was identified using *in vitro* translation systems (Viale, Courseaux et al. 2000). However, many questions remained unsolved regarding the region of insertion of the retrogene, the fine structure of both genes (exact boundaries, described repeated elements found inside the duplicon, complete exon/intron structure...), as well as their expression patterns (relative abundance and tissue-specificity of processed and unprocessed transcripts). In particular, an important challenge was to find a suitable non-human primate model to test *PMCHL* gene regulations and functions.

In this paper, we established the structure of *PMCHL1/PMCHL2* genes and regional organization at both loci on human chromosome 5. We revealed a very dense cluster of repetitive sequences that seemingly contribute to the high rate of sequence rearrangement of these genes when introduced in bacteria. Furthermore, varying mutation rates along the sequence and among the primate species are also documented. Finally, PMCHL mRNAs expressions are compared in macaque and human brains, and macaque appears to be a good and accessible *in vivo* model to study these genes.

MATERIALS and METHODS

Tissues

Different human tissues from adults were provided by the National Research Specimen Bank (Los Angeles, Calif.). Dr. D. Jordan (Faculté de medecine Laennec, Lyon) provided other newborn and fetal tissues. The reference, age, gender, cause of death and postmortem delay for each sample were described elsewhere (Viale, Courseaux et al. 2000).

Cortex and cerebellum samples from two adult macaques (*Macacca fascicularis*) were delivered by Dr. E. Bezard (Basal Gang lab,CNRS UMR 5584, Bordeaux 2).

Genomic DNAs

Genomic DNAs were collected from *Cebus capucinus* (gift from B. Dutrillaux, cytogénétique moléculaire et oncologie, CNRS, Institut Curie, Paris, France), *Tarsius syrichta*, *Saguinus oedipus*, *Chlorocebus aethiops*, *Hylobates lar*, *Pan paniscus*, *Pan troglodytes* and *Homo sapiens* (gift from Ph. Dijan, CEREMOD,Meudon, France).

Genomic DNA was isolated from the occipital cortex of a *Macaca fascicularis* (provided by Dr. E. Bezard) according to the Blin and Stafford's method.

RNA extraction and reverse-transcription

Whole cell RNA was extracted according to standard guanidium phenol method (Chomczynski) or using a fastprep apparatus (FP220A Thermo instrument, Qbiogene, France).

cDNAs were synthetised during the reverse-transcription (RT) of 2 to 3 µg of total RNAs, using the reverse-transcriptase Superscript II (Invitrogen) and according to the manufacturer's protocol.

PCR amplification

Primers

Oligonucleotides were purchased from Eurogentec (Belgium). Sequences, Tm and PCR conditions are provided in table 1 found as online supplement.

PCR conditions

One hundred to two hundred nanograms of genomic DNA or 2•L of RT samples were added to PCR reaction solution following the supplier's protocol (LA TAQ,TaKaRa, except with the macaque RNA samples, for which the Eppendorf Hotmaster was used), in a total volume of 25 •l. Thirty-five cycles of amplification were carried out as follows: 30 s at 96 °C (denaturation), 30 s at a variable temperature, depending on the oligonucleotides Tm (annealing), 1 to 10 min at 72 °C (extension). A final extension step of 7 min at 72 °C was performed.

When necessary, nested PCR was performed with internal primers using 2 •L of the pure first round products.

To reveal DNA contamination, RT of RNA was performed in absence of the enzyme, and PCR was carried out under standard conditions.

Southern blotting

PCR products were electrophorized on an (1%) agarose gel containing ethidium bromide and visualized under UV. The gel and samples were then treated and transferred on a cellulose membrane (Biodyne A, Polylabo) as described previously (Viale, Ortola et al. 1998). The samples were finally cross-linked.

^{32}P-labeled specific fragments were prepared according to the manufacturer's protocol (Prime-a-gene labelling system, Promega) and applied to the membrane (5.10^5 dpm.mL^{-1}). Pre-hybridization, hybridization and washing conditions were described elsewhere (Viale, Ortola et al. 1998). Densitometric analysis was performed with a Fujifilm phosphoimager (BAS-1500).

DNA sequencing

The DNA sequence of PCR-amplified fragments was determined from double-strand DNA using the Ampli Taq Polymerase FS and the Big Dye Terminator sequencing kit (Applera) and a Perkin Elmer ABI PRISM 377 sequencer.

Sequence alignments and analyses

Sequences obtained from the public databases (EMBL/GenBank/DDBJ) and fragments sequenced by PCR were aligned manually using SEAVIEW (Galtier, Gouy et al. 1996). The matrices of distances produced by DNADIST (see below) using the Kimura correction were used to estimate sequences similarities.

Repeated sequences were identified using the censor web server, available at GIRI.

Phylogenetical analysis

Phylogenetic dendrograms were reconstructed according to three different methods: Neighbour Joining (BIONJ), Maximum Likelihood (ML, using the Global option), and Maximum Parsimony (MP). For the Neighbour Joining (NJ) analysis, a distance matrix was calculated according to the Kimura two parameters correction. Bootstraps were done using 1000 replications, BIONJ and Kimura two parameters correction. BIONJ was according to Gascuel (1997)(Gascuel 1997), ML and MP were from PHYLIP (Phylogeny Inference Package, version 3.573c, distributed by J. Felsenstein, Department of Genetics, UW, Seattle, WA, USA). Phylogenetic analyses were done excluding domains that were not common to every sequence as well as low complexity domains that could not be properly aligned. The phylogenetic dendrograms were drawn using NJPLOT (Perrière & Gouy, 1996).

Studying the nucleotidic variability

In order to estimate sequence variability we took the simple approach to estimate entropy along the sequences, according to a Shannon equation where indels are seen as a 5^{th} character. In order to obtain "readable" curves, a mean was calculated according to a sliding window (30 to 50 nucleotides). These estimates are not correct from a phylogenetic point of view as long indels are likely the result of a single event, but from a structural point of view this allowed to spot out domains that were similar among sequences from domains that had many mutations or deletions. These domains were then extracted and

distances between each pair of sequences were calculated using DNADIST and the Kimura two parameters correction (additional tables)(Yang 1997).

RESULTS

1. Structural analysis of PMCHL genes

We previously proposed (Courseaux and Nahon 2001) that *PMCHL2* was created from a duplication of a large, but undefined in size, genomic DNA fragment comprising *PMCHL1*, "jumping" from ancestral hominid chromosome 5p14 to 5q13. The limits of the duplicon were now precisely determined by similarity using a BLAST search and it appeared that a fragment of 92kb, encompassing 17kb upstream and 65kb downstream the 10kb of *PMCHL1* was duplicated (see figure 1A). This 92kb duplicon corresponds mainly to the 5' portion of the third intron of *CDH12* gene encoding the brain cadherin (as defined in http://www.ensembl.org), but also encompasses the cdh12 exon 3 and the last 4kb of intron 2 (see Figure 1A).

No other exon (found elsewhere in the genome) than those derived from the *MCH* gene were found in the duplicon. However, three EST (identified from IMAGE clones), two of them deriving from the same gene, mapped to the duplicon region, on the same strand than the *CDH12* gene (i.e. opposite strand compared to the *PMCHL1* gene). These are ENSESTT00000060506, ENSESTT00000060502 and ENSESTT00000060504 (according to the Ensembl nomenclature). The percentages of identity between the 5p and 5q elements were equivalent all along the duplicon (i.e. in the *PMCHL* genes, and in the 5' and 3' flanking regions), and are close to 98%.

We also aimed to determine the fine structure of the *PMCHL1/L2* genes. Amplification from human foetal brain (Marathon-Ready cDNAs; Clontech Laboratories) and sequencing of messenger RNAs allowed the discovery of a novel exon, located between exon 2 and former exon 3. It was named exon 3.1 and exon 3 became exon 3.2. The *PMCHL1* gene was also precisely analyzed to identify known repeated sequences (figure 1B). It is worth noting that a complete Alu sequence mapped in exon 2, appeared to be an AluSg, and not an Alu Sq as previously described)(Viale, Courseaux et al. 2000). A high density of parts of known repeated sequences was also found in

intron C, between exon 3.2 and 4, which could explain the difficulty to sequence this part of the gene. This high number of repeated sequences could also contribute to the extraordinary instability of this gene, as observed in various bacterial systems while attempting to clone the gene. Indeed, PMCHL1 fragments encompassing the full 10-15 kb sequence (case of the phMCHL37 clone (AF 227988) corresponding to a 12kb fragment) or various fragments encompassing 5'UTR to exon 2' (amplified from a BAC) were conspicuously recombined either in DY 380 strains (having recombining properties), and even in DH10B- or SURE2 bacteria (Fig. 1B and data not shown).

Primates *PMCHL* sequences were obtained by PCR amplification of genomic DNA and direct sequencing, which allowed the determination of their fine structure in terms of presence/absence of repeated elements, and especially to date the insertion of the Alu Sg sequence inside exon 2 of *PMCHL1*. This insertion was concomitant to the retroposition. We actually did not find any primate species carrying the *PMCHL1* exon 2 without the Alu sequence. Each event was thus dated after the divergence of *Cebus* species (*C. appella* and *C. capucinus*), and before the divergence of the cercopithecoids, approximately between 35 and 25 Mya (see Figure 1C). Other repeated elements found in the human *PMCHL1* gene were already present in the Cercopithecoids, such as LINE2 (in the 5'UTR region) and the CATA repeat (inside intron B1). (not shown). Furthermore, numerous short indels (less than 30 bp) were found, some being fixed in hominoid lineage (red arrows), other present only in Cercopithecoids (pink arrows) or in *Cebus MCH* gene (blue arrowheads).

A nucleotidic phylogenetic analysis was also performed to date more precisely the retroposition event. As *PMCHL1* encompasses some *MCH* gene sequences (sense and antisense), it was possible to align primates sequences of these specific parts of the *PMCHL1* and *PMCH* genes. A phylogenetic analysis was therefore performed using the parcimony, maximum-likelihood and neighbour-joining methods and with the rat and mouse sequences as outgroups. All phylogenetic methods led to congruent data (cf. Fig 2)(with high bootstrap values with the neighbour joining method). The tree showed an apparent

aberration relative to our present knowledge concerning the relation between species and the creation of the *PMCHL1* gene (circled in figure 2A). Thus, we expected the *PMCH* sequences of *Cebus capucinus* and *Saguinus oedipus* to be grouped with the *MCH* sequence of *Tarsius syrichta* rather than with the *PMCHL* sequences. However, a noteworthy low bootstrap value (34 %) was found for this branching. The position of the *PMCHL* sequences as well as the uncertainty for positioning the *Cebus capucinus* and *Saguinus oedipus MCH* sequences suggest that the duplication leading to the *PMCHL* genes occurred very shortly (likely within 5 My) after the split of Platyrrhini/Cathyrrhini

When focusing on the *PMCHL* genes the nucleotidic phylogenetic tree (Figure 2B) fully corresponded to an accepted species tree, indicating that no particular and global (since there the entire gene sequences were used) evolutionary event interfered. However, we observed a difference in the *Pan troglodytes*/ *Pan paniscus*/ *Homo sapiens* positioning between the two phylogenetic trees. This simply corresponds to an inherent irresolution, the three species diverging at the same time. All three species should probably be grouped under the *Pan* clade as previously suggested (Goodman 1999).

We then focused on local variability. Estimation of sequences variability among a set of sequences is often done using tools such as PAML (Yang 1997), REVCOM (Bordner and Abagyan 2005) or WebVar (Mignone, Horner et al. 2004). Unfortunately these approaches are geared toward protein coding sequences and do not take into account indels. A specific algorithm measuring changes in variability of the sequences between species was developed and used. First, an analysis of the pattern of substitution of the *PMCH* gene in primates revealed, as expected (Cardinaud, Darre-Toulemonde et al. 2004), constrained regions selectively within protein coding sequences of the three exons (see supplementary data, figure 1A). Thereafter we extended our analysis to the *PMCHL* genes and identified regions exhibiting either apparent hyper variability or strong conservation by comparison with neutral sequence (supplementary data, figure 1B).

2. Expression of the *PMCHL* genes in human and macaque.

We have shown previously (Viale, Courseaux et al. 2000; Courseaux and Nahon 2001) that both sense and antisense unspliced *PMCHL1* transcripts overlapping the retrotransposed MCH sequence could be found in human foetal brain and testis. Sense spliced *PMCHL* gene products were also identified but appeared rather weakly expressed.

Here, we aimed first to characterize further the spliced transcripts and their relative abundance. Nested PCR on cDNA (home-made or marathon library), followed by a sequencing step, allowed the characterisation of novel alternative splicings, with splicing feets inside the exons, and a novel exon in intron B (named exon 3.1) (Figure 3). The relative variations in detection of each mRNA with the same RT and PCR conditions gave a crude estimate of their relative abundance. The most abundant spliced messengers in the foetal brain harboured exons 1-2' and 4-5a, thus missed exon 2, and corresponded selectively to the *PMCHL1* gene. Moreover, no human EST encompassing exons 4-5 but not exons 1-2 were found in the EST databases (data not shown). Testis presented a rather high diversity of splice transcripts that appeared to be equally distributed within at least seven classes. All these alternatively spliced mRNAs lead to a variety of putative ORFs but of short length (less than 75 amino-acids): five putative ORFs were described in two regions of the *PMCHL* genes (see Figure 3).

Thereafter, we characterized the expression pattern of the *PMCHL1* gene in human and macaque brain. First, RT-PCR and Southern blotting were performed with RNAs extracted from cortex and cerebellum of adult human and macaque in order to identify the main spliced and unspliced messenger RNAs, containing exons 1-2, 4-5 and 1-5 respectively (Figure 4A). As shown in Fig 4B, unspliced RNAs encompassing exons 1-2 and 4-5 were both detected in human and macaque cortex, while only exons 4-5-containing unspliced mRNAs were found in the cerebellum of both species. Interestingly, under the RT-PCR conditions we used, no spliced mRNAs encompassing exons 1 to 5 could be identified in the cortex and cerebellum of adult human and macaque. Therefore, these data suggest that two

separate transcriptional units produce unspliced mRNAs containing respectively exons 1-2 and 4-5, displaying similar expression patterns in the adult human and macaque cortex and cerebellum.

The expression pattern of *PMCHL1* gene was further characterized in human and macaque cortex by looking for transcriptional initiation in the 5'flanking region, i.e. upstream the site of retrotransposition (fig.5). Previous RACE-PCR analyses using a Marathon fetal brain cDNA library have suggested the location of a major cap site about 300 bp upstream from the insertion site of the retrotransposon (Viale, Courseaux et al. 2000; Courseaux and Nahon 2001). To determine the extend of PMCHL1 transcripts we performed RT-PCR with five pairs of primers named a/b to g/h which identified regions covering up to 6.8 kb upstream this putative cap site (fig 5A). In agreement with previous RACE-PCR experiments, we found amplification with a/b couple but not with other pairs of primers in RNA extracted from human cortex (Fig. 5B). Surprisingly, PCR products were identified with all primer couples in RNA samples extracted from macaque cortex (Fig. 5B). All corresponded to unspliced PMCHL1 transcripts, some of them were characterized in detail and overlapped up to 3.0 kb of the 5' flanking region upstream from the originally defined cap site. Further characterization of RT-PCR products amplified with h/g and f/e primers revealed alternative start sites in the region covered by probes 4 and 5 (not shown). Even though we cannot exclude very low expression of similar transcripts in the human cortex, this suggests that different starting points of transcription operate in the adult macaque and human cortex.

DISCUSSION

We provide here new data concerning the evolution of the *PMCHL* genes family offering a substantial improvement of the creation model proposed previously (Courseaux and Nahon 2001). The insertion of the Alu element was more precisely dated, and the concomitance of this event with the retrotransposition event was proposed. These structural data and the positioning of the specific indels also point out on lineage specific mutations. Cercopithecoids-specific mutations appear particularly clearly.

Phylogenetic analyses and difficulties to resolve a specific node indicate that *PMCHL1* was created rapidly after the divergence of platyrhini species, i.e. 30-35 Mya (Glazko and Nei 2003; Enard and Paabo 2004). On the other hand, the irresolution between the two *Pan* species and *Homo sapiens* simply corresponds to an inherent issue to Hominidae speciation. Indeed, the three species diverged at the same time and followed subsequently a complex speciation with interbreeding before final separation about 2.9-4.3 Mya (Wall 2003; Patterson, Richter et al. 2006). All three species should probably be grouped under the *Pan* genus (Goodman 1999; Enard and Paabo 2004)((Wildman, Uddin et al. 2003).

It is often stated that phylogenetic analyses may be misleading when large variations in evolutionary rates occur in the different lineages. However, the phylogenetic trees inferred from data sets simulated under realistic, observed levels of heterogeneity for mammalian genes are reconstructed with accuracy comparable to those simulated with homogeneous nucleotide frequencies (Rosenberg and Kumar 2003). For the primate species, it is now clear that local molecular clocks do exist, and that different lineages accumulate nucleotide substitutions at different rates (Hacia 2001; Enard and Paabo 2004). Taking into consideration these remarks, and having shown that no global selective constraint or positive selection could be detected, we chose to study varying substitution rates along the sequences. Thus, novel data on local and variable evolution speeds in *PMCHL* genes were

revealed, but they unfortunately seem to be due to technical problems. Many recent papers indeed mentioned human-specific acceleration of protein sequence changes in genes involved in the control of brain development (Dorus, Vallender et al. 2004), neurotransmission (Burki and Kaessmann 2004) or lifespan (Grossman, Wildman et al. 2004), suggesting strong Darwinian positive selection. However, the notion of positive selection implies a coding potential, as it is well described in (Zhang, Rosenberg et al. 1998) with the use of Ka/Ks ratio. In our case, as no coding potential was ascertained, we did not calculate Ka/Ks ratios, but chose to study the sequence variability among primate species and along the sequences. We did not compare distance matrices for the domains highlighted with the entropy tool, due to the absence of reference distance values for primate species and to the shortness of the sequence domains when excluding indels and repeated regions. Therefore, the functional relevance of these evolutionary changes remains unknown but it is tempting to speculate that they might play significant roles in the primate-lineage specific variability in PMCHL1 transcript production, processing and/or stability.

The *PMCHL1/L2* system, which combines the retroposition/exon shuffling and the segmental duplication models, has been one of the first hominoid-specific gene creation model described (Viale, Ortola et al. 1998; Viale, Ortola et al. 1998; Viale, Courseaux et al. 2000; Courseaux and Nahon 2001; Courseaux, Richard et al. 2003; Long, Deutsch et al. 2003; Nahon 2003). A question concerning the creation of novel genes, particularly in the case of retroposition/exon shuffling and segmental duplication, is relative to the acceptor sites. Which kind of sequences favours the original retrotransposition? Which kind of sequences allows invasion by foreign sequences? Which kind of sequences can we find at the segmental duplication boundaries? No clear boundaries specificities are described for segmental duplication except for a significant enrichment in young Alu Y and Alu S sequences and other repeats similar to these involved in Ig heavy chain recombination in pericentromeric and interstitial Segmental Duplications (SDs) (Samonte and Eichler 2002; Frazer, Chen et al. 2003; Bailey

and Eichler 2006). Alu mediated DNA duplications have exceptionally been reported in eukaryotes (Jurka, Kohany et al. 2004). These duplications appeared however to affect mainly hyper recombinogenic chromosomal regions, and particularly for secondary duplications (Samonte and Eichler 2002). Line 1 elements were also directly (i.e. not only favouring Alu sequences duplication) implicated in exon recombination and have been proposed to mediate exon shuffling (Long 2001), but none of the previously described human chimeric genes (Eichler, Johnson et al. 2001) harbour this kind of element at its boundaries. In the case of *PMCHL1/L2*, no particular sequence could be found at the boundaries, neither at the first insertion site of the MCH antisense retrotransposon (in an intron of the *Brain Cadherin* gene at the 5p14 locus), nor at the 5q13 locus when creating *PMCHL2*. Many repeated sequences, and in particular Alu sequences, are however present inside the genes. These primate specific transposable elements are also known to play important role in genome evolution by affecting the gene structure and expression (Gagneux and Varki 2001; Hacia 2001; Frazer, Chen et al. 2003). In this context, we observed an extraordinary instability of the *PMCHL* genes when we tried to propagate the whole (exons 1 to 5) or part (5'UTR to exon 2) of the gene in bacteria. Interestingly, a truncated Line 2 sequence was identified at the boundaries of the rearrangement originally observed in the phMCH-L37 clone (Viale, Courseaux et al. 2000), suggesting that repetitive sequences could promote recombination in bacteria. However, this needs to be directly addressed. When considering *PMCHL1/PMCHL2* gene rearrangements at the chromosome level, it is worth noting that *PMCHL2* stands very close to the SMA locus in 5q13.3 that contained a very complex arrangement of duplicated modules (Schmutz, Martin et al. 2004). Differences in size, content and organization of these duplicated modules were found in human populations and generated distinct haplotypes (Nahon 2003; Schmutz, Martin et al. 2004). The various genomic deletions associated with the SMA syndrome likely resulted from extensive variation in the segmental duplications (Courseaux, Richard et al. 2003). Interestingly, the SMA region was highly reluctant to standard

cloning and sequencing strategies, many DNA segments being unclonable in current bacterial vectors (Schmutz, Martin et al. 2004). A situation highly reminiscent of the difficulties we had to face with the cloning of *PMCHL1/PMCHL2* genes.

An important aspect of the work presented here is the demonstration that PMCHL1 transcripts can be identified in the brain of a non-human primate, e.g. *Macaca fascicularis*. While the relative expression pattern of exon 1-2 and exon 4-5 containing transcripts appeared quite similar in human and macaque brain, detailed characterization of PMCHL1 mRNAs in prefrontal cortex (PFC) revealed both quantitative and qualitative differences. Indeed, using identical RT-PCR conditions *PMCHL1* gene expression levels were always found higher in macaque than in human brain, as indicated by the detection on ethidium bromide-stained gels of PCR products derived from macaque but not human brain (see Fig.5). Interestingly, the apparent down-regulation of *PMCHL1* gene expression in human brain concerned the two classes of unspliced transcripts covering exons 1-2 and exons 4-5, respectively. In contrast, spliced PMCHL1 mRNAs covering exons 1-5 were identified in human but not macaque brain, albeit they represent rare transcripts compared with unspliced mRNAs.

For mRNAS containing exons 4-5, our data fitted with the sequence analysis of *PMCHL1* gene that predicted the lack of functional splice sites in intron D in the cercopithecoid lineage and their creation in the hominoid lineage (Courseaux and Nahon 2001). In addition, we failed to identify unspliced transcripts linking exons 1-2 to exons 4-5 in the macaque or human brain (our data) or in cDNAs or EST libraries. Marked difference in expression of exons 1-2 and exons 4-5 containing mRNAs was also noted in human and macaque cerebellum. This suggested therefore that the synthesis of unspliced exons 4-5 containing transcripts depend on a promoter distinct than the one used to produce exons 1-2 containing mRNAs. Further RACE-PCR experiments should clarify this issue.

For exons 1-2 containing mRNAs, distinct cap sites were found in macaque and human when analysing the transcriptional capacity of the region covering 7 kb upstream from the site of insertion of the MCH antisense retrotransposon. In human, a

putative *PMCHL1* cap site was mapped close to the insertion site, in agreement with previous RACE-PCR experiments (Viale, Ortola et al. 1998; Courseaux and Nahon 2001). In macaque, PMCHL1 mRNA containing exons 1-2 extended clearly up to 2.5 kb upstream from the cap site found in human, some transcripts encompassing a region located 7 kb upstream from the insertion site. The comparative sequence analysis of the putative TATA box associated with the human *PMCHL1* cap site revealed a mutation that predicted binding of polymerase II complex at this site only in *Pan* and *Homo* (Fig.6). This putative cap site would have evolved from a non-functional sequence in a common ancestor to Hominids.

Collectively, our gene expression results are therefore consistent with the hypothesis that the promoter regions of the *PMCHL1* transcription unit diverge in the cercopithecoid and hominid lineages resulting in species-specific differential expression levels but with similar regional expression pattern of major unspliced transcripts in macaque and human brain. Relative evolutionary conservation in the expression pattern of exons 1-2 and exons 4-5 containing mRNAs is of prime importance when considering the macaque as a Non-human Primate (NHP) model to investigate the regulation of *PMCHL1* gene expression. This gene expression profile is compatible with the nearly neutral theory for transcriptome evolution (Khaitovich, Enard et al. 2006) that predicted that functional constraint resulting from multiple gene network interactions is the primary force that shape gene expression changes in the primate brain. Furthermore, it has been noted recently that the rate of promoter evolution was significantly higher in the primate lineage compared to other mammals (Taylor, Kai et al. 2006). The differences in the *PMCHL1* gene promoter use we observed in macaque and human cortex could reflect this general trend. Indeed, primate-specific divergent expression patterns associated with sequence variations in the regulatory regions have been reported for the foetal γ globin gene in *Galago* and human (TomHon, Zhu et al. 1997), the class I γADH gene in baboon and human (Cheung, Holmes et al. 1999), the *apo(a)* gene in chimpanzee and human (Huby, Dachet et al.

2001) and the α•G gene in Platyrrhines and Catarrhines (Maston and Ruvolo 2002). A Cytosine to Thymidine substitution at the first base of a putative TATA box sequence that occurred selectively in the hominoid lineage could be a potential determinant of the switch observed in the human *PMCHL1* promoter use by comparison with the macaque. However, experiments involving expression in cell lines and site-directed mutagenesis should be carried out to test this hypothesis.

A puzzling question concerns the function of PMCHL1 transcripts in the primate brain. In particular, what might be the role(s) of the major unspliced PMCHL1 mRNAs in the cortex of human and macaque. Previous in vitro translation and transfected cell studies have suggested that the exon 1-2 containing transcript would encode a putative nuclear protein of 72 amino acids, named VMCH-p8, albeit this protein appeared weakly expressed and under restricted conditions (Viale, Courseaux et al. 2000). Antibodies raised against the nuclear localization sequence (NLS) and the NEI domain of VMCH-p8 failed to detect the expected polypeptide in protein extracts from developing human brain or adult macaque cortex (unpublished data). Because of the poor protein coding potential in *in vitro* systems and absence of corresponding protein in primate brain it is tempting to propose that these PMCHL1 transcripts work mainly as a non-protein coding RNA (ncRNA). Since the realization that 98 % of the transcriptional output in mammals consisted of ncRNAs, the importance of this new class of regulatory RNAs has grown tremendously (Mattick and Gagen 2001; Storz 2002). Small ncRNAs (including snoRNAs, miRNAs, siRNAs) played extensive roles in developmental and physiological pathways in multicellular organisms (Mattick and Makunin 2005). Incidentally, it appeared in our tests that *PMCHL1* is neither a support nor a target for siRNAs. Other ncRNAs of large size (ranging from 1 to more than 100kb) are associated with chromosome inactivation such as Xist and antisense Tsix transcripts involved in X inactivation in female mammals (Panning, Dausman et al. 1997; Johnston, Nesterova et al. 1998), mammalian development at imprinted loci such as the H19

and Igf2 RNAs (Arney 2003) or the Air RNA (Sleutels, Zwart et al. 2002). The *Air* gene is particularly interesting in providing some clues about how PMCHL1 transcript could act as a ncRNA. Indeed, Air RNA is a very long, unspliced, repeat-rich transcript (all features shared with the PMCHL1 transcripts) that appeared responsible for imprinted repression of nearby genes (including *Igf2r* gene) through an antisense-mediated mechanism (Sleutels, Zwart et al. 2002). Sense/antisense PMCHL1 transcripts were conspicuously found in developing human brain (Viale, Courseaux et al. 2000; Courseaux and Nahon 2001) and several adult macaque brain areas (our unpublished data). Is the sense/antisense transcripts ratio in any way crucial for *PMCHL1* function? Are they controlling the expression of neighbouring genes in cis (an obvious candidate is the *brain cadherin* gene) or multiple genes in trans through RNA-RNA duplexes. We are now addressing these intriguing questions.

In conclusion, we provided here compelling evidence that *PMCHL1* gene was created soon after divergence of Platyrrhini and Catharrini (i.e. 30-35 Mya) and that it acquired very rapidly a transcriptional activity in the brain of Catarrhine ancestors that can be recognized today in the overlapping expression pattern of unspliced transcripts in the macaque and human cortex and cerebellum. It is tempting to speculate that MCH antisense retrotransposon fortuitously has been inserted near brain-specific promoter elements, a favourable genomic environment that allowed selective pressure to maintain expression in specific brain areas. Incidentally, the macaque therefore appears to constitute a useful and accessible model in which to analyse these regulatory elements as well as the regulation of the major PMCHL1 transcripts shared among Catharrines. Thereafter, the molecular evolutionary history of *PMCHL1* became complex with mutations selectively in the hominidae lineage that contribute to the emergence of new exons (Viale, Courseaux et al. 2000; Courseaux and Nahon 2001) and a new putative transcriptional starting site (see Fig.6) that allowed production of both spliced and unspliced PMCHL1 transcripts. This increase in complexity was accompanied by an overall decrease in PMCHL1 mRNA synthesis, a down-regulation

that resulted likely from a switch in promoter use but modifications in mRNA stability should also be considered. Overall, the *PMCHL1* gene represents an obvious candidate gene to look for genetic differences that might be associated with the emergence of new brain structures and functions in the Catarrhine lineage. A puzzling question that we should address now is the relative contribution of the *PMCHL1* gene in relation to the multiple genic changes (reviewed in (Varki 2004; Sikela 2006)) that operated in the primate lineage leading to brain reorganization and cognitive implementation in human.

ACKNOWLEDGMENTS

We thank very much Dr. .....for helpful advice and critical reading of the manuscript. We are grateful to C. Payré for excellent technical assistance. Brain samples from human fetus and new-borns were kindly provided by Dr D. Jordan (Faculté de médecine A. Carrel, Lyon, France). The adult human brain structures were generously offered by the National Neurological Research Specimen Bank (Los Angeles, CA) which is sponsored by NINDS/NIMH, National Multiple Sclerosis Society, Hereditary Disease Foundation, and Veterans Health Services and Research Administration, Department of Veterans Affairs. This work was strongly supported by the Centre National de la Recherche Scientifique (CNRS programme OHLL 2002-2004 ; crédits exceptionnels Biothèque Primate) and by a 6th FP EU STREPS/NEST (APES project n° 28594). Proposals (submitted by J.-L. N.) concerning this work were rejected in 2005 and 2006 by the Agence Nationale de la Recherche (Programme « Neurosciences,neurologie et psychiatrie"). F. D.-T. was a recipient of a fellowship from the French Education Minister (Allocation couplée MENRS/ENS). A.C. and A.D.–A. were supported by the CNRS (crédits exceptionnels Biothèque Primate) and A.C. is presently supported by the APES project (n° 28594).

LEGENDS

Figure 1 : Schematic map of the *PMCHL1/PMCHL2* loci and genes
A, the position and orientation of *PMCHL1* inside *CDH12*, *PMCHL2* and documented EST inside the duplicon as well as the length of each fragment are indicated. EST506, ST504 and EST502 stand for ENSESTT00000060506, ENSESTT00000060504and ENSESTT00000060502 respectively.
B, the *PMCHL1* gene and its immediate environment are particularly rich in repeated sequences, mainly on the antisense strand according to the standard orientation of *PMCHL1*. The phMCH-L37 clone is shown as an example of gene rearrangement after cloning/expansion in bacteria; the transposed sequences are indicated as well as repeated sequences.
C, Alignment of the PMCHL region encompassing exons 1-2' in different species with corresponding MCH gene domains. The *PMCHL1(L1)* and *PMCHL2(L2)* genes resulted from AROM messenger retrotransposition and thus derived from the MCH sequence (sense strand). A break in exon 1 (concomitant to PMCHL1 creation) and the insertion of the remaining part in the opposite orientation downstream from the region homologous to exon III of the MCH gene is represented at the bottom of the figure. Insertion of the Alu element is figured, as well as punctual indels (dark red for indels conserved in the human lineage, light red for indels found in Cercopithecoids, pink for indels appearing in Cebus *MCH* gene only).

Figure 2: Phylogenetic studies of the MCH and PMCHL genes
A, phylogenetic analysis of the aligneable MCH and PMCHL sequences, B,phylogenetic analysis of a longer PMCHL gene portion. Both, Unrooted trees obtained using a neighbor-joining (NJ) method. Branches also found by maximum likelihood (G option) and parsimony are indicated respectively with ** ($p<0.01$) and +. Percentages are given for a 1000 bootstrap replications with the NJ method.

Figure 3: Schematic representation of PMCHL spliced transcripts in human tissues:
Several mRNAs were identified through nested PCR followed by sequencing. The oligonucleotide couples used are F1/R1a or F1/R1b for the first PCR, and F2/R2 for the nested PCR (sequences available in the supplemental table). Tissue expression, exons involved in alternative splicing and characteristics of the PCR amplification are indicated, as well as putative ORFs.

Figure 4: Comparison of PMCHL expression in human and maccaca brain
A, positions of the oligonucleotides couples and probes used for the PCR amplification and southern blotting detection (sequences available in the supplemental table. Note that F1-2/F1-5 and R1-5/R4-5 correspond to F1 and R1a shown in figure 4 respectively).
B, RT-PCR analysis of RNA extracted from human or macaque cortex and cerebellum. Ethidium Bromide (EB)-stained gels and southern blotting are shown for each PCR amplification that joined exons 1-2, 4-5 or 1-5. The probes used for the southern blotting are indicated above. The probes used for the southern blotting are indicated above. RT :Reverse-Transcribed ; NRT : Non Reverse-Transcribed. The size of PCR products are shown when identified on EB-stained gels. NA stands for Not Available, CX and Cb for Cortex and Cerebellum.

Figure 5: Differential expression pattern of *PMCHL* 5' region in human and macacca cortexes
A, Organization of cDNA obtained with hMCH7 and P9AS specific RT-PCR. Locations of primers and length of expected PCR_amplified products are represented by black double arrows, probes to detect PCR products are represented by color boxes and color thick lines. The white arrowhead represents the human putative capsite and the black arrowheads represent the location of primers used for specific RT-PCR. Putative TATA box sequences in different primate species are shown and sequence identity is boxed.

B, Specific RT-PCR analysis of RNA extracted from human and maccaca cortex. Southern blotting experiments are shown for several 5’ PCR amplifications on a length of 7000 pbs upstream exon 1 of PMCHL. Primers used for specific RT-PCR and PCR amplifications are indicated above and probes used for southern blotting are indicated below. RT stands for reverse-transcribed, and NRT for non reverse-transcribed.

Supplementary data, Figure 1: structural constraints on DNA micro domains in the *PMCHL1* gene

A, B, Evolution of the variability between primates sequences along the *MCH* (A) and *PMCHL1* (B) genes, respectively (to scale). The variability is expressed in arbitrary unit, according to the nucleotidic variability Program. Indels are indicated, they corresponds to indels between human and the most distant species (i.e. Cercopithecoids for the 5' part of the *PMCHL* gene, and Cebus capucinus for the 3' part of the same gene).

REFERENCES

Arney, K. L. (2003). "H19 and Igf2--enhancing the confusion?" Trends Genet **19**(1): 17-23.

Bailey, J. A. and E. E. Eichler (2006). "Primate segmental duplications: crucibles of evolution, diversity and disease." Nat Rev Genet **7**(7): 552-64.

Bordner, A. J. and R. Abagyan (2005). "REVCOM: a robust Bayesian method for evolutionary rate estimation." Bioinformatics **21**(10): 2315-21.

Britten, R. J. (2002). "Divergence between samples of chimpanzee and human DNA sequences is 5%, counting indels." Proc Natl Acad Sci U S A **99**(21): 13633-5.

Britten, R. J., L. Rowen, et al. (2003). "Majority of divergence between closely related DNA samples is due to indels." Proc Natl Acad Sci U S A **100**(8): 4661-5.

Burki, F. and H. Kaessmann (2004). "Birth and adaptive evolution of a hominoid gene that supports high neurotransmitter flux." Nat Genet **36**(10): 1061-3.

Caceres, M., J. Lachuer, et al. (2003). "Elevated gene expression levels distinguish human from non-human primate brains." Proc Natl Acad Sci U S A **100**(22): 13030-5.

Cardinaud, B., F. Darre-Toulemonde, et al. (2004). "Comparative analysis of melanin-concentrating hormone structure and activity in fishes and mammals." Peptides **25**(10): 1623-32.

Cheung, B., R. S. Holmes, et al. (1999). "Evolution of class I alcohol dehydrogenase genes in catarrhine primates: gene conversion, substitution rates, and gene regulation." Mol Biol Evol **16**(1): 23-36.

Clark, A. G., S. Glanowski, et al. (2003). "Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios." Science **302**(5652): 1960-3.

Courseaux, A. and J. L. Nahon (2001). "Birth of two chimeric genes in the Hominidae lineage." Science **291**(5507): 1293-7.

Courseaux, A., F. Richard, et al. (2003). "Segmental duplications in euchromatic regions of human chromosome 5: a source of evolutionary instability and transcriptional innovation." Genome Res **13**(3): 369-81.

Dorus, S., E. J. Vallender, et al. (2004). "Accelerated evolution of nervous system genes in the origin of Homo sapiens." Cell **119**(7): 1027-40.

Eichler, E. E., M. E. Johnson, et al. (2001). "Divergent origins and concerted expansion of two segmental duplications on chromosome 16." J Hered **92**(6): 462-8.

Enard, W., P. Khaitovich, et al. (2002). "Intra- and interspecific variation in primate gene expression patterns." Science **296**(5566): 340-3.

Enard, W. and S. Paabo (2004). "Comparative primate genomics." Annu Rev Genomics Hum Genet **5**: 351-78.

Evans, P. D., J. R. Anderson, et al. (2004). "Adaptive evolution of ASPM, a major determinant of cerebral cortical size in humans." Hum Mol Genet **13**(5): 489-94.
Frazer, K. A., X. Chen, et al. (2003). "Genomic DNA insertions and deletions occur frequently between humans and nonhuman primates." Genome Res **13**(3): 341-6.
Gagneux, P. and A. Varki (2001). "Genetic differences between humans and great apes." Mol Phylogenet Evol **18**(1): 2-13.
Galtier, N., M. Gouy, et al. (1996). "SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny." Comput Appl Biosci **12**(6): 543-8.
Gascuel, O. (1997). "BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data." Mol Biol Evol **14**(7): 685-95.
Glazko, G. V. and M. Nei (2003). "Estimation of divergence times for major lineages of primate species." Mol Biol Evol **20**(3): 424-34.
Goodman, M. (1999). "The genomic record of Humankind's evolutionary roots." Am J Hum Genet **64**(1): 31-9.
Grossman, L. I., D. E. Wildman, et al. (2004). "Accelerated evolution of the electron transport chain in anthropoid primates." Trends Genet **20**(11): 578-85.
Hacia, J. G. (2001). "Genome of the apes." Trends Genet **17**(11): 637-45.
Hsieh, W. P., T. M. Chu, et al. (2003). "Mixed-model reanalysis of primate data suggests tissue and species biases in oligonucleotide-based gene expression profiles." Genetics **165**(2): 747-57.
Huby, T., C. Dachet, et al. (2001). "Functional analysis of the chimpanzee and human apo(a) promoter sequences: identification of sequence variations responsible for elevated transcriptional activity in chimpanzee." J Biol Chem **276**(25): 22209-14.
Johnston, C. M., T. B. Nesterova, et al. (1998). "Developmentally regulated Xist promoter switch mediates initiation of X inactivation." Cell **94**(6): 809-17.
Jurka, J., O. Kohany, et al. (2004). "Duplication, coclustering, and selection of human Alu retrotransposons." Proc Natl Acad Sci U S A **101**(5): 1268-72.
Khaitovich, P., W. Enard, et al. (2006). "Evolution of primate gene expression." Nat Rev Genet **7**(9): 693-702.
Khaitovich, P., B. Muetzel, et al. (2004). "Regional patterns of gene expression in human and chimpanzee brains." Genome Res **14**(8): 1462-73.
King, M. C. and A. C. Wilson (1975). "Evolution at two levels in humans and chimpanzees." Science **188**(4184): 107-16.
Long, M. (2001). "Evolution of novel genes." Curr Opin Genet Dev **11**(6): 673-80.
Long, M., E. Betran, et al. (2003). "The origin of new genes: glimpses from the young and old." Nat Rev Genet **4**(11): 865-75.
Long, M., M. Deutsch, et al. (2003). "Origin of new genes: evidence from experimental and computational analyses." Genetica **118**(2-3): 171-82.
Maston, G. A. and M. Ruvolo (2002). "Chorionic gonadotropin has a recent origin within primates and an evolutionary history of selection." Mol Biol Evol **19**(3): 320-35.

Mattick, J. S. and M. J. Gagen (2001). "The evolution of controlled multitasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms." Mol Biol Evol **18**(9): 1611-30.
Mattick, J. S. and I. V. Makunin (2005). "Small regulatory RNAs in mammals." Hum Mol Genet **14 Spec No 1**: R121-32.
Mekel-Bobrov, N., S. L. Gilbert, et al. (2005). "Ongoing adaptive evolution of ASPM, a brain size determinant in Homo sapiens." Science **309**(5741): 1720-2.
Mignone, F., D. S. Horner, et al. (2004). "WebVar: A resource for the rapid estimation of relative site variability from multiple sequence alignments." Bioinformatics **20**(8): 1331-3.
Nahon, J. L. (2003). "Birth of 'human-specific' genes during primate evolution." Genetica **118**(2-3): 193-208.
Panning, B., J. Dausman, et al. (1997). "X chromosome inactivation is mediated by Xist RNA stabilization." Cell **90**(5): 907-16.
Patterson, N., D. J. Richter, et al. (2006). "Genetic evidence for complex speciation of humans and chimpanzees." Nature **441**(7097): 1103-8.
Rosenberg, M. S. and S. Kumar (2003). "Heterogeneity of nucleotide frequencies among evolutionary lineages and phylogenetic inference." Mol Biol Evol **20**(4): 610-21.
Samonte, R. V. and E. E. Eichler (2002). "Segmental duplications and the evolution of the primate genome." Nat Rev Genet **3**(1): 65-72.
Schmutz, J., J. Martin, et al. (2004). "The DNA sequence and comparative analysis of human chromosome 5." Nature **431**(7006): 268-74.
Sikela, J. M. (2006). "The jewels of our genome: the search for the genomic changes underlying the evolutionarily unique capacities of the human brain." PLoS Genet **2**(5): e80.
Sleutels, F., R. Zwart, et al. (2002). "The non-coding Air RNA is required for silencing autosomal imprinted genes." Nature **415**(6873): 810-3.
Storz, G. (2002). "An expanding universe of noncoding RNAs." Science **296**(5571): 1260-3.
Taylor, M. S., C. Kai, et al. (2006). "Heterotachy in mammalian promoter evolution." PLoS Genet **2**(4): e30.
TomHon, C., W. Zhu, et al. (1997). "Evolution of a fetal expression pattern via cis changes near the gamma globin gene." J Biol Chem **272**(22): 14062-6.
Uddin, M., D. E. Wildman, et al. (2004). "Sister grouping of chimpanzees and humans as revealed by genome-wide phylogenetic analysis of brain gene expression profiles." Proc Natl Acad Sci U S A **101**(9): 2957-62.
Varki, A. (2004). "How to make an ape brain." Nat Genet **36**(10): 1034-6.
Viale, A., A. Courseaux, et al. (2000). "Structure and expression of the variant melanin-concentrating hormone genes: only PMCHL1 is transcribed in the developing human brain and encodes a putative protein." Mol Biol Evol **17**(11): 1626-40.
Viale, A., C. Ortola, et al. (1998). "Emergence of a brain-expressed variant melanin-concentrating hormone gene

during higher primate evolution: a gene "in search of a function"." Mol Biol Evol **15**(2): 196-214.
Viale, A., C. Ortola, et al. (1998). "Structure, expression, and evolution of the variant MCH gene in primates." Ann N Y Acad Sci **839**: 214-8.
Wall, J. D. (2003). "Estimating ancestral population sizes and divergence times." Genetics **163**(1): 395-404.
Wang, Y. Q. and B. Su (2004). "Molecular evolution of microcephalin, a gene determining human brain size." Hum Mol Genet **13**(11): 1131-7.
Wildman, D. E., M. Uddin, et al. (2003). "Implications of natural selection in shaping 99.4% nonsynonymous DNA identity between humans and chimpanzees: enlarging genus Homo." Proc Natl Acad Sci U S A **100**(12): 7181-8.
Yang, Z. (1997). "PAML: a program package for phylogenetic analysis by maximum likelihood." Comput Appl Biosci **13**(5): 555-6.
Zhang, J., H. F. Rosenberg, et al. (1998). "Positive Darwinian selection after gene duplication in primate ribonuclease genes." Proc Natl Acad Sci U S A **95**(7): 3708-13.

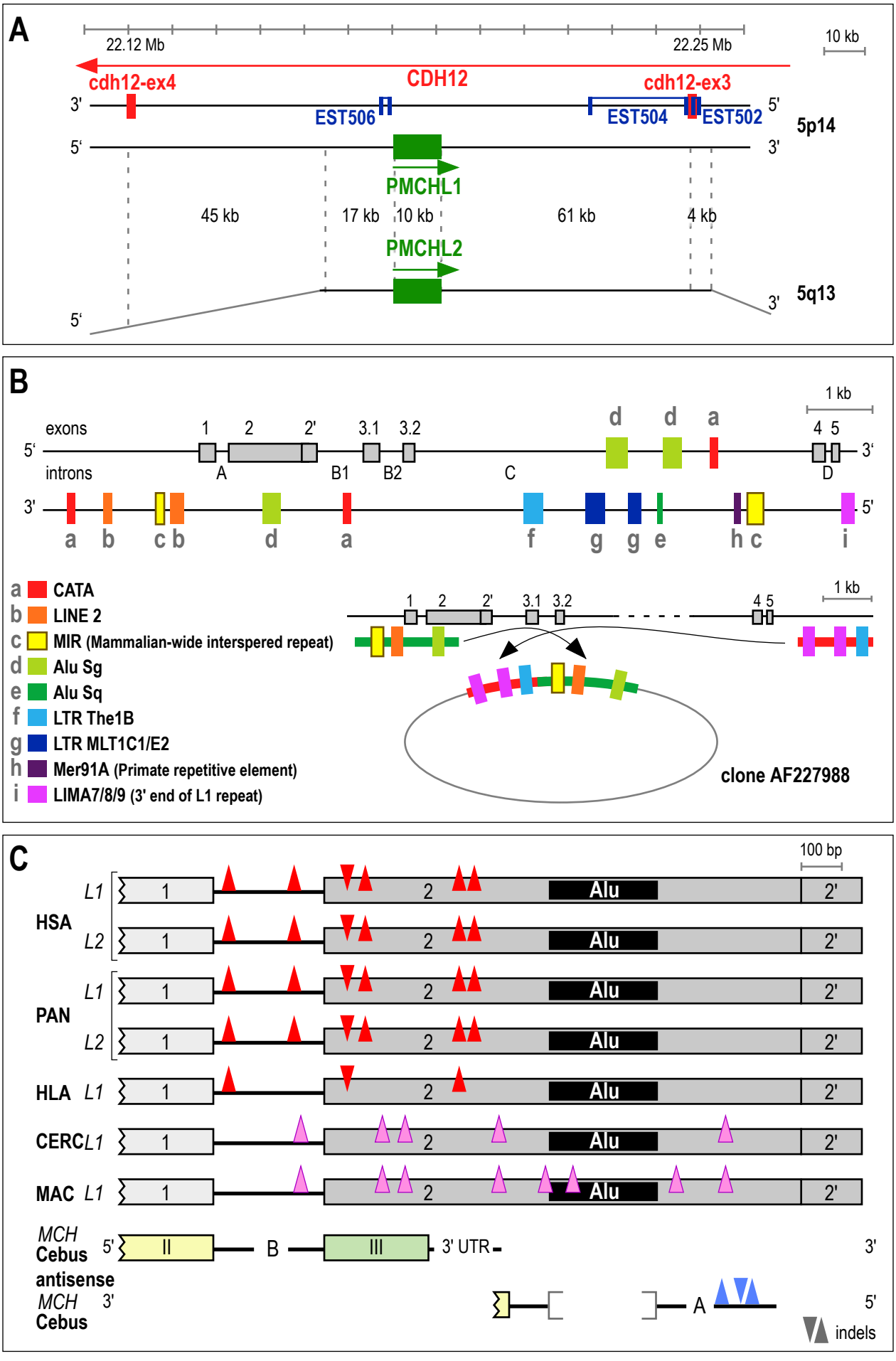

A
22.12 Mb
22.25 Mb
10 kb
cdh12-ex4
CDH12
cdh12-ex3
3'
EST506
EST504
EST502
5'
5p14
5'
3'
PMCHL1
45 kb
17 kb
10 kb
61 kb
4 kb
PMCHL2
5q13
B
1 kb
exons
introns
1
2
2'
3.1
3.2
d
d
a
4 5
A
B1
B2
C
D
a b c b d a f g g e h c i
a CATA
b LINE 2
c MIR (Mammalian-wide interspered repeat)
d Alu Sg
e Alu Sq
f LTR The1B
g LTR MLT1C1/E2
h Mer91A (Primate repetitive element)
i LIMA7/8/9 (3' end of L1 repeat)
clone AF227988
C
100 bp
L1
L2
HSA
PAN
HLA L1
CERC L1
MAC L1
1
2
Alu
2'
MCH Cebus
II
B
III
3' UTR
antisense MCH Cebus
A
indels

Fig. 1

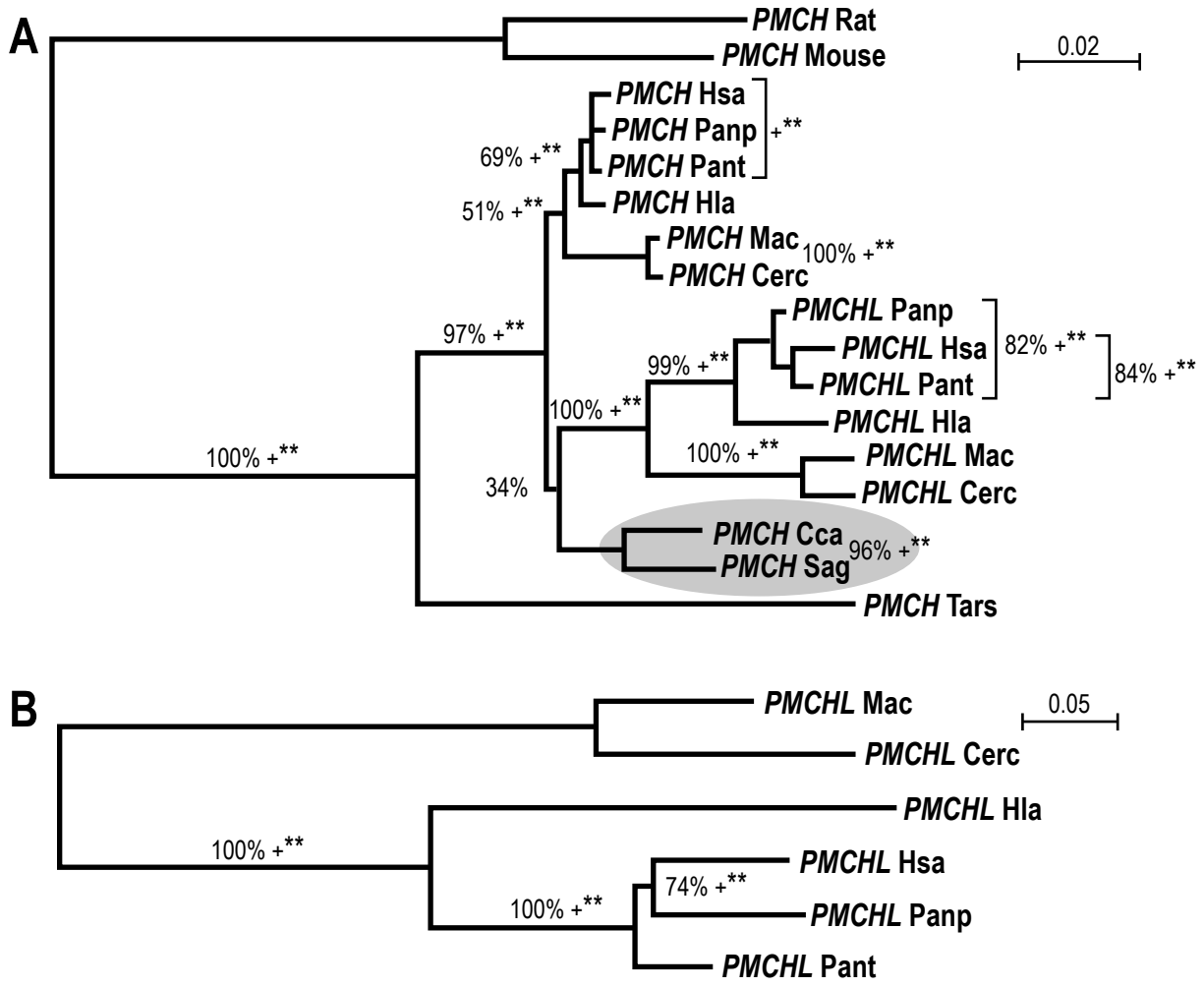

A
PMCH Rat
PMCH Mouse
0.02
PMCH Hsa
PMCH Panp
PMCH Pant
+**
69% +**
PMCH Hla
51% +**
PMCH Mac
PMCH Cerc
100% +**
PMCHL Panp
PMCHL Hsa
PMCHL Pant
82% +**
84% +**
97% +**
99% +**
100% +**
PMCHL Hla
100% +**
PMCHL Mac
PMCHL Cerc
34%
PMCH Cca
PMCH Sag
96% +**
PMCH Tars
B
PMCHL Mac
0.05
PMCHL Cerc
PMCHL Hla
100% +**
PMCHL Hsa
74% +**
100% +**
PMCHL Panp
PMCHL Pant

Fig. 2

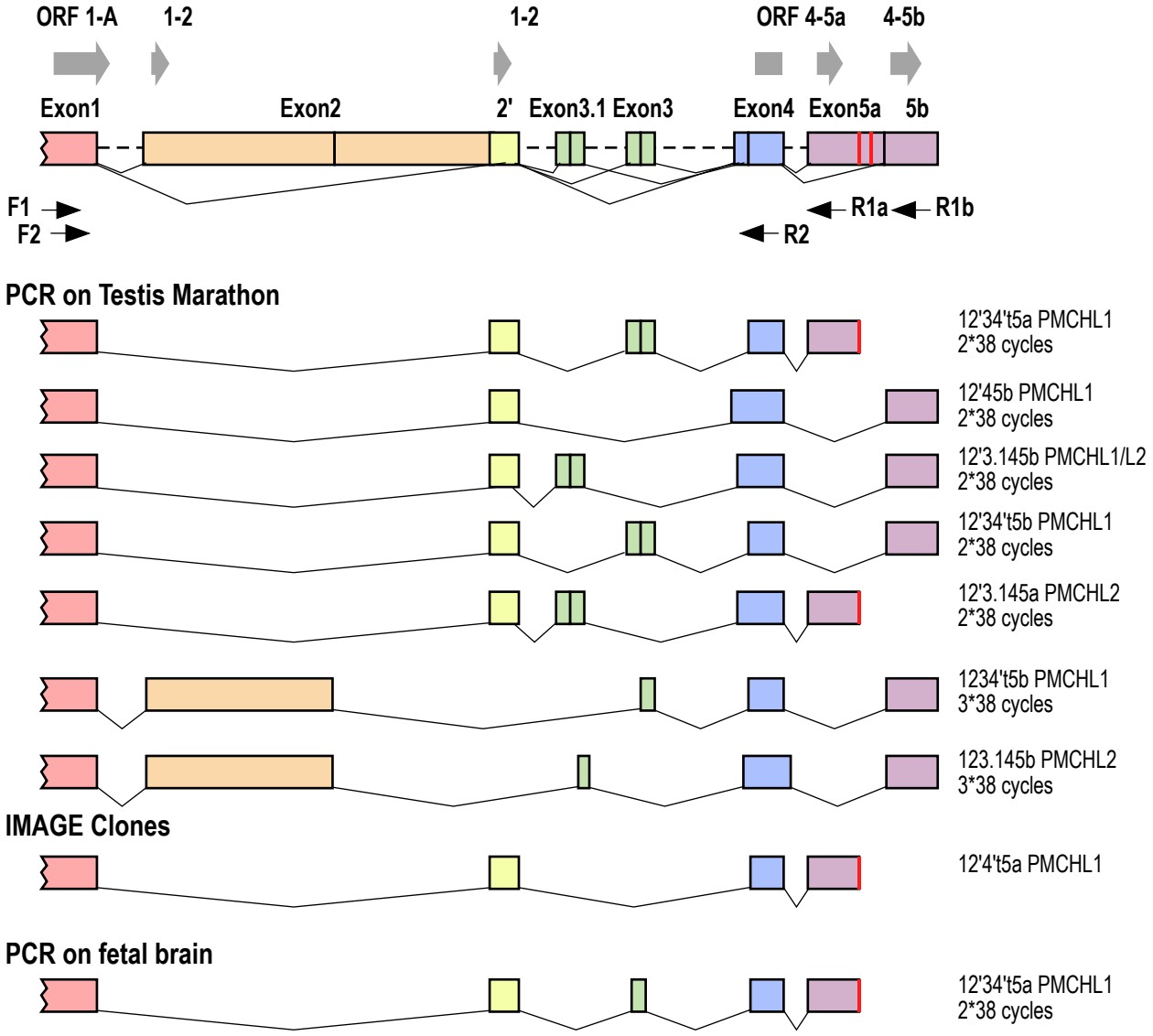

ORF 1-A
1-2
1-2
ORF 4-5a
4-5b
Exon1
Exon2
2'
Exon3.1
Exon3
Exon4
Exon5a
5b
F1
F2
R1a
R1b
R2
PCR on Testis Marathon
12'34't5a PMCHL1
2*38 cycles
12'45b PMCHL1
2*38 cycles
12'3.145b PMCHL1/L2
2*38 cycles
12'34't5b PMCHL1
2*38 cycles
12'3.145a PMCHL2
2*38 cycles
1234't5b PMCHL1
3*38 cycles
123.145b PMCHL2
3*38 cycles
IMAGE Clones
12'4't5a PMCHL1
PCR on fetal brain
12'34't5a PMCHL1
2*38 cycles

Fig. 3

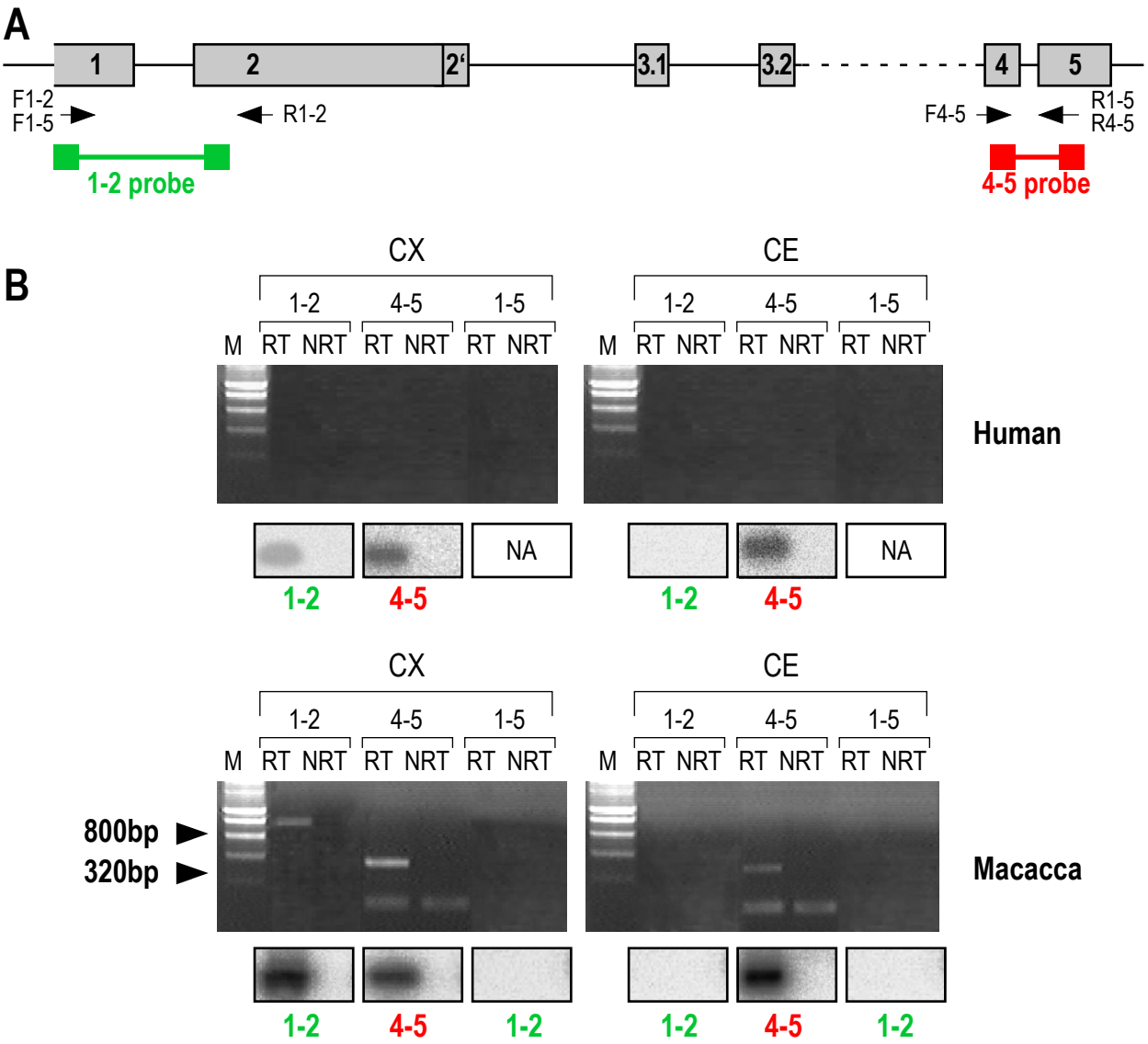

A
1
2
2'
3.1
3.2
4
5
F1-2
F1-5
R1-2
F4-5
R1-5
R4-5
1-2 probe
4-5 probe
B
CX
CE
1-2
4-5
1-5
M
RT NRT RT NRT RT NRT
Human
NA
1-2
4-5
800bp
320bp
Macacca

Fig. 4

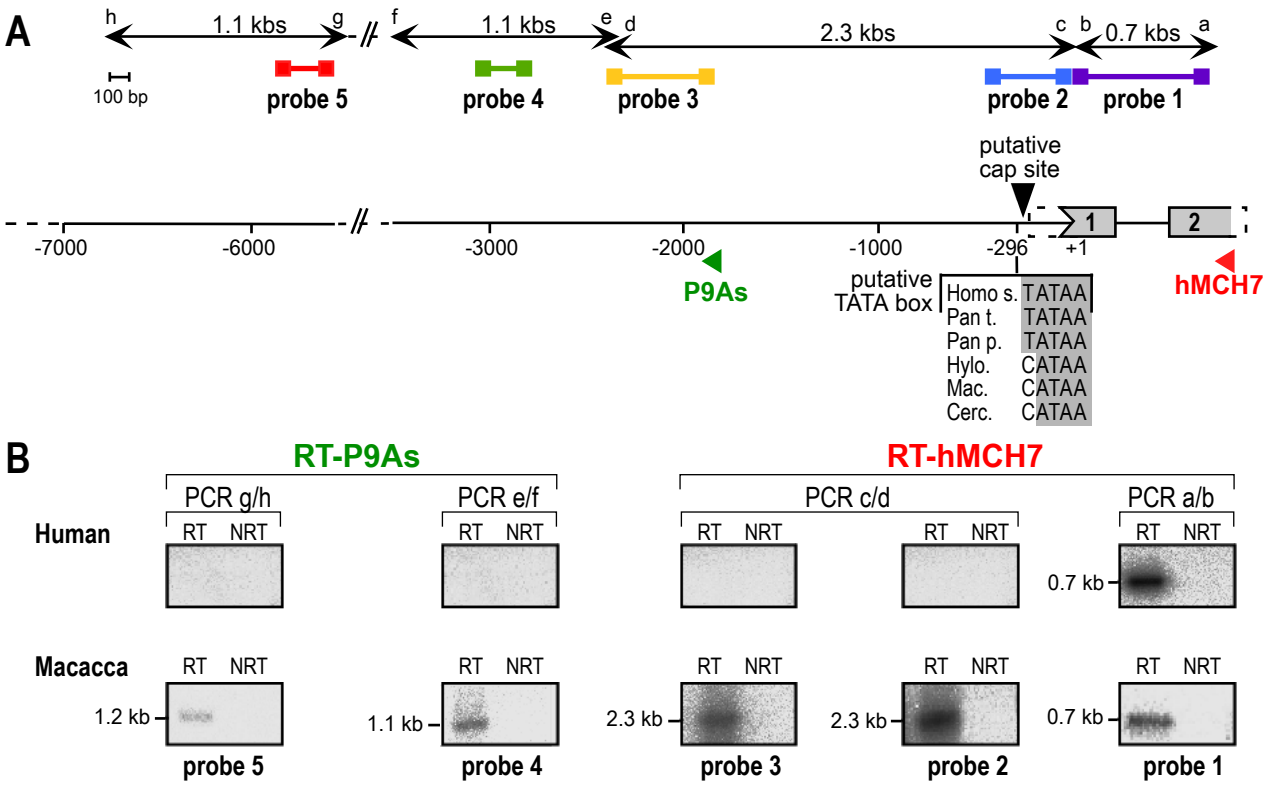

A
h
1.1 kbs
g
f
1.1 kbs
e
d
2.3 kbs
c
b
0.7 kbs
a
100 bp
probe 5
probe 4
probe 3
probe 2
probe 1
putative
cap site
-7000
-6000
-3000
-2000
-1000
-296
+1
1
2
P9As
hMCH7
putative
TATA box
Homo s. TATAA
Pan t. TATAA
Pan p. TATAA
Hylo. CATAA
Mac. CATAA
Cerc. CATAA
B
RT-P9As
RT-hMCH7
PCR g/h
PCR e/f
PCR c/d
PCR a/b
RT NRT
Human
Macacca
0.7 kb
1.2 kb
1.1 kb
2.3 kb
2.3 kb
0.7 kb
probe 5
probe 4
probe 3
probe 2
probe 1

Fig. 5