



**HAL**  
open science

# ETUDE ANTHROPOGENETIQUE DE LA POPULATION COMORIENNE DE MARSEILLE

Jacques Chiaroni

► **To cite this version:**

Jacques Chiaroni. ETUDE ANTHROPOGENETIQUE DE LA POPULATION COMORIENNE DE MARSEILLE. Anthropologie biologique. Université de la Méditerranée - Aix-Marseille II, 2003. Français. NNT: . tel-00011766

**HAL Id: tel-00011766**

**<https://theses.hal.science/tel-00011766>**

Submitted on 8 Mar 2006

**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 la Méditerranée – Aix-Marseille II  
*Faculté de Médecine de Marseille*

**THESE**

de

**Doctorat de l'Université de La Méditerranée**  
Spécialité : Anthropologie biologique

**ETUDE ANTHROPOGENETIQUE DE LA POPULATION  
COMORIENNE DE MARSEILLE**

*Soutenue publiquement le 31.10.03 par :*

*Jacques CHIARONI*

Né le 1<sup>er</sup> MAI 1959 à Marseille

*Composition du Jury*

**Gilles BOËTSCH**, Directeur de Recherche au CNRS, Marseille, *Directeur de thèse.*

**Philippe de MICCO**, Professeur des Universités, Université de la Méditerranée.

**Jean-Michel DUGOUJON**, Chargé de Recherches au CNRS, *Rapporteur.*

**Olivier DUTOUR**, Professeur des Universités, Université de la Méditerranée.

**Emma RABINO-MASSA**, Professeur à l'Université de Turin, *Rapporteur.*

Laboratoire d'accueil : UMR 6578 CNRS-Université de la Méditerranée  
Faculté de Médecine de Marseille

Ecole Doctorale : Sciences de l'Environnement

# SOMMAIRE

<b><u>INTRODUCTION</u></b> .....	<b>4</b>
<b><u>I. Problématique</u></b> .....	<b>5</b>
<b><u>II. Les paramètres linguistiques, historiques et culturels.</u></b> .....	<b>10</b>
<u>II.1 Les données linguistiques.</u> .....	10
<u>II.2 Les données historiques</u> .....	15
<u>II.2.1. Début du 1<sup>e</sup> millénaire : Etablissement des sociétés Swahilies sur la côte orientale de l’Afrique.</u> .....	18
<u>II.2.2. Entre les V<sup>e</sup> et VII<sup>e</sup> siècles : le premier peuplement des Comores et de Madagascar.</u> .....	21
<u>II.2.3. Au VIII<sup>e</sup> siècle : Mise en place d’une société plus structurée, dite “DEMBENT”.</u> .....	22
<u>II.2.4. Entre les IX<sup>e</sup> et X<sup>e</sup> siècles : Les contacts Austronésiens.</u> .....	22
<u>II.2.5. Entre les XI<sup>e</sup> et XV<sup>e</sup> siècles : Islamisation des Comores et l’influence de l’élite Schirazienne.</u> .....	25
<u>II.2.6. Au XVI<sup>e</sup> siècle : Mise en place d’une Aristocratie Schirazienne.</u> .....	26
<u>II.2.7. Entre les XVII<sup>e</sup> et XVIII<sup>e</sup> siècles : Guerres inter-sultanats et raids Malgaches.</u> .....	27
<u>II.2.8. En 1843 débute l’histoire Française aux Comores.</u> .....	27
<u>II.3 Les données culturelles.</u> .....	27
<u>II.4 L’histoire récente.</u> .....	30
<b><u>III. Les marqueurs génétiques utilisés.</u></b> .....	<b>33</b>
<u>III.1. L’apport de la génétique en Anthropologie.</u> .....	33
<u>III.2. Le polymorphisme génétique et la notion de groupes sanguins chez l’homme.</u> .....	38
<u>III.3. Bases de la sélection des marqueurs génétiques.</u> .....	45
<u>III.3.1. Généralités.</u> .....	45
<u>III.3.2. Les systèmes de groupes sanguins érythrocytaires.</u> .....	47
<u>III.3.3. Le polymorphisme du locus <i>HLA-DRBI</i>.</u> .....	49
<u>III.3.4. Le polymorphisme des gènes <i>KIR</i>.</u> .....	49
<u>III.3.5. Le polymorphisme du chromosome Y.</u> .....	50
<u>III.4. Description des marqueurs génétiques sélectionnés.</u> .....	50
<u>III.4.1. Les systèmes de groupes sanguins érythrocytaires.</u> .....	50
<u>III.4.1.1. Le système <i>ABO</i>.</u> .....	52
<u>III.4.1.2. Le système <i>Rh</i>.</u> .....	55
<u>III.4.1.3. Le système <i>Kidd</i>.</u> .....	73
<u>III.4.1.4. Le système <i>Duffy</i>.</u> .....	75
<u>III.4.1.5. Le système <i>Kell</i>.</u> .....	83
<u>III.4.1.6. Le système <i>MNS</i>.</u> .....	88
<u>III.4.2. Le polymorphisme <i>HLA-DRBI</i>.</u> .....	90
<u>III.4.3. Le polymorphisme des gènes <i>KIR</i>.</u> .....	93
<u>III.4.4. Le polymorphisme du chromosome Y.</u> .....	97

<b><u>MATERIEL et METHODES.</u></b> .....	<b>109</b>
<b><u>I.Obtention et caractéristiques des Echantillons.</u></b> .....	<b>110</b>
<b><u>II. Méthodes d’exploration et d’analyse du polymorphisme érythrocytaire.</u></b> .....	<b>115</b>
II.1 Détermination des groupes sanguins érythrocytaires.....	115
II.2 Détermination du génotypage Duffy.....	115
II.3 Recherche des allèles rares du gène RHCE par biologie moléculaire.....	116
II.4 Analyse des données.....	117
<b><u>III.Méthodes d’exploration et d’analyse du polymorphisme HLA DRBI</u></b> .....	<b>125</b>
III.1 Génotypage HLA-DRBI.....	125
III.2 Analyse des données.....	125
<b><u>IV.Méthodes d’exploration et d’analyse du polymorphisme KIR.</u></b> .....	<b>127</b>
IV.1 Population.....	127
IV.2 Génotypage KIR.....	127
IV.3 Les méthodes statistiques.....	127
IV.4 Les autres populations.....	128
<b><u>RESULTATS.</u></b> .....	<b>130</b>
<b><u>I. Résultats du polymorphisme érythrocytaire (gènes et antigènes courants)</u></b> .....	<b>131</b>
I.1 Les fréquences phénotypique et allélique.....	131
I.2 Les distances génétiques.....	136
I.3 L’analyse multidimensionnelle.....	141
I.4 La classification ascendante hiérarchique.....	143
I.5 L’admixture.....	145
<b><u>II. Résultats des fréquences phénotypiques des donneurs de sang de l’EFS.</u></b> .....	<b>146</b>
<b><u>III. Résultats de la détection des allèles « rares » du gène RHCE.</u></b> .....	<b>148</b>
<b><u>IV. Résultats du polymorphisme HLA-DRBI.</u></b> .....	<b>150</b>
IV.1 Les fréquences alléliques.....	150
IV.2 Les distances génétiques.....	153
IV.3 L’analyse multidimensionnelle.....	153
<b><u>V. Résultats du polymorphisme des gènes KIR.</u></b> .....	<b>156</b>
V.1 Les fréquences géniques.....	156
V.2 Les génotypes KIR.....	158
V.3 Comparaison du polymorphisme des gènes KIR dans les populations.....	160

<b><u>DISCUSSION.</u></b> .....	<b>163</b>
<b><u>I. Analyse de la contribution d’Afrique subsaharienne.</u></b> .....	<b>164</b>
I.1 Les données du polymorphisme érythrocytaire. ....	164
I.2 Les données du polymorphisme HLA-DRBI .....	167
I.3 Les données du polymorphisme des gènes KIR. ....	168
I.4 Conclusions relatives à la contribution Africaine. ....	168
<b><u>II. Analyse de la contribution « Arabe ».</u></b> .....	<b>169</b>
II.1 Les données du polymorphisme érythrocytaire. ....	169
II.2 Les données du polymorphisme HLA-DRBI .....	170
II.3 Les données du polymorphisme des gènes KIR. ....	171
II.4 Conclusions relatives à la contribution « Arabe ». ....	171
<b><u>III. Contribution de l’Asie du Sud (Inde et Pakistan).</u></b> .....	<b>172</b>
<b><u>IV. Contribution de l’Asie du Sud-Est.</u></b> .....	<b>173</b>
<b><u>V. Influence Européenne.</u></b> .....	<b>174</b>
<b><u>VI. Evaluation quantitative des diverses contributions.</u></b> .....	<b>174</b>
<b><u>VII. Les premiers résultats du polymorphisme du chromosome Y.</u></b> .....	<b>176</b>
<b><u>VIII. L’exploration du polymorphisme KIR en Anthropologie.</u></b> .....	<b>176</b>
<b><u>IX. Analyse de l’effet de la sélection naturelle.</u></b> .....	<b>176</b>
<b><u>X. Communauté Comorienne migrante et impacts transfusionnels.</u></b> .....	<b>177</b>
X.1 Les antigènes et les phénotypes courants. ....	177
X.2 Les allèles « rares » et les phénotypes « rares ». ....	178
<b><u>CONCLUSION.</u></b> .....	<b>181</b>
<b><u>PERSPECTIVES.</u></b> .....	<b>185</b>
<b><u>BIBLIOGRAPHIE.</u></b> .....	<b>188</b>
<b><u>ANNEXES</u></b> .....	<b>209</b>

# **INTRODUCTION**

## I. Problématique

Les Comores sont situées à l'entrée du canal du Mozambique, à 400 km au Nord-Ouest de l'île de Madagascar et à 300 km des côtes africaines (Carte 1 p.8). L'archipel comprend quatre îles : Grande-Comore (N'Gazidja), Mohéli (Mwali), Anjouan (Ndzuwani) et Mayotte (Mahoré). C'est sur l'île de la Grande-Comore qu'est implantée la capitale, Moroni. Alors que Mayotte est une Collectivité Territoriale Française d'Outre-Mer, les trois autres îles appartiennent à la République Islamique des Comores. La superficie totale du pays (Mayotte exclue) est de 1862 km<sup>2</sup>. Les Comores sont devenues membre de la Francophonie en 1977 et membre de la Ligue des États Arabes en 1993.

Depuis le V<sup>e</sup> siècle et probablement avant, ces îles ont été peuplées par des arrivées successives de populations de cultures différentes (Carte 2 p.9). De grands groupes "ethniques" et linguistiques (bantous, sémitiques, austronésiens, voir indo-européens) ont apporté leur contribution au peuplement de ce secteur. L'analyse des données de différentes disciplines fait considérer le canal du Mozambique comme la limite la plus Occidentale de la pénétration Austronésienne, la limite la plus Orientale de la diaspora Bantoue et la limite la plus Boréale de l'expansion de l'Islam dans l'hémisphère Sud. La population des Comores, apparaît « bantouisée » (Allibert C, 1993) par les langues, les traditions et pour une bonne partie par les hommes puis islamisée par les Arabes et enfin, elle semble encore montrer des traces du monde austronésien. Les preuves de ces différents contacts sont essentiellement archéologiques, historiques, linguistiques, socio-culturelles et religieuses.

L'analyse de la littérature ne faisant état d'aucune donnée anthropogénétique, il nous a semblé intéressant de rechercher si ces contacts avaient laissé des traces dans le profil génétique de cette population et dans quelles proportions.

Nous proposons donc, de décrire le profil génétique de la population comorienne à partir d'individus résidant à Marseille et d'essayer, en comparant ces données aux populations historiquement impliquées dans le peuplement de cette région, de détecter et éventuellement quantifier cette triple contribution (Bantoue, Arabe et Austronésienne).

La population des Comores compte un peu plus de 615 000 habitants (estimation 2002). Cette population, enracinée dans une tradition migratoire ancienne, voit près de 1/10 de ses habitants vivre à Marseille. Or, des migrations humaines naît une problématique transfusionnelle (Issitt PD, 1994). En effet, les populations migrantes peuvent être considérées comme des populations de receveurs à haut risque d'allo-immunisation, en raison d'une forte variation du polymorphisme érythrocytaire avec la population de donneurs de sang « habituels » du pays d'accueil. Cet aspect peut être considéré comme une véritable pathologie du migrant directement liée à ce que Jean Bernard et Jacques Ruffié appellent la « géographie des conflits immunologiques » (Bernard J, 1966).

Les sujets originaires d'Afrique subsaharienne présentent des combinatoires phénotypiques ou des phénotypes rares dans certains systèmes érythrocytaires (RH, FY et MNS). La transfusion de ces sujets avec des hématies de phénotypes courants, peut induire la production d'anticorps, potentiellement dangereux, nécessitant le recours à des unités de sang compatibles. Les incompatibilités transfusionnelles qui en résultent ont abouti parfois au décès de patients par manque d'unités de sang (Noizat-Pirenne F, 2001).

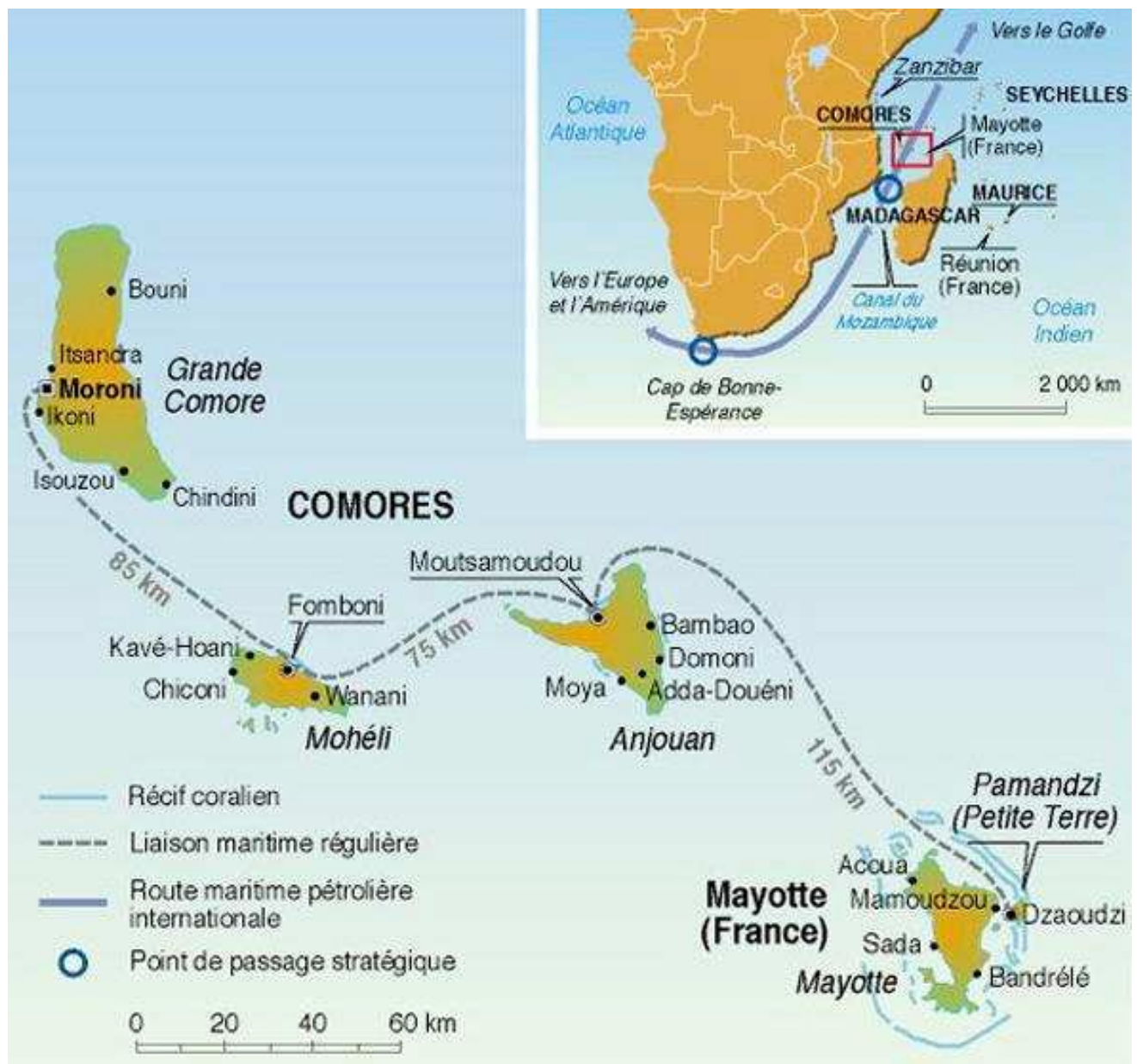
Ainsi, parallèlement à notre problématique anthropologique s'est ajoutée une problématique biomédicale liée au fait que la population étudiée représente une importante communauté d'immigrés (70 000 individus) résidant à Marseille.



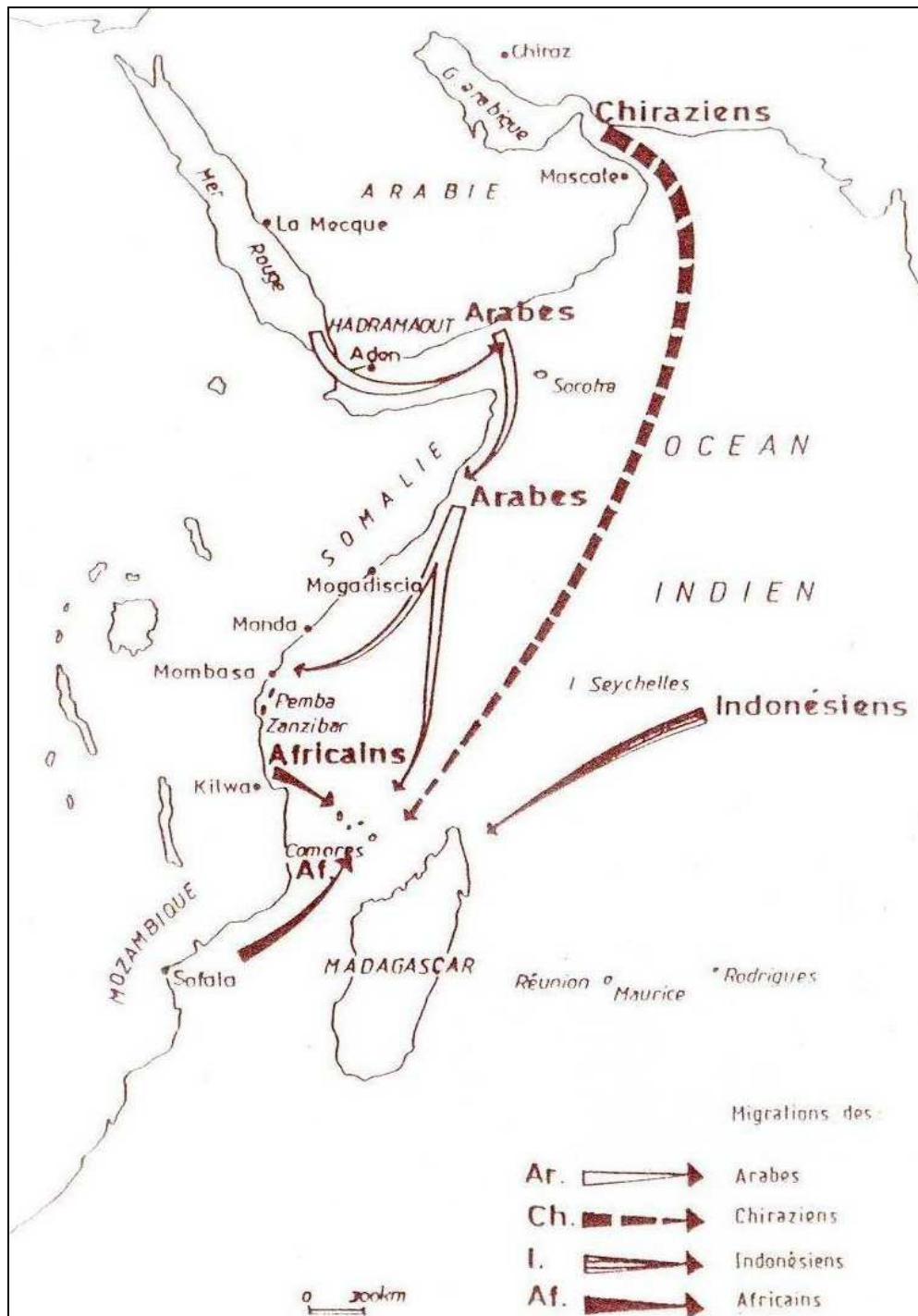
Nous proposons donc, sur la base de critères anthropobiologiques, d'explorer le polymorphisme de cette population, afin de définir les fréquences des phénotypes pouvant aboutir à des difficultés transfusionnelles (difficultés de diagnostic ou d'approvisionnement en unités compatibles).

Ces données permettront d'identifier les individus porteurs de ces marqueurs dans le but de les sensibiliser au don de sang et d'approvisionner la « Banque Nationale de Sangs Rares ». Par ailleurs, l'analyse de certains allèles « rares », véritables marqueurs identitaires, complétera l'approche anthropobiologique de cette population.

Afin de répondre à cette double problématique nous avons, dans un premier temps, analysé les données linguistiques, historiques et culturelles afin de définir les populations impliquées dans les événements historiques qui ont abouti au peuplement des Comores. Ces données nous ont permis de retenir les populations qui nous ont paru les plus pertinentes pour des analyses comparatives et de confronter nos résultats génétiques aux autres disciplines. Dans un deuxième temps, nous avons sélectionné, parmi l'ensemble des systèmes génétiques "disponibles", les marqueurs qui nous paraissaient utiles et dont l'exploration était accessible à nos différents laboratoires. Nous présenterons, pour chacun d'eux, la justification de ce choix ainsi qu'un rappel concernant les bases moléculaires de leur polymorphisme. Nous aborderons ensuite dans le chapitre « Matériel et Méthodes » les modalités d'obtention et les caractéristiques de notre échantillon, ainsi qu'une partie spécifique à chaque système ou groupes de systèmes étudiés. Chacune de ces parties comportera, les modalités d'exploration mises en oeuvre, les méthodes d'analyses des données et les populations retenues pour les analyses comparatives. Le chapitre « Résultats » présentera les données obtenues qui seront interprétées dans le chapitre « Discussion » où l'analyse des diverses contributions génétiques tentera de retracer l'histoire du peuplement de ce secteur. Enfin, nous définirons les « Perspectives » qu'ouvre notre étude.



Carte 1 : Localisation géographique de l'archipel des Comores.



Carte 2 : Les migrations à l'origine du peuplement des Comores (D'après Vérin P, 1994)

## II. Les paramètres linguistiques, historiques et culturels.

### II.1 Les données linguistiques.

Elles représentent le premier indice sur l'origine africaine de la population comorienne et sur ses contacts avec le Moyen Orient.

- **La rencontre Arabo-Bantoue.**

Le langage « SABAKI » apparaît historiquement sur la côte Orientale de l'Afrique. Il naît de la différenciation des langues Bantoues initiales ayant subi dès le début du premier millénaire les premières influences Arabes pré-Islamiques. Le « SABAKI » se différencie ultérieurement en trois branches : le « SWAHILI » qui s'étend sur toute la côte Est Africaine, le « POKOMO » qui reste restreint à l'embouchure du fleuve Tana en Ethiopie et vers 850 après JC au fur et à mesure de « l'insularisation », le « COMORIEN » qui est parlé par les populations DEMBENI, les ancêtres des habitants des Comores (Vérin P, 1994). Ainsi, les migrations parties de la côte africaine vers les Comores avaient subi la même influence arabe initiale que la langue et la culture Swahilies (Vérin P, 1994).

- **La marque de l'Afrique.**

L'étude de la structure grammaticale du « COMORIEN » (le shimasiwa) le rattache à la famille des langues Bantoues ([www. Ethnologue.com](http://www.Ethnologue.com)) (Table 1 p.12). Le « COMORIEN » présente deux groupes de locuteurs : le groupe Est (le « SHIMAORE » et le « SHINDZAWANI ») utilisé à Mayotte et à Anjouan et le groupe Ouest (le « SHINGAZIDJA » et le « SHIMWALI ») utilisé en Grande Comore et à Mohéli (Vérin P, 1994 ; Laffon M, 1991). Toutefois, ces langues sont inter-compréhensibles et présentent, bien entendu, des similitudes avec les langues de la côte africaine. Le groupe Est présente des similitudes avec le « SWAHILI » de MOMBASSA et le groupe Ouest avec les langues usitées au MOZAMBIQUE comme le « ESAKAJI » (Vérin P, 1994).

Le « COMORIEN » est donc différent du « SWAHILI » : la confusion est née du fait que le « SWAHILI » a été utilisé aux Comores dans les actes officiels des Sultans, jusqu'à la venue des Français.

- **La marque de la péninsule Arabique.**

De nombreux éléments linguistiques confirment par ailleurs les influences de la Péninsule Arabique comme :

→ L'utilisation de la calligraphie et de l'alphabet arabes pour écrire le « COMORIEN ».

→ L'emprunt de près de 30% de vocabulaire à l'Arabe (Laffon M, 1991).

→ L'Arabe classique, avec le Français, représente la langue officielle de la République Islamique des Comores. Le Comorien étant essentiellement la langue parlée.

- **Absence d'influence austronésienne sur le Comorien.**

Contrairement à Madagascar, on ne retrouve pas dans le « COMORIEN » d'empreinte austronésienne (Laffon M 1974, 1991; Vérin P, 1994). Toutefois, à Mayotte en plus du Comorien et du Français, est parlée une langue d'origine Austronésienne, le « KIBUSHI », qui présente des similarités avec la langue « SAKALAVA » de Madagascar. Cette langue représente le témoin de l'influence du royaume Ouest Malgache, le royaume SAKALAVA d'Andriandahifosty, qui émergea autour de 1400 après JC et qui établit une principauté à Mayotte (Robinau C, 1966).

Les langues Malgaches et celle parlée, en partie, à Mayotte, témoignent de l'influence Austronésienne dans l'océan Indien. L'étude grammaticale les assimile aux langues « BARITO » parlées en Indonésie au Sud Est de Bornéo enrichies de terme Javanais et Sumatrais. La langue parlée à l'extérieur de Madagascar qui présente le plus de similarités avec la langue Malgache semble être le « MA'AMYAN » du Sud de Bornéo dans la région de TAMIANGLAYANG drainée par la rivière PATAI ([www.Ethnologue.com](http://www.Ethnologue.com)) (Table 2 p.12)



Table 1 : Evolution linguistique du Shimasiwa

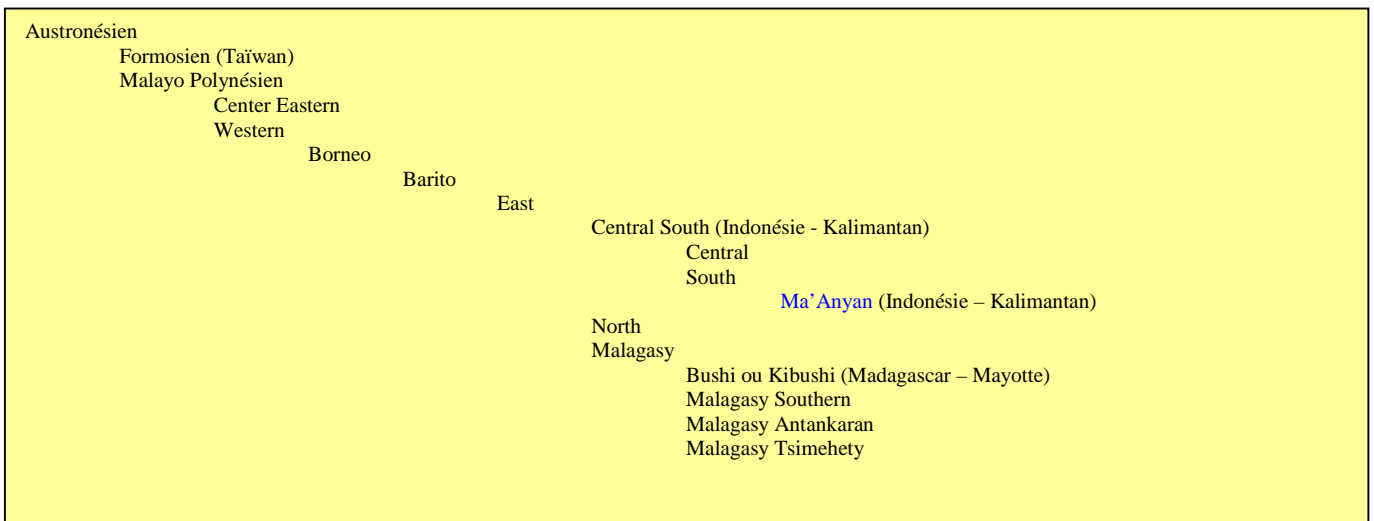


Table 2 : Evolution du langage du Sakalava parlé à Mayotte

- **Le comorien aujourd'hui.**

Le statut des langues est défini très brièvement dans la Constitution de 1985 : «Les langues officielles sont le Français et l'Arabe.» Étant donné que la population a été fortement islamisée depuis près de dix siècles, c'est la langue *écrite* de la population, l'arabe classique, et non la langue *parlée*, le Comorien, qui a été choisie comme langue co-officielle avec le Français.

Au Parlement, les députés s'expriment presque toujours en français, rarement en comorien, et jamais en Arabe. Les lois sont rédigées et promulguées en français seulement.

En matière de justice, la cour et les justiciables s'expriment en français, mais le comorien est également utilisé dans les communications orales. Dans les documents écrits, c'est encore massivement le français, très rarement l'arabe, qui est employé. Le juge rend généralement sa sentence en français, mais peut le faire en premier lieu en comorien (en français ensuite). Dans les cours d'appel, il arrive que l'on parle le comorien.

Dans l'administration gouvernementale, les services oraux sont fournis indifféremment en français ou en comorien, mais la tendance est à favoriser le français considéré comme la langue de l'État. Les documents écrits sont généralement rédigés en français, mais certains sont distribués en comorien, notamment dans le cas des municipalités et des organismes reliés à la justice, à la santé et à la police.

En matière d'éducation, la politique appliquée est fonction du niveau d'enseignement. A la maternelle, on enseigne en comorien et en français. Au primaire, seul le français est enseigné. Au secondaire, les élèves apprennent comme seconde langue l'arabe (pour lire le Coran) et parfois l'anglais. On inculque aux enfants les rudiments de l'écriture arabe, puis on les initie à la lecture du Coran, avant de les familiariser avec les subtilités des dogmes et des rites musulmans.

Enfin, dans la vie économique, le français occupe une place exclusive dans les communications écrites. Même les produits de consommation courante ne portent que des étiquettes ou des modes d'emploi en français.

Les médias écrits paraissent en français; la radio diffuse en français et en comorien et la télévision, en français.

Au total, le français occupe la place qu'il a toujours occupé depuis la seconde moitié du XIX<sup>e</sup> siècle: les fonctions officielles de l'État et les affaires. Le comorien, pour sa part, occupe les fonctions qui lui sont dévolues depuis une dizaine de siècles: la vie familiale et les communications informelles. Quant à l'arabe, son rôle reste strictement symbolique: on l'apprend à l'école pour pouvoir lire le Coran.



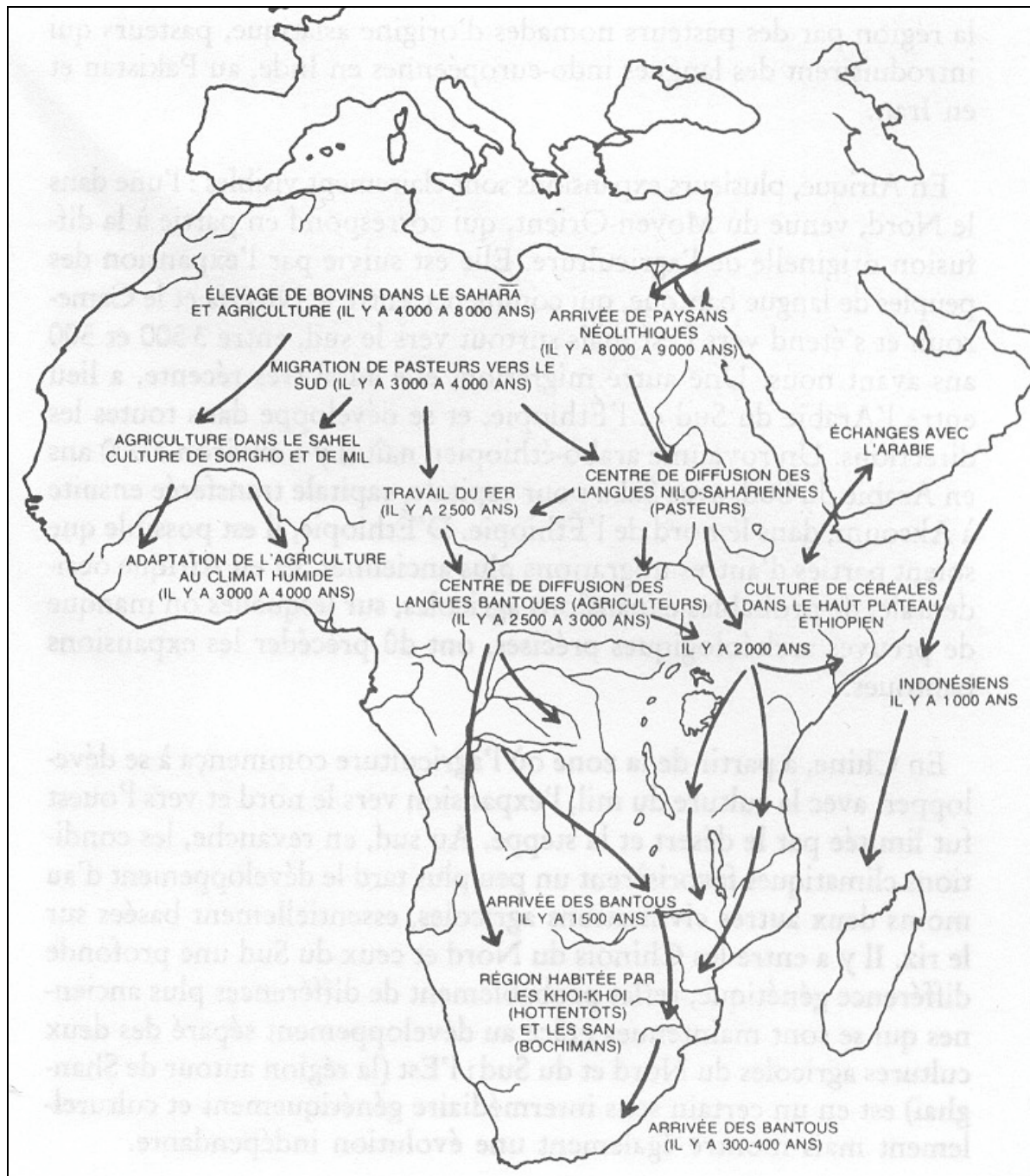
## **II.2 Les données historiques**

Au contact des mondes africain, arabe et austronésien, les Comores et Madagascar offrent une exceptionnelle diversité ethnique, linguistique et culturelle expliquée par l'histoire du peuplement de ces îles. Outre l'étude linguistique, son analyse repose sur la confrontation des données archéologiques, sur la réinterprétation des textes anciens comme le « Périple de la Mer Erythrée » ou les « Merveilles de l'Inde », sans oublier les données relevant de la tradition orale conservée dans des grandes familles comme celles d'Anjouan (Al Mandra, Al Mosela) dont d'éminents notables sont encore capables de donner une version de cette histoire ancienne. Certaines de ces versions ont été traduites, comme celle de Said Ahmed, ancien Cadi d'Anjouan, ou de Said Ali Amir (Robinau C, 1966).

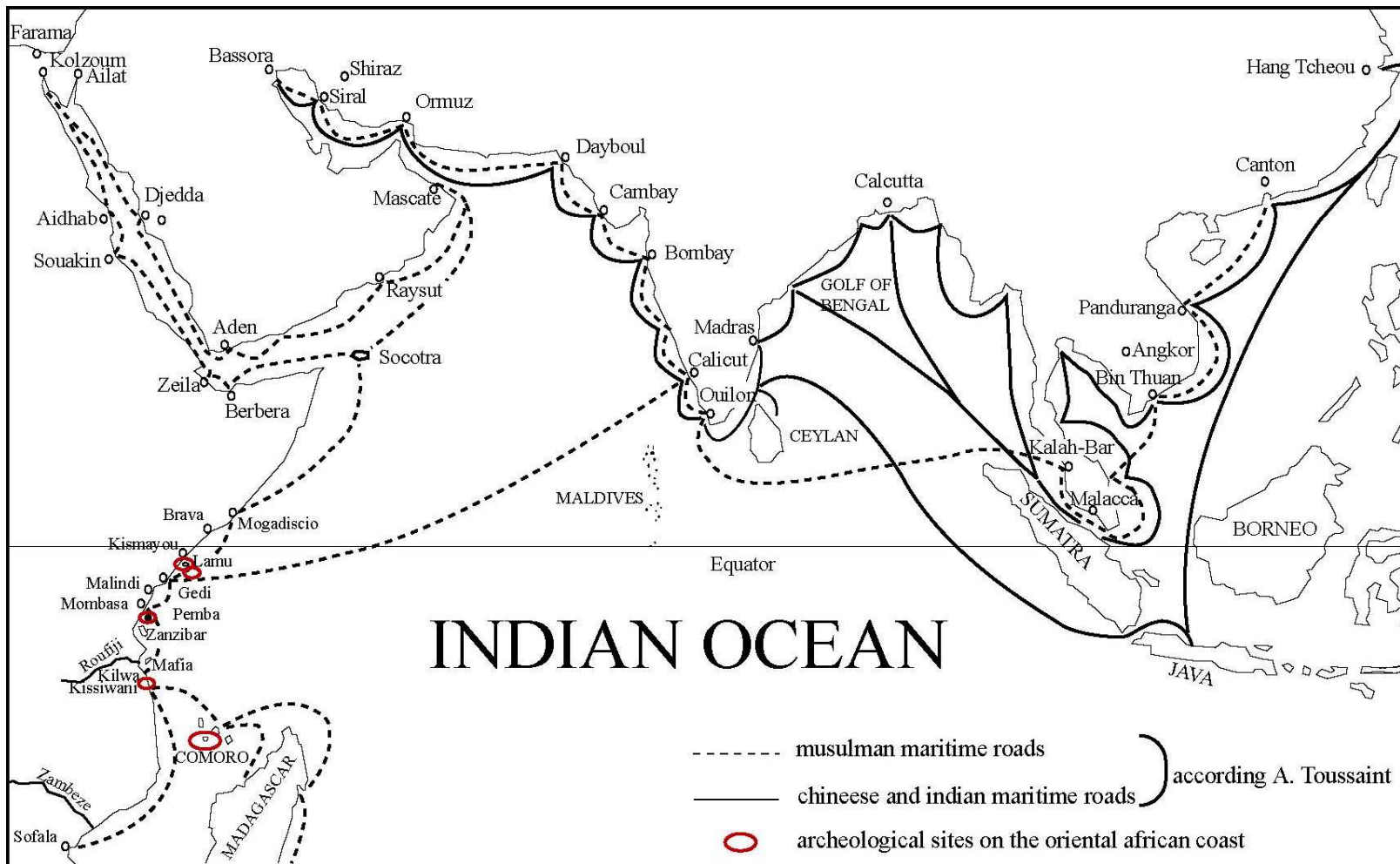
L'histoire du peuplement de l'archipel des Comores, comme celle de Madagascar, ne peut être séparée de celle de l'ensemble de l'Océan Indien où au moins trois éléments ont interféré :

- L'arrivée sur la côte orientale de l'Afrique de populations Bantoues métallurgistes (carte 3 p.16).
- L'expansion des commerçants Moyen Orientaux islamiques sur le littoral Africain (carte 4 p.17).
- La présence, dans la région de l'Océan Indien Occidental, de communautés Austronésiennes.

C'est à Neville Chittick (Chittick N, 1965) que l'on doit la meilleure analyse de ces influences arabo-africaines qui vont donner dans un premier temps naissance aux cultures swahilies de la côte africaine puis comorienne dans les îles.



Carte 3 : Expansions récentes probables en Afrique. « L'expansion bantoue » s'accorde assez bien avec les données linguistiques et les données archéologiques. (D'après Cavalli-Sforza, 1994b)



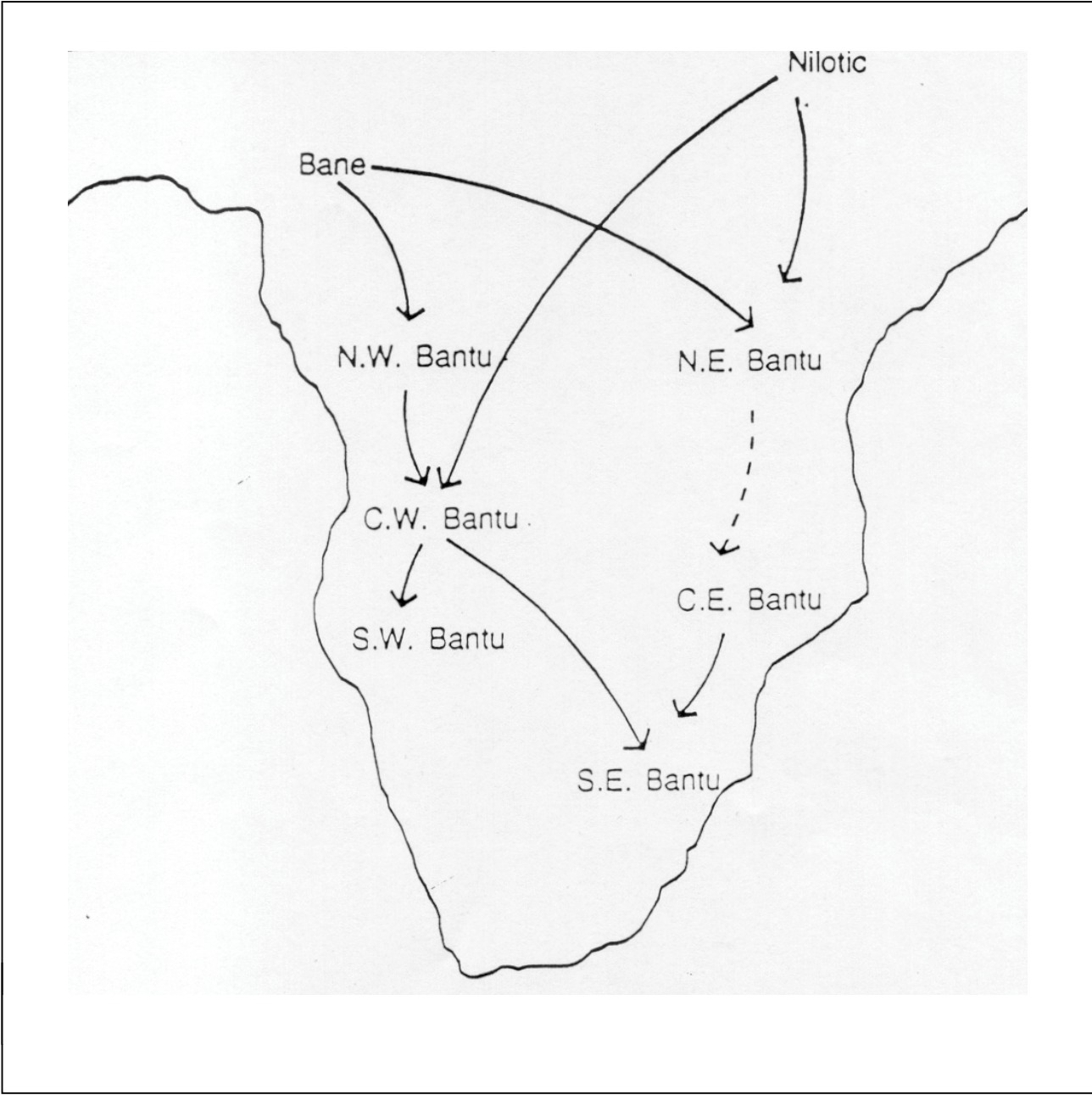
Carte 4 : Les échanges maritimes dans l'Océan Indien au Moyen Age (Cornevin M, 1998 modifiée par Gibert M, 2003).

### **II.2.1. Début du 1<sup>e</sup> millénaire : Etablissement des sociétés Swahilies sur la côte orientale de l'Afrique.**

La mise en place de la société Swahilie sur la côte Est Africaine est née de la rencontre de peuples Africains (Bantous et Couchitiques) et de navigateurs en provenance du Golfe Persique.

- **Les populations Bantoues sur la côte orientale de l'Afrique.**

Elles représentent la limite Est de l'expansion Bantoue qui a débuté environ 3000 ans avant notre ère dans l'Ouest Africain, avant l'introduction de la métallurgie du fer (Hammond - Tooke W.D, 1974). Les données génétiques, rapportant un faible taux de diversification interne, confirment l'hypothèse du caractère récent de cette dispersion (Excoffier, 1987 et 1991 ; Poloni, 1997). Les données linguistiques indiquent que l'origine se situe dans la région de la vallée de Benue, proche du centre d'extraction du fer de Nok, au Sud Est du Nigéria (région attestant de la plus grande diversité linguistique). Cette expansion se serait faite tout d'abord vers le Sud Ouest (à travers la forêt) puis vers l'Est à partir d'un ou plusieurs centres de dispersion situés dans le bassin du Congo [Les Nucleus congolais] (Heine, 1977 ; Ehret, 1982). L'arrivée des peuples Bantouphones sur la côte Est de l'Afrique est à situer il y a près de 1800 à 2000 ans (Cavalli-Sforza, 1994 ; Ehret, 1999 ; Vansina, 1990). Ces populations Bantoues apparaissent comme des groupes bien distincts par rapport aux autres populations Africaines (Afrique de Ouest non Bantoue, Khoisan, Pygmées et Afrique de l'Est) qu'ils ont remplacé dans certaines régions et avec qui ils ont parfois établi des échanges. A la suite de cette dispersion, il est classique de séparer les Bantous en 6 grands groupes : Nord-Ouest (NW), Centre-Ouest (CW), Sud-Ouest (SW), Nord-Est (NE), Centre-Est (CE) et Sud-Est (SE). [Carte 5 p.19]



Carte 5 : L'expansion Bantoue (Cavalli-Sforza, 1994a).

- **Les populations Arabes sur la côte orientale d'Afrique.**

Dès le début du 1<sup>e</sup> millénaire de notre ère, la côte Orientale de l'Afrique apparaît comme une dépendance totale du monde Arabe. En effet, cet Empire est tributaire de l'Ancien Monde d'où il reçoit, or, argent et fer Est Africain, transformé en acier en Inde méridionale ouvrage en Syrie et en Espagne avant d'être réexporté dans tout le monde connu. De même, les importations de bois, de tissus et surtout d'esclaves (Bantous Zanj) sont nécessaires à son urbanisme et ses constructions navales. Ces esclaves Africains Bantous sont très tôt, en Basse Mésopotamie, associés à la culture de la canne à sucre. Leur révolte dans les marécages du delta de l'Euphrates en 869-883 ap JC représente une preuve de leur présence dans cette région (Ottino P, 1974). Par ailleurs, l'installation des Arabes Yéménites sur la côte Africaine est signalée par le Périple de la Mer Erythrée dès le I<sup>er</sup> siècle (Freeman – Grenville GSP, 1966). Dès le V<sup>e</sup> siècle, après l'occupation du Yémen par les Ethiopiens d'Axum, les Arabes de la Péninsule Arabique (Oman, Arabie Saoudite), de l'Irak et des provinces Iraniennes du Golfe Persique vont maintenir pendant plusieurs siècles le monopole dans cette partie de l'Océan Indien. Ils y fondent Pate et Kilwa (Cerulli E, 1957). Par la suite, après les prédications du prophète, les expéditions de 658-705 ap JC sur l'Oman, considéré comme païen par le 5<sup>e</sup> Calife des Omayyades, vont grossir les communautés déjà installées en Afrique Orientale (Mantran R, 1969) (Hourani GF, 1963).

- **Etablissement des sociétés swahilies sur la côte orientale de l'Afrique.**

Ces arabes non musulmans d'abord puis musulmans, originaires de l'Asie du Sud Ouest, se sont donc installés sur la côte orientale d'Afrique et se sont métissés avec les groupes environnants Bantous ou Couchitiques. Ils ont constitué, avec une langue particulière, le swahili (Prins A.H.J, 1961), une culture originale fortement influencée par les croyances, les rites, les coutumes et par la suite, l'organisation de l'Islam.

## **II.2.2. Entre les V<sup>e</sup> et VII<sup>e</sup> siècles : le premier peuplement des Comores et de Madagascar.**

On ne sait pas exactement à quelle époque les îles ont commencé à être habitées. Neville Chittick estime que le premier document à prendre en compte est « le Périples de la Mer Erythrée », texte d'un marchand grec d'Alexandrie du III<sup>e</sup> siècle. Ce document, qui décrit notamment la navigation et le commerce le long de la côte orientale de l'Afrique, ne mentionne jamais les Comores. Aucune preuve n'est disponible aujourd'hui pour attester d'une éventuelle occupation des îles du canal du Mozambique en des temps si reculés. Pour Said Ahmed (Robinau C, 1966) les premiers habitants des Comores étaient des Africains de la côte du Mozambique abandonnés par les navigateurs Arabes à l'occasion de la traite des esclaves à laquelle ils se livraient dans cette partie de l'Océan Indien. En effet, les négriers Arabes achetaient leurs esclaves au Mozambique puis, à l'occasion d'arrêt d'approvisionnement d'eau douce dans ces îles, ils débarquaient une partie de leur cargaison. Il semble que l'implantation des premiers hommes en provenance de l'Afrique de l'Est Bantoue se soit réalisée entre le V<sup>e</sup> et le VII<sup>e</sup> siècle (Vérin P, 1994) durant une période pré-islamique, comme l'indiquent un certain nombre d'éléments :

- Les "chroniques de Pate" rappellent que dès le VII<sup>e</sup> siècle, les OMANI essaient sur toute la côte y compris aux Comores (Freeman-Grenville GSP, 1966).
- Les fouilles archéologiques ont permis de retrouver au Sud Ouest de Ngazidja une sablière qui contenait une sépulture où le défunt avait été déposé en position non islamique (Vérin P, 1994).
- La mise en évidence de restes de porcs, de lémuriers et de tortues évoquant des habitudes de consommation non autorisées par l'Islam (Vérin P, 1994)
- Les naturalistes notent des transformations de la faune et de la flore de Madagascar sous l'influence humaine estimée au V<sup>e</sup> siècle (Vérin P, 1971).
- Le Livre des Zen'ji rapporte que, depuis les temps les plus anciens, les Africains Zanj accompagnent les Arabes et que leurs relations sont institutionnalisées (Cerulli E, 1957)

### **II.2.3. Au VIII<sup>e</sup> siècle : Mise en place d'une société plus structurée, dite 'DEMBENI'.**

Depuis la seconde moitié du VIII<sup>e</sup> siècle et jusqu'au X<sup>e</sup>, de gros villages se sont développés aux Comores. Les premiers de ces établissements humains ont été découverts à Dembeni (Mahoré). Ce terme de Dembeni désigne alors cette période de mise en place des premiers Comoriens sur l'ensemble des îles, puisque les mêmes types d'établissements ont été découverts à Mro Dewa (Mwali), Sima (Ndzواني) et Mbashile (Ngazidja). Ainsi, une civilisation appartenant à la famille des peuples parlant les langues bantoues s'est transposée depuis le littoral de l'Afrique Orientale jusqu'à l'archipel pour aboutir au VIII<sup>e</sup> siècle aux ancêtres des Comoriens : les DEMBENI. Cette composante majoritairement africaine et pré-islamique est la source des coutumes et traditions qui se sont perpétuées dans la société Comorienne (Direche-Slimani K, 2002 ; Vérin P, 1994).

### **II.2.4. Entre les IX<sup>e</sup> et X<sup>e</sup> siècles : Les contacts Austronésiens.**

Le commerce soutenu entre l'Afrique, le monde Arabo-Persan et la Mer de Chine (Ottino P, 1974) ainsi que les techniques de navigation rendent plausibles des contacts anciens avec l'Asie du Sud et l'Asie du Sud Est dès 950 ap JC. En effet, les voies commerciales partaient d'Afrique et de la péninsule Arabique vers l'Inde, puis d'Inde vers le Royaume Sri Vajaya (Sumatra) et enfin vers la Péninsule Malaisienne et vers la Chine (Carte 4 p.17). De même, les Indiens et les Arabes apparaissaient comme les intermédiaires nécessaires aux exportations de Chine vers l'Afrique (Cornevin M, 1998).



### **Les éléments relatifs aux contacts avec l'Asie du Sud.**

- D'anciennes chroniques font état de l'existence de migrants indiens du Nord, les WaDebuli, provenant de Dabohl au Sud de Bombay. Selon les fouilles réalisées sur la côte orientale de l'Afrique et du Pakistan, Chittick note une correspondance entre le nom de ces cités qui représentaient les centres commerciaux les plus actifs de l'Océan Indien et la dénomination de cette population qui s'était établi en plusieurs points sur la côte Est Africaine et en particulier à Pemba et Zanzibar. Compte tenu des données de la tradition orale, cet auteur estime l'arrivée des WaDibuli entre le VIII<sup>e</sup>-IX<sup>e</sup> siècle et le XII<sup>e</sup> siècle, avant l'influence Schirazienne (Ottino P, 1974). La question relative à l'atteinte des Comores par ces peuples après qu'ils furent chassés de la côte africaine reste encore sans réponse.
- Une autre vague de migrants, appelés WaDiba, vint des Maldives vers la côte Est Africaine. Selon les "chroniques de Lamu", les bateaux WaDiba arrivèrent sur l'île de Kiwayu (Nord de Lamu) à la suite d'une tempête à la fin du XV<sup>e</sup> siècle et ce avant l'arrivée des Portugais (Ottino P, 1974). Ces WaDiba ont introduit de nombreuses techniques de navigation, mais rien n'est établi sur d'autres contacts avec la côte Est Africaine et les Comores.

### **Les éléments relatifs aux contacts avec l'Asie du Sud Est.**

Les états Dembéni ont été en contact avec des populations Indonésiennes utilisant les Comores comme têtes de ponts en vue de renforcer leur implantation à Madagascar. Il semblerait que ces Austronésiens soient des Indonésiens en provenance surtout de Bornéo comme l'évoque l'absence de trace d'Hindouisme dans la culture Malgache ainsi que les données linguistiques.

- En effet, on a supposé, en l'absence de toute trace d'hindouisme dans la culture traditionnelle malgache, qu'ils étaient antérieurs à l'indhouisation de l'Indonésie, c'est-à-dire au III<sup>e</sup> siècle après Jésus-Christ. En fait, les îles de Bali, Java et Sumatra ont subi l'impact de l'hindouisme. Donc, si les Proto Malgaches sont originaires des îles non hindouisées de l'Indonésie ainsi qu'on l'a prétendu – îles Célèbes (Sulawesi), Bornéo (Kalimantan), îles de la Sonde\_, leur départ de l'Austronésie pourrait être beaucoup plus récent, soit aux alentours du X<sup>e</sup> siècle de notre ère. On avance aujourd'hui que ces premières circumnavigations indonésiennes, déjà liées au commerce des épices, auraient pu commencer dès le VIII<sup>e</sup> ou le IX<sup>e</sup> siècle.
- De même, les données linguistiques de Madagascar et de Mayotte (Bushi) évoquent des contacts avec les populations du Sud de Bornéo dont la langue le Ma'Myan comporte le plus de similitude avec les langues Malgaches ([www.Ethnologue.com](http://www.Ethnologue.com)).

Ottino (Ottino P, 1974) évoque aussi, qu'entre les X<sup>e</sup> et XII<sup>e</sup> siècles des Malais indianisés s'installent aux Comores et dans la région du Nord de Madagascar, et avec eux des idées comme les conceptions indiennes de la royauté, des techniques comme celles de la production du riz de montagne et du riz de repiquage en rizières irriguées.

Les Comores n'ont pas été épargnées par ces contacts. Les données archéologiques évoquent, en effet, des activités de métallurgie à Mayotte (Allibert C, 1993) avec l'utilisation de batteries de fours métallurgiques du X<sup>e</sup>-XI<sup>e</sup> siècle dont le système de ventilation à puits, pourrait faire penser au système à cylindres verticaux utilisés à Madagascar et en Indonésie. Par ailleurs, si contrairement à Madagascar, l'apport indonésien n'a laissé aucune trace linguistique, la culture comorienne a gardé néanmoins des éléments de ces contacts austronésiens comme la pirogue à balanciers, la culture du bétel, du cocotier, du riz, de

l'arbre à pain, de la canne à sucre, du gingembre qui coexistent avec les plantations africaines (igname et mil) (Laffon M, 1974).

#### **II.2.5. Entre les XI<sup>e</sup> et XV<sup>e</sup> siècles : Islamisation des Comores et l'influence de l'élite Schirazienne.**

Entre le XI<sup>e</sup> et le XV<sup>e</sup> siècle, l'histoire des rois de Kilwa ("chronique de Kilwa") donne une place essentielle à la venue des Schirazi (Iran actuel) dont elle présente l'épopée comme l'évènement principal de l'histoire de la côte Orientale de l'Afrique et de ses régions. Il convient ici de faire la part entre l'influence du mythe Schirazien dont sont imprégnés aujourd'hui beaucoup de populations Africaines orientales et Comoriennes, de la réalité notamment en ce qui concerne l'origine de cette élite.

##### ***Le Mythe fondé d'après la tradition orale.***

Le Sultan de Schiraz, Ali Bin Al Hassan, avec ses 6 fils fit voile vers la côte Africaine dans 7 bateaux dont chacun fit escale dans un endroit différent. Le Sultan débarqua aux Comores et un de ses fils Ali à Kilwa. Ainsi, la ville de Kilwa fut érigée en 975 en sultanat et les Comores se trouvaient dans sa dépendance au même titre que Zanzibar et une partie de Madagascar.

##### ***Les apports de l'archéologie.***

S'appuyant sur les résultats des fouilles archéologiques, Neville Chittick a établi que, contrairement à leurs propres traditions, les Schirazi du Sud de la côte Orientale d'Afrique ne viendraient pas directement du Golfe Persique mais de la côte autour de Mogadiscio, ville qui, au X<sup>e</sup> siècle, commence à prendre beaucoup d'importance pour devenir au début du XII<sup>e</sup> le grand emporium d'Afrique de l'Est. L'expansion Schirazi vers le Sud au XII<sup>e</sup> siècle est immédiatement marquée par le développement de Kilwa qui s'assure le monopole du marché de l'or de Zimbabwe et dont l'aire d'influence s'étend, au travers du Canal de

Mozambique, sur les Comores et la côte Occidentale et Septentrionale de Madagascar (Le Bourdieu F, 1969 ; les implantations des comptoirs arabes à Madagascar, Atlas de Madagascar, planche 19, commentaires. Battistini P et Vérin P).

Ce mythe Schirazien, issu de la tradition orale, doit être interprété comme un renforcement des liens entre l'Afrique Orientale, les Comores, le Moyen Orient et l'Islam et matérialisé par des vagues successives de migrants provenant de Schiraz avec un détour préalable plus ou moins long sur la côte Africaine.

Ainsi, à leur arrivée aux Comores, ces Schiraziens imposent naturellement une domination intellectuelle, religieuse et politique aux Comoriens qui apparaissent en position d'infériorité vis à vis de ces pieux musulmans qui arrivent des grands foyers religieux de l'Islam, à savoir les « Cités Etat » d'Afrique de l'Est . C'est ainsi que ces bantous se convertissent progressivement à l'Islam et l'archipel est intégré au commerce swahili entre la côte Orientale de l'Afrique, le Proche Orient et l'Orient. On peut estimer que l'archipel est totalement musulman à l'arrivée des portugais au XV<sup>e</sup> siècle (Vérin P, 1994).

#### **II.2.6. Au XVI<sup>e</sup> siècle : Mise en place d'une Aristocratie Schirazienne.**

A partir du XVI<sup>e</sup> siècle, les Arabo-Schirazi mettent en place une aristocratie Schirazienne qui contrôle les activités portuaires puis, tout en nouant des alliances avec les chefs locaux (ils n'ont pas de femme et épousent les filles de ces chefs), ils forment des clans qui deviennent peu à peu des Sultanats. La société est organisée en classes avec les Wakaboula (les nobles), la classe d'hommes libres (agriculteurs et pêcheurs) et les esclaves.

Ils construisent la première mosquée de pierre de Mayotte à Chingoni et d'Anjouan à Sima en 1566. En Afrique Orientale, comme aux Comores, les Schirazi introduisent une architecture développée utilisant la chaux et le ciment, la charpenterie, le tissage du coton, le calendrier solaire persan et de autres nombreux produits venus du Golfe Persique (Trimingham JSP, 1964).

### **II.2.7. Entre les XVII<sup>e</sup> et XVIII<sup>e</sup> siècles : Guerres inter-sultanats et raids Malgaches.**

Au XVIII<sup>e</sup> siècle, les quatre îles sont la cible de pirates malgaches poussant les sultans à rechercher une protection au près de grandes puissances. Ces “îles des sultans batailleurs” prennent à leur service des pirates européens chassés de la mer des Antilles, ouvrant ainsi leur territoire à l’influence de mercenaires et d’aventuriers qui ne cessent de s’impliquer dans la vie politique locale. En 1790, les pirates malgaches dévastent Anjouan.

### **II.2.8. En 1843 débute l’histoire Française aux Comores.**

*Voir le chapitre relatif à « L’histoire récente » page 30.*

## **II.3 Les données culturelles**

Les données culturelles actuelles évoquent la coexistence de l’Islam avec des coutumes témoignant d’une imprégnation africaine persistante.

Toute la culture comorienne est marquée par la vision mythique de l’arrivée de l’Islam et de ses valeurs arabo-persanes. En comorien, civilisation se dit “ustaarabu”, ce qui signifie « devenir arabe ».

Toutefois, les Comores sont une nation islamique dont l’Islam n’a pas remplacé, mais a imprégné des coutumes Bantoues Africaines. C’est un Islam sunnite, orthodoxe, sous rite Chafiste (imam Al Chafi mort au Caire en 820 fondateur de l’une des quatre écoles sunnites) avec de nombreuses confréries religieuses, accommodée de coutumes locales d’origine Africaine.

La plus frappante est celle du grand mariage, qui règle la place de l'homme dans la société Comorienne. La célébration de cette tradition coutumière, totalement différente de l'union légale musulmane, obligatoire pour les enfants aînés (et les personnes qui en ont les moyens) donne au mari son statut "d'homme accompli" et lui accorde le droit à la parole dans les lieux publics et le respect des anciens. C'est un impératif social aux Comores, la consécration absolue dans le cycle de vie d'un comorien. Cette tradition cloisonne la société comorienne entre les mariés ayant le pouvoir politique, ceux qui doivent accomplir le grand mariage et les autres. Cette coutume repose sur une spécificité matriarcale d'origine africaine. La propriété foncière indivise appartient traditionnellement aux femmes d'un même lignage maternel et la maison du couple appartient à l'épouse qui l'a reçue de ses parents pour le mariage.

Cet aspect donne aux femmes une autonomie rare en milieu musulman, le mari étant avant tout considéré comme un géniteur. Les maris peuvent se succéder pour une même femme et ils sont partagés entre les obligations de leur propre foyer et celles de leur famille maternelle.

A côté de cette dévolution matrilineaire des droits fonciers on note d'autres éléments "africains" comme l'importance de la vie associative à l'intérieur de classes d'âges ou de quartier, la responsabilité de l'oncle maternel sur les enfants de sa sœur ou l'association de la possession du bétail au pouvoir. En effet, l'oncle maternel représente la figure masculine de référence devant prioritairement protection et assistance à ses sœurs et ses nièces.

Au total, les impacts non génétiques de ces divers contacts peuvent être résumés dans la Table 3 p.29.

<b>AFRIQUE</b>	<b>PENINSULE ARABIQUE</b>	<b>INDONESIE</b>
<b>Langue :</b>	<b>Langue :</b>	<b>Langue :</b>
Le Comorien est une langue Bantoue.	Alphabet arabe. Emprunt de près de 30% des mots à l'Arabe. L'Arabe classique est la langue officielle avec le Français.	Absence d'impact sur la langue comorienne. Toutefois, à Mayotte, parlé du BUSHI, similaire aux langues malgaches appartenant au Malayo-Polynésien.
<b>Elevage et Agriculture :</b>		<b>Elevage et Agriculture :</b>
Association du bétail au pouvoir. Culture du coton, de l'igname et du mil.		Culture du riz dans rizières irriguées et culture en terrasse. Culture du bétel, du cocotier, de la canne à sucre et du gingembre.
<b>Culture et traditions :</b>	<b>Culture et traditions :</b>	<b>Culture et traditions :</b>
Grand Mariage. Dévolution matrilineaire des droits fonciers. Importance de la vie associative. Responsabilité de l'oncle maternel sur les enfants de sa sœur.	Religion musulmane. Monument islamique de type mosquée. Architecture avec chaux et ciment. Charpenterie. Tissage du coton. Calendrier solaire persan. Mythe Schirazien.	Pirogue à balancier.

**Table 3** : Contribution des différentes cultures dans la société comorienne actuelle.

#### ***II.4 L'histoire récente.***

La France s'intéresse aux Comores dès 1810, en compensation de la perte de l'île Maurice au profit de la Grande Bretagne. Elle achète Mayotte en 1848 à un prince malgache pour 1000 piastres. A partir de cette date, des colons Réunionnais et Nantais s'y installent pour cultiver la canne à sucre. Les autres îles sont ensuite occupées et acquièrent le statut de protectorat français en 1886, de colonie française en 1912 puis de TOM en 1946. Le référendum de 1975 verra les trois îles devenir, en 1978, République Islamique des Comores et Mayotte demeurée collectivité territoriale française.

- **La migration vers la France.**

La migration comorienne vers la France est récente et s'enracine dans une tradition migratoire plus ancienne orientée vers les pays voisins de l'archipel : l'Océan Indien, l'Afrique de l'Est, Zanzibar, le Kenya, la Tanzanie, Madagascar et la Réunion. Certaines caractéristiques demeurent inchangées et, en particulier, on observe que la Grande Comore reste l'île d'origine d'une grande majorité de migrants. Ainsi, la migration Comores-France se révèle plutôt une migration Grande Comore-France qui prend source dans une organisation sociale originale fondée sur l'accomplissement de la coutume du "Grand Mariage". Néanmoins, des navigateurs (employés sur les navires à partir de la seconde moitié du XIX<sup>e</sup> siècle) aux migrants actuels, le nouveau profil de cette population en France est façonné par le passage d'une migration temporaire, concernant des hommes seuls, à une migration plus durable, avec l'arrivée croissante des femmes et la constitution de familles. En effet, la colonisation Française associée au développement de la marine marchande a fait de la France, et surtout de ses villes portuaires, l'un des principaux lieux de passage puis de destination de cette communauté. La migration Comorienne est entretenue par le fait que l'économie comorienne reste sous perfusion de la rente migratoire. En effet, l'émigré comorien devient un acteur économique essentiel à la survie de son pays.



Pauvre à l'origine, il devient avec l'aventure migratoire quelqu'un de respecté et d'accompli notamment par la possibilité de réaliser son Grand Mariage au pays (Direche-Slimani K, 2002).

- **La migration vers Marseille.**

Depuis toujours Marseille a conservé sa capacité d'intégration et d'absorption de nombreuses populations immigrantes. Marseille, par sa situation géographique, s'est construite au rythme des vagues migratoires et de l'histoire coloniale africaine qui a déplacé des individus qui vont lier leur destin à celui de la cité phocéenne et contribuer à l'écriture de son histoire. Les Comoriens arrivent à Marseille dès 1940. Il s'agit de navigateurs recrutés par les compagnies de messageries maritimes ou la compagnie des Indes. Ces marins débarquèrent en migrants temporaires et pour des séjours brefs. Ils représentent, en 1950, une faible proportion de la population Africaine. Par la suite, le remplacement du charbon par le mazout a pour conséquence la diminution de leur embauche car les Comoriens occupaient les métiers à faible niveau de qualification (soutiers). Ils deviennent alors dockers par défaut. Ces navigateurs sont les individus fondateurs de l'immigration Comorienne à Marseille. Originaire dans leur majorité de la Grande Comore, ils ont quitté cette île alors qu'elle était toujours Française. Dès 1970, on assiste aux premiers regroupements familiaux qui initient une migration massive durant toutes les années 80. Ainsi, parmi les nombreuses populations présentes à Marseille, la communauté Comorienne est en train de devenir l'une des plus importantes. Dans le département des Bouches-du-Rhône, la population Comorienne est actuellement évaluée à près de 70 000 individus organisés selon une structure sociale calquée sur leur système originel de regroupement en "villages". Si les échanges au sein de cette population sont courants, il n'en est pas de même, pour le moment, en ce qui concerne les échanges avec les autres populations présentes à Marseille (Direche-Slimani K, 2002).

- **Les patronymes.**

Le nom se prête particulièrement à une approche pluridisciplinaire. Il s'agit d'un élément universel, tout individu est porteur d'un nom qui le désigne au sein de la communauté. Mais, il est aussi le reflet de pratiques culturelles, le nom ne sera pas le même, il ne sera pas transmis selon le même mode et selon les différents groupes humains. Le passage du nom individuel au nom de famille ne s'est pas réalisé de la même manière selon les sociétés. En Europe, dans de nombreux pays, le nom de famille est ou a été un « patronyme », mais parfois un « matronyme », et parfois aussi un composé du nom du père et de celui de la mère. Le nom, présent dans toute population historique, depuis des siècles, sous toutes les latitudes, peut être sujet d'étude pour le sociologue, l'historien, le démographe, le généticien, l'anthropologue. Il se prête à des études qui prennent en considération la dimension spatiale, la profondeur chronologique, la dimension sociale ou culturelle, ou peut être considéré comme un marqueur biologique (Brunet G, 2001). Les Comores ont une particularité intéressante : les noms patronymiques ne sont composés que d'un prénom.

La notion de nom ou de prénom n'existe pas pour les comoriens. Elle leur a été imposée. Pour compliquer la chose, le prénom du fils aîné est très souvent celui du père, celui du cadet de l'oncle paternel, celui du benjamin du grand père ou d'un oncle maternel. Dès 1820, pendant la colonisation des Comores, des fonctionnaires français débarqués dans les 4 îles ont bien tenté d'y imposer un état civil. Sans trop de succès, étant donné la complexité des filiations. La transmission des patronymes comoriens n'étant pas rigoureuse, nous n'avons pas pu les utiliser comme marqueur biologique.

### **III. Les marqueurs génétiques utilisés.**

Nous aborderons, au cours de ce chapitre, quatre points :

- L'apport de la génétique en Anthropologie.
- La notion de polymorphisme humain afin de présenter un panorama des différents systèmes disponibles et de resituer, au sein de la variabilité génétique humaine, ceux que nous aurons sélectionnés.
- Les éléments pris en compte pour la sélection des systèmes étudiés dans ce travail.
- La description génétique, phénotypique, fonctionnelle et évolutive des différents systèmes retenus.

#### ***III.1. L'apport de la génétique en Anthropologie.***

Pour comprendre l'évolution de l'homme et sa diversité actuelle, l'Anthropologie dispose de deux outils génétiques : l'exploration du polymorphisme génétique humain qui fut complétée, à partir des années 30, par les modélisations fournies par la génétique des populations (Modèle de Hardy Weinberg). L'exploration du polymorphisme permet de saisir la diversité génétique entre et à l'intérieur des différents groupes humains et, compte tenu que l'Anthropologue d'aujourd'hui se préoccupe moins de classer que d'expliquer la diversité, la génétique des populations lui donne les outils de réflexion pour comprendre les mécanismes évolutifs. Ces mécanismes sont gouvernés par des paramètres qui ne sont pas tous d'ordre génétique puisque, outre la mutation la sélection ou la dérive génétique, ils intègrent les écarts à la panmixie, les migrations et la taille de la population (Loi de Hardy-Weinberg). A l'échelon de l'individu, la source de la diversité génétiquement transmissible est la mutation dans les cellules germinales : le polymorphisme naît de la mutation et est

entretenu, par la méiose. A l'échelle de la population, l'évolution de la diversité génétique et de son génome collectif dépend de l'échelle d'observation.

La comparaison intra- ou inter-populationnelle donne, en effet, les clés pour cette analyse : la diversité intra-populationnelle va diminuer sous l'action de la sélection naturelle ou de la dérive génétique mais pourra s'accroître en cas de migration dont l'impact sera d'autant plus important que la population sera de faible effectif. A contrario, la diversité inter-populationnelle diminuera par l'existence d'un phénomène de migration entre les populations considérées alors que dérive génétique et sélection l'accentueront.

Au sein d'une population, la variabilité entre les individus peut provenir de leur diversité génétique et de la diversité des milieux dans lesquels s'expriment leurs gènes. Cela nécessite de prendre en compte la notion d'interaction « gène-environnement » afin de distinguer son impact de celui qui est directement lié à l'histoire des migrations humaines. Cette interaction intègre la notion de sélection naturelle qui représente un tri des gènes les plus adaptés aux conditions du moment, parmi les gènes disponibles et accessibles à ce phénomène. Ceci est dirigé par une fécondité différentielle du plus apte qui augmente le nombre de descendants porteurs des formes alléliques adéquates.

C'est un processus dynamique dans un écosystème donné qui favorise les individus qui apparaissent les plus en adéquation avec l'interface « gène-environnement ». La sélection ne crée pas la variabilité. La variation génétique est aléatoire par rapport à son utilité éventuelle laquelle peut être multiple et changeante. Il n'y a pas de mutation dirigée. La sélection opère un tri parmi la variabilité existante et disponible sous contrainte du milieu.

L'approche génétique d'une population ne doit pas être interprétée isolément mais elle doit être confrontée aux autres disciplines dont l'Anthropologie représente le carrefour. L'une de ces approches est représentée par la linguistique.

Bien qu'il y ait une absence de lien direct entre les gènes d'une personne et sa langue, il existe une forte corrélation entre les familles de langues définies par les linguistes et les populations humaines définies par les généticiens (Figure 1 p.36). En effet, les familles de langues (Figure 2 p.37) et les populations résultent conjointement d'événements historiques

expliquant le fait que les évolutions génétiques et linguistiques peuvent concorder (Ruhlen M, 1991 ; Cavalli-Sforza L, 1994a).

La raison principale en est que les deux évolutions suivent en principe la même histoire c'est à dire une série de scissions. Deux populations qui viennent de se séparer entament en même temps un processus de différenciation de leurs gènes et de leurs langues. Sans qu'il soit nécessaire que ces deux processus aient un rythme d'évolution parfaitement constant, il est raisonnable de s'attendre, dans les deux cas, à une différenciation quasi proportionnelle au temps écoulé. Ils doivent donc concorder, sauf dans le cas d'évènements extérieurs, tels un apport génétique (remplacement de gènes) ou un changement de langue (remplacement de langue) imposé par des envahisseurs (Renfrew C, 1987). Ceux-ci peuvent estomper le tableau génétique et linguistique sans toutefois le masquer complètement.

Ainsi, en partant du principe que les mots évoluent à un rythme bien spécifique comme le font les gènes, les linguistes proposent une chronologie de l'apparition des grandes familles linguistiques. Toutefois d'autres linguistes, sont critiques sur un travail ne prenant en compte que l'évolution des mots qui, tous les jours, entrent et sortent d'une langue. Il conviendrait plutôt de s'attacher aux structures phonologiques et ou syntaxiques de chaque langue qui apportent des informations plus fiables dans le cadre des études phylogéniques. C'est ce qu'a fait J.H.Greenberg (Greenberg JH, 1963) pour faire entrevoir l'existence d'une très ancienne langue ancestrale commune qui s'est ramifiée au cours des millénaires sur le continent eurasiatique. Ainsi, les grandes subdivisions de nature linguistique sont les paramètres les plus fortement corrélés aux variations génétiques des populations, par rapport aux composantes spatiales (Excoffier L, 1987). Les études conjointes gènes et langues complétées par les données archéologiques conduisent à une meilleure compréhension des origines et de la répartition des êtres humains sur la planète.

***Quels sont donc « les gènes » qui sont à la disposition des anthropologues pour appréhender le profil génétique d'une population ? Comment accéder à leur diversité dont le cumul définit la notion de polymorphisme génétique humain et l'expression "sanguine", la notion de groupes sanguins chez l'Homme ?***

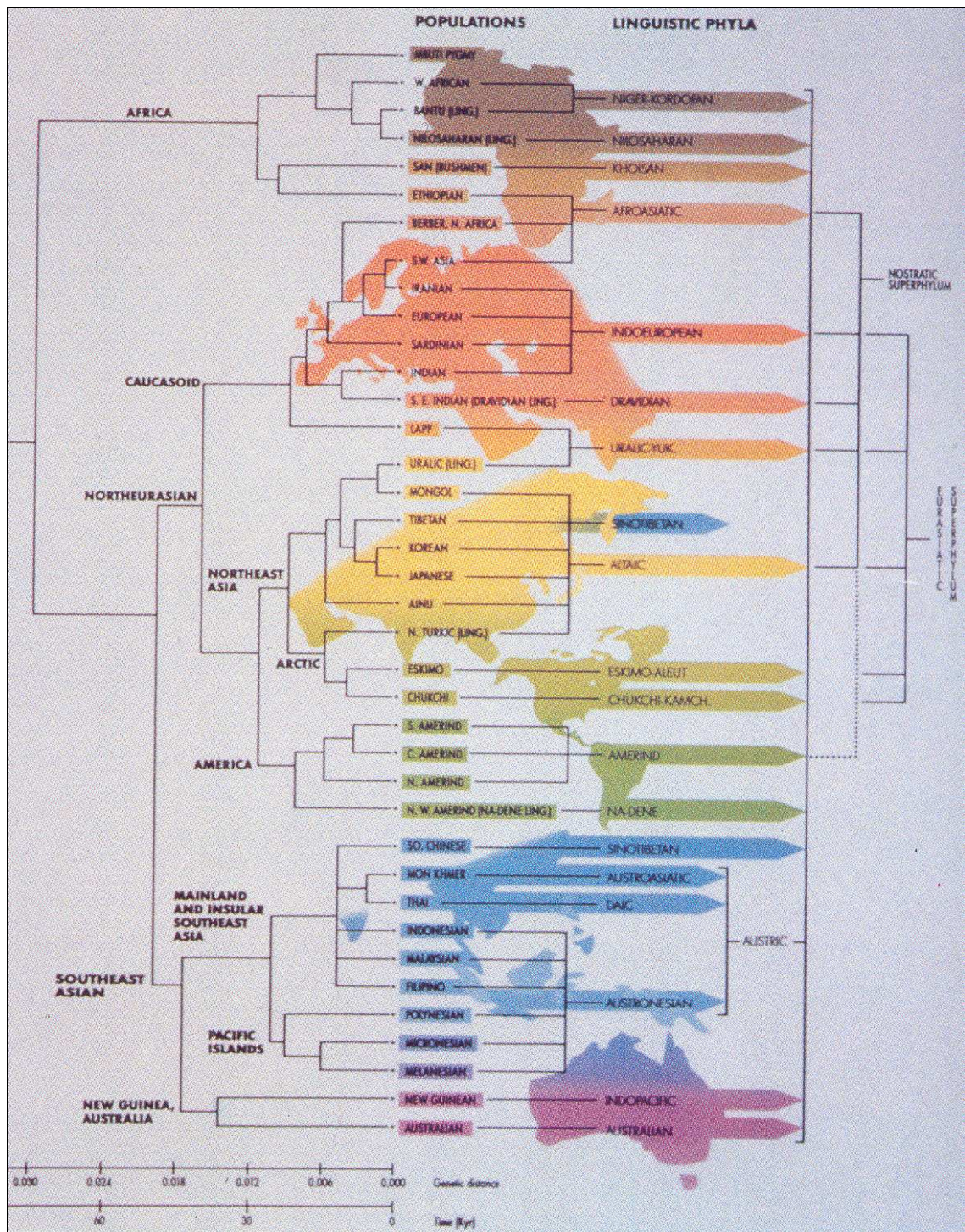
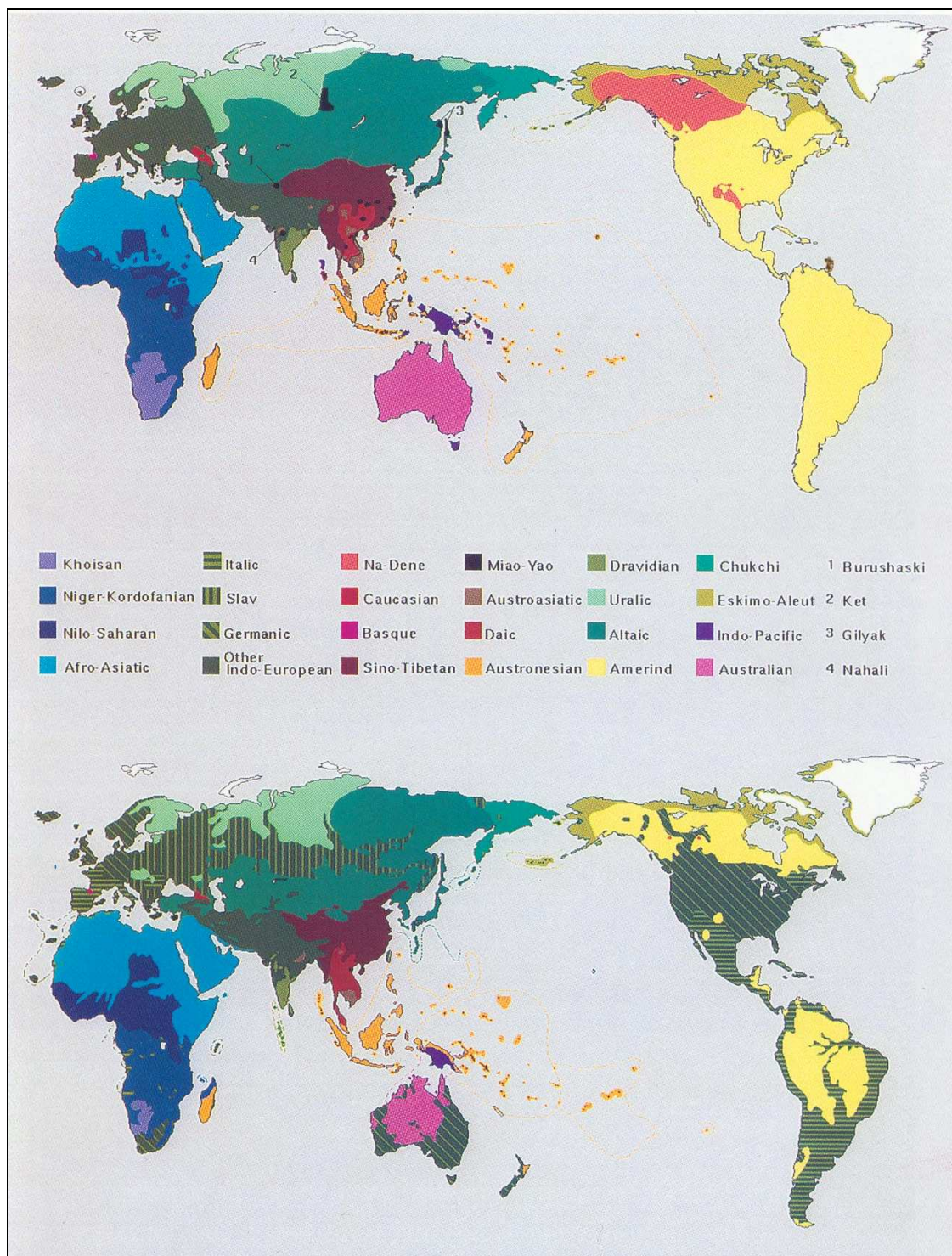


Figure 1 : Comparaison des arbres génétiques et des familles de langues. Les différentes couleurs définissent les 5 continents et les couleurs à l'intérieur des continents représentent les phylums linguistiques. (D'après Cavalli-Sforza L, 1994a)



**Figure 2 :** Distribution géographique des familles de langues selon Ruhlen M. La répartition initiale des langues est représentée en haut de la figure et la représentation actuelle en bas. (D'après Cavalli-Sforza L, 1994a)

### **III.2. Le polymorphisme génétique et la notion de groupes sanguins chez l'homme.**

La notion de groupe sanguin est littéralement définie sur deux termes qui sont « GROUPE » et « SANGUIN ». La notion de groupe est caractérisée par un ensemble d'individus ayant un ou plusieurs caractères en commun. Dans le cadre des groupes sanguins ce(s) caractère(s) en commun est (sont) exprimé(s) dans le sang. Bien entendu, il convient d'entendre par « caractères » des spécifications qui sont génétiquement transmises.

On peut donc définir la notion de groupe sanguin comme toute expression de la variabilité génétique humaine détectable dans le sang.

Cette notion extrêmement vaste recouvre à priori la totalité du polymorphisme humain qui représente l'ensemble des molécules exprimant une variabilité génétique dont l'immunogénicité conditionne son champ d'application et sa méthodologie d'analyse.

Le polymorphisme immunogène regroupe des antigènes portés par des cellules [érythrocytaires : groupes sanguins érythrocytaires (Figure 3 p.39), leucocytaires : groupes sanguins leucocytaires, plaquettaires : groupes sanguins plaquettaires, toutes cellules nucléées : groupe HLA] ou par des molécules libres dans le torrent circulatoire [immunoglobulines : marqueurs Gm (Figure 4 p.40), Am, Km, protéines du complément : groupes Chiddo Rodgers].

Son exploration est basée sur la mise en œuvre de techniques immunologiques : utilisation d'anticorps ; si l'anticorps se fixe, c'est que l'antigène est là et donc que l'allèle correspondant est présent. Son champ d'application intéresse surtout le médecin puisque le polymorphisme immunogène s'oppose à la transfusion, la greffe, la transplantation et même la grossesse incompatible. De même, ces marqueurs peuvent être la cible du propre système immunitaire de l'individu, base des pathologies auto-immunes.



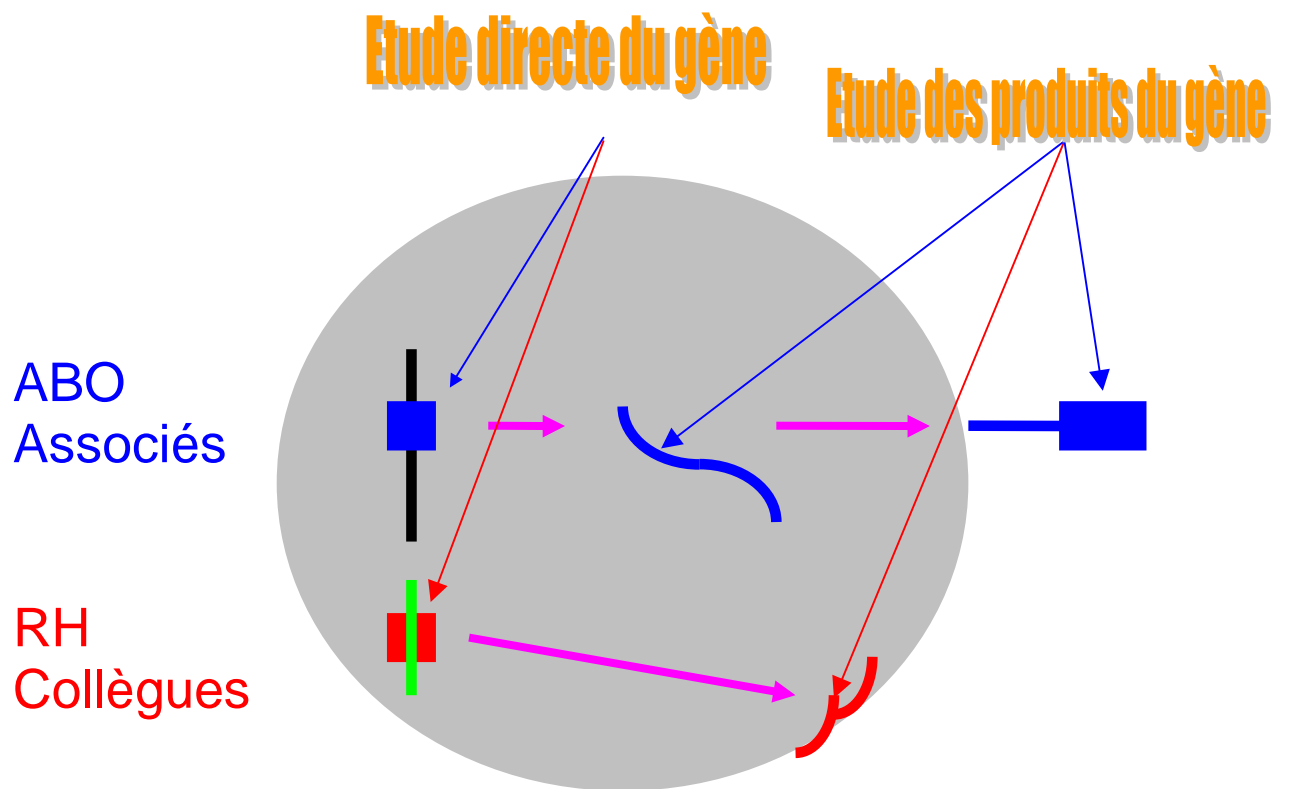


Figure 3 : Les niveaux d'analyse du polymorphisme génétique humain avec comme exemple ici les groupes érythrocytaires

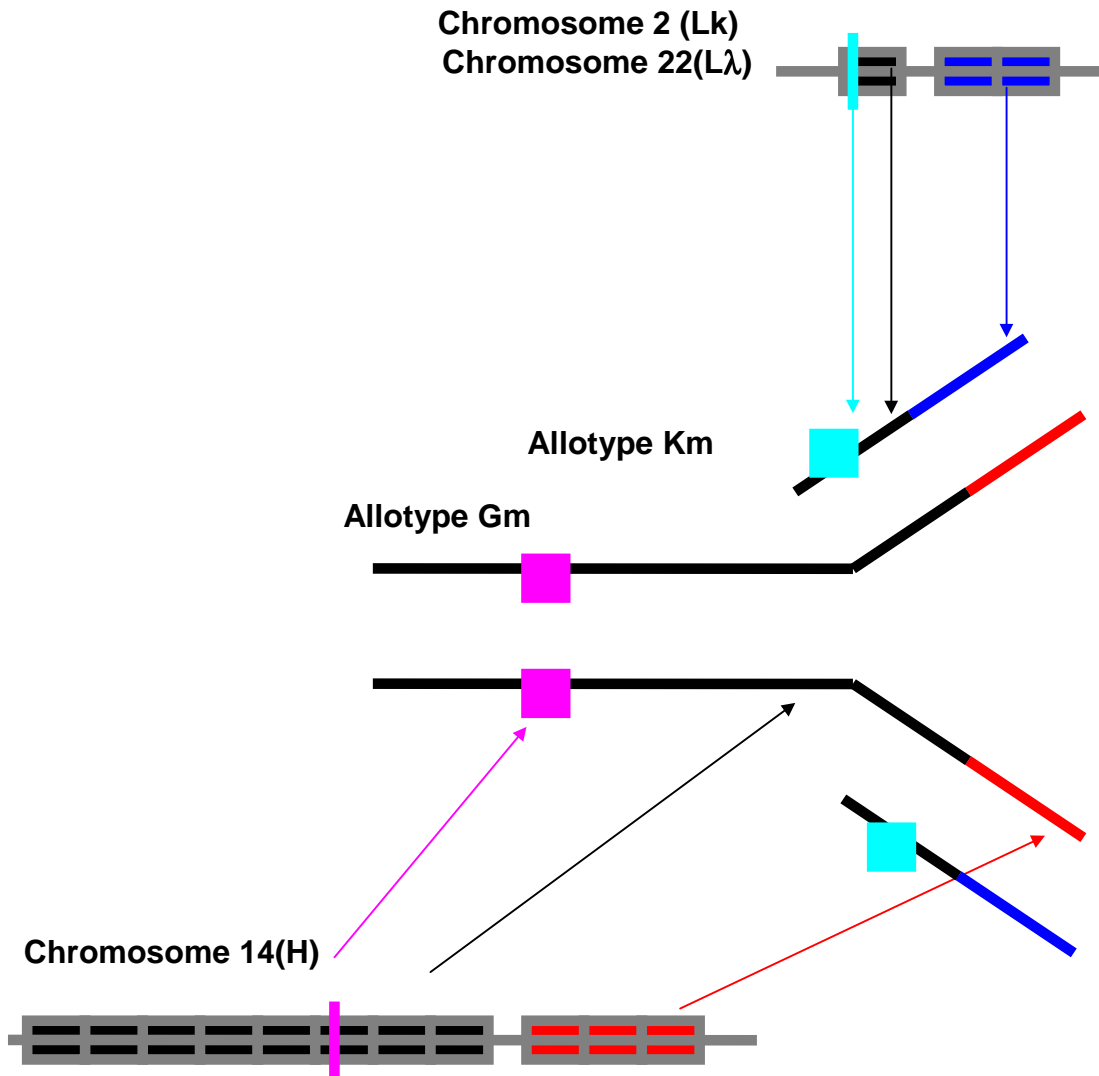


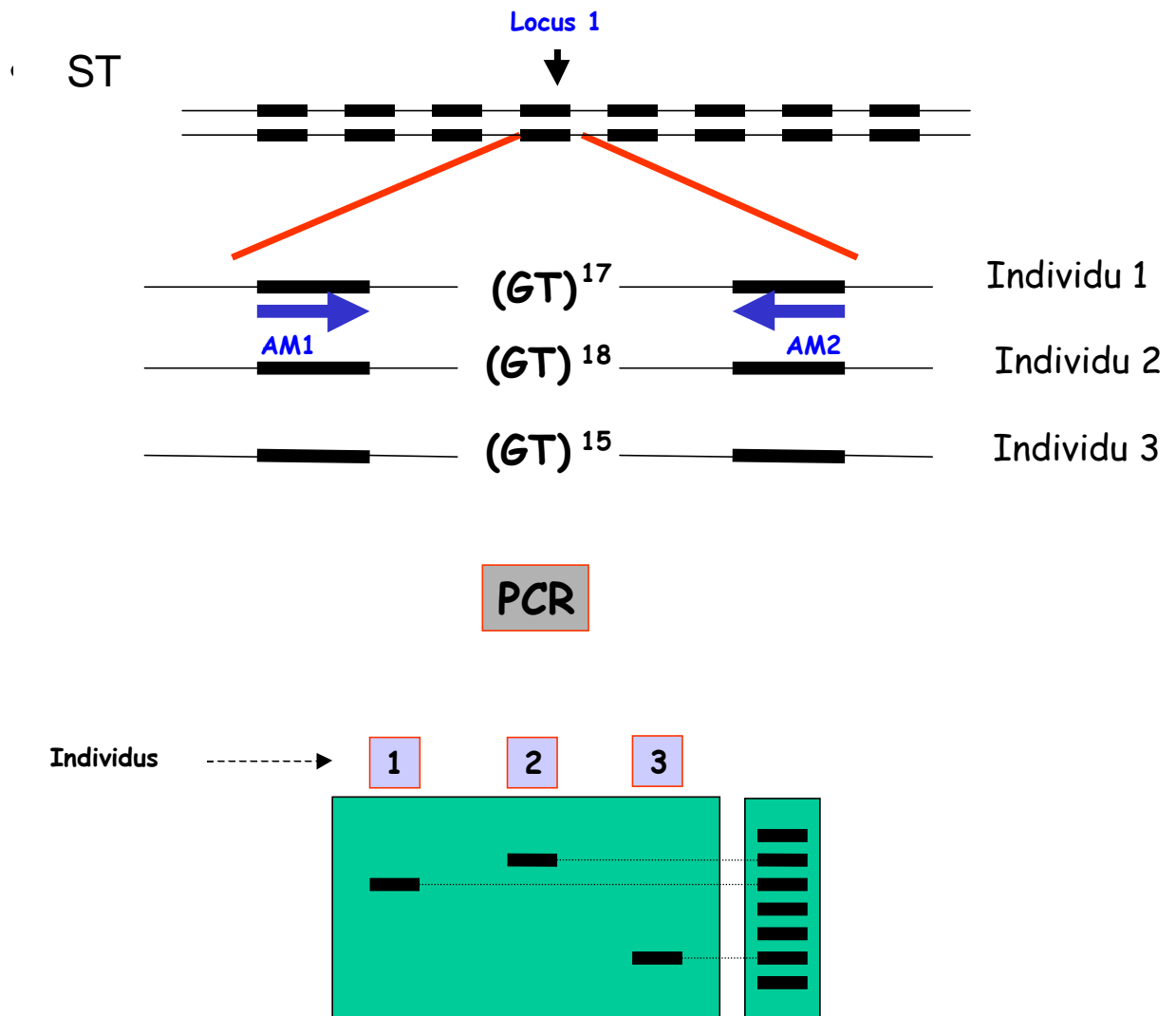
Figure 4 : Bases moléculaires du polymorphisme des Immunoglobulines G.

H : chaîne lourde  
 Lk : chaîne légère kappa  
 Lλ : chaîne légère lambda

Le polymorphisme non-immunogène regroupe des molécules plasmatiques (haptoglobine, alpha 2 antrypsinase...) ou intra-cellulaires (phospho-glucomutase érythrocytaire ...). L'absence d'immunogénicité impose le recours à des méthodes d'analyse de type biochimique basées notamment sur la propriété amphotère des protéines (iso-électro focalisation). Son champ d'application intéresse surtout le généticien, qui, en complément du polymorphisme immunogène, pourra définir génétiquement un individu, ou une population.

Ces deux types de polymorphisme permettent l'analyse de la variabilité génétique humaine par l'étude des produits des gènes. Cette exploration du polymorphisme humain peut être complétée par une analyse directe du matériel génétique reposant sur la mise en œuvre de techniques de biologie moléculaire. L'étude peut concerner alors des parties codantes ou non codantes de l'ADN.

L'analyse des parties codantes (Figure 3 p.39) vient compléter les données de l'analyse des produits des gènes en expliquant les bases moléculaires du polymorphisme phénotypique et en détectant des variants génétiques inaccessibles à l'observation phénotypique. L'étude des parties non codantes de l'ADN (Figure 5 p.42), les microsatellites par exemple, recherche un polymorphisme de répétition de séquences de nucléotides transmis selon les lois mendéliennes. Ces analyses directes du matériel génétique peuvent concerner non seulement l'ADN nucléaire (autosomes et chromosome Y), mais aussi l'ADN mitochondrial de transmission maternelle exclusive.



**Figure 5 :** Exploration des microsatellites (STR) humains (Représentation haploïdique). Dans la réalité chaque individu peut être représenté par deux bandes s'il est hétérozygote et une seule bande s'il est homozygote pour le locus considéré.

AM1 et AM2 désignent les amorces spécifiques. La séquence répétée est ici constituée de deux nucléotides ; G et T. Le nombre qui leur est affecté représente le nombre de répétitions sur le chromosome considéré de chaque individu.

L'étendue de ce polymorphisme, base fondamentale de capacité adaptatrice en terme d'évolution, avance l'idée que nous sommes, en dehors des jumeaux monozygotes, tous unique et évoque le concept de « solitude biologique ». Ainsi, avec quelques millilitres de sang le biologiste sera capable d'étudier près de 50 systèmes génétiques distincts et de donner la formule biologique qui caractérise chaque individu. Si pour un individu donné, on étudie chacun de ces systèmes de groupes sanguins et que l'on calcule après chaque système étudié, la probabilité de rencontrer un autre individu identique, on constate, en cumulant au fur et à mesure les différents systèmes, que cette probabilité s'amenuise. Au terme de la détermination de tous les groupes sanguins du globule rouge, des globules blancs, des plaquettes, des protéines plasmatiques, des enzymes du globule rouges et de l'ADN, il apparaît que cet individu n'existe qu'à un seul exemplaire sur 100 millions de milliards, chiffre vertigineux dépassant le nombre d'individus qui se sont succédés sur Terre depuis l'aube de l'humanité. Nous disposons donc là, comme le souligne Charles Salmon, de plus de 50 systèmes *d'étiquettes* pour définir avec une précision redoutable un individu et démontrer que le plus banal d'entre nous est un être unique (Salmon C, 1995).

De cette performance sont nées des applications en génétique humaine. Lors des études de filiation, réalisées dans le cadre de procédures judiciaires, les recherches de paternité permettent de ne pas confondre le vrai père avec un autre homme, accusé à tort, pris au hasard dans la population. De même, comme il est possible de définir un individu, il sera possible de définir le profil génétique d'une population et d'en déduire son histoire en termes de migration, d'invasion, d'isolement ou de consanguinité (génétique des populations). L'analyse de ces marqueurs a en effet engendré un ensemble de connaissances sur la génétique des populations humaines, donnant des informations précieuses sur la variation héréditaire dans notre espèce.

De ces différences est née une problématique de médecine transfusionnelle qui est celle de trouver des donneurs dont le sang est compatible avec celui des receveurs. Cette recherche de compatibilité pouvant s'avérer plus problématique en situation de transfusion « inter-populationnelle ».

Ainsi, cette « solitude » biologique par sa capacité *discriminatoire* est l'alliée du généticien et par l'impossibilité de trouver l'*identique*, l'ennemie du transfuseur.

*Nous constatons donc, au travers de ce panorama, que s'offrent à nous une multitude de marqueurs permettant d'accéder à l'analyse de la variabilité génétique individuelle et populationnelle. L'étude anthropogénétique d'une population doit donc débiter par une sélection des marqueurs génétiques et de leur méthodologie d'exploration.*

### **III.3. Bases de la sélection des marqueurs génétiques.**

#### **III.3.1. Généralités.**

Les critères de sélection des systèmes génétiques doivent prendre en compte l'étendue de leur polymorphisme qui conditionne leur capacité à différencier les populations humaines entre elles, la connaissance validée de leur modalité de transmission génétique, l'action éventuelle de la sélection naturelle sur certains d'entre eux, la possibilité de mettre en œuvre des méthodes d'exploration validées et maîtrisées, et enfin l'existence de données bibliographiques "exploitables" pour de nombreuses populations en vue de comparaison. Il est bien entendu possible d'y associer de nouveaux systèmes génétiques afin d'évaluer leurs possibilités d'utilisation en Anthropologie génétique.

Parmi les systèmes génétiques polymorphes aucun d'entre eux, pris isolément, n'est parfait. On peut classer globalement les systèmes polymorphes en quatre groupes. Le premier est représenté par les systèmes de groupes sanguins érythrocytaires. Compte tenu d'un polymorphisme moins étendu, ils peuvent apparaître moins performants que d'autres en terme de définition génétique d'une population. Toutefois, ceux-ci ayant été les premiers à avoir été utilisés dans cette discipline, ils représentent une information importante aussi bien d'un point de vue quantitatif que qualitatif en raison de la profondeur historique des échantillons analysés. Un deuxième groupe de systèmes est représenté par les systèmes HLA et Gm des Immunoglobulines. Ceux-ci allient aujourd'hui une capacité discriminante plus importante que les systèmes érythrocytaires ainsi que de nombreuses données disponibles pour un grand nombre de populations mondiales. Un troisième groupe est représenté par le polymorphisme du chromosome Y et de l'ADN mitochondrial. Tous deux sont de transmission uni-parentale ce qui permet, par l'étude des mutations qui se sont accumulées sans recombinaison, de reconstituer des phylogénies maternelle et paternelle. Ces deux systèmes extrêmement informatifs en Anthropogénétique manquent encore, surtout pour le chromosome Y, de données disponibles dans de nombreuses populations.

Enfin, un dernier groupe peut être représenté par des systèmes polymorphes, récemment décrits ou en cours d'exploration, dont l'évaluation vis à vis des problématiques Anthropologiques est nécessaire. L'un des exemples est représenté par le système des gènes *KIR*.

Quoiqu'il en soit, la performance de la définition de la structure génétique d'une population repose avant tout sur l'analyse de plusieurs systèmes génétiquement indépendants les uns des autres. Alors que les marqueurs classiques (Groupes sanguins érythrocytaires, protéines plasmatiques.) sont étudiés depuis presque 80 ans, un temps suffisant ne s'est pas encore écoulé pour permettre de recenser autant de données pour les marqueurs d'ADN (Cavalli-Sforza L, 1998).

Compte tenu des compétences et des capacités d'exploration de nos laboratoires, nous avons sélectionné quatre types de polymorphisme pour étudier notre échantillon Comorien. Deux sont classiquement utilisés depuis plusieurs années en Anthropologie. Il s'agit des groupes sanguins érythrocytaires et du système HLA. Si l'exploration moléculaire du polymorphisme érythrocytaire a nécessité une mise au point préalable, son exploration sérologique ainsi que l'exploration moléculaire du polymorphisme HLA classe II sont réalisées en routine à l'EFS Alpes-Méditerranée (EFS AM). Les deux autres sont représentés par le polymorphisme des gènes *KIR* et celui du chromosome Y. Bien que son exploration soit de plus en plus utilisée en Anthropologie, les techniques d'analyse du polymorphisme porté par le chromosome Y ont nécessité des mises au point en collaboration avec un laboratoire extérieur (Unité INSERM 491, Génétique moléculaire de la spermatogénèse, M.Mitchel) à l'EFS AM et son étude fait partie d'un travail de thèse plus global analysant ce polymorphisme dans plusieurs populations d'Afrique et du Moyen-Orient réalisé par Axel Ducourneau. En ce qui concerne le polymorphisme des gènes *KIR*, son utilisation en anthropologie est très limitée à ce jour compte tenu de l'absence de description complète de ses allèles. La mise au point technique, la description de son polymorphisme dans la population Comorienne et l'évaluation de sa



capacité à être utilisé en anthropologie génétique ont été réalisés dans le cadre d'un travail de DEA en Anthropologie (C.Frassati, 2003). Nous prévoyons, en collaboration avec des laboratoires extérieurs à notre unité, de réaliser l'exploration des deux autres systèmes précédemment cités à savoir les marqueurs Gm qui ont montré de grande corrélation avec les données linguistiques et l'ADN mitochondrial (JM.Dugoujon - Toulouse).

*Quels sont donc les éléments qui ont été pris en compte pour retenir chacun d'eux ?*

### **III.3.2. Les systèmes de groupes sanguins érythrocytaires.**

Six systèmes distincts (ABO, RH, KEL, FY, JK, MNS) ont été étudiés par des méthodes immunologiques. Ces systèmes ont été sélectionnés pour trois raisons ; (i) leur capacité, surtout lorsqu'ils sont cumulés, à définir différentes populations humaines, (ii) pour les nombreuses données bibliographiques existant dans de multiples populations et enfin (iii) pour l'existence possible d'une pression sélective sur certains d'entre eux, notamment pour le système Duffy (Hamblin MT, 2000, 2002). En effet, ce mécanisme semble impliqué dans l'apparition de variants caractéristiques de régions particulières comme les zones impaludées à *Plasmodium vivax* d'Afrique subsaharienne, ce qui les fait considérer comme des marqueurs pertinents pour l'étude des migrations des populations humaines. L'analyse sérologique nous apparaît ici nécessaire pour deux raisons principales. La première est liée aux données concernant les autres populations qui sont essentiellement obtenues avec cette méthodologie, la deuxième est basée sur la nécessité d'explorer le polymorphisme du produit du gène qui est le seul à être accessible à une éventuelle action de la sélection naturelle, nous permettant ainsi d'aborder la signification biologique de ce polymorphisme. En effet, le gène n'est en interface, avec les pressions environnementales, que par l'intermédiaire de la molécule qu'il code. Pour deux de ces systèmes (RH et FY), les données sérologiques ont été complétées par une exploration directe des gènes par des méthodes de biologie moléculaire afin de détecter des polymorphismes non ou difficilement accessibles à la sérologie ou afin d'expliquer certains phénotypes sérologiques.

L'exploration moléculaire du locus *FY* a pour objectifs d'analyser le mécanisme génétique d'un éventuel phénotype *Fy(a-b-)* qui est fréquent dans les populations originaires d'Afrique subsaharienne et de détecter des allèles difficilement décelables par des analyses purement sérologiques comme l'allèle *FYX*.

L'exploration moléculaire du système RH concerne la recherche d'allèles "rares" du locus *RHCE* et l'analyse des mécanismes génétiques aboutissant au phénotype RHD négatif dans ces régions. Cette exploration permettra tout d'abord de répondre à l'aspect biomédical de notre problématique puisque la présence en double dose de certains de ces allèles peut aboutir à des blocages transfusionnels surtout en situation migratoire. Cette exploration a aussi un objectif Anthropologique lié à l'existence exclusive de ces allèles dans les populations originaires d'Afrique subsaharienne (Noizat-Pirenne F, 2001) et à la possibilité de retrouver plusieurs mécanismes génétiques impliqués dans l'expression d'un même phénotype D négatif (Singleton, 2000). L'exploration des allèles rares du gène *RHCE* est en cours et l'analyse des bases moléculaires du phénotype D négatif dans la population comorienne représentera l'une de nos perspectives. L'exploration érythrocytaire a fait l'objet d'un article soumis à *Human Biology* (Annexe Article, Chiaroni J, 2003). La recherche des allèles "rares" du gène *RHCE* a fait l'objet d'une subvention de recherche allouée par le conseil scientifique de l'Etablissement Français du Sang en 2003 et sur une année.

### **III.3.3. Le polymorphisme du locus *HLA-DRBI*.**

Le système HLA constitue l'un des systèmes les mieux caractérisés tant au niveau populationnel que moléculaire ou fonctionnel. L'existence d'un polymorphisme extrêmement étendu et de déséquilibres de liaison constituent les principaux avantages de ce système. L'étendue de son polymorphisme ainsi que les nombreuses données bibliographiques disponibles au niveau mondial (123 populations) nous ont incité à analyser le locus *HLA-DRBI*. Son étude dans de nombreuses populations a d'ailleurs fait l'objet d'un travail de thèse au sein de notre Etablissement (M.Gibert, 2001) afin de développer une méthodologie (basée sur la fonctionnalité de la molécule et ses associations avec les pathologies) permettant d'analyser les effets de la sélection naturelle. Cette méthode permet ainsi d'explorer l'extrême polymorphisme du système HLA en distinguant les effets de la sélection naturelle de ceux de l'Histoire démographique et des migrations humaines. Ce polymorphisme a été exploré par des méthodes de biologie moléculaire. L'exploration du polymorphisme *HLA-DRBI* dans la population comorienne a fait l'objet d'un article actuellement soumis (M. Gibert, 2003).

### **III.3.4. Le polymorphisme des gènes *KIR*.**

Les études en génétique des populations ainsi que les analyses comparatives interspécifiques ont montré que les gènes *KIR*, comme les gènes *HLA classe I*, sont hautement polymorphes. Au sein des populations humaines, les gènes constituant les haplotypes *KIR* varient d'un point de vue quantitatif et fonctionnel. Ainsi, certaines combinaisons géniques sont retrouvées plus fréquemment que d'autres. Compte tenu de ces particularités et bien que son polymorphisme soit toujours en cours de description, il nous paraît intéressant d'évaluer la possibilité d'utiliser ce marqueur polymorphe en Anthropogénétique et d'analyser les résultats de son exploration au sein de la population Comorienne. L'exploration de ce polymorphisme a fait l'objet d'un mémoire de DEA et d'un article en cours de rédaction (C. Frassati, 2003).

### **III.3.5. Le polymorphisme du chromosome Y.**

La découverte d'un grand nombre de marqueurs spécifiques sur le chromosome Y, la bonne connaissance de sa structure et son séquençage complet ont ouvert un nouveau champ d'investigation. L'état haploïde de ce chromosome permet d'exploiter son polymorphisme dans les tentatives de reconstruction de structure de population en dépit de données populationnelles encore limitées. L'exploration du polymorphisme du chromosome Y est en cours dans le cadre d'une thèse de troisième cycle réalisée par Axel Ducourneau.

Nous présenterons, ici, uniquement des résultats préliminaires.

### **III.4. Description des marqueurs génétiques sélectionnés.**

#### **III.4.1. Les systèmes de groupes sanguins érythrocytaires.**

Les groupes sanguins ont été découverts au début du XX<sup>e</sup> siècle lorsque Landsteiner constata que le plasma de certains individus agglutinait les hématies d'autres individus. Durant les 45 années suivantes, uniquement ceux détectés par des anticorps agglutinants ont pu être étudiés (ABO, MNS, P). Avec le développement du test de Coombs, les découvertes des autres antigènes vont s'enchaîner pour aboutir aujourd'hui à 270 authentiques antigènes de groupes sanguins regroupés en 26 systèmes (Table 4 p.51), 5 collections, 13 séries 901 et 32 séries 700.

Nom du système	Symbole	Antigène	Molécules	Locus	Chromosome
ABO	ABO	4	Carbohydrates	ABO	9
MNS	MNS	43	GPA / GPB	GYP A / GYP B	4
P	P1	1	Carbohydrates	P1	22
Rh	RH	46	Protéine D et Protéine CcEe	RHD / RHCE	1
Lutheran	LU	18	IgSF	LU	19
Kell	KEL	24	Endopeptidase	KEL	7
Lewis	LE	6	Carbohydrates	FUT3	19
Duffy	FY	6	Récepteur de Chémokines	FY	1
Kidd	JK	3	Echangeur d'urée	SLC14A1	18
Diego	DI	21	Bande 3 / Echangeur d'anions	SLC4AE1	17
Yt	YT	2	Acétylcholinestérase	ACHE	7
Xg	XG	2	Glycoprotéines	XG / MIC2	X/Y
Scianna	SC	3	Glycoprotéines	SC	1
Dombrock	DO	5	ADP-Ribosyl Transférase	DO	12
Colton	CO	3	Aquaporin-1	AQP1	7
Landsteiner – Wiener	LW	3	ICAM-4 / IgSF / CD242	LW	19
Chido/Rodgers	CH/RG	9	C4A / C4B	C4A / C4B	6
Hh	H	1	CD173 Type 2 H	FUT1	19
Kx	XK	1	Protéine	XK	X
Gerbich	GE	7	GPC / GPD	GYP C	2
Cromer	CROM	10	CD55 / DAF C Regulator	DAF	1
Knops	KN	7	CD35 / CR1 C Regulator	CR1	1
Indian	IN	2	CD44	CD44	11
Ok	OK	1	CD147 / EMMPRIN / IgSF	CD147	19
Raph	RAPH	1	Glycoprotéine	MER2	11
John Milton Hagen	JMH	1	CDw108 / Semaphorin	SEMA7A	15

Table 4: Les systèmes de groupes sanguins érythrocytaires (les systèmes étudiés sont surlignés)

### **III.4.1.1. Le système ABO.**

#### **a) Génétique formelle.**

Le locus *ABO*, situé sur le chromosome 9, comporte trois allèles courants *A*, *B* et *O*. Les allèles *A* et *B* sont codominants. L'allèle *O* est récessif par rapport aux allèles *A* et *B*. Ainsi, on ne peut pas toujours déduire le génotype à partir du phénotype et le calcul des fréquences géniques repose sur une estimation basée sur la formule de Bernstein (modifiée par certains auteurs Yasuda N, 1968) adaptée à des systèmes à plusieurs allèles dont un est « silencieux ».

#### **b) Génétique moléculaire.**

Les allèles *A* et *B* diffèrent par 7 nucléotides dont 4 sont responsables de la substitution de 4 acides aminés dans la séquence de la glycosyl transférase (Yamamoto M, 1990a, 1990b). L'allèle *O* (*O<sup>I</sup>*) possède une séquence identique à l'allèle *A<sup>I</sup>* avec une délétion débutant au nucléotide 261, aboutissant à une enzyme tronquée sans site catalytique. Les glycosyl transférases synthétisées par ces allèles sont les produits primaires des gènes (Figure 3 p.39). Dans l'érythrocyte, de manière synergique et séquentielle, ces enzymes complètent l'action des gènes du système H pour aboutir à la synthèse des antigènes A et B. Si l'allèle *H* (système Hh) est défaillant, les antigènes A et B ne peuvent être synthétisés même si les allèles *A* et *B* sont présents. Ce phénotype, caractérisé par un déficit en antigène H et donc en antigènes A et B, est le phénotype Bombay dont la fréquence en Inde de l'Ouest est estimée à 1/6000. Compte tenu de la présence de communautés Indiennes, ce type de phénotype est aussi retrouvé à la Réunion, où il coexiste avec un autre phénotype H déficient de type « Réunionnais » caractérisé par un affaiblissement de l'allèle H (Lependu J, 1983). Ces deux phénotypes ne semblent pas liés d'un point de vue phylogénétique.

### **c) Les aspects phénotypiques.**

Le phénotype ABO est caractérisé par le type d'antigène présent sur les hématies et les anticorps naturels présents dans le plasma. Quatre phénotypes de base sont ainsi définis. Le phénotype A caractérisé par la présence de l'antigène A (GalNac), le phénotype B par la présence de l'antigène B (Gal), le phénotype AB par la présence des deux antigènes et enfin le phénotype O caractérisé par l'absence de ces deux antigènes.

L'étude des groupes ABO, pour des besoins transfusionnels, démontra très tôt que leur répartition variait en fonction des peuples. Ainsi, en Europe, on trouve en moyenne 45% de A, 43% de O, 9% de B et 3% de AB. L'étude de la répartition des groupes chez les Indiens d'Amérique, révèle qu'à l'exception de certaines tribus du Canada, ils sont tous de groupes O. Deux hypothèses peuvent être formulées pour expliquer cette absence de groupe A et B sur le continent Américain. La première est basée sur un effet fondateur lié au fait qu'un petit groupe d'Homo Sapiens tous de groupe O ont franchi le détroit de Béring il y a 15 000 ans. La deuxième hypothèse était que la sélection naturelle avait fait disparaître les autres groupes. Malgré tout, à l'exception des Amérindiens, les fréquences des groupes ABO sont assez constantes dans le monde et varient moins d'un endroit à l'autre que d'autres gènes (Cavalli-Sforza L, 1994b).

### **d) Les aspects fonctionnels et évolutifs.**

Les gènes ABO ont été bien conservés au cours de l'évolution (Kominato Y, 1992, Saitou N 1997). Ils présentent 95% d'homologie de séquence avec les autres Primates. De plus, les substitutions différentielles entre A et B sont survenues avant la séparation des lignées aboutissant à l'Homme, au Chimpanzé, au Gorille et à l'Orang-Outan (Kominato Y, 1997). Ce polymorphisme semble donc très ancien. De même, le gène *ABO* de la souris présente une organisation similaire, mais abouti à la synthèse d'une enzyme ayant une double activité A et B transférase (Yamamoto M, 2001).

De nombreuses études ont rapporté un lien statistique entre le polymorphisme ABH et la susceptibilité aux maladies.

D'un point de vue synthétique, les sujets de groupe A présentent une susceptibilité aux infections bactériennes, alors que ceux du groupe O aux infections virales. Les cancers sont plus fréquemment associés aux sujets de groupe A (*des sujets authentiquement O expriment des antigènes A sur les cellules tumorales gastriques, dont le but est l'élimination par les anticorps naturels anti-A présents. Chez le sujet de groupe A, ce mécanisme protecteur ne peut être mis en œuvre*) alors que les pathologies auto-immunes et hémorragiques aux sujets de groupe O (Henry, 2000).

Enfin, l'incompatibilité fœto-maternelle ABO possède une action protectrice vis à vis de l'allo-immunisation maternelle RH1 (D) grande pourvoyeuse de Maladie Hémolytique du Nouveau Né.

D'un point de vue fonctionnel, le rôle des antigènes ABH est inconnu. De nombreuses molécules membranaires expriment les antigènes ABH :  $10^6$  molécules transporteurs d'anions (Bande 3),  $10^5$  molécules transporteurs de glucose,  $10^6$  glycolipides (Viitala J, 1985). L'ensemble de ces molécules participe à l'organisation du glycocalix membranaire protégeant l'hématie contre des agressions mécaniques ou microbiennes (Viitala J, 1985 ; Koscielak J, 1986).

Si la signification biologique du polymorphisme A/B demeure un mystère, son apparition il y a environ 13 MA et son maintien au cours du temps, est probablement en rapport avec une susceptibilité aux infections microbiennes en particulier (Martinko JM, 1993).



### **III.4.1.2. Le système Rh.**

Le système Rh est le système de groupe sanguin le plus complexe. Il comporte 46 antigènes de RH1 à RH53 dont 7 sont obsolètes.

#### **a) Génétique formelle.**

Les antigènes Rh sont codés par deux gènes homologues étroitement liés sur le bras court du chromosome 1. Le gène *RHD* produit la protéine RhD porteuse de l'antigène D et le gène *RHCE* produit la protéine RhCE porteuse, en fonction des formes alléliques, des antigènes C ou c et E ou e (Figure 6 p.56). Compte tenu de l'absence de recombinaison, les deux gènes, donc les trois antigènes, sont transmis en bloc sous forme d'haplotypes qui sont notés *DCE*, *DcE*, *dce*, *Dce*, *dCe*, *dcE*, *DCE* et *dCE* où *d* représente l'allèle *RHD* en délétion ou inactif. Au niveau de ces haplotypes, les formes alléliques *C* et *c* d'une part et *E* et *e* d'autre part sont codominantes. Conformément à l'analyse de Fischer, 8 haplotypes peuvent être identifiés avec des fréquences variables d'une population à l'autre (Mourant AE, 1976). En Europe l'haplotype *r* est plus fréquent au Nord et *R2* au Sud. L'haplotype *R0* caractérise l'Afrique subsaharienne. En Asie de l'Est, dans le Pacifique et en Amérique les haplotypes manquant du gène *RHD* sont exceptionnels. L'haplotype *ry* est extrêmement rare dans toutes les populations (Table 5 p.57).

Ces huit haplotypes peuvent être appariés deux à deux pour former  $(8/2) (8+1) = 36$  génotypes. Toutefois, en utilisant les réactifs de routine, anti-D, -C, -E, -c, -e, seulement 18 phénotypes différents peuvent être distingués. Seuls 8 de ces phénotypes n'ont qu'un génotype possible les 10 autres présentent 2, 3 ou 6 génotypes possibles. Parmi ces génotypes, l'un est le plus probable (statistiquement le plus fréquent), les autres sont les possibles. Il est théoriquement possible de définir le génotype avec l'utilisation de réactifs spécifiques des antigènes composés (Table 6 p.56). Le calcul des fréquences haplotypiques est donc complexe et peut être basé sur une méthode de maximum de vraisemblance comme l'a fait Fischer (Fisher RR, 1946 a et b, 1947) ou sur d'autres méthodes (Mourant AE, 1976).

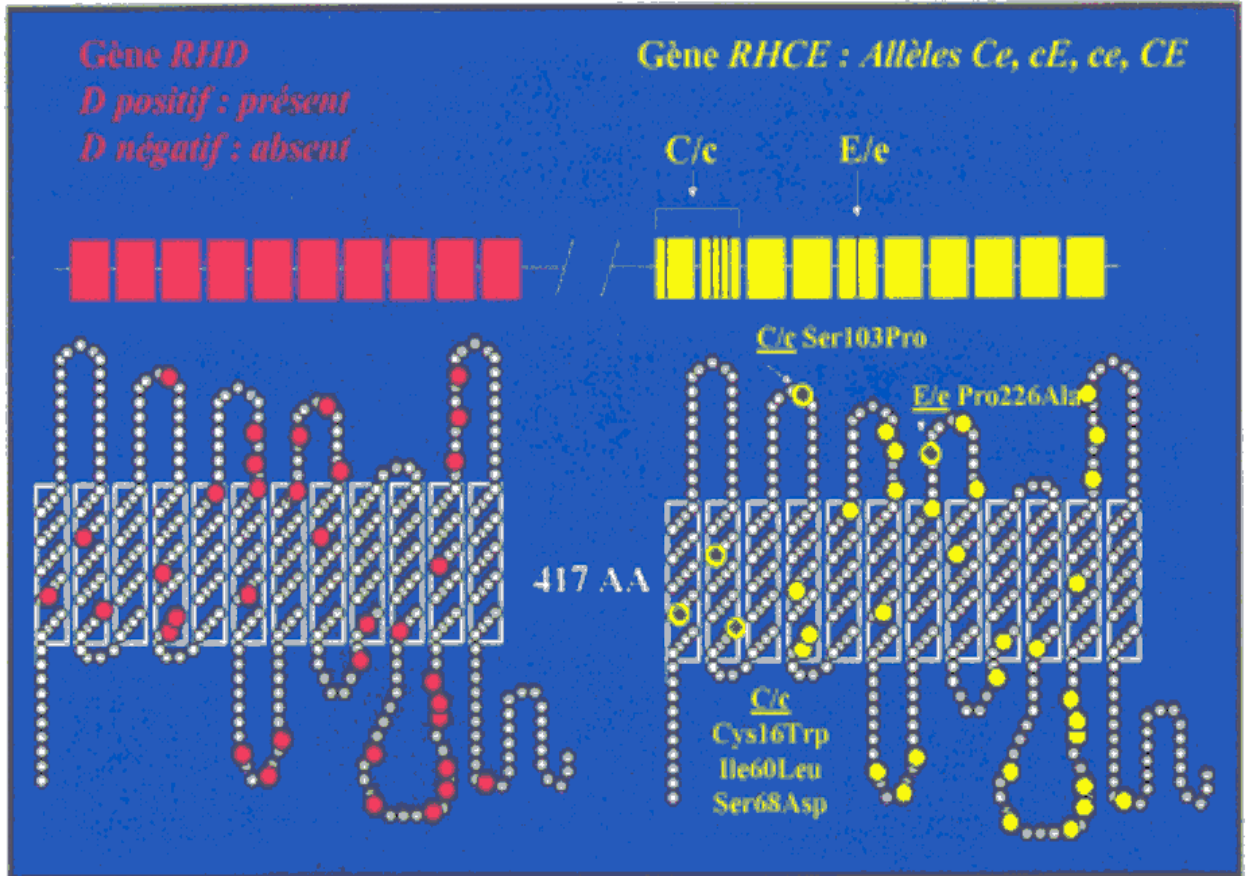


Figure 6 : Organisation des gènes *RHD* et *RHCE* et structure des molécules correspondantes (Noizat-Pirenne F, 2002).

CDE	Fischer	English (Race RR, 1948)	Nigerian (MourantAE, 1976)	Chinese (Mackay J, 1969)
<i>DCe</i>	$R^1$	0.4205	0.0602	0.7298
<i>Dce</i>	$R$	0.3886	0.2028	0.0232
<i>DcE</i>	$R^2$	0.1411	0.1151	0.1870
<i>Dce</i>	$R^0$	0.0257	0.5908	0.0334
<i>DcE</i>	$r''$	0.0119	0	0
<i>DCe</i>	$r'$	0.0098	0.0311	0.0189
<i>DCE</i>	$R^Z$	0.0024	0	0.0041
<i>DCE</i>	$r^y$	0	0	0.0036

Table 5 : Fréquence des différents haplotypes *RH* (Daniels G, 2002)

1 Phénotype Sérologique					6 Génotypes Possibles	1 Génotype Probable	Antigènes Composés				
-D	-C	-E	-c	-e			ce	Ce	CE	Ce	
					<i>DCe/DcE</i>	<i>R1/R2</i>	<b>11.87</b>	<b>0</b>	+	<b>0</b>	+
+	+	+	+	+	<i>DCe/dcE</i>	<i>R1/r''</i>	1	<b>0</b>	+	<b>0</b>	+
					<i>DcE/dCe</i>	<i>R2/r'</i>	0.28	<b>0</b>	+	<b>0</b>	+
					<i>DCE/dce</i>	<i>RZ/r</i>	0.19	+	<b>0</b>	+	<b>0</b>
					<i>Dce/DCE</i>	<i>R0/RZ</i>	0.01	+	<b>0</b>	+	<b>0</b>
					<i>Dce/dCE</i>	<i>R0/ry</i>	<0.01	+	<b>0</b>	+	<b>0</b>

Table 6 : Exemple de phénotypes et génotypes *RH* (Daniels G, 2002)

## b) Génétique moléculaire.

Les deux gènes comportent 10 exons et possèdent 93.8% d'homologie. Les différences les plus notables sont situées dans l'intron 4 où *RHD* a une délétion de près de 600 pb par rapport à *RHCE* (Arce MA, 1993; Okuda H, 1997) et dans la région 3' de l'exon 10 où le *RHD* possède une séquence spécifique (Figure 7 p.62). Ces différences entre le gène *RHD* et le gène *RHCE* représentent la base du génotypage *RHD*. Compte tenu des bases moléculaires, la détection de la présence ou de l'absence du gène *RHD* peut reposer sur :

- La recherche de la séquence de la région 3' de l'exon 10 spécifique du gène *RHD* où l'obtention d'un amplifiat signe la présence du gène alors que la non-détection du produit atteste de son absence.
- L'amplification d'une région de l'intron 4 qui donne un produit de 600 pb pour le gène *RHD* et 1200 pb pour le gène *RHCE*. L'absence du produit de 600 pb signant l'absence du gène *RHD*.
- Enfin, l'amplification d'une région de l'exon 7 commune aux gènes *RHD* et *RHCE* sert de témoins.

Chacune de ces méthodes, détectant une seule région du gène *RHD*, peut donner des résultats faussement négatifs et parfois faussement positifs en cas de variant :

- Les hématies DVI et DVa, sont D positif, alors que manquant de l'intron 4 de *RHD*, elles donneront un résultat négatif.
- Le gène hybride *RHCE-D-CE* responsable de DBT, manquant de l'exon 7, aboutira à une réaction négative avec les techniques utilisant cet exon.
- Le gène hybride *RHD-CE-D-* associé au complexe *d(C)ce<sup>s</sup>* ne produit pas de D sur l'hématie mais possède l'exon 10 dérivé de *RHD* (risque de faux positif).
- Le pseudogène *RHD $\Psi$*  des populations Africaines possède un gène *RHD* qui est inactif (risque de faux positif).

Ainsi, la technique décrite par Singleton (Singleton BK, 2000) permet de détecter, lors d'une PCR multiplex, l'exon 7, l'exon 10 et l'intron 4 du gène *RHD* et le pseudo gène *RHDΨ*. Toutes ces méthodes ne permettent pas, toutefois, de distinguer un individu homo- ou hétéro-zygote pour le gène *RHD*.

### **Bases moléculaires du polymorphisme D.**

*Voir Annexe 5 p.224*

En Europe, les 2 gènes homologues (*RHD* et *RHCE*) sont présents chez le sujet D positif, alors qu'il n'en existe qu'un (*RHCE*) chez le sujet D négatif. L'homozygotie de la délétion est le mécanisme majeur du phénotype D négatif chez les sujets Européens. Par contre, dans les populations Africaines trois mécanismes moléculaires aboutissant au phénotype D négatif sont possibles :

- Le mécanisme courant (67%) repose sur la présence (Aubin JT, 1997; Singleton BK, 2000) d'un gène *RHD* inactif, nommé pseudogène *RHDΨ*.
- Un autre mécanisme (15%) est lié à la présence d'un gène hybride *RHD-CE-D<sup>s</sup>* (Faas BHW, 1997; Daniels G, 1998) qui ne produit pas de D, mais produit un C anormal.
- Enfin, dans 18% des cas, on note un mécanisme de délétion du gène *RHD* tel qu'il a été décrit chez les Européens.

### **Bases moléculaires des polymorphismes C/c et E/e.**

*Voir Annexe 5 p.220*

Le polymorphisme C/c est associé à la substitution de 6 nucléotides entre les deux allèles aboutissant à 4 substitutions d'acides aminés : Cys16Trp (exon1), Ile60Leu, Ser68Asn et Ser103Pro (exon2).

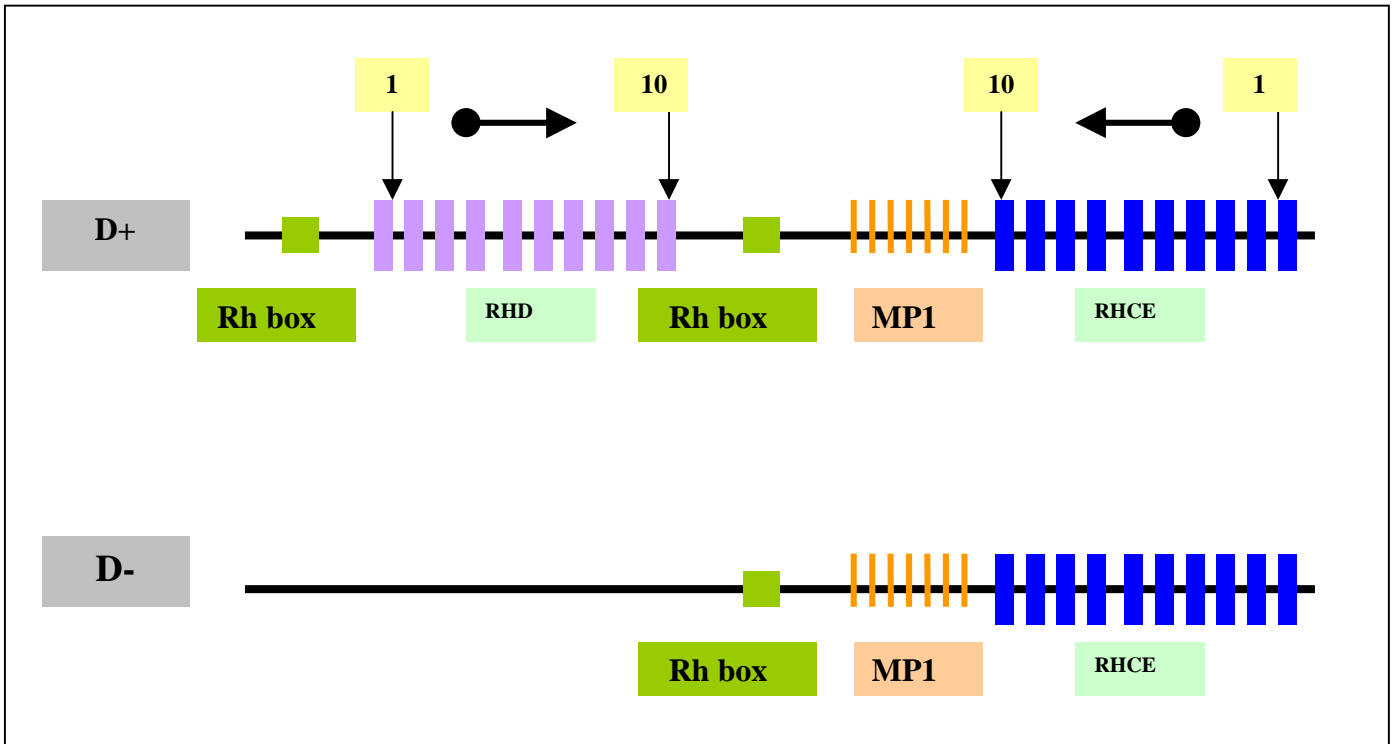
Seule la substitution en 103 est extra-membranaire sur la deuxième boucle. Le polymorphisme E/e est associé à la substitution C676G dans l'exon 5 aboutissant à Pro226Ala dans la quatrième boucle extra-cellulaire de la protéine CcEe. La détermination du statut c et E des fœtus par PCR a été décrite par Le Van Kim en 1994.

C et e sont des antigènes sensibles à la conformation :

Ala226 est aussi synthétisé par *RHD* qui n'exprime pas e.

Ser103 est aussi synthétisé par *RHD* mais les anticorps capables de se fixer sur les deux déterminants (Ser 103) ne sont pas des anti-C mais des anti-G

La Pro226 de E aboutit à un affaiblissement de C en position cis dans le rare haplotype *RZ*.



**Réarrangement génomique des gènes *RHD* et *RHCE*.**

Des mécanismes de conversion géniques entre les deux gènes homologues aboutissent à la création de gènes hybrides *RHD-CE-D* et *RHCE-D-CE*. Les segments échangés peuvent être de tailles différentes. De plus ces phénomènes de conversion sont souvent associés à des mutations. *RHD* et *RHCE* étant en orientation opposée, il semble probable que ces phénomènes de génération de gènes hybrides se fassent entre les deux gènes en cis.

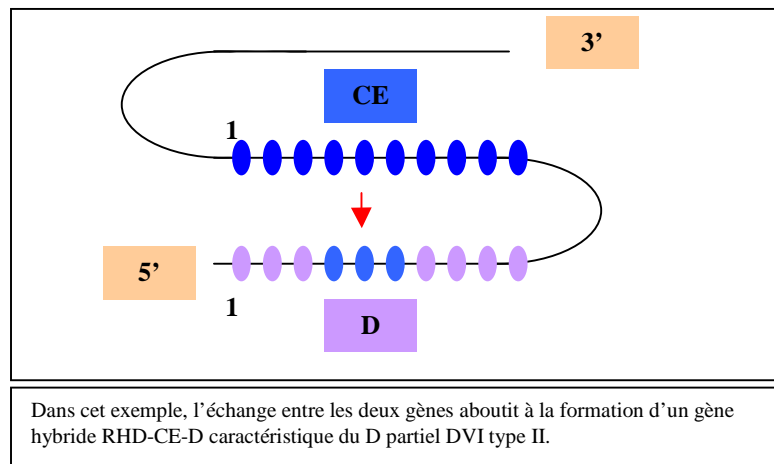


Figure 7 : Organisation et bases du polymorphisme des gènes *RH* (Daniels G, 2002).

### c) Les aspects phénotypiques.

Ces antigènes sont portés par des protéines hautement hydrophobes, non glycosylées, qui s'insinuent dans la membrane érythrocytaire à 12 reprises (Figure 6 p.56).

**Le phénotype D positif** est lié à la présence de la protéine RhD. Le phénotype D négatif est lié à son absence sur la membrane érythrocytaire. La proportion D négatifs / D positifs présente un gradient Est → Ouest.

**Le phénotype D négatif** atteint en moyenne 15% en Europe de l'Ouest, alors qu'il est presque inexistant en Extrême Orient.

**Les antigènes C et c d'une part et E et e d'autre part** représentent deux paires d'antigènes antithétiques dont le polymorphisme est contrôlé par le gène *RHCE*. Ils sont portés par la protéine RhCE. En Europe, l'antigène C est présent chez 68% des sujets et l'antigène c chez 81% ce qui abouti à des fréquences géniques de l'allèle C à 43.3% et c 56.7%. En Afrique, la fréquence de l'antigène c est beaucoup plus élevée que l'antigène C. En extrême Orient, l'antigène C approche les 100% et l'antigène c est un antigène de faible fréquence. Dans toutes les populations, l'antigène e apparaît significativement plus fréquent que l'antigène E avec des fréquences géniques assez homogènes de l'ordre de 15.5% pour l'allèle E et 84.5% pour l'allèle e (Fisher RA, 1946b).

**Le phénotype Rh<sub>null</sub>** n'exprimant aucun antigène du système RH est lié à deux mécanismes. L'un est lié à la présence en double dose d'une délétion *RHD* associée à un gène *RHCE* inactif, l'autre repose sur la présence en double dose, sur le chromosome 6, d'une mutation empêchant la synthèse d'une protéine RhAG nécessaire à la stabilisation membranaire du complexe RH. De plus, les hématies Rh<sub>null</sub> présentent des anomalies d'expression d'autres antigènes n'appartenant pas au système RH ainsi que des anomalies membranaires aboutissant à une diminution de la durée de vie des hématies.



**Les allèles “rares” du gène *RHCE*** sont l’apanage des populations d’Afrique subsaharienne.

- Certains phénotypes RH rares sont caractérisés par l’absence d’antigènes de grande fréquence (Noizat-Pirenne F, 2002).

**Le Phénotype RH:-46** (Figure 8 p.65) D’un point de vue phénotypique, les sujets sont apparemment D CC ee, avec un affaiblissement des antigènes RH2 et RH5, et limité aux ethnies Peul (Le Pennec PY, 1989). L’anti-RH46 reconnaît toutes les hématies exprimant les produits du gène *RHCE*. D’un point de vue moléculaire, il s’agit d’un allèle hybride (Ce-D(4)-Ce) qui peut, à l’état homozygote produire ce phénotype. Une PCR-ASP (allèle spécifique) détectant un gène *RHCE* hybride peut détecter cet allèle. Le dépistage de cet allèle n’est pratiqué que chez les sujets exprimant l’antigène C (RH2) et e (RH5).

**Le Phénotype RH:-18** (Figure 9 p.66) D’un point de vue phénotypique, les sujets sont apparemment D cc ee avec ou sans affaiblissement de RH5 (e). En cas d’allo-immunisation, l’anti-RH18 présente un profil d’anticorps anti-public reconnaissant toutes les hématies portant un produit du gène *RHCE*. Après absorption sur des hématies R2R2 (DCcEe), on retrouve une apparente spécificité anti-e ou anti-ce. Par ailleurs, cet anti-RH18 peut ressembler, en début d’immunisation, à un anti-e et peut être fréquemment associé à un anti-D. D’un point de vue moléculaire, trois allèles *RHce* (ceAR, ceEK et ceBI) peuvent, à l’état homozygote ou hétérozygote composite, produire ce phénotype. Ces trois allèles ont en commun une conversion génique en position 712 de l’exon 5 de l’allèle *RHce*. Une PCR-ASP détectant cette mutation permet de dépister ces 3 allèles. Si cette PCR-ASP de dépistage est positive un séquençage de l’exon 5 est nécessaire pour préciser l’allèle présent. Le dépistage de ces allèles n’est pratiqué que chez les sujets exprimant les antigènes c (RH4) et e (RH5).

**Le Phénotype RH:-34** (Figure 10 p.67) D'un point de vue phénotypique, les sujets sont apparemment dd Cc ee avec ou sans affaiblissement de RH3 (C). En cas d'immunisation, l'anti-RH34 reconnaît toutes les hématies exprimant un produit du gène *RHCE*. Après absorption sur des hématies R2R2, on retrouve un anti-C ou un anti-Ce. Un haplotype (C)*ce*<sup>s</sup> peut, à l'état homozygote produire ce phénotype. Cet haplotype est caractérisé par un gène *RHD* hybride (*D-CE(3-8)-D*) ne produisant pas d'antigène D (RH1) mais un antigène C (RH2) partiel et un allèle *RHce* muté en 2 positions : 733 (conversion génique) et 1006 (mutation ponctuelle sur un site non polymorphe). Le dépistage de cet haplotype est basé sur la réalisation de 3 PCR-ASP qui détectent respectivement le gène *RHD* hybride, la mutation 733 et la mutation 1006. Le dépistage de cet haplotype n'est pratiqué que chez les sujets exprimant l'antigène C (RH2) et c (RH4) et e (RH5). Ces phénotypes, en cas d'allo-immunisation obstétrico-transfusionnelle peuvent être associés à la production d'anticorps (anti-RH18, anti-RH34 et anti-RH46) hautement significatifs d'un point de vue clinique dont certains ont été responsables de décès de patients (Noizat Pirenne F, 2002).

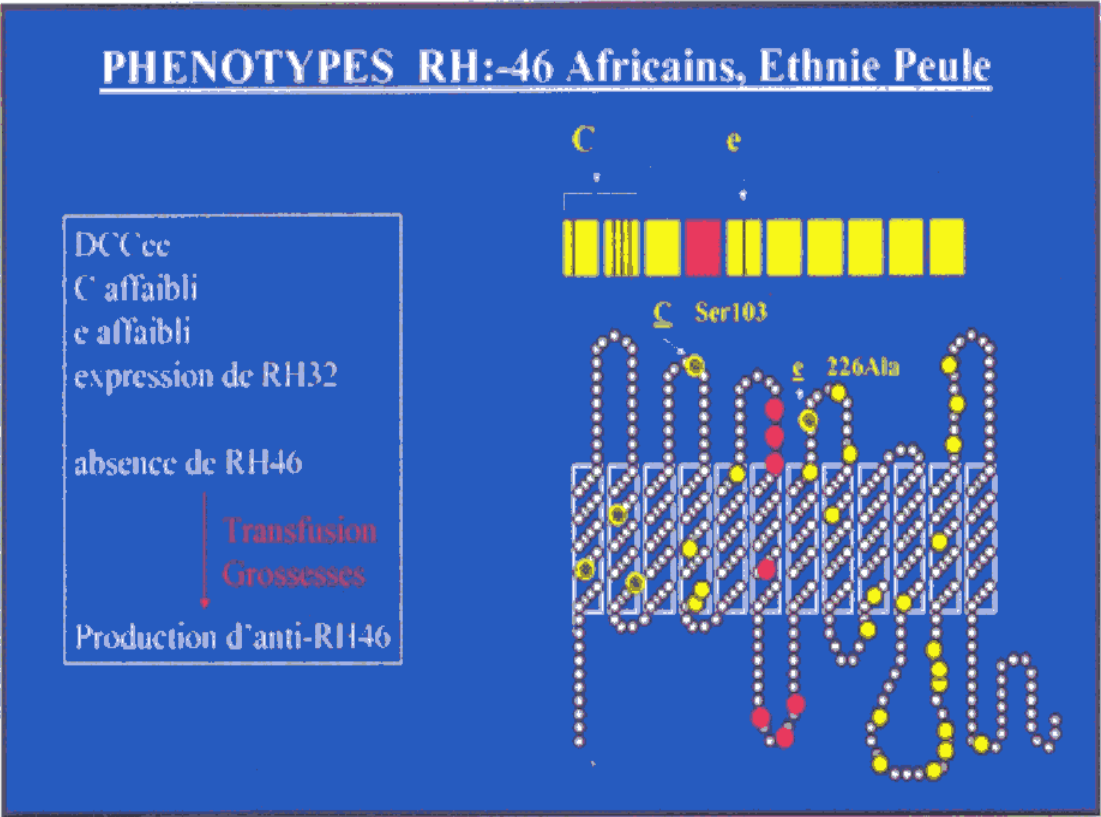


Figure 8 : Bases moléculaires du phénotype RH : -46 (Noizat-Pirenne, 2002)

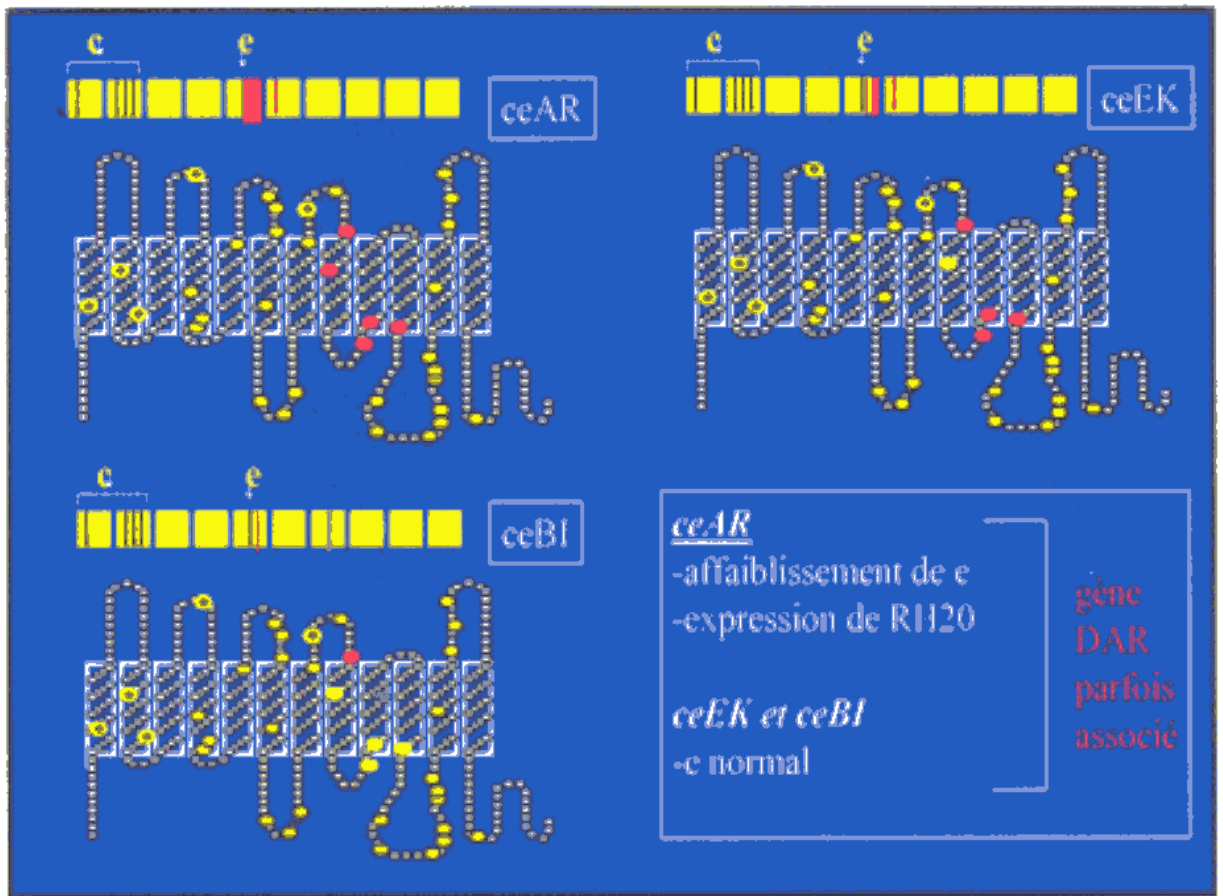


Figure 9 : Bases moléculaires du phénotype RH : - 18. (Noizat-Pirenne F, 2002).

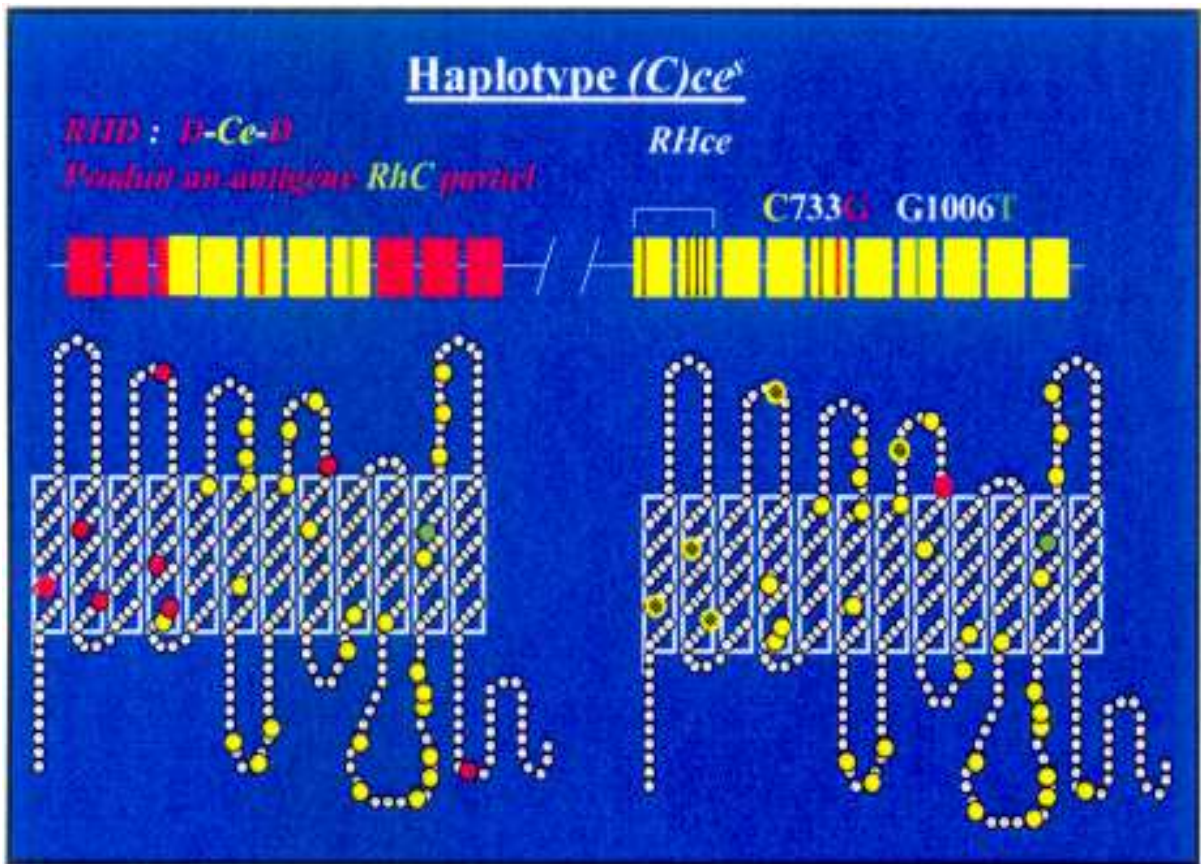


Figure 10 : Bases moléculaires du phénotype RH : - 34 (Noizat-Pirenne, 2002)

- D'autres phénotypes RH sont caractérisés par des variants de l'antigène RH5 aboutissant à des antigènes partiels. L'allèle *ceMO* (Figure 11 p.69) est caractérisé par une mutation de type de conversion génique dans l'exon 5 en position 667. Cet allèle est dépisté par une PCR-ASP spécifique de la mutation 667. L'allèle *ce<sup>s</sup>(340)* (Figure 11 p.69) est caractérisé par une mutation en position 340 dans l'exon 3 et en position 733 dans l'exon 5 par conversion génique. Cet allèle est dépisté par 2 PCR-ASP qui détecte respectivement les mutations 733 et 340. Comme dans le cas des antigènes D (RH1) partiels, l'allo-immunisation dirigée contre l'épitope manquant est possible (Issit P, 1994 ; Noizat Pirenne F, 2001).

Enfin, ces populations sont caractérisées par une présence significative d'antigènes de basse fréquence. Deux antigènes sont caractéristiques : RH10 et RH20. La mutation en 733 dans l'exon 5 d'un allèle *RHce* est associée à l'expression des antigènes de basse fréquence RH10 et RH20. La PCR-ASP détectant cette mutation n'est réalisée que chez les sujets exprimant l'antigène e (RH5). Bien que les conséquences transfusionnelles liées aux antigènes de faible fréquence ne soient pas du même ordre que celles liées à l'absence d'un antigène de grande fréquence, leur prise en compte peut s'avérer nécessaire en contexte de transfusion "intra-populationnelle" (voir Discussion).

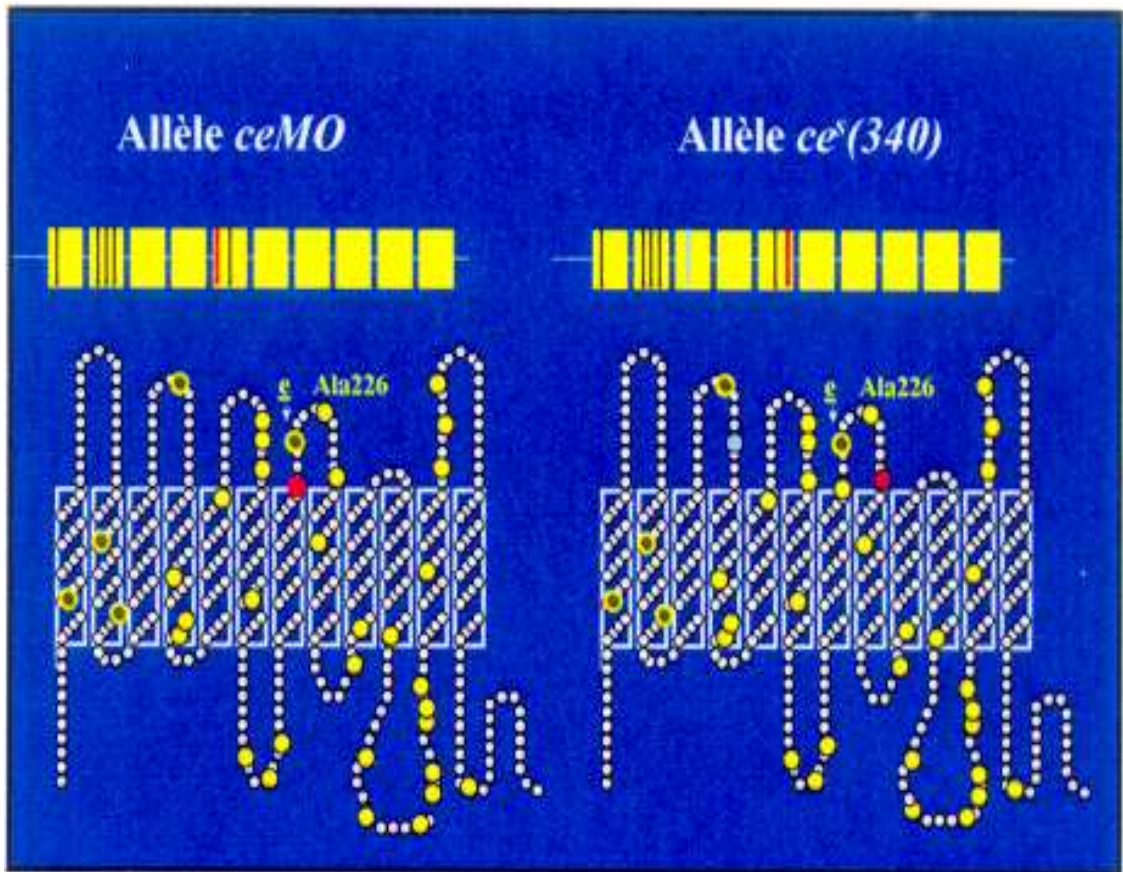


Figure 11 : Bases moléculaires des antigènes RH5 partiels (Noizat-Pirenne F, 2002)

#### **d) Les aspects évolutifs.**

La présence de la protéine RhAg dans la membrane érythrocytaire conditionne l'expression des antigènes Rh. Le gène contrôlant l'expression de la protéine RhAg est localisé sur le chromosome 6. Ce gène est très similaire aux gènes *RHD* et *RHCE* (Matassi G, 1998; Huang CH, 1998; Liu Z, 2000). Les ancêtres des gènes de la famille Rh apparaissent être primitifs, homologues non érythroïdes de *RHAG* et largement répandus dans d'autres espèces animales. Les gènes ancestraux *RH* ont été formés par duplication d'un gène ancestral *RHAG Like*. Les analyses d'homologie entre *RHD* et *RHAG* chez l'Homme, le Macaque, la Souris et le Rat évoquent une pression sélective sur les deux gènes avec toutefois un gène *RHAG* plus conservé que les gènes *RH* qui ont évolué trois fois plus vite (Kitano T, 1998; Matassi G, 1999) suggérant que *RHAG* possède une signification fonctionnelle plus importante. La duplication d'un gène *RHAG like* en un gène *RH* ancestral est estimée à 240/346 MA durant la divergence entre les Mammifères et les Oiseaux (Kitano T, 1998; Matassi G, 1999). Toutefois une étude plus récente (Kitano T, 2000) estime cette duplication plus précoce à 510 MA. Des gènes homologues de *RH* ont été découverts chez de nombreux Animaux. Des transcrits de ARNm ont été détectés dans le tissu hématopoïétique de Chimpanzés de Gorilles et de Macaque (Salvignol I, 1995; Mouro I, 1994) et les séquences d'ADNc démontrent un haut degré d'homologie avec les gènes humains et la possibilité de produire une protéine de 417 aa. Comme les Humains, les Chimpanzés et les Gorilles ont deux gènes *RH* alors que les autres Singes (Orang-Outan compris) n'en possèdent qu'un (Blancher A, 2000). Ainsi, la duplication entre le gène *RH* ancestral et les deux gènes *RHD* et *RHCE* est survenue chez l'ancêtre commun aux Gorilles, Chimpanzés et Humains avant la divergence estimée à 11 MA. (Carritt B, 1997 ; Blancher A, 1997). L'analyse du mécanisme impliqué dans cette duplication est complexe compte tenu de l'orientation inversée des gènes *RHD/RHCE* sur l'haplotype.

Des antigènes similaires aux antigènes RH ont été retrouvés sur des hématies de Primates non humains. L'anti-c et l'anti-D détectent un polymorphisme sur les hématies de



Chimpanzés Rc, de Gorille Dgor et d'Orang-Outan Rh<sub>0</sub>Or. Par contre, il n'existe pas de données relatives à la présence de C, E et e sur les hématies de Primates non humain (Blancher A, 2000, 1977; Carritt B, 1997). Un certain nombre d'évènements sont supposés avoir généré les haplotypes RH actuels (Figure 12 p.72, Carritt B, 1997). Une duplication d'un gène ancestral, suivie par des mutations et des recombinaisons complexes ont généré un gène ressemblant au gène *RHD* et l'allèle *ce* du gène *RHCE*. *Dce* est le point de départ du polymorphisme Rh chez l'Homme. C'est ce même haplotype qui apparaît comme le plus fréquent en Afrique subsaharienne. La délétion ou l'inactivation du gène *RHD* créa par la suite l'haplotype *dce*, une recombinaison non réciproque de la séquence de l'exon 2 du gène *RHD* dans l'allèle *ce* du gène *RHCE* pourrait avoir produit l'allèle *DCE* et une mutation ponctuelle dans l'allèle *ce* pourrait avoir produit *DcE*. En accord avec la thèse de Fischer et Race (Fisher RA, 1946b) de 1946, Carrit (Carritt B, 1997) proposa que *dCe* apparut à la suite d'une recombinaison entre *DCE* et *dce*. De même, *dcE* serait apparu après une recombinaison entre *DcE* et *dce*. L'haplotype *DCE* serait issu d'une recombinaison entre *DCE* et *DcE*. Le très rare haplotype *dCE* serait apparu à la suite d'une recombinaison de *dCe* et *dcE*.

Le système RH est hautement polymorphe, particulièrement en Afrique où il y a trois supports moléculaires relativement communs du D négatif. Si le complexe RH possède une fonction de transporteur d'ammonium, des changements complexes par mutation des gènes pourraient altérer la concentration ionique intra-cellulaire, ce qui pourrait affecter la susceptibilité de la cellule à être un hôte adéquat pour le paludisme par exemple. Cette fine balance évolutive entre le parasite et son hôte pourrait être responsable de ce haut niveau de polymorphisme du système RH dans ces régions.

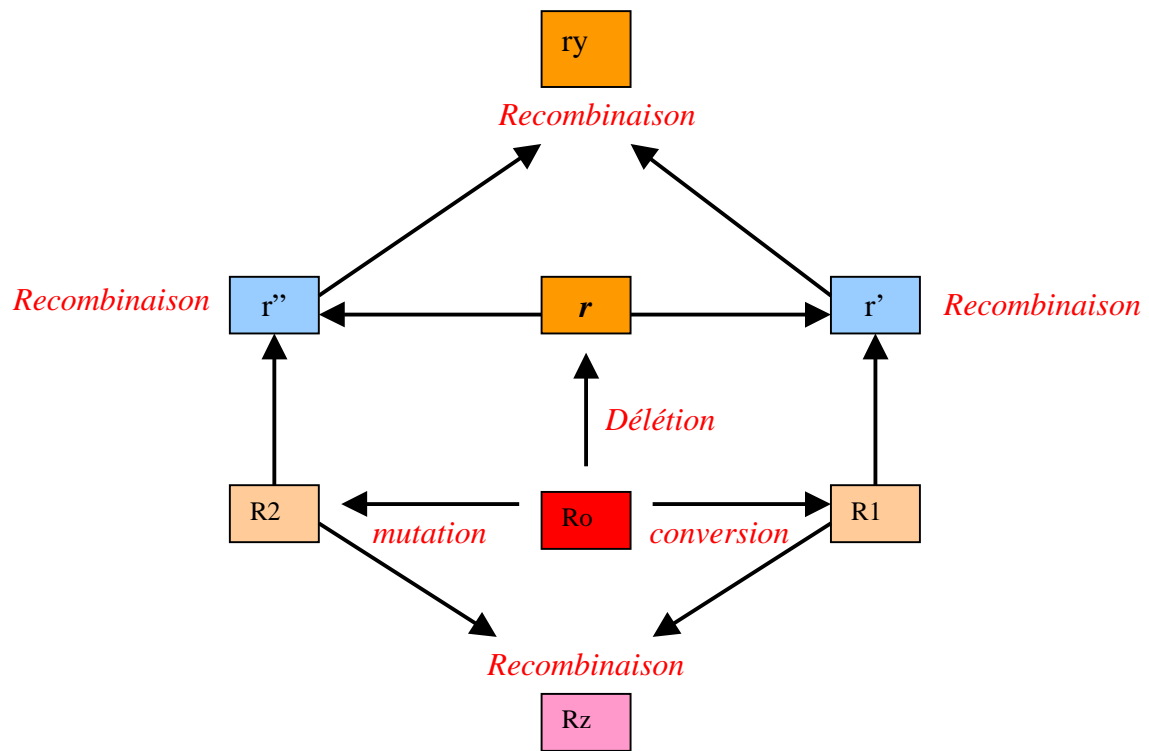


Figure 12.: Evolution des gènes *RH*. (D'après Carrit B, 1997)

### ***III.4.1.3. Le système Kidd.***

#### **a) Génétique formelle.**

Le locus *JK* (*SLC14A1*), situé sur le chromosome 18, comporte trois allèles. Deux allèles courants *JK1* (*Jka*) et *JK2* (*Jkb*) et un allèle rare silencieux *JK*. Les allèles *JK1* et *JK2* sont codominants et dominant sur l'allèle silencieux. L'existence de cet allèle, plus fréquent dans certaines populations (Polynésie) peut imposer un recours à la formule de Bernstein pour le calcul des fréquences géniques.

#### **b) Génétique moléculaire.**

Les allèles *JK1* et *JK2* se différencient par 3 substitutions nucléotidiques. Seule la substitution G838A aboutit, au niveau protéique, à une substitution d'acide aminé Asp280Asn (Olives B, 1997). Par ailleurs, cinq types de mutation peuvent aboutir à un phénotype silencieux JK :-1,-2 :

- Une mutation G/A dans l'intron 5 abouti à la perte de l'exon 6, sur un allèle *JK2*. Ce mécanisme a été décrit principalement dans les populations Polynésiennes (Lucien N, 1998 ; Irshaid NM, 2000). La fréquence génique de 8.7% de cette allèle, permet une estimation d'une fréquence phénotypique à 0.76% dans cette population.
- Une mutation T871C sur un allèle *JK2* abouti à la substitution d'acide aminé Ser291Pro qui compromet la glycosylation de la protéine Kidd et sa fixation à la membrane. Cette mutation est décrite dans la population Finlandaise (Sidoux Walter F, 2000).
- Un cas de translation G/T dans l'exon 7 d'un allèle *JK2* a été décrit chez une Française (Lucien N, 1998).
- Une délétion de 1.6 kb impliquant les exons 4 et 5 d'un allèle *JK1* ont été décrit chez une Anglaise (Irshaid NM, 2002) et une Tunisienne (Lucien N, 2002).
- Une mutation non sens dans l'exon 7 d'un allèle *JK1* a été décrit chez 3 sœurs Suisses (Irshaid NM, 2002).

**c) Les aspects phénotypiques.**

Les deux allèles principaux codent pour deux antigènes qui définissent 3 phénotypes courants dont les fréquences en Europe (Race RR, 1975) sont les suivantes : Jk (a+b-) 26.4%, Jk (a+b+) 50%, Jk(a-b+) 23.6%. Le phénotype silencieux Jk (a-b-) est caractérisé par l'absence, outre des antigènes courants, d'un antigène de grande fréquence, JK3. Le mécanisme de ce phénotype peut être lié à la présence en double dose de l'une des 5 mutations ou bien à la présence d'un gène inhibiteur dominant *In (JK)* (Okubo Y, 1986). Dans ce cas, de faibles quantités d'antigènes JK1, JK2 et JK3 peuvent être détectées par fixation élution. Un phénotype silencieux transitoire a été décrit chez une patiente Russe atteinte d'une hémopathie maligne (Issitt P, 1990) réversible après rémission. Ces phénotypes silencieux sont caractérisés par une augmentation de la résistance de lyse par l'urée.

**d) Les aspects fonctionnels.**

Les molécules Kidd sont des transporteurs d'urée au travers de la membrane érythrocytaire. Elles présentent des homologies de structure avec une autre molécule transporteur d'urée présente uniquement dans les cellules tubulaires rénales la HUT2. L'absence de molécules Kidd diminue de 1000 fois la pénétration de l'urée à l'intérieur de la cellule. Ces molécules semblent avoir des fonctions de protection osmotique cellulaire lors du passage des hématies dans les vaisseaux de la médullaire rénale. Compte tenu d'absence d'anomalie chez les sujets de phénotype silencieux, cette fonction semble être compensée par d'autres molécules transporteurs (Olives B, 1996).

#### **III.4.1.4. Le système Duffy.**

##### **a) Génétique formelle.**

Le locus *FY*, localisé sur le bras long du chromosome 1, comporte dans les populations Européennes deux allèles courants (*FYA*, *FYB*) et un allèle plus rare (*FYX* codant pour un antigène *FYB* faible). Dans les populations d'Afrique subsaharienne ce locus comporte trois allèles courants (*FYA*, *FYB*, *FY0*). Les allèles *FYA*, *FYB* et *FYX* sont codominants. L'allèle *FY0* est récessif. L'existence de cet allèle impose un recours à la formule de Bernstein pour le calcul des fréquences géniques.

##### **b) Génétique moléculaire.** (Table 7 p. 79, Figure 13 p.78)

- **Le polymorphisme antigénique *FYA*/*FYB*.**

Les deux antigènes *FYA* et *FYB* ne diffèrent que par un acide aminé Gly42Asp (Mallinson G, 1995) et les allèles par une substitution nucléotidique G125A. La substitution Ala100Thr (G298A) dans le deuxième domaine intra-membranaire, parfois observée avec l'antigène *FYB*, n'apparaît pas affecter l'expression antigénique (Chaudhuri A, 1993; Mallison 1995). Cette substitution n'a pas été retrouvée chez les sujets originaires d'Afrique subsaharienne (Olsson ML, 1998a, 1998b). Des homologues de glycoprotéines Duffy ont été retrouvés chez des primates non humains avec des homologies atteignant 99% chez le Chimpanzé et 93% chez le *Macacus rhésus* (Chaudhuri A, 1995). Tous ces Primates non-humains expriment le gène codant pour Asp42, ce qui suggère que *FYB* est le gène ancestral (Chaudhuri A, 1995; Li J, 1997).

- **Le polymorphisme antigénique FYB/FYX.**

Le polymorphisme *FYB* / *FYX* est basé sur une substitution nucléotidique C286T aboutissant à la substitution d'un seul acide aminé Arg89Cys (Olsson ML, 1998b).

- **Le phénotype FY:-1,-2 (Fya-b-) Africain.**

Le phénotype FY:-1,-2 des populations Africaines est lié à la présence, en double dose, d'une mutation dans le promoteur érythroïde du gène *Duffy* empêchant la fixation du facteur de transcription GATA-1. Cette mutation aboutit à l'absence d'expression des molécules Duffy sur l'érythrocyte (Nichols ME, 1988) mais pas sur les autres cellules (Peiper SC, 1995) dont le promoteur est situé à 82 paires de bases en amont (Iwamoto S, 1996). Les individus FY:-1,-2 sont résistants à l'infection par le *Plasmodium vivax* et leurs hématies sont réfractaires à sa pénétration in vitro. La séquence codante du gène *FY0* est identique (Chaudhuri A, 1993; Mallinson G, 1995; Peiper SC, 1995) à l'allèle *FYB* mais avec un changement d'une base T46C, au niveau du promoteur du gène. La présence des molécules FYB et FY3 exprimées sur les cellules endothéliales explique le fait que les sujets FY:-1,-2 d'origine Africaine ne s'immunisent que très rarement après stimulation par des hématies de phénotypes Duffy courants.

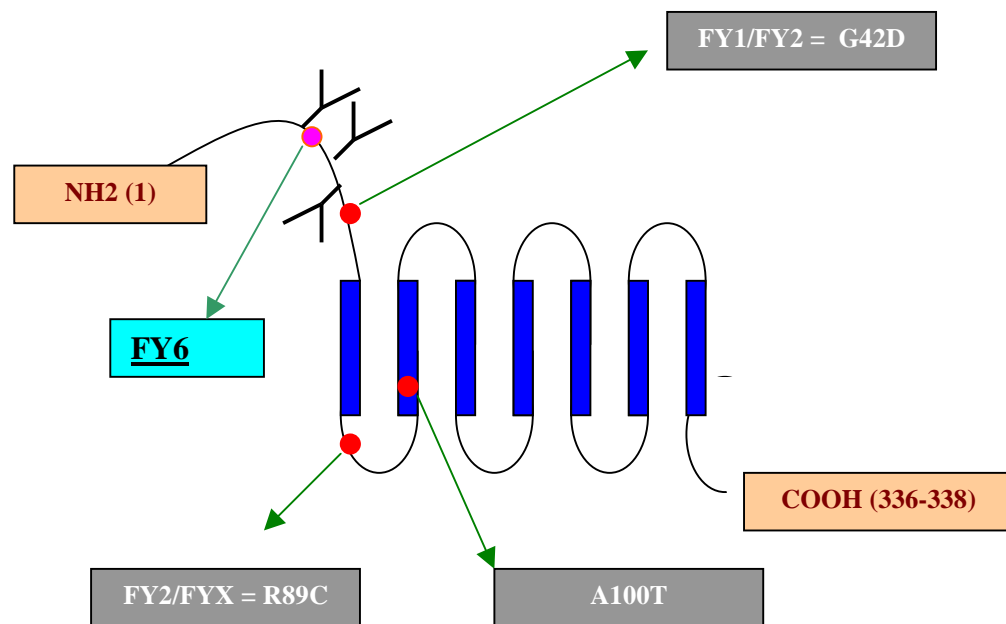
- **La mutation GATA en Papouasie Nouvelle Guinée (PNG).**

Cette mutation GATA a été retrouvée sur un allèle *FYA* en Papouasie Nouvelle Guinée. Une étude rapporte que 33 individus parmi 1062 testés ont été retrouvés hétérozygotes pour un allèle *FY0* (mutation GATA) sur un gène *FYA* (Gly42) (Zimmerman PA, 1999). Aucun sujet homozygote n'a, à ce jour, été mis en évidence. Cet allèle est probablement silencieux dans la lignée érythroïde et les cellules de ces sujets présentent une moindre expression de FY6 que celles qui sont homozygotes pour l'allèle *FYA* normal. A ce titre, ces sujets semblent plus résistants au *Plasmodium vivax*.

L'apparition de cette mutation en zone de forte endémie de *Plasmodium vivax* pourrait être le témoin d'un début d'adaptation par maintien de l'apparition à priori récente de cette mutation.

- **Le phénotype FY:-1,-2 non africain.**

Quelques rares exemples de phénotypes FY:-1,-2 non africains ont été découverts par le biais de la détection d'un anti-FY3. Les bases moléculaires de ce phénotype ont été définies chez cinq sujets non africains dont quatre avec anti-FY3 apparu à la suite d'une transfusion et ou d'une grossesse. Il s'agit ici de délétions ou mutation non sens qui aboutissent à l'absence d'expression des antigènes Duffy, non seulement dans le tissu érythroïde mais aussi dans les tissus extra hématopoïétiques. L'absence totale de ces molécules explique l'allo-immunisation systématique des phénotypes FY:-1,-2 non Africain en cas d'exposition à des hématies de phénotypes courants.



La protéine DARC présente une topologie à 7 domaines transmembranaires typique des récepteurs couplés aux protéines G. C'est une protéine N glycosylée de 338 aa pour sa forme  $\alpha$  et de 336 aa pour sa forme  $\beta$ . La substitution en G42D est associée au polymorphisme antigénique FY1/FY2. L'allèle FYX se différencie de l'allèle FY2 par deux substitutions. L'une en R89C, qui est responsable du défaut d'insertion de la protéine DARC sur la membrane cellulaire. La deuxième substitution A100T est toujours associée à la première dans l'allèle FYX mais ne reflète qu'un polymorphisme de population puisqu'elle peut être retrouvée avec les allèles FY1 et FY2 normaux. Le rapprochement dans l'espace des premier et quatrième domaines transmembranaires de DARC définit une poche de fixation pour les chimiokines. La partie N-terminale extracellulaire portant l'épitope FY6 est nécessaire à la fixation de mérozoïtes de *Plasmodium vivax*.

Figure 13 : Structure de la molécule Duffy (Daniels G, 2002)



Allèles	Nt 46	Nt 125/aa 42	Nt 286/aa 89	Nt 319/aa 100
<i>FY1</i>	T	G / Gly	C / Arg	G / Ala
<i>FY2</i>	T	A / Asp	C / Arg	G/Ala ou A/Thr
<i>FYX</i>	T	A / Asp	T / Cys	A / Thr
<i>FYO</i>	C	A / Asp	C / Arg	G / Ala

Table 7 : Polymorphisme moléculaire des gènes Duffy (Daniels, 2002)

Phénotypes	Génotypes	Afrique Subsaharienne	Europe	Alleles	Afrique Subsaharienne	Europe
FY:1,-2	<i>FY1 / FY1</i>	<b>0</b>	<b>21</b>	<i>FY1</i>	0.003	0.410
	<i>FY1 / FYO</i>	<b>4</b>	<b>0</b>	<i>FY2</i>	0.175	0.590
FY:1,2	<i>FY1 / FY2</i>	<b>2</b>	<b>40</b>	<i>FYO</i>	0.795	0
FY:-1,2	<i>FY2 / FY2</i>	<b>2</b>	<b>39</b>			
	<i>FY2 / FYO</i>	<b>29</b>	<b>0</b>			
FY:-1,-2	<i>FYO / FYO</i>	<b>63</b>	<b>0</b>			

Table 8 : Fréquences génotypiques dans différentes populations (Daniels G, 2002)

### **c) Les aspects phénotypiques.**

#### **Les antigènes et phénotypes courants.**

Le système Duffy comporte 2 antigènes antithétiques FY1(Fya) et FY2(Fyb) déterminant, dans les populations européennes, 3 phénotypes principaux : FY: 1,2 / FY: 1,-2 / FY: -1,2. Dans les populations Africaines, le phénotype le plus commun est FY:-1,-2, résultat de l'homozygotie pour un gène apparemment silencieux *FY0*. Ce phénotype FY:-1,-2 est exceptionnel dans les populations non Africaines. Les fréquences de *FY1* et *FY2* en Afrique sont variables mais faibles comparativement à celles observées en Europe en raison de la fréquence élevée de l'allèles *FY*. En Europe, *FY2* est plus commun que *FY1* et en Extrême Orient *FY1* dépasse largement *FY2* (Mourant AE, 1976). La table 8 p.79 présente les fréquences phénotypiques et génotypiques déterminées après analyse par biologie moléculaire.

#### **L'Antigène FY2 faible.**

L'allèle *FYX* (codominant / *FY1* et récessif / *FY2*) code pour un antigène FY2 faible surtout décrit dans les populations Européennes avec une fréquence estimée à 1.5% (Lewis M, 1972). Cette diminution de la réactivité antigénique est liée à une diminution quantitative du nombre de sites plutôt qu'à une altération ou à une néo-conformation de l'antigène FY2.

#### **L'Antigène FY1 faible.**

Dans les populations du Sud-Est Asiatique, de forte endémie en *Plasmodium vivax* et *Plasmodium falciparum*, ont été décrits des antigènes Fya et Fyb faibles pouvant être la cause d'erreur de phénotypage [individus apparaissant faussement Fy (a-b-)]. Les analyses en biologie moléculaire ont démontré qu'il s'agissait d'allèles *FY1* et *FY2* normaux. Le maintien de cette diminution d'expression antigénique pourrait être le résultat d'un mécanisme adaptatif en vue de diminuer les chances d'invasion par *Plasmodium vivax*. (Shimizu Y, 2000).

**Les antigènes FY3, FY5 et FY6** sont des antigènes de grande fréquence dans les populations Européennes et « privés » dans certaines parties de l’Afrique de Ouest. L’antigène FY3 est aussi présent sur les hématies de primates non humains.

**d) Les aspects fonctionnels.**

Le polymorphisme Duffy est localisé sur des molécules qui fixent certaines variétés de chémokines. Ces molécules sont nommées DARC (Duffy Antigen Receptor for Chemokine). La fonction exacte des glycoprotéines Duffy est inconnue et certains ont suggéré un rôle d’épurateur des chémokines. Quoiqu’il en soit ce rôle semble limité compte tenu que les sujets Fy (a-b-) africains et non africains sont tous apparemment sains. Toutefois, le maintien de ce gène à travers les espèces souligne l’importance de cette protéine. Aussi, si les molécules Duffy possèdent une fonction importante, il semble que d’autres structures assument le même rôle en assurant ainsi leur suppléance.

**e) Les antigènes Duffy et Paludisme.**

Le *Plasmodium vivax* est responsable de la forme la plus répandue de paludisme et dont les formes cliniques sont moins sévères que celles liées aux *Plasmodium falciparum*. Les sujets homozygotes *FYO* sont résistants à l’infection et les hématies ne peuvent être infestées in vitro (Miller LH, 1975). Il apparaît probable que la haute incidence de l’allèle *FYO* en Afrique soit le résultat d’une pression sélective liée à l’avantage que confère l’homozygotie vis à vis de cette infection. De même, les hématies hétérozygotes *FYO* sont moins infestées que celles dépourvues d’allèle *FYO* ce qui confère aussi un avantage à l’hétérozygote dans les régions endémiques au *Plasmodium vivax*. Dans certaines parties d’Afrique de l’Ouest, la fréquence de cet allèle, s’élevant à 100%, a abouti à la disparition totale du parasite dans ces régions (Welch SG, 1977) [en raison de la rupture de son cycle de reproduction]. L’analyse de l’invasion chez les Primates non humain par *Plasmodium vivax* et par *Plasmodium knowlesi* démontre que FY6 est un épitope important pour l’invasion.

**f) Le locus Duffy : un lieu d'action complexe de la sélection naturelle.**

Le locus Duffy a depuis longtemps été considéré comme une cible probable de l'action de la sélection naturelle en raison de l'extrême degré de différenciation géographique de ses trois allèles majeurs (*FYB*, *FYA*, *FYO*). En absence de migration, en cas de neutralité évolutive, la variation des fréquences alléliques au travers de populations est uniquement déterminée par la dérive génétique. Ainsi, tous les loci neutres du génome ont la même potentialité de différenciation. Lorsque le locus *FY* est comparé avec un locus supposé neutre, le degré de différenciation géographique de ses trois allèles est clairement le plus élevé dans l'espèce humaine. La répartition géographique la plus frappante est celle de l'allèle *FYO* qui s'est fixé, il y a 33 000 ans (Hamblin MT, 2002), en Afrique subsaharienne et pas ailleurs. De plus, la résistance qu'il confère vis à vis du *Plasmodium vivax* représente un élément supplémentaire en faveur d'une pression sélective positive. L'allèle *FYA* présente aussi un niveau de différenciation géographique inhabituel, principalement, par sa fixation récente en Asie de l'Est et dans le Pacifique. Enfin, l'allèle ancestral *FYB* tend à disparaître dans de nombreuses populations humaines pour se réduire à Asie de l'Ouest, l'Europe et les Amériques.

La fixation d'un allèle neutre par le biais d'un processus aléatoire de dérive génétique est un processus lent (Kimura M, 1983). Etant donné qu'un allèle de grande fréquence, fixé par ce mécanisme, est supposé être plus ancien et, ainsi, apparaître dans un environnement génétique de variabilité élevé, il est surprenant de trouver une faible variabilité moléculaire entourant l'allèle *FYA* (Hamblin MT, 2000). Le faible niveau de variation entourant un tel allèle courant, suscite la question de l'existence d'une pression sélective méconnue sur l'allèle *FYA* (ou sur un allèle qui lui serait fortement lié). Alors que, contrairement à l'allèle *FYO*, l'allèle *FYA* n'est pas associé à un phénotype particulier.

Ces données ont été confirmées par Hamblin MT (2002) qui a réalisé une étude de la variabilité moléculaire des régions entourant le gène *FY* à partir d'échantillons provenant de populations d'Afrique subsaharienne (Hausa) et d'Asie de l'Est (Chinois Han) qui ont respectivement fixé les allèles *FYO* et *FYA*. La variabilité de ces échantillons est comparée à

celle observée dans des populations d'Europe (Italie) et d'Asie du Sud (Pakistan) dans lesquelles il n'apparaît pas de fixation de ces allèles. Les données de la région *FY* sont comparées aux variations obtenues pour des loci putativement neutres, ce qui évite la nécessité d'une prise en compte des données historiques sous jacentes. Les résultats montrent distinctement une signature de la sélection naturelle dans la population Africaine et sont cohérents avec une sensible, mais plus complexe, signature dans les populations non-Africaines et d'Asie de l'Est en particulier.

#### ***III.4.1.5. Le système Kell***

Ce système est composé de 24 antigènes et au moins 2 autres systèmes indépendants sont en relation étroite avec lui : le système Kx et le système Gerbich.

##### **a) Génétique formelle.**

Le locus *KEL*, situé sur le chromosome 7, comporte 3 couples d'allèles courants codominants: *K* et *k* ; *Kp<sup>a</sup>* et *Kp<sup>b</sup>* ; *Js<sup>a</sup>* et *Js<sup>b</sup>*. L'haplotype originel apparaît être *k, Kpb, Jsa*. Il n'y a pas plus d'un mutant par haplotype retrouvé au niveau des haplotypes *KEL*.

##### **b) Génétique moléculaire.**

Les bases moléculaires du polymorphisme Kell sont connues et reposent sur des substitutions, au niveau génétique, d'un seul nucléotide et au niveau protéique d'un seul acide aminé (Table 9 p.86)

##### **c) Les aspects Phénotypiques.**

Les antigènes Kell sont localisés sur une glycoprotéine membranaire (CD238) de 93Kd (Figure 14 p.87).

### **Les antigènes K/k.**

K et k sont deux antigènes antithétiques bien développés à la naissance. K est présent chez 9% des sujets Européens, moins fréquent chez les Africains et extrêmement rare en Asie de l'Est et dans les populations natives du continent américain. Cet antigène atteint les plus hautes valeurs mondiales dans la Péninsule Arabique avec près de 25%. Du point de vue des fréquences géniques, la répartition est la suivante : allèle *K* est surtout présent dans les populations européennes de l'Ouest avec 4.6% (Race, 1975), à 2% chez les Finlandais (Furuhjel U, 1968), 0.7% chez les Africains (Stroup M, 1965) et 0.01% chez les Japonais (Hamilton HB, 1971). L'antigène *k* (Cellano) est un antigène de grande fréquence dans toutes les populations.

### **Les antigènes Kpa, Kpb et Kpc.**

Allen décrit en 1957 Kpa et son antigène antithétique Kpb. Les deux allèles codant pour ces antigènes apparaissent liés aux allèles *K/k*. Des études familiales ont démontré cette forte liaison puisque, parmi les sujets *K+Kpa+*, aucun ne possédaient les deux allèles sur le même haplotype (Dichupa PJ, 1969). L'haplotype commun apparaît être *k,Kpb*. L'antigène Kpa est présent chez 2 à 3% des sujets, dans la population Européenne et 1 à 2% des sujets *K+* apparaissent Kpa (Race RR, 1975). L'antigène Kpb est un antigène public dans toutes les populations. En 1979, un nouvel antigène Kell (Kpc) est découvert au Japon chez un sujet apparemment *Kp (a-b-)* (Yamaguchi H, 1979). Cet antigène présent chez 0.3% de la population japonaise n'a été décrit que chez un seul sujet en dehors du Japon (sujet espagnol *Kp (a-b-c+)* avec anti-Kpb).

### **Les antigènes Jsa et Jsb.**

L'association de ces antigènes antithétiques, au système Kell est liée à leur absence systématique sur les hématies *Ko*. Il apparaît que Jsa est confiné aux populations d'origine Africaine où il peut atteindre près de 16%.

**d) Les aspects évolutifs.**

Les Chimpanzés possédant les antigènes k, Kpb et Jsa (Redman CM, 1989 ; Blancher A, 2000) leurs gènes respectifs sont considérés comme les allèles ancestraux.

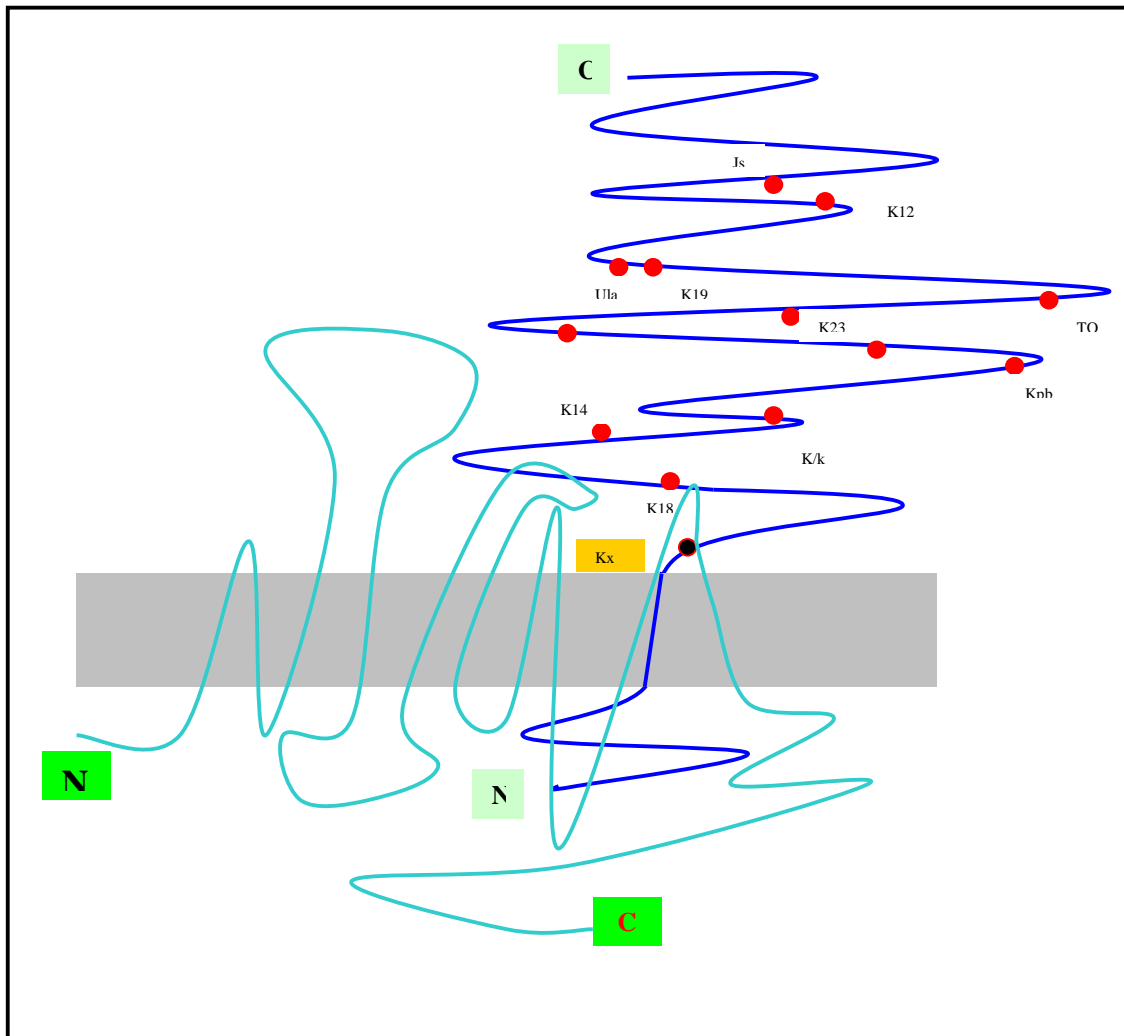
**e) Les aspects fonctionnels.**

La protéine Kell porte un motif pentamérique de fixation du Zinc (HELLH) permettant la fixation d'endopeptidases zinc dépendant (Lee S, 1991). La fonction de la molécule Kell n'est pas claire au niveau érythrocytaire.

<b>ISBT</b>	<b>Nom</b>	<b>Fréquence relative</b>	<b>Allèle</b>	<b>Bases moléculaires</b>
KEL1	K	Faible	<i>K</i>	Met193Thr
KEL2	K	Haute	<i>K</i>	Thr193Met
KEL3	Kp <sup>a</sup>	Faible	<i>Kp<sup>b</sup>, Kp<sup>c</sup></i>	Trp281Arg ou Gln
KEL6	Js <sup>a</sup>	Faible	<i>Js<sup>b</sup></i>	Pro597Leu
KEL7	Js <sup>b</sup>	Haute	<i>Js<sup>a</sup></i>	Leu597Pro

Table 9 : Nomenclature des antigènes KEL.





### La glycoprotéine Kell.

Les antigènes Kell sont portés par une glycoprotéine de 732 aa. Cette protéine de 93 Kd est absente des hématies de phénotype Ko. La portion N terminale intracytoplasmique comporte 47 aa, la portion C terminale extra cytoplasmique en comporte 665 aa dont 15 résidus de cystéine suggérant la présence de 7 ponts disulfure assurant un repliement de la molécule dans l'espace. La protéine Kell présente des homologies structurales avec la famille des endopeptidases fixant le zinc. La glycoprotéine Kell est fortement associée dans la membrane à la protéine Kx de 120 Kd par un pont disulfure entre Cys72 de Kell et Cys347 de Kx.

Figure 14 : Structure de la molécule Kell (Daniels G, 2002)

### **III.4.1.6. Le système MNS.**

#### **a) Génétique formelle.**

C'est un système d'une grande complexité qui comporte 48 antigènes. Deux couples d'allèles courants, *M/N* et *S/s*, sont situés sur 2 loci très fortement liés sur le chromosome 4. L'association des deux couples d'allèles fait l'objet d'un déséquilibre de liaison.

#### **b) Les aspects phénotypiques.**

Les phénotypes courants en Europe sont les suivants : MM 28%, MN 50%, NN 22%, SS 11%, Ss 44% et ss 45% (Race, 1975). En Afrique, 1% des sujets sont S-s- et manquent d'un antigène de grande fréquence U (Wiener AS, 1954) ou présentent une faible expression de celui-ci (Race RR, 1975).

#### **c) Les aspects moléculaires.**

Les antigènes M et N sont portés par une glycoprotéine : la glycophorine A (GPA). M diffère de N par deux acides aminés Ser1Leu et Gly5Glu. Les chaînes glucidiques latérales participent à l'expression antigénique de M et de N. Les antigènes S et s sont portés par une autre glycoprotéine : la glycophorine B (GPB). Ils ne diffèrent que par un seul acide aminé Met29Thr. Les 26 premiers acides aminés de la GPB sont identiques aux 26 premiers de GPA<sup>N</sup>. Cet antigène 'N' peut être détecté par certains anti-N chez des sujets homozygotes *M/M*.

#### **d) Glycophorines et paludisme.**

Les sujets Ena-, déficitaires en GPA, sont plus résistants à l'invasion palustre (Miller, 1977). La protéine EBA-175 de *Plasmodium falciparum* se fixe spécifiquement sur la GPA et non sur la GPB (Sim, 1994). Par ailleurs, les acides sialiques de la GPA sont essentiels à la pénétration du parasite (Templeton TJ, 1998). Les sujets S-s-, déficitaires en GPB, sont aussi plus résistants mais moins que les Ena- déficitaires en GPA et GPB (Facer CA, 1983).

**e) Glycophorines et autres pathogènes.**

La GPA et surtout la GPA<sup>M</sup> est le récepteur de nombreuses bactéries comme *E. Coli* (Jokinen M, 1985) ou de toxines comme celle de *E.Coli* (Cortajanera AL, 2001) ou bien de *Vibrio cholerae*. Enfin GPA apparaît comme le récepteur de virus comme les *Influenza*, (Kathan RH, 1961) par exemple.

**f) Les aspects fonctionnels et évolutifs.**

La principale caractéristique des glycophorines est l'importante glycosylation de leur domaine extracellulaire qui les affecte d'une forte charge négative empêchant toute agglutination spontanée des hématies. Ces molécules contribuent aussi à structurer le glycocalix, véritable manteau de protection contre les agressions mécaniques ou microbiennes de l'hématie.

La GPA pourrait avoir des fonctions de ligand induisant des signaux de transduction (Chasis JA, 1992), de transporteur d'anions (Groves JD, 1992), d'inhibition de la fonction complémentaire C5b-7 (Tominta A, 1993). Par ailleurs, la GPA joue un rôle important dans la fixation et la pénétration du *Plasmodium falciparum*. Les sujets totalement dépourvus de GPA étant extrêmement rares, ces molécules doivent avoir ou avoir eu des fonctions importantes.

D'un point de vue évolutif, le locus  $GPA^N$  semble être le gène ancestral qui, après duplication, a produit le gène  $GPB/E$  qui donnera ensuite les gènes ancestraux de GPB et GPE.  $GPE$  intègrera l'exon 2 de  $GPA^M$ . Cela explique notamment pourquoi GPB possède 'N' et GPE possède la séquence  $GPA^M$ . GPA est présente chez tous les Primates non humains alors que GPB et GPE plus récentes se retrouvant uniquement chez les Chimpanzés et les Gorilles (Xie S, 1997; Blancher A, 2000).

### **III.4.2. Le polymorphisme HLA-DRBI.**

[Synthèse Thèse D. Reviron, 1999]. Le complexe majeur d'histocompatibilité (CMH), nommé système HLA (Human Leukocyte Antigen) chez l'Homme, comporte plus de 100 gènes portés par le bras court du chromosome 6 (Figure 15 p.91 ).

#### **a) Organisation et polymorphisme des gènes HLA.**

Les gènes HLA sont organisés en trois sous régions : classe I, classe II et classe III.

Les gènes HLA de classe I sont situés en position télomérique et les gènes HLA de classe II en position centromérique. Ils codent pour les molécules HLA, respectivement glycoprotéines HLA de classe I et glycoprotéines HLA de classe II. Les gènes de classe III sont situés entre les gènes de classe I et de classe II. Ils codent notamment pour des protéines du système du complément, les facteurs TNF et des protéines de choc thermique (HSP 70).

#### **b) Organisation et polymorphisme des gènes de classe II.**

La région HLA de classe II s'étend sur 900 kb et comprend environ 30 gènes. Parmi ceux-ci, les gènes principaux codant pour les 6 familles différentes de molécules de classe II, sont HLA DRA, DRB1. Ces gènes (indiqués par la lettre A ou B) codent pour les chaînes  $\alpha$  et  $\beta$  des molécules HLA de Classe II. Ces chaînes sont hautement polymorphes et sont exprimées sous formes hétéro dimères  $\alpha/\beta$  [Nepom GT. 1991, Radley E. 1994].

#### **c) Organisation et polymorphisme des gènes de la région DR.**

La région DR est la plus complexe des régions de classe II. Elle comprend un locus DRA et plusieurs locus DRB. Le locus DRA est peu polymorphe (2 allèles décrits : DRA\*0101, DRA\*0102) dont les allèles codent pour la chaîne  $\alpha$ . Le locus DRB1 est le plus polymorphe. Ces allèles DRB1 codent pour les chaînes  $\beta$ 1 qui, associées aux chaînes DR  $\alpha$ , constituent les molécules de spécificité DR1 à DR18.

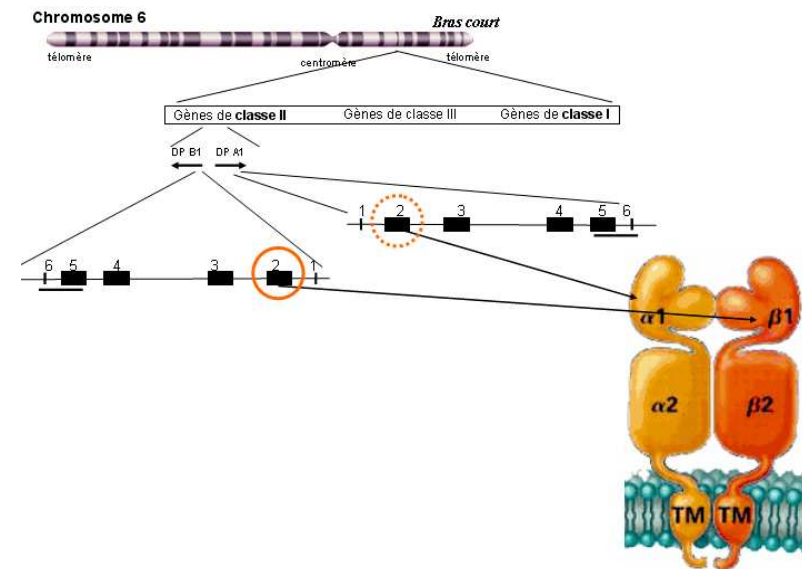
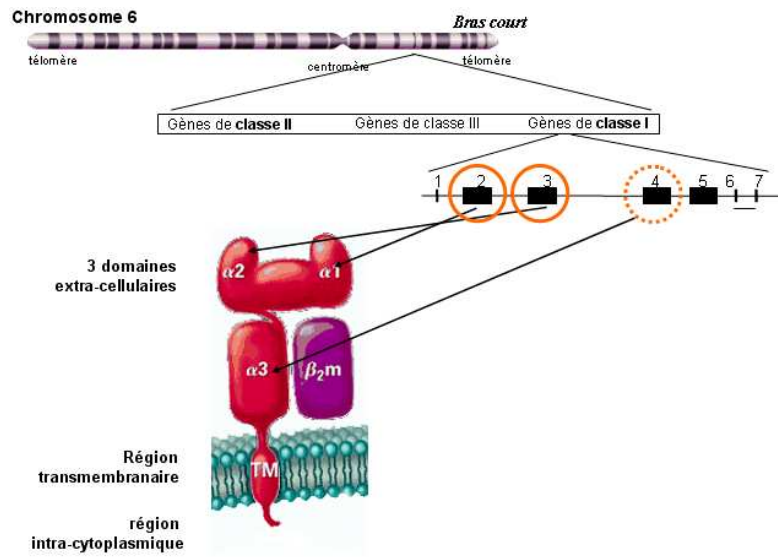


Figure 15 : Organisation des gènes du système HLA  
 D'après A. Dormoy, 2003, Communication personnelle

#### **d) Les aspects fonctionnels.**

La fonction principale des molécules HLA de classe I et de classe II est d'assurer la fixation et le transport de fragments peptidiques générés dans les cellules et de les présenter, au niveau de la membrane cytoplasmique, aux lymphocytes T. Grâce à ce mécanisme, ces derniers peuvent ainsi exercer une immunosurveillance constante du milieu intracellulaire représenté par la diversité des peptides liés aux molécules HLA. La reconnaissance spécifique par les lymphocytes T met alors en jeu une interaction tri moléculaire impliquant : le récepteur T (TcR), la molécule HLA et le peptide associé. Ce mécanisme de reconnaissance constitue la base de la réponse immunitaire de type cellulaire. Il confère à l'organisme une protection spécifique et adaptée contre divers agents pathogènes, mais il peut aussi jouer un rôle lors de processus auto-immuns. Les molécules HLA assument donc essentiellement une fonction de présentation de peptides aux lymphocytes T. Cette fonction étant réalisée dans le cadre de la réponse cellulaire dirigée contre les antigènes du soi et du non-soi ainsi que dans l'éducation thymique des lymphocytes T. Les peptides présentés par les molécules de classe II sont issus de la dégradation de protéines exogènes ou endogènes. Les protéines exogènes parviennent dans les compartiments cellulaires des cellules présentant l'antigène (CPAg) grâce à un processus d'endocytose. Celui-ci peut être spécifique, médié par les immunoglobulines de surface des lymphocytes B, ou non spécifique comme la pinocytose, la phagocytose ou l'autophagie [Sandvig K. 1994]. La fixation de ces peptides est réalisée grâce à des poches situées dans le sillon de la molécule HLA de classe II. Les molécules HLA de classe II (poches P1, P4, P6, P7, P9) présentent les peptides aux lymphocytes T CD4+.

#### **e) Suggestion de l'intervention d'une pression sélective (sélection balancée)**

*[Synthèse Thèse M. Gibert, 2001]*

- Le nombre d'allèles attendu sous l'hypothèse de neutralité est inférieur à celui observé (Potts WK, 1993).
- Le nombre de différences nucléotidiques inter-alléliques attendu sous l'hypothèse de neutralité est inférieur à celui observé (Takahata N, 1990). Cette différenciation inter allélique est le témoin d'une ancienneté des allèles (Lawlor DA, 1988 ; Adams EJ, 2000 ; Klein J, 1998).
- L'hétérozygotie attendue sous hypothèse de neutralité est inférieure à l'observée (Hédrick W, 1983 ; Charron D, 1997 ; Valdes AM, 1999 ; Salamon H, 1999).
- La concentration de la variation du polymorphisme est située dans les régions de la molécule impliquée dans la présentation peptidique (Valdes AM, 1999). Avantage procuré par une augmentation de la gamme de peptides pouvant être présentés.
- Le taux de substitution non-synonymes est supérieur au taux de substitutions synonymes au niveau des sites de reconnaissance des molécules classe I et II.
- L'existence de forme de protection et de susceptibilité vis à vis de maladies infectieuses.
- Le nombre de déséquilibres de liaisons entre les loci HLA sous hypothèse de neutralité est inférieur à l'observé (Robinson WP, 1991).

#### **III.4.3.Le polymorphisme des gènes KIR.**

*[Synthèse DEA C. Frassati, 2003]*

Les Natural Killer (NK) n'expriment pas comme les lymphocytes T ou B de récepteurs spécifiques de l'antigène issus d'évènements de recombinaison mais un ensemble complexe de récepteurs qui fournissent des signaux inhibiteurs ou activateurs à la cellule, en fonction

de la présence du ligand à la surface de la cible. Parmi ces récepteurs, on trouve la famille des KIR (Killer cell Immunoglobulin-like Receptor) qui comportent des récepteurs inhibiteurs et activateurs. Un même lymphocyte NK exprime un répertoire de récepteurs KIR activateurs et inhibiteurs dont la résultante module sa cytotoxicité. La distinction entre une cellule cible potentielle et une cellule normale est liée au recensement des molécules Classe I sur cette cellule. Les molécules classe I sont en effet les ligands spécifiques des KIR inhibiteurs qui, par l'intermédiaire de motifs ITIM, empêche toute action cytolytique (Figure 16a p.96). Ainsi, les molécules KIR régulent la réponse des cellules NK humaines au travers de la reconnaissance des molécules HLA classe I dont l'expression peut être absente ou altérée dans les cellules tumorales et les cellules infectées par des pathogènes intracellulaires (Algarra I, 2000; Wagtmann N, 1995). D'un point de vue fonctionnel, les molécules KIR 3D interagissent avec les molécules HLA-A et HLA-B, alors que les molécules KIR 2D peuvent interagir avec les molécules HLA-C et HLA-G.

Les KIR inhibiteurs possèdent une partie intra-cytoplasmique longue (KIR-L, pour « long »), ils sont appelés KIR2DL ou 3DL en fonction du nombre de domaines de type Immunoglobuline (Ig) qu'ils possèdent (Andre P, 2001). Les récepteurs KIR activateurs possèdent une région intra-cytoplasmique courte (KIR-S, pour « short »), ils sont appelés KIR2DS ou 3DS en fonction du nombre de domaines Ig comme leurs équivalents inhibiteurs. La nomenclature des KIR-S est très proche de celle des KIR-L et répond à l'existence d'une homologie très importante de leur partie extracellulaire, bien que les ligands des récepteurs activateurs ne soient pas clairement identifiés à ce jour (figure 16c p .96).

Les récepteurs KIR sont codés par des gènes situés dans une région complexe de récepteurs leucocytaires (chromosome 19q13.4) (figure 16b p.96) qui montre une très grande variabilité haplotypique aussi bien du point de vue du nombre de gènes que des membres contenus dans le complexe génique (Martin AM, 2000 ; Wilson MJ, 2000, 2001 ; Barten R, 2001 ; Wende H, 1999).



L'étude phylogénétique de ce système traduit une évolution rapide et récente par duplications successives d'un gène ancestral pour former une famille multigénique comprenant 14 loci et 2 pseudogènes (Hughes AL, 2002). Ils sont présents uniquement chez les Primates et ils semblent être apparus en même temps que HLA-C qui est leur principal ligand. Il est aussi intéressant de souligner que la localisation chromosomique de ces gènes sur un chromosome distinct de celui du locus du système HLA (chromosome 6) exclut une ségrégation conjointe de ces deux systèmes.

L'étude des génotypes KIR sur plusieurs individus a montré la présence de déséquilibres de liaison entre certains de ces gènes faisant émerger la notion d'haplotype (Uhrberg M, 1997). On distingue deux groupes d'haplotypes ; l'haplotype A qui contient les gènes inhibiteurs 2DL1, 2DL3, 3DL1 et un seul gène activateur, le 2DS4 (Uhrberg M, 1997) et l'haplotype B qui possède les gènes inhibiteurs 2DL2, 2DL5 et une combinaison variable de gènes activateurs à l'exception du 2DS4 (Vilches C, 2000 ; Uhrberg M, 1997). Les gènes 3DL3, 2DL4, et 3DL2 sont présents dans les deux groupes d'haplotypes et ils sont dits «gènes d'ancrage » car ils sont situés aux extrémités et au milieu de la région codant pour les gènes KIR.(Wilson MJ, 2000 ; Barten R, 2001). Les haplotypes KIR sont caractérisés par une grande diversité qualitative et quantitative au niveau des gènes qu'ils contiennent (Wilson MJ, 2000 ; Uhrberg M, 1997).

Les quelques études populationnelles menées à ce jour ont montré que certaines combinaisons de gènes KIR sont retrouvées plus fréquemment que d'autres (Norman PJ, 2001, 2002; Rajalingam R 2002; Uhrberg M, 1997; Witt, 1999).

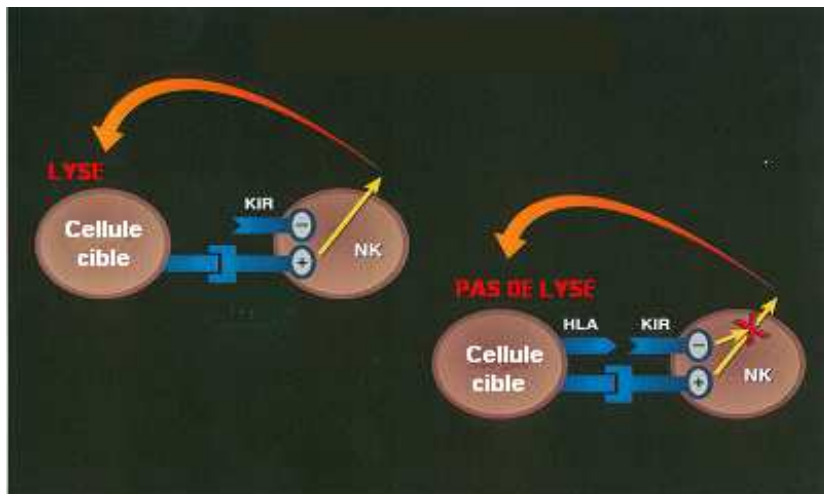


Figure 16a : Fonction des récepteurs KIR situés à la surface des cellules NK (D'après Carrington M, communication personnelle)



Figure 16b : organisation des gènes KIR (D'après Carrington M, communication personnelle)

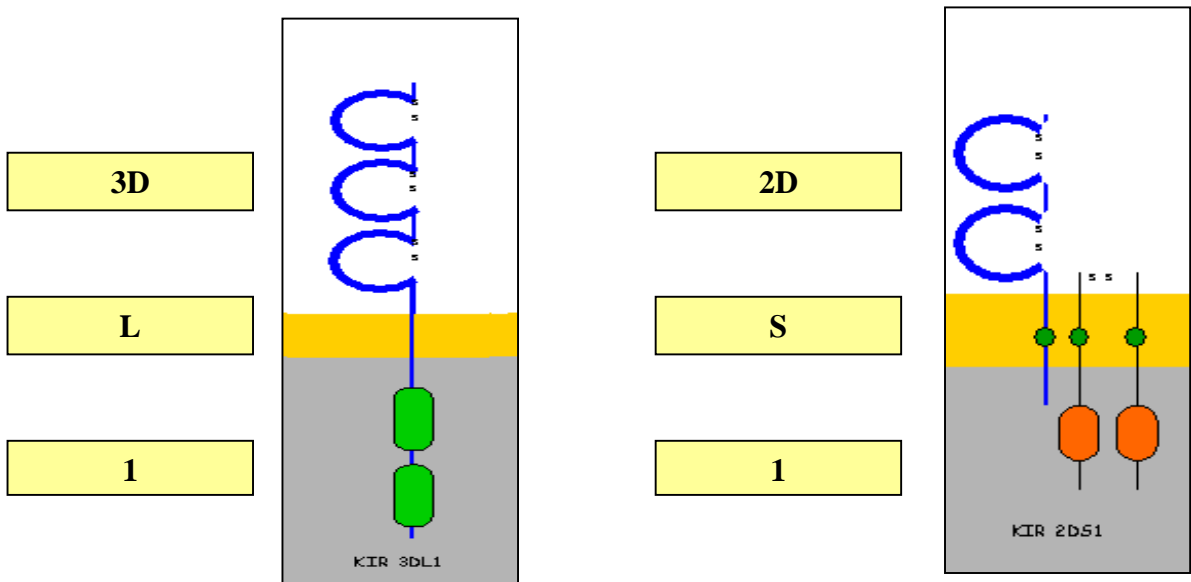


Figure 16c : Nomenclature des récepteurs KIR (D'après Trowsdale, 2001)

#### **III.4.4. Le polymorphisme du chromosome Y.**

De récentes découvertes montrant l'existence de recombinaison à l'intérieur de la portion anciennement dénommée NRY ou NRPY, ont poussé certains auteurs à renommer cette région, devenue MSY (male-specific region) (Figure 17 p.99). La région centrale MSY est bordée de deux régions pseudo-autosomales, PAR1 et PAR2, dont le taux de recombinaison méiotique est comparable à celui du génome autosomal. La portion euchromatique de MSY constitue la portion « active » de MSY. La région PAR1 est localisée sur la région terminale du bras court (Yp) et la région PAR2 sur la pointe du bras long (Yq). Alors que les régions PAR1 et PAR 2 représentent 5% de la totalité du chromosome, les 95% restants sont appelés : « portion non recombinante du Y » (NRY).

##### ***III.4.4.a. Les marqueurs polymorphes.***

Au sein de sa partie non recombinante, le polymorphisme du chromosome Y comporte trois types de mécanismes moléculaires. Le premier est représenté par les **SNP** (Single Nucleotide Polymorphism) qui sont caractérisés par la mutation d'une seule base à un locus donné. Les études interspécifiques permettent d'identifier le marqueur qui est ancestral et celui qui est dérivé. Le très faible taux de mutation constaté au sein des SNP, permet de considérer que la survenue d'un tel événement est unique. Ces SNP permettent de définir 18 haplogroupes majeurs (figure 18 p.105). Par exemple, l'allèle ancestral du marqueur M91 permet de définir l'haplogroupe A alors que l'allèle dérivé caractérise les autres. L'allèle ancestral du locus M181 caractérise l'haplogroupe B alors que l'allèle dérivé caractérise les autres. Par ailleurs les analyses interspécifiques permettent de démontrer que l'haplogroupe A est plus ancien que B. Les haplogroupes A et B sont uniquement retrouvés en Afrique où ils constituent le fond africain de ce polymorphisme. Cela signifie que les migrations « out of Africa » sont survenues après l'apparition des mutations des deux allèles ancestraux caractérisant ces deux haplogroupes. Le deuxième type de polymorphisme est représenté par la présence d'une **Insertion Alu**. A un locus donné, chez certains individus, on note la présence d'une séquence supplémentaire d'ADN de 500 pb probablement dérivé d'un ARN ribosomal.

Ce polymorphisme est un phénomène général et peut être constaté à divers endroits du génome. Sur le chromosome Y, il est noté YAP (Y Alu Polymorphism). Il est détecté à un locus déterminé qui est encadré par deux amorces spécifiques et sa présence est révélée par une bande qui migre moins loin et moins vite sur le gel en raison de la présence de cette séquence supplémentaire. La détection de YAP permet de suspecter la présence de l'haplogroupe D ou E dont la distinction est basée sur l'analyse du marqueur SNP SRY 4064 dont l'allèle ancestral caractérise l'haplogroupe D et le dérivé l'haplogroupe E. La forme ancestrale de ce SNP, représentant un site de restriction, explique que l'haplogroupe D soit, lors d'une RFLP, caractérisée par deux bandes courtes alors que l'haplotype E est caractérisé par une seule bande plus longue. Le marqueur YAP est fréquent en Afrique mais non exclusif (fréquence D + fréquence E). L'haplogroupe E est caractéristique de l'expansion Bantoue. Enfin, le dernier type de polymorphisme est caractérisé par des **microsatellites** dont l'analyse permet d'affiner le polymorphisme au sein d'un haplogroupe. Le taux de mutation de ces microsatellites étant beaucoup plus rapide que celui des SNP.

Les analyses des marqueurs polymorphes du chromosome Y ont permis d'évoquer les principales vagues migratoires au niveau planétaire. Nous présentons dans le chapitre suivant une partie de ces hypothèses migratoires évoquées par S.Wells ainsi que quelques populations « étapes » caractérisant ce processus.

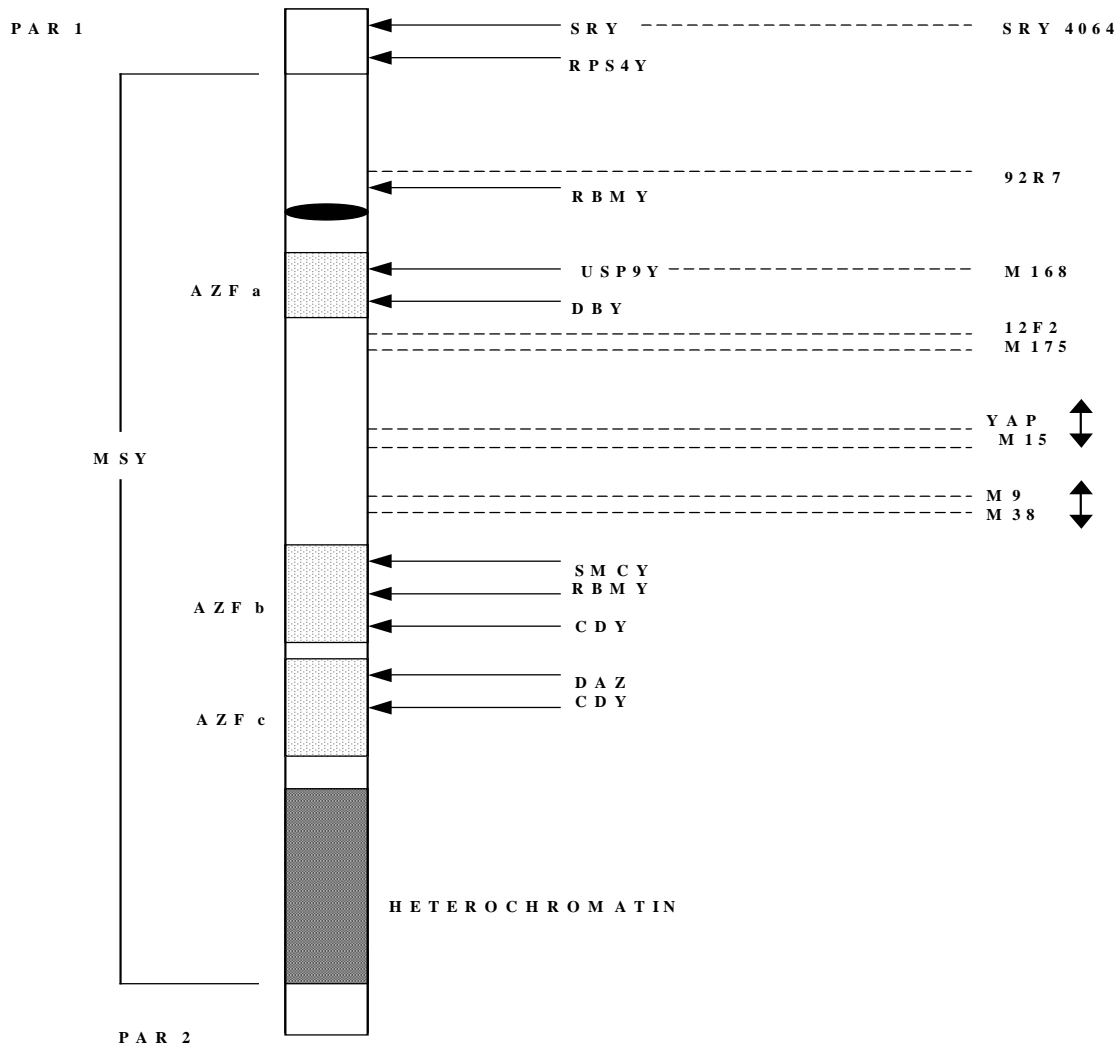


Figure 17 : Organisation du chromosome Y (Jobling MA, 2003)

*III.4.4.b. Quelques populations représentant les principales étapes migratoires (Wells S, 2003).*

- **Les SAN.**

Les SAN sont une population indigène d'Afrique du Sud. Ils vivent aujourd'hui au Botswana, en Namibie et au Nord West de l'Afrique du Sud. On pense qu'ils sont des descendants directs des premiers Homo Sapiens qui apparurent il y a 100 000 ans. Les colons hollandais les baptisèrent les « Bushmen » (Hommes de la brousse). D'un point de vue génétique les marqueurs retrouvés chez les SAN sont très différents de ceux retrouvés hors d'Afrique. Dans l'arbre génétique de l'humanité leur branche est la première à se détacher du tronc. D'un point de vue morphologique, leur visage comporte une multitude de caractères appartenant aux autres populations : forme des yeux moyen-orientaux, pommettes hautes des mongoles, la peau brune pouvant devenir plus ou moins sombre. Les recherches montrent qu'ils quittèrent l'Afrique et mirent en mouvement l'arbre génétique de l'homme.

- **Les PIRAMALAI KALLAR (PK).**

La tribu des PK habite les provinces isolées de Madurai, une ville du Sud de l'Inde. Leurs caractéristiques physiques diffèrent de celles de leurs voisins (leur peau est par exemple plus sombre) et donnent à penser qu'ils pourraient descendre du tout premier groupe d'humains venu d'Afrique il y a 50 000 ans. Comme nombre de communautés de la région, les PK, sont fermiers et guerriers de génération en génération. Leur ADN est similaire à celui des Aborigènes Australiens. Les premiers Africains semblent donc s'être arrêtés en Inde sur la route de l'Australie et leurs descendants sont les PK. Si les premiers émigrants venus d'Afrique sont passés par ici alors leurs gènes, dont le furtif marqueur australien, pourrait subsister quelque part. Il n'existe aucun témoignage archéologique de la présence des premiers hommes dans le sud de l'Inde. Ces villageois sont un échantillon de choix pour rechercher l'antique marqueur car ils vivent dans cette région depuis des générations et très peu d'étrangers ont été admis à épouser des membres de la communauté.

- **Les ABORIGENES.**

Les Aborigènes d’Australie arrivèrent sur le continent il y a environ 50 000 ans. Ils venaient d’Afrique et avaient traversé l’Inde avant de franchir l’Océan qui les séparait de l’Australie. Quand les Européens colonisèrent cette région à la fin du XVII<sup>e</sup> siècle, il y avait 300 000 à 1 Million d’Aborigènes divisés en 500 groupes tribaux et parlant 200 langues différentes. Avant la colonisation les Aborigènes vivaient de chasse de cueillette. Ils conçurent leurs propres mythes de la création, transmis oralement à travers les récits chantés de génération en génération jusqu’à nos jours.

- **Les OUÏGOURES.**

Les hommes arrivèrent dans la région du KAZAKHSTAN environ 40 000 ans avant notre ère. Ils venaient d’Afrique via le Proche Orient. Si l’homme est né en Afrique, l’Asie Centrale semble bien avoir été son jardin d’enfants. De fait, presque toutes les populations qui ne sont ni d’origine Africaine ni d’origine Aborigène remontent à des ancêtres d’Asie Centrale.

- **Les CHUKCHIS.**

Les CHUKCHIS vivent dans l’extrême Nord Est de la Sibérie. Ils descendent des premiers hommes qui arrivèrent d’Asie Centrale il y a 20 000 ans. Le nom de CHUKCHIS est une forme de CHACHU qui, en langage CHUKCHI, signifie « qui possède de nombreux rennes ». La survie des CHUKCHIS dépend de leur capacité à chasser, domestiquer et conduire leurs troupeaux de rennes. Le renne assure le transport de leur tente et le déplacement de ces populations nomades. Aujourd’hui ils sont de moins en moins nombreux à suivre les rennes, passant les rudes hivers arctiques en ville. On resense actuellement 14 000 CHUKCHIS.

- **Les NAVAJOS.**

La tribu des Navajos descend des premiers hommes qui passèrent de la Sibérie à l'Alaska il y a 13 000 ans. Ce groupe ne comptait pas plus de 10 à 20 voyageurs. Ils sont descendus du Canada jusqu'au Sud Ouest des Etats Unis où ils ont vécu en Nomades de chasse et de culture. Quand les Espagnols introduisirent les moutons au début du XVII<sup>e</sup> siècle ils se sédentarisèrent et devinrent bergers. A cette époque, on recensait 9 000 Navajos et , en 1868, ils se virent attribuer un territoire couvrant des régions du Nouveau Mexique, de l'Arizona et de l'Utha. Aujourd'hui ils sont 170 000 et on retrouve dans leurs gènes les traces des premiers voyageurs.

*III.4.4.c.Schéma migratoire selon les marqueurs du chromosome Y. (Figure 19bis p.108 d'après Wells S, 2003)*

**M168** : marqueur « Out of Africa ».

Il est présent chez tous les hommes qui vivent hors d'Afrique. Il est apparu en Afrique de l'Est il y a 50 à 60.000 ans. Il a engendré 3 grands lignages M130, M89 et le M1.

**M130** : marqueur «côtier ».

Ce marqueur asiatique côtier est fréquent chez les Australiens. On le retrouve au Nouveau-Mexique où l'ADN des Navajos témoigne du voyage entrepris il y a 50.000 ans.

**M89** : marqueur « Moyen-Orient ».

Il trace des populations qui chassaient en Afrique du Nord et au Proche-Orient il y a 45.000 ans. Il caractérise également 90% des populations d'Asie Centrale parties d'Afrique via le Proche-Orient. Le développement de l'agriculture les auraient conduits jusqu'au Nord-Ouest de l'Europe.



**M9** : marqueur « Eurasien ».

Le marqueur des « steppes » caractérise une grande majorité des populations de l'hémisphère Nord. Il est apparu en Iran ou en Asie centrale il y a 40.000 ans.

**M20** : marqueur « Indien ».

Il est présent chez 50% des habitants du Sud de l'Inde. Il identifie le premier groupe qui, venant du Proche-Orient, s'est installé en Inde il y a 35.000 ans.

**M45** : marqueur d' « Asie Centrale ».

Il caractérise des populations qui vivaient dans les steppes asiatiques il y a 35.000 ans. Disséminés partout dans le monde les descendants du M 45 vivent en Europe (M 173), en Amérique (M 242) et en Inde (autres lignages du M 45)

**M173** : marqueur d' « Europe de l'Ouest ».

Il est le principal marqueur européen. Il caractérise la principale colonie européenne qui vivait pendant le paléolithique ; on le surnomme le marqueur « Cro-Magnon ». Apparu il y a environ 30.000 ans, probablement en Asie Centrale, il s'est déplacé vers l'Ouest en passant par les steppes, porté par les chasseurs de Mammouths, les hommes des cavernes et les chasseurs-cueilleurs. 70% des Britanniques et 95% des Irlandais et des Basques le portent.

**M242** : marqueur « Sibérien ».

Apparu il y a 15 à 20.000 ans dans le Sud de la Sibérie ou en Asie Centrale et s'est déplacé vers le Nord- Est de l'Asie au cours des millénaires qui ont suivi.

**M3** : marqueur « Américain ».

Il est largement répandu du Nord au Sud du continent Américain. Il est apparu il y a environ 15.000 ans.

**M175** : marqueur d' « Asie Orientale ».

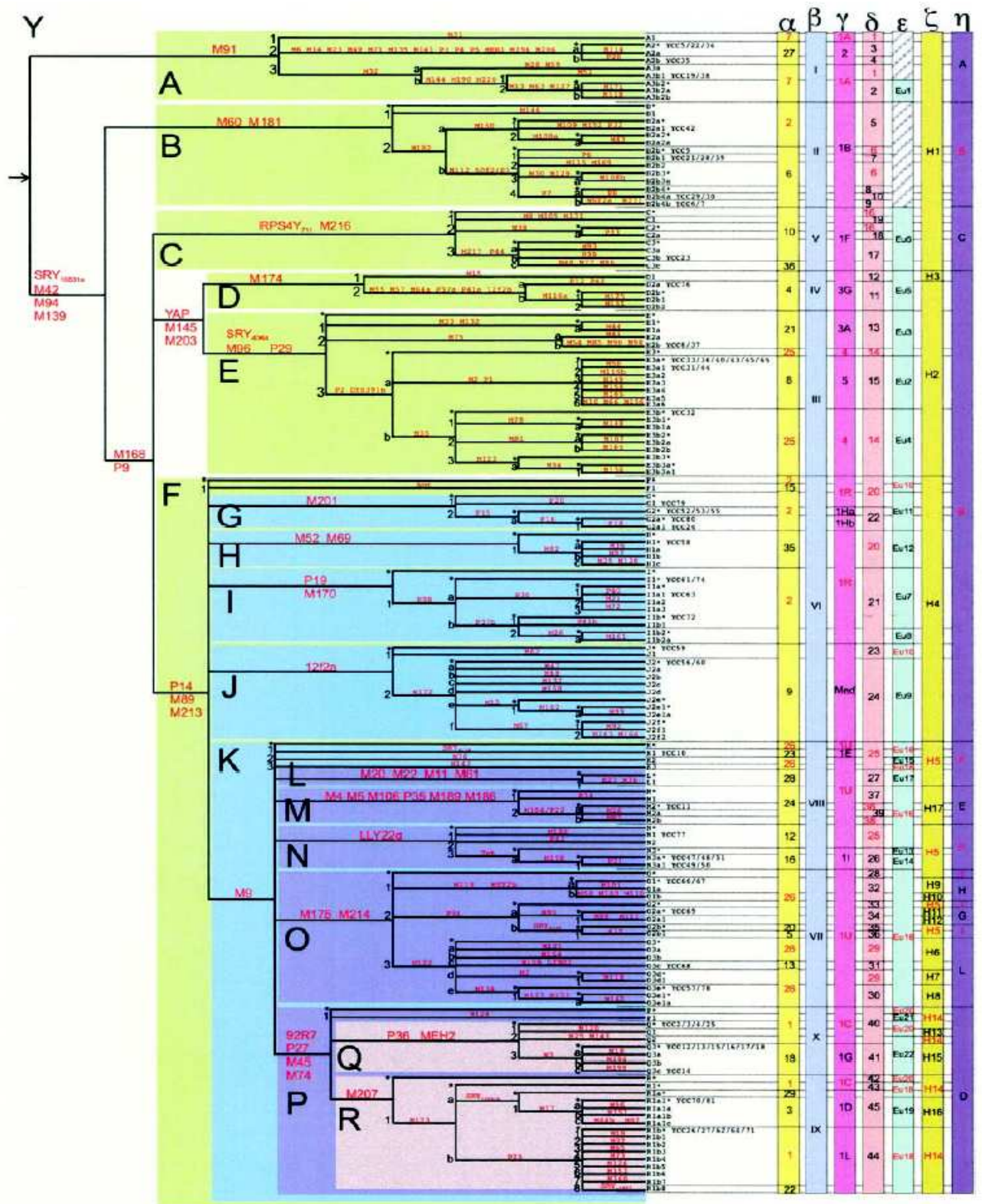
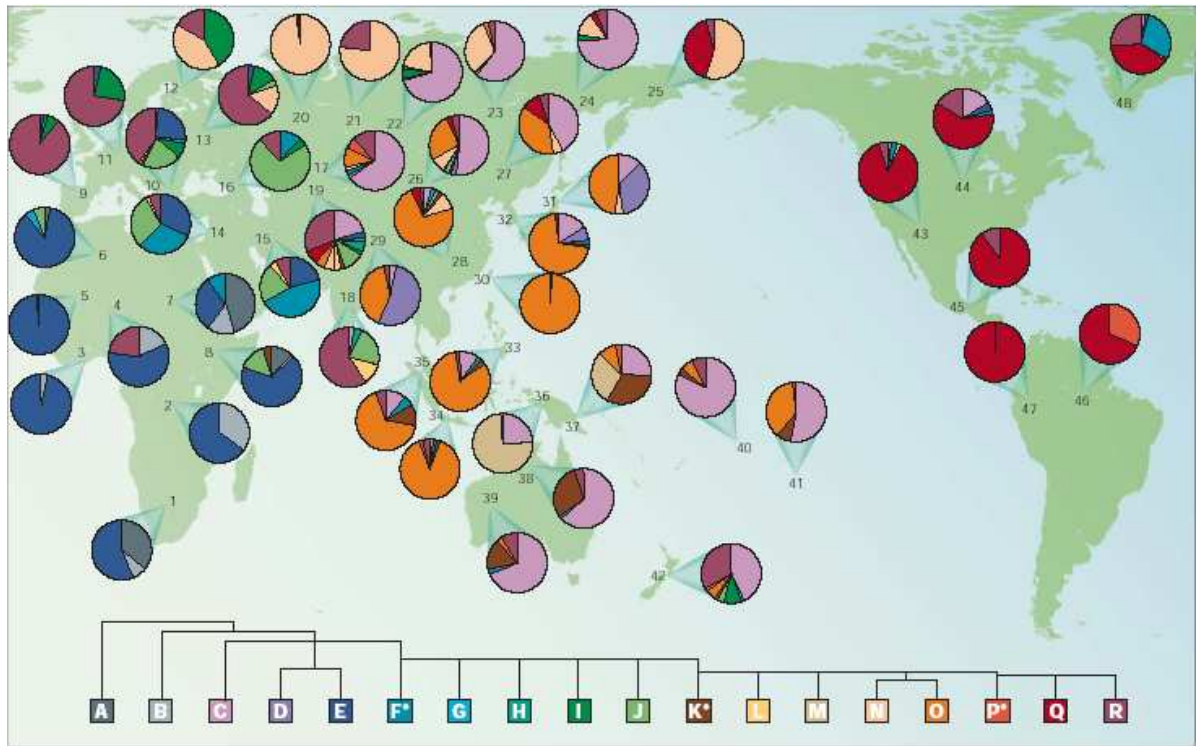


Figure 18 : Organisation des différents haplogroupes (Y Chromosome Consortium, 2002)

Les haplogroupes A et B sont presque exclusivement Africains (Figure 19 p.107) et représentent encore aujourd'hui la majorité des chasseurs-cueilleurs ou de leurs descendants. Toutefois, en Afrique, les évènements du Paléolithique sur la diversité du chromosome Y ont été estompés par une importante et récente expansion de l'haplogroupe E. La dispersion de cet haplogroupe est probablement en rapport avec l'expansion des fermiers Bantous débutée en Afrique de l'Ouest 2 à 3 000 ans auparavant. C'est la raison pour laquelle, aujourd'hui, les haplogroupes A et B sont relativement rares et la majorité des chromosomes Y africains comportent l'haplogroupe E. Un autre exemple migratoire récent est caractérisé par un retour en Afrique subsaharienne de populations ayant apporté, via l'Afrique du Nord, l'haplogroupe R dans le Nord du Cameroun.

En dehors de l'Afrique, on retrouve les haplogroupes H et L qui restent localisés en Inde, une partie de l'haplogroupe C qui est retrouvée en Océanie, le reste en Mongolie, Sibérie, et pour une partie au Nord-Ouest du continent Américain (locuteurs Na-Dene). L'haplogroupe D se retrouve en Asie du Sud-Est et au Japon. L'expansion des populations de l'Asie du Nord subsiste essentiellement en Asie de l'Est avec l'haplogroupe O, tandis que la branche N s'est propagée principalement au Nord-Est de l'Europe parmi les locuteurs Ouraliens. L'haplogroupe I, provenant du Nord de l'Asie a été à l'origine de ce qui est probablement la première grande expansion Paléolithique en direction de l'Europe centrale. Les haplogroupes G et J sont retrouvés actuellement au Moyen-Orient et se sont ensuite étendus à l'Europe, essentiellement dans le Sud avec les fermiers du Néolithique. L'haplogroupe R est trouvé en Europe, en Inde, au Pakistan et en Amérique même si une branche antérieure semble être retournée dans la partie centrale du Sahel via l'Afrique du Nord. L'haplogroupe Q définit majoritairement la population Amérindienne, à l'exception des Eskimos et des locuteurs Na-Dene. L'haplogroupe I est également retrouvé dans le Nord de l'Europe Centrale où il a probablement émergé il y a environ 20 000 ans (Cavalli-Sforza L, 2003).



**Figure 19** : Distribution schématique des haplogroupes Y. Chaque cercle représente une population et les fréquences des 18 haplogroupes majeurs. (D'après Joinbling MA, 2003)



Figure 19 bis : Les premières migrations de l'homme d'après les marqueurs du chromosome Y (d'après Wells S, 2003)

## **MATERIEL et METHODES.**

## I. Obtention et caractéristiques des Echantillons.

L'obtention des échantillons a nécessité plusieurs étapes :

**a) Elaboration du protocole.**

L'analyse de la littérature nous ayant fait porter un regard critique sur les modalités de définition des échantillons, nous avons comme objectif d'identifier avec précision un échantillonnage exploitable et significatif d'un point de vue anthropologique.

Notre objectif était donc d'obtenir un nombre suffisant de sujets d'origine comorienne sans lien de parenté avec un sexe ratio équilibré. Pour cela, nous devions donc préciser l'origine de chaque individu, de ses parents et grands-parents, leur lieu de naissance avec positionnement sur une carte (Carte 6 p.114), leur langue et bien entendu leur identité. Il convenait par ailleurs de préciser l'absence de parenté et d'obtenir un effectif minimum suffisant que nous avons fixé à 150 individus.

**b) Elaboration du consentement et de la notice d'information préalable** (Annexe 1 p .210).

**c) Elaboration du questionnaire** relatif aux critères d'inclusion et au recueil d'informations relatives à chaque individu (Annexe 2 p.211).

**d) Demande d'autorisation CCPPRB** conformément à la réglementation relative à la protection des personnes participant à la recherche biologique. Nous avons présenté à la Commission les garanties nécessaires à la mise en œuvre d'une telle étude, en particulier la confidentialité des données. Ce point a fait l'objet d'une attention particulière de la part des membres de la commission en raison du caractère sensible de la gestion de données nominatives associant profil génétique et



spécifications populationnelles. La première demande est datée de Octobre 2000 et l'autorisation a été obtenue en Mai 2001 (Annexe 3 p.213).

**e) Etablissement des contacts avec la communauté Comorienne de Marseille.** Le premier contact s'est réalisé par une approche de type médical. Connaissant les habitudes médicales en terme de lieu de consultation, nous avons inclus dans la liste des investigateurs deux médecins responsables des Mutuelles de Provence (Dr I.Koné, Dr G.Drouin). De plus nous avons pu organiser deux séances de présentation du protocole à certains responsables d'association de Comoriens. Cette première approche n'a eu qu'une efficacité limitée (30 échantillons en 1 an). L'une des causes repose sur le foisonnement associatif de cette communauté avec parfois inhibition mutuelle des différentes actions. Compte tenu de ces premiers résultats, il nous est apparu nécessaire d'envisager une autre approche plus individuelle passant par l'identification d'une personne connaissant ce réseau communautaire complexe, capable de comprendre et d'explicitier la problématique, devant être largement connue et appréciée par la communauté en vue d'apporter les garanties de confiance nécessaire y compris vis à vis des autorités religieuses locales. Grâce à nos premiers contacts « mutualistes » nous avons donc rencontré un étudiant d'origine comorienne (M. K.Papa) connaissant parfaitement les méandres communautaires de Marseille et bien entendu apprécié de l'ensemble de la communauté. Nous avons donc pu obtenir, en 4 séances de prélèvements organisées sur le site de Marseille de l'Etablissement Français du Sang, près de 200 individus correspondants aux critères préalablement définis. De plus, sa bonne connaissance des personnes sensibilisées, nous a permis de nous assurer de l'absence de liens de parenté directe entre des différents individus.

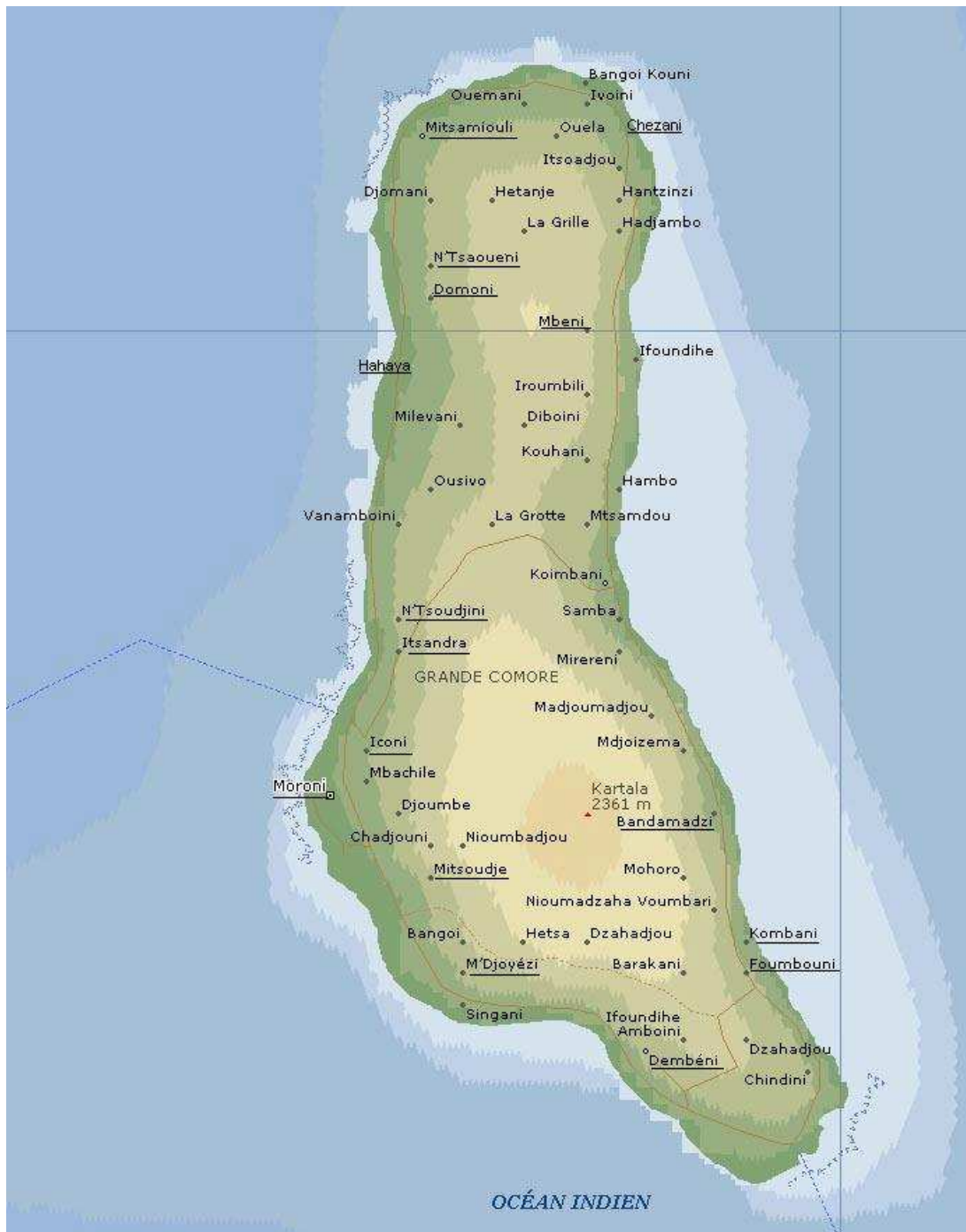
Au total, 200 individus répondant aux critères d'inclusion suivants ont été prélevés :

- prélèvements aléatoires de sujets,
- sans lien de parenté entre eux,
- dont les parents et les quatre grands-parents sont originaires des Comores
- dont la langue maternelle ainsi que celle de leurs parents et grands-parents est le Comorien.

L'échantillon, que nous avons étudié, peut être considéré comme représentatif de la population comorienne et notamment de la Grande Comore pour les raisons suivantes :

- Importance de la communauté dans laquelle s'est effectué l'échantillonnage. Marseille compte en effet la plus importante communauté comorienne émigrée, soit environ 70.000 sujets.
- L'entretien préalable des individus qui nous a permis de confirmer l'axe migratoire Grande Comore → France et de montrer une répartition homogène de leur origine sur l'ensemble de l'île (Carte 6 p.110).
- Le caractère massif et récent (1970) de l'émigration comorienne vers la France qui concerne 10% de la population avec l'arrivée de familles entières (Direche-Slimani K, 2002).
- Enfin, compte tenu de la transposition fidèle de leurs coutumes et traditions dans le pays d'accueil, les unions continuent de se réaliser à l'intérieur de leur communauté et il y a encore très peu d'unions avec des individus de la population Française ou d'autre communauté résidant en France.

Les prélèvements ont été réalisés sur EDTA. Trois échantillons de 6 ml ont été prélevés par personne ; un pour le laboratoire d'immuno-hématologie « sérologie », un pour le laboratoire immuno-hématologie « biologie moléculaire » et un pour le laboratoire HLA. Les analyses sérologiques d'immuno-hématologie sont pratiquées dans les 3 jours et le reste des hématies est congelé en azote liquide à  $-190^{\circ}\text{C}$ . Les analyses de biologie moléculaire (HLA et érythrocytaires) sont différées. En ce qui concerne le laboratoire HLA, dès réception des échantillons → Lyse cellulaire (SLR) / Récupération des leucocytes / Purification de l'ADN génomique / Quantification de l'ADN par spectrophotométrie / Congélation de l'ADN à  $-80^{\circ}\text{C}$ . En ce qui concerne le laboratoire de Biologie Moléculaire érythrocytaire, dès réception des échantillons → Congélation à  $-40^{\circ}\text{C}$  du sang total (2 x 200  $\mu\text{l}$  et 2 x 1 ml) et du plasma.



Carte 6 : Répartition de l'origine géographique des individus prélevés (villages soulignés).

## **II. Méthodes d'exploration et d'analyse du polymorphisme érythrocytaire.**

*Cette partie a fait l'objet d'un article actuellement soumis à Human Biology (Chiaroni J, 2003).*

### **II.1 Détermination des groupes sanguins érythrocytaires.**

L'ensemble des analyses a été réalisé dans les mêmes conditions techniques de routine appliquées à la qualification biologique des dons sur le plateau technique de Marseille de l'EFS Alpes-Méditerranée. Les groupages ABO.D ainsi que les phénotypages Rh.K (C, E, c, e, K) ont été réalisés en technique automatique micro plaque sur Olympus PK 7200. La détermination des antigènes Fya, Fyb, Jka, Jkb et S a été réalisée en technique automatique sur colonne de filtration (Scangel. Biorad. France). Enfin, la détermination des antigènes M, N, s, k (Cellano) et Kpa a été réalisée en technique manuelle tube ou filtration. Chaque antigène a été déterminé à deux reprises en utilisant deux lots (ou deux clones pour les réactifs monoclonaux) différents. Les réactifs ont été utilisés conformément aux instructions du fournisseur (Diagast, Biotest et Biorad). Les échantillons non utilisés ont été congelés pour partie (hématies) en azote liquide à  $-190^{\circ}\text{C}$  pour partie (ADN et plasma) à  $-40^{\circ}\text{C}$ .

### **II.2 Détermination du génotypage Duffy**

Le génotypage Duffy a été déterminé par PCR avec utilisation d'amorces allèles spécifiques (PCR-ASP). L'ADN génomique a été extrait de 200 $\mu\text{l}$  de sang total en utilisant le kit QIAamp blood DNA MiniKit (Qiagen, Hilden, Germany). La méthode de PCR – Allèle spécifique mise en œuvre pour la détection simultanée des quatre allèles *FY* (*FY*\*A, *FY*\*B, *FY*\*X et *FY*\*Fy) a été développée et validée dans notre laboratoire à partir de la méthode publiée par Olsson ML (Olsson ML 1998a, 1998b). Les modalités techniques sont décrites en Annexe 4 p.221

### **II.3 Recherche des allèles rares du gène RHCE par biologie moléculaire**

Cette recherche concerne à ce jour la détection des allèles liés au phénotype RH :-34 et aux phénotypes RH5. La détection des allèles liés aux phénotypes RH :-18 et RH :-46 est toujours en cours de validation technique. Les modalités techniques sont décrites en Annexe 5 p.224. La mise en œuvre des différentes PCR a suivi le plan d'exploration suivant :

#### **Détection des sujets RH :-34 (Hr<sup>B</sup> négatif ou (C) ce<sup>S</sup>).**

*Cet haplotype est caractérisé par la présence d'un gène RHD hybride et de deux mutations au niveau du gène RHCE. L'une en position 733C>G dans l'exon 5 (allèle ce<sup>S</sup>) et l'autre en position 1006G>T dans l'exon 7. Afin d'estimer la fréquence de cet haplotype, l'ensemble des PCR sont réalisées chez tous les individus RH : 2, 4, 5. Toutefois, en ce qui concerne, la détection des sujets de phénotype RH :-34 les 3 PCR ne sont mises en œuvre que chez les sujets de phénotype ddCee*

Réalisation d'une PCR détectant l'allèle **733G** chez les sujets RH : 2, 4, 5.

Si négative : stop - Si positive → PCR de détection de l'allèle **1006T**

Si négative : stop - Si positive → PCR de détection du gène **Hybride D (3-8) CE- D**

Si négative : stop - Si positive → PCR de détection des allèles sauvages : **733C,1006G,Ex3.**

Si l'une est positive : le sujet est hétérozygote – Si toutes sont négatives : **sujet RH :-34.**

#### **Détection des sujets RH:-46**

*Cet haplotype est caractérisé par un gène RHCE hybride CE (5)-D (4)-CE. L'estimation de la fréquence de cet allèle repose sur la mise en œuvre d'une PCR chez tous les sujets RH : 2,5. S'il s'agissait de dépister uniquement les sujets RH :-46, cette PCR ne serait réalisée que chez les sujets RH :1,2,-3,-4,5 (R1R1). Toutefois, en ce qui concerne, la détection des sujets de phénotype RH :-46 les 3 PCR ne sont mises en œuvre que chez les sujets de phénotype DDCCee*

Réalisation d'une PCR détectant le gène **RHCE Hybride** chez les sujets RH : 2, 5.

Si négative : stop - Si positive → PCR de détection de l'allèle sauvage : **Exon 5 RHCE**

Si positive : le sujet est hétérozygote - Si négative : **sujet RH:-46.**

### **Détection des sujets RH5 partiels.**

*L'antigène RH5 partiel peut être produit par deux types de mécanismes. L'un est représenté par l'allèle 733G qui comporte en plus une mutation en position 340C>T (MAYET) dans l'exon 3 du gène RHCE. L'autre est représenté par l'allèle ceMO (MORI) qui est caractérisé par une mutation en position 667T>G dans l'exon 5 du gène RHCE. L'estimation de la fréquence de ces allèles repose sur la mise en œuvre de PCR chez les sujets RH5.*

Réalisation d'une PCR détectant l'allèle **733G** chez les sujets RH : 5.

Si négative : stop - Si positive → PCR détectant l'allèle **340T**.

Si négative : stop - Si positive → PCR allèle sauvage **340C**.

Si positive : le sujet est hétérozygote - Si négative : **sujet RH5 partiel ce<sup>s</sup> (340)**

Réalisation d'une PCR détectant l'allèle **667G**.

Si négative : stop - Si positive → Séquençage pour nouvel allèle et PCR allèle sauvage **667T**

Si positive : le sujet est hétérozygote - Si négative : **sujet est RH5 partiel ceMO (MORI)**

### **II.4 Analyse des données.**

**a) Les fréquences géniques** ont été déterminées par la méthode du maximum de vraisemblance basée sur l'hypothèse d'équilibre de Hardy Weinberg, en utilisant le programme Arlequin (Schneider S, 2000).

**b) Les distances génétiques** ont été calculées à partir des 23 fréquences alléliques en utilisant la méthode de Reynolds (Reynolds J, 1983) avec le logiciel Phylip 3.5c (Felsenstein J, 1989). Les arbres phylogénétiques ont été construits à l'aide du logiciel Neighbor du package Phylip et Treview (Page, R. D. M. 1996 ; Saitou N, 1987).

**c) Les analyses multidimensionnelles** ou nonmetric multidimensional scaling (NMDS ou MDS) ont été réalisées à l'aide du logiciel SPSS 10.0. Ces analyses cherchent à représenter, dans un nombre réduit et préimposé d'axes, les relations entre les populations. Pour un nombre donné et petit d'axes (2 ou 3), le MDS permet de représenter une matrice de distances avec moins de déformations qu'une analyse en composante principale.

**c) La classification ascendante hiérarchique (CAH)** a pour objectif principal de répartir les éléments d'un ensemble en groupes, c'est à dire d'établir une partition de cet ensemble. De plus, on réalise une hiérarchie de partitions qui constitue un arbre appelé le dendrogramme. Dans un dendrogramme, la hauteur des nœuds a une signification (plus le nœud est bas plus les éléments regroupés sont proches). Par contre, la position verticale est arbitraire. La CAH repose sur un algorithme itératif à deux étapes : à la première étape, on réunit les deux objets les plus proches pour former un nouveau groupe, et à la seconde étape, on recalcule la distance entre le groupe qui vient d'être créé et tous les autres objets. On recommence ensuite jusqu'à ce qu'il n'y ait plus qu'un seul objet constitué de la réunion de tous les éléments initiaux de l'ensemble de départ. La méthode que nous avons utilisé pour le calcul de la distance entre deux objets est la méthode du lien moyen, ou distance moyenne (average linkage). Elle correspond à la méthode de classification automatique la plus employée, aussi appelée UPGMA (Unweighted Pair Group Method of Aggregation).

**d) Les calculs des contributions génétiques (admixture)** ont été réalisés en utilisant le logiciel ADMIX95 selon la méthode décrite par Chakraborty (Chakraborty R, 1985). Ces calculs ont pris en compte le modèle tri-hybride basé sur l'utilisation des populations originelles putatives et sur les valeurs des fréquences génétiques.



### **e) Les populations comparées.**

Les calculs de distances génétiques ont pris en compte les 23 fréquences alléliques décrites dans différentes populations Africaines et du pourtour de l'Océan Indien. Ces populations sont présentées en (Table 10 p.121). Compte tenu des données historiques nous avons retenu la majorité des populations Africaines comme les Bétéké (Bantous, Gabon), les Bédik (Sénégal), les Béja (Ethiopie), les Pygmées Biaka (Congo), les Bitonga (Bantous, Mozambique), les Hadza (Tanzanie), Kavango (Bantous, Afrique du Sud), les Kgalagadi (Bantous, Bostwana) , les Libyens (Fezzan, Libye), les San (Afrique du Sud), les Somalis (Somalie) et les Zéruru (Bantous, Zimbawe). Pour la Péninsule Arabique, nous avons inclu une population de Saoudiens (Arabie Saoudite). Pour le continent Indien, nous avons retenu une population de langue Dravidienne les Gadaba (Andhra Pradesh). Enfin, pour le Sud-Est Asiatique nous avons inclus les Indonésiens (Indonésie) et les Toba-Batak (Sumatra Nord).

En ce qui concerne l'origine des données, le travail a débuté par une analyse de la littérature au cours de laquelle nous avons été confrontés à deux problèmes. L'un concernait les modalités de définition et de description des populations étudiées, l'autre concernait la nécessité d'avoir les mêmes loci étudiés dans plusieurs populations afin de mener des analyses comparatives multiloci. En ce qui concerne le premier problème, l'analyse des données originales révèle une grande disparité et des manques de précision relatifs à l'origine géographique et ethnique. Certaines études publient des fréquences géniques qui sont directement affectées à une population occupant une zone géographique donnée. D'autres sont imprécises quant à la population étudiée qui peut être définie comme « Africains » par exemple ou dont les résultats sont basés sur l'analyse de donneurs de sang d'une grande ville. Par ailleurs, beaucoup de données transgressent les règles de la démarche scientifique qui supposent que l'on sélectionne des échantillons préalablement définis, alors que bon nombre d'études ont été menées en rétrospectif. En ce qui concerne les analyses multiloci, la problématique est représentée par la nécessité de trouver les 6 loci érythrocytaires étudiés dans les populations sélectionnées en vue de comparaison.

La difficulté est liée au fait que les loci étudiés ne sont pas forcément les mêmes dans chaque population et le nombre de gènes étudiés est limité.

Nous avons donc analysé 77 articles originaux concernant l'exploration sérologique du polymorphisme érythrocytaire de près de 120 populations (Table 11 p.124) d'Afrique, de la Péninsule Arabique, de l'Inde et de l'Asie du Sud Est. Sur ces 120 populations seule une vingtaine comportait les données relatives aux 6 loci érythrocytaires que nous avons étudiés. Parmi ces 20 populations nous avons sélectionné les populations précédemment décrites (Table 10 p.121).

#### **f) Cas particulier de l'analyse comparative des phénotypes entre population de Donneurs de Sang et population Comorienne de Marseille.**

Notre problématique biomédicale étant de trouver du sang compatible en vue de transfusion, une analyse comparative a été menée entre les fréquences phénotypiques des donneurs de sang de l'EFS Alpes-Méditerranée et les fréquences phénotypiques retrouvées dans notre échantillon de la communauté comorienne résidant à Marseille.

Populations	Localisation	Langue	Auteurs	Année
Somali	Somalia	Afro-Asiatic	Sistonen J	1987
Bitonga	Mozambic	Niger Congo-Bantu	Matznetter T	1969
Béja	Sudan	Afro-Asiatic	El Hassan AM	1968
Zéruru	Zimbawee	Niger Congo-Bantu	Lowe RF	1979
Sandawee	Tanzania	Khoisan	Godber	1976
Bétéké	Gabon	Niger Congo-Bantu	Languillat G	1980
			Gueguen A	1980
Kgalagadi	Botswana	Niger Congo-Bantu	Jenkins T	1987
Kavango	South Africa	Niger Congo-Bantu	Nurse GT	1977b
Hadza	East Tanzania	Khoisan	Tills D	1982
Biaka	Congo	Niger Congo-Bantu	Cavalli-Sforza L	1986
			Rouger P	1982
Bédik	Senegal	NC-West-Atlantic	Bouloux C	1972
			Gessain R	1965
Saoudiens	Saudia	Afro-Asiatic	Abdelaal MA	1999
			Chaabani H	2000
San	South Africa	Khoisan	Nurse GT	1977a
Libyens	Fezzan - Libya	Afro-Asiatic	Mourant AE	1976
			Simonovic B	1968
			Woodfield G	1970
			Corrain C	1968
Gadaba	Andhra Pradesch – India	Dravidian	Rao PM	1999
Indonésiens	Indonesia	Austric	Jacob T	1974
Toba Bataks	North Sumatra	Austric	Hawkins B	1973

**Table 10** : Les populations sélectionnées en vue de comparaison (polymorphisme érythrocytaire).

Populations	Auteurs	Année	Systèmes étudiés						Localisation
			ABO	FY	K	JK	MNS	RH	
Somali AFE	Sistonen	1987	x	x	x	x	x	x	Somalie
Bitonga BSE	Spielmann	1970	x	x	x	x	x	x	Mozambique
Sandawe KHO	Godber	1976	x	x	x	x	x	x	Tanzanie
Nyaturu BNE	Godber	1976		x	x	x	x	x	Tanzanie
Kumana AFE	Mourant	1974	x	A	x		x	x	Erythrée/Ethiopie
Baria AFE	Mourant	1974							Erythrée/Ethiopie
Kisi BCE	Hauvast	1971	x						Tanzanie
Hadza W KHO	Tills	1982	x	x	x	x	x	x	Tanzanie
Hadza E KHO	Tills	1982	x	x	x	x	x	x	Tanzanie
Afar AFE	Fourquet	1969	x				MN	x	Ethiopie
Soudanais AFE	Wasfi	1980	x	x	x		x	x	Soudan
Kenya AFE	Mwangi	1999	DDS						Nairobi
Kikuyu	Beecher	1967	x						Kenya
Wakamba	Beecher	1967	x						Kenya
Gusii	Beecher	1967	x						Kenya
Abaluhya	Beecher	1967	x						Kenya
Luo	Beecher	1967	x						Kenya
Nandi	Beecher	1967	x						Kenya
Saoudien ARA	Abdelaal	1999	x	x	x	x	x		Arabie Saoudite
Yéménite ARA	Chaabani	2000						x	Yémen
Berbere AFN	Chaabani	1988	?						Tunisie
Saoudien ARA	Bashwari	2001	x						Arabie Saoudite
Shia ARA	Al-Hilli	1985	x					D	Barhein
Suni ARA	Al-Hilli	1985	x					D	Barhein
Saoudien ARA	Maranjan	1966							Arabie Saoudite
Iran SW MO	Bajtzadeh	1966							Iran
Iran Golf MO	Nijenhuis	1964							Iran
Liban	Kfoury	2001	DDS						
Jordanien MO	Irshaid	2002	DDS						Jordanie
Dacca	Rahman	1973	DDS						Bengladesh
Inde sud	Das	2001	DDS						
Gadaba Drav	Rao	1999	x	x	x	x	x	x	Andhra Pradesh
21 tribues	Babu	2001	x					D	Andhra Pradesh
Rajaka Inde	Parvatheesam	1997	x					D	Andhra Pradesh
Mali Inde	Babu	1999	x				MN	x	Andhra Pradesh
Birmanie	Mya-Tu	1971	x		x		x	x	Birmanie
Tai/Bangkok	Nathalang	2001	DDS						Thaïlande
Red Karen	Shimizu	1997						x	Thaïlande
White Karen	Shimizu	1997						x	Thaïlande
Black Lahu (54)	Shimizu	1997						x	Thaïlande
Shan (62)	Shimizu	1997						x	Thaïlande

Lisu (74)	Shimizu	1997							x	Thaïlande
Manni (30)	Shimizu	1997							x	Thaïlande
Bali	Breguet	1982	x	x	x	x	x	x	x	Bali
Indonésie	Jacob	1974	x	x	x	x	x	x	x	Bali
Bali	Adipura	1977	DDS							Bali
Toba-Bataks	Hauwkins	1973	x	x	x	x	x	x	x	Sumatra Nord
Asie SE/Océanie	Shimizu	2000		x						ASE/océanie
Filipinos	Windhof	1983	x	x	x	x	x	x	x	Philippines
Singapore	Mohan	1987	x							Singapoure
Malaisie	Lewis	1988		x						Malaisie
Malaisie	Chattapadhyay	1977	x							Malaisie
Chinois/Malay	Chattapadhyay	1977	x							Malaisie
Seychelles	Welch	1975	x	x				x	x	Seychelles
Réunion	Fernandez	1998	H							Réunion
Réunion	Lependu	1983	H							Réunion
Réunion	Gérard	1982	H							Réunion
Réunion	Lependu	1983	H							Réunion
Mossi AFW	Broussal	1979	x							Hte-Volta
Gourounsi	Broussal	1979	x							Hte-Volta
Bissa	Broussal	1979	x							Hte-Volta
Samo	Broussal	1979	x							Hte-Volta
Peulh	Broussal	1979	x							Hte-Volta
San	Nurse	1977	x	x		x	x	x	x	Bostwana
ZeZuru	Lowe	1971			x			x		Zimbabwe
Swazi	Trangle	1979	x						D	Swaziland
Kwangali BSW	Nurse	1977	x	x		x	x	x	x	Kavengo AFS
Sambuyu BSW	Nurse	1977	x	x		x	x	x	x	Kavengo AFS
Gciriku BSW	Nurse	1977	x	x		x	x	x	x	Kavengo AFS
Mbukushu BSW	Nurse	1977	x	x		x	x	x	x	Kavengo AFS
Kwengo BSW	Nurse	1977	x	x	x	x	x	x	x	Kavengo AFS
BSE/Zeruru	Lowe	1971	x	x	x	x	x	x	x	Zimbabwe
BSE/Kgalagadi	Jenkins	1987	x	x	x	x	x	x	x	Bostawana
CUS/Mbugu	Spedini	1981	x	x				x	x	RCA
Congo	Empana	1982	DDS							Congo
Hutu	Fraser	1966	x					x	x	Congo
Shi	Fraser	1966	x					x	x	Congo
Yaka	Fraser	1966	x					x	x	Congo
Mixed Bantus	Fraser	1966	x					x	x	Congo
Mixed Bantus	Fraser	1966	x					x	x	Congo
Tutsi	Fraser	1966	x					x	x	Congo
Ngbaka	Fraser	1966	x					x	x	Congo
Pygmees/Efe	Fraser	1966	x					x	x	Congo
BNE/Tonga										
BCW/Njinga	Nurse	1979	x	x	x			x	x	Angola
AFW/Berba	Biondi	1996	x	x	x	x		x	x	Bénin

Zambian	Gibson	1973	x	x	x	x	x	x	Zambie
Bantu/Kena	Beaumont	1979	x	x	x		x	x	Lesotho
Nigéria SE	Udoh	1997	DDS				x		Nigéria SE
Cameroon	Malesieux	1969	DDS						Cameroon
Nigéria	Falusi	2000	DDS						Nigéria
Sara/Majingay	Hiernaux	1976	x	A	x		MN	x	Tchad Sud
Bedik	Bouloux	1972	x		x	x		x	Sénégal
Bedik	Gessain	1965	x	x	x	x	MN	x	Sénégal
Shangan	Matznetter	1969							Mozambique
Bitonga	Matznetter	1969							Mozambique
Tswana	Jenkins	1973	x				MN	x	Afrique Sud
Pedi	Jenkins	1973	x				MN	x	Afrique Sud
Kung	Jenkins	1973	x				MN	x	Afrique Sud
Koikoi	Jenkins	1973	x				MN	x	Afrique Sud
Amhara	Harrison	1969	x				MN	x	Ethiopie
Wolof	Blavy	1986	x						Sénégal
Serere	Blavy	1986	x						Sénégal
Toucouleur	Blavy	1986	x						Sénégal
Diola	Blavy	1986	x						Sénégal
Peulh	Blavy	1986	x						Sénégal
Mande	Blavy	1986	x						Sénégal
Bateke	Languillat	1980	x	x	x	x	x	x	Gabon
Obamba	Languillat	1980	x	x	x	x	x	x	Gabon
Bawandji	Languillat	1980	x	x	x	x	x	x	Gabon
Pygmées (29)	Languillat	1980	x	x	x	x	x	x	Gabon
Bandjabi	Languillat	1980	x	x	x	x	x	x	Gabon
Powi (28)	Languillat	1980	x	x	x	x	x	x	Gabon
10 Tribues	Ssebabi	1973	x					D	Uganda

**Table 11** : revue de la littérature relative à l'étude du polymorphisme érythrocytaire

### **III.Méthodes d'exploration et d'analyse du polymorphisme *HLA DRB1***

*Cette partie a fait l'objet d'un article actuellement soumis (Gibert M, 2003).*

#### **III.1 Génotypage *HLA-DRB1*.**

Parmi les 200 échantillons prélevés, 117 ont été analysés. Le génotypage HLA a été réalisé par PCR–SSO conformément au protocole de référence défini par la XII<sup>o</sup> Workshop internationale (Charron D, 1997) et / ou par PCR–SSP (Olérup O, 1992) (Annexe 6 p.230).

#### **III.2 Analyse des données.**

**a) Les fréquences géniques** ont été déterminées en utilisant le programme Arlequin (Schneider S, 2000).

**b) Les distances génétiques** ont été calculées en utilisant la méthode Fst à l'aide du logiciel Arlequin (Schneider S, 2000).

#### **c) Les analyses multidimensionnelles**

Programmes utilisés: SPSS 10.0.

#### **d) Les populations comparées.**

Elles sont issues de la XII<sup>o</sup> Workshop (Clayton J, 1997) complétée par les données de la littérature. Cette analyse comparative est limitée aux populations pouvant être, par hypothèse, impliquées dans l'histoire du peuplement des Comores. Nous retiendrons les populations d'Afrique subsaharienne, de l'Asie du Sud Ouest, de l'Asie du Sud, de l'Asie du Sud Est et enfin celle de Madagascar (Mérina). L'ensemble des populations retenues est présenté en Table 12 p.122. Le regroupement linguistique est donné par les auteurs ou réalisé après consultation du site spécifique [www.Ethnologue.fr](http://www.Ethnologue.fr).

Localisation	Langue	Populations	Auteurs	Année
Congo	NC-Bantu	CONGOLAIS	Renquin	2001
Iran	AF-Sémitique	IRANIENS	Armirezgar	2001
Egypte	AF-Sémitique	EGYPTIEN	Krokowski	1998
Gabon	NC-Bantu	GABONAIS	Migot-Nabias	1999
Israël	AF-Sémitique	ISRAELIENS	Amar	1999
Burkina Fasso	NC-West Atlantic	RIMAIBE	Clayton	1997
Egypte	AF-Sémitique	SIWA	Clayton	1997
Pakistan	Indo-Européen	PATHAN	Mohyuddin	2002
Burkina Fasso	NC-Voltaic	MOSSI	Clayton	1997
Soudan	NC-Kordofanien	NOBA	Clayton	1997
Cameroon	NC-Bantu	CAMEROUNAIS	Pimtanothai	2001
Pakistan	Indo-Européen	SINDHI	Mohyuddin	2002
Pakistan	Indo-Européen	KALASH	Mohyuddin	2002
Liban	AF-Sémitique	LIBANAIS	Clayton	1997
Inde	?	INDE NORD	Rani	1998
Rep. Centre. Afriq	NC-Bantu	PYGMES AKA	Renquin	2001
Thaïlande	?	BANKOK	Chandanayingyong	1997
Israël	AF-Sémitique	PALESTINIENS	Arnaiz-Villena	2001
Pakistan	Indo-Européen	BURUSCHO	Mohyuddin	2002
Madagascar	AU-Malayo.Polynésien	MERINA	Migot	1995
Jordanie	AF-Sémitique	JORDANIENS	Sanchez-Velasco	2001
Thaïlande	Tai.Kadai-Tai	BUYI	Chandanayingyong	1997
Guinée équatorial	NC-West Atlantic	BUBI	De Pablo	1997
Yemen	AF-Sémitique	YEMENITE	Amar	1999
Malaisie	AU-Malayo.Polynésien	MALAY	Mack	2000
Sénégal	NC	MANDENKA	Tiercy	1992
Ethiopie	AF-Cushitic	AMHARA	Clayton	1997
Ethiopie	AF-Sémitique	OROMO	Clayton	1997
Thaïlande	Tai.Kadai-Tai	DAI LUE	Chandanayingyong	1997
Burkina Fasso	NC-West Atlantic	FULANI	Clayton	1997
Pakistan	Dravidien	BRAHUI	Mohyuddin	2002
Pakistan	Indo-Européen	BALOCH	Mohyuddin	2002
Océanie	AU-Malayo.Polynésien	BORNEO	Zimdahl	1999
Indonésie	AU-Malayo.Polynésien	MOLUQUES	Mack	2000
Indonésie	AU-Malayo.Polynésien	NUSA-TENGARRA	Mack	2000
Océanie	AU-Malayo.Polynésien	JAVA	Mack	2000

Table 12 : Les populations sélectionnées dans les études HLA (Gibert M, 2003).



## **IV.Méthodes d'exploration et d'analyse du polymorphisme *KIR*.**

### **IV.1 Population.**

Parmi les 200 individus comoriens prélevés, 45 ont été analysés.

Nous avons également analysé 45 individus originaires du Sud Est de la France sélectionnés parmi le registre des donneurs de moelle osseuse de l'EFS Alpes-Méditerranée. Cette population a été prise en compte comme témoin.

### **IV.2 Génotypage *KIR*.**

Le génotypage *KIR* complet de chaque individu est basé sur la recherche des 14 gènes et des 2 pseudogènes. Il a été réalisé par PCR-SSP en utilisant des amorces spécifiques de chaque gène. Les modalités techniques sont décrites en Annexe 7 p.232

### **IV.3 Les méthodes statistiques**

Les fréquences observées ( $f$ ) ont été déterminées par le comptage des individus trouvés positifs pour un gène donné sur le nombre de sujets testés. La fréquence estimée de chaque gène ( $F$ ) a été calculée à partir de la formule suivante :

$$F = 1 - \sqrt{1 - f}$$

Les différences entre les fréquences des gènes *KIR* dans les deux populations étudiées ont été testées par un Chi-deux,  $\chi^2$  à l'aide du logiciel EPI INFO 6.

Nous avons calculé l'équilibre de Hardy-Weinberg pour la distribution des génotypes pour vérifier que nos deux populations étaient à l'équilibre.

Nous avons réalisé une analyse en composante principale (ACP) à l'aide du logiciel SPSS version 10.0, afin de comparer la répartition des populations en fonction du polymorphisme des gènes *KIR*.

Le seuil de signification statistique retenu pour l'ensemble des tests est  $p < 0.05$ . En cas de comparaison multiple, la correction de Bonferroni ( $p$  corrigé) a été appliquée en multipliant le  $p$  par le nombre de comparaisons effectuées (Haldane JB, 1955).

#### ***IV.4 Les autres populations.***

Ces gènes ont été découverts très récemment et de fait, seule une dizaine de populations ont été étudiées à l'échelle mondiale (Table 13 p.129).

<b>Populations</b>	<b>Effectif</b>	<b>Référence</b>
Afrique	62	
Trinidad Asie Sud	108	Norman PJ, 2002
Karachi Asie Sud	78	
Inde	72	Rajaingam R, 2002
Allemagne	99	Becker S, 2002
Japon	41	Yawata et al.,2002
Angleterre	136	
Palestine	105	Norman PJ, 2001
Thaïlande	119	
Vietnam	59	Toneva, 2001

**Table 13** : Les populations étudiées pour le polymorphisme des gènes KIR.

## **RESULTATS.**

## I. Résultats du polymorphisme érythrocytaire (gènes et antigènes courants)

### I.1 Les fréquences phénotypique et allélique

La répartition phénotypique ainsi que l'estimation des fréquences alléliques sont présentées dans la Table 14 p.134 L'hypothèse d'équilibre de Hardy Weinberg est acceptée à un niveau de 5%.

- Le système ABO.

Nous retrouvons une prédominance de l'allèle O (57.8%) par rapport aux allèles A (14.1%) et B (6.9%). Aucun phénotype déficitaire en antigène H n'a été détecté qu'il soit de type Indien (classique phénotype Bombay) ou Réunionnais. Le phénotype Bombay est originaire de la côte Ouest de l'Inde et coexiste dans l'île de la Réunion avec le type Réunionnais qui ne semble pas lié d'un point de vue phylogénétique. Compte tenu de la fréquence basse de ces phénotypes (1/6000) (Lependu, 1983) et de la taille de notre échantillon, nous ne pouvons pas affirmer l'absence totale de ce type de variant de l'allèle *H* dans la population des Comores.

- Le système RH.

L'haplotype *R0* (*Dce*) apparaît comme le plus fréquent avec une fréquence de 48%. Cet haplotype avec les trois autres que sont *R1* (*DCe*), *r* (*dce*) et *R2* (*DcE*) représentent les haplotypes les plus communs dans la population Comorienne. Parmi les haplotypes rares (*r'* et *r''*), seul *r'* est retrouvé avec une fréquence à 4.6%. Parmi les haplotypes considérés comme exceptionnels (*RZ* et *ry*), aucun n'a été mis en évidence comme cela est souvent le cas dans la majorité des populations.

Par ailleurs toutes les réactions sérologiques ayant été contrôlées visuellement, nous n'avons retrouvé aucune diminution d'expression des antigènes RH testés.

- Le système KEL.

Nous notons la faible fréquence de l'allèle *K* (1.2%) et l'absence totale d'antigène  $Kp^a$  ou de phénotype Cellano (*k*) négatif. L'absence de disponibilité de réactif anti-Jsa, ne nous a pas permis de rechercher cet antigène qui est caractéristique des populations d'Afrique subsaharienne.

- Le système Duffy.

Le phénotype Duffy apparaissant le plus fréquent en Grande Comore est le phénotype Fy (a-b-) avec une fréquence de 86.2%, ce qui nous donne une estimation de la fréquence de l'allèle FY\*0 à 92.87%.

Parmi les sujets exprimant des antigènes Duffy sur leurs hématies, l'allèle FY\*A apparaît avec une fréquence (6.1%) largement supérieure à celle de l'allèle FY\*B (0.9%). L'analyse en biologie moléculaire a permis de valider, par comptage direct des allèles, la fréquence de chacun d'eux. La répartition des différents génotypes observés est donnée dans la Table 15 p.135. On constate que l'estimation des fréquences par comptage directe donne des résultats très similaires (FY\*A 5.9%, FY\*B 0.9% et FY\*0 93.1%) aux estimations obtenues avec la formule de Bernstein. Les résultats obtenus en biologie moléculaire précisent les bases génétiques de chaque phénotype Duffy. Le phénotype Fy(a-b-) en Grande Comore repose sur la présence en double dose d'un allèle FY\*0 qui apparaît identique à un allèle FY\*B mais avec une substitution nucléotidique T46C dans la région promotrice du gène (Mallinson G 1995).

L'allèle FY\*A diffère de l'allèle FY\*B par une seule substitution (G125A) (Iwamoto S, 1995). L'allèle FY\*X qui est caractérisé par la substitution nucléotidique C286T sur un

allèle FY\*B (Tournamille C, 1998) n'a pas été retrouvé dans cette population. Nous n'avons pas retrouvé de modification dans la région promotrice associée à un allèle FY\*A comme nous l'avions avec FY\*B. Enfin, comme pour l'ensemble des systèmes, le contrôle visuel de chaque réaction sérologique n'a détecté aucune diminution de réactivité antigénique avec les deux antigènes Fya et Fyb.

- Le système Kidd.

L'allèle JK\*A est le plus fréquent (presque 70%) et aucun phénotype silencieux Jk (a-b-) n'a été détecté. Nous notons à nouveau une absence de diminution de réactivité antigénique.

- Le système MNS.

L'allèle M (58.7%) est plus fréquent que l'allèle N et l'allèle S (25.3%) est largement en dessous de l'allèle s. Aucun phénotype S-s- n'a été détecté dans cet échantillon. Enfin, compte tenu de la non-disponibilité de sérum rare de type anti-MNS5 (U) l'absence de réactivité U n'a pu être recherchée.

Systèmes	Fréquences phénotypiques (n)	Fréquences alléliques		
ABO <i>N=164</i>	A	25% - (41)	<i>A</i>	14.1%
	B	12.2% - (20)	<i>B</i>	6.9%
	O	61.6% - (101)	<i>O</i>	78.5%
	AB	1.2% - (2)		
	Dccee	38.4% - (63)	<i>DCe (R1)</i>	19.5%
	DCcee	38.4% - (63)	<i>DcE (R2)</i>	7.6%
	DCCee	1.8% - (3)	<i>Dce (Ro)</i>	47.9%
	DCcEe	3.6% - (6)	<i>dce (r)</i>	20.3%
RH <i>N=164</i>	DccEe	11.6% - (19)	<i>dCe (r')</i>	4.6%
	ddccee	3.6% - (6)		
	dCcee	2.4% - (4)		
	KK	0% - (0)	<i>K</i>	1.25%
KEL <i>N=160</i>	Kk	2.5% - (4)	<i>k</i>	98.7%
	kk	97.5% - (156)	<i>Kp<sup>a</sup></i>	0%
	Kp <sup>a</sup> positive	0% - (0)		
FY <i>N=160</i>	Fy (a+b+)	0% - (0)	<i>FYA</i>	6.1%
	Fy (a+b-)	11.9% - (19)	<i>FYB</i>	0.9%
	Fy (a-b+)	1.9% - (3)	<i>FYO</i>	92.9%
	Fy (a-b-)	86.2% - (138)		
JK <i>N=160</i>	Jk (a+b+)	45% - (72)	<i>JKA</i>	69.4%
	Jk (a+b-)	46.9% - (75)	<i>JKB</i>	30.6%
	Jk (a-b+)	8.1% - (13)	<i>JKO</i>	0%
	Jk (a-b-)	0% - (0)		
MNS <i>N=162</i>	MMSS	1.2% - (2)	<i>M</i>	58.6%
	MMSs	16.04% - (26)	<i>N</i>	41.3%
	MMss	19.7% - (32)	<i>S</i>	25.3%
	MNSS	3.7% - (6)	<i>s</i>	74.7%
	MNSs	13.6% - (22)		
	MNss	26% - (42)		
	NNSS	3.7% - (6)		
	NNSs	3.7% - (6)		
	NNss	12.3% - (20)		

**Table 14 :** Fréquences phénotypique et allélique des groupes sanguins érythrocytaires dans la population Comorienne.



Phénotypes	Fréquences génotypiques (n)		Fréquences alléliques	
Fy(a+b-)	<i>FYA/FYA</i>	0	<i>FYA</i>	5.93%
	<i>FYA/FYO</i>	11.87% - (19)	<i>FYB</i>	0.93%
Fy(a+b+)	<i>FYA/FYB</i>	0	<i>FYX</i>	0%
	<i>FYA/FYX</i>	0	<i>FYO</i>	93.12%
Fy(a-b+)	<i>FYB/FYB</i>	0		
	<i>FYX/FYX</i>	0		
	<i>FYB/FYX</i>	0		
	<i>FYB/FYO</i>	1.88% - (3)		
Fy(a-b-)	<i>FYX/FYO</i>	0		
	<i>FYO/FYO</i>	86.25 % - (138)		

Table 15 : Résultats du génotypage Duffy dans la population Comorienne.

## 1.2 Les distances génétiques

Les distances génétiques calculées d'après la formule de Reynolds, avec les marqueurs érythrocytaires, entre la population Comorienne et les différents groupes populationnels retenus pour les analyses comparatives sont données dans la Table 17 p.138. Cette analyse révèle une proximité entre la population Comorienne et celle du continent Africain et de la Péninsule Arabique. Par contre, la population Comorienne apparaît éloignée des populations du Sud-Est Asiatique. L'analyse des valeurs obtenues permet de distinguer trois catégories de populations (Table 18 p.139).

Celles dont la distance génétique par rapport à la population comorienne est inférieure à 0.1 (Somali, Bitonga, Béja, Zéruru, Sandawee, Bétékés, Kgalagadi, Kavango).

Celles dont les distances génétiques sont inférieures à 0.2 (Hadza, Biaka, Bédik, Saoudiens, San).

Enfin, celles dont les distances génétiques sont supérieures à 0.2 (Libyens, Indonésiens, Toba Bataks, Gadaba).

Ces résultats démontrent une proximité génétique de la population Comorienne avec les populations d'Afrique de l'Est (2 populations (Somali et Béja) parmi les 8 plus proches) et avec les populations Bantoues (5 populations Bantoues (Bitonga, Zéruru, Bétékés, Kgalagadi, Kavango) parmi les 8 plus proches). La proximité des Sandawee sera analysée dans le chapitre relatif à la Discussion. Les populations Arabes (Saoudiens) apparaissent avec des valeurs de distances inférieures à celles obtenues avec d'autres populations Africaines comme l'Afrique du Nord et les San. Enfin, les valeurs les plus élevées sont affectées aux populations d'Asie du Sud-Est.

La représentation des distances sous forme d'arbre phylogénétique (Figure 20 p.140) confirme le "cluster" Africain avec les San à part, la position intermédiaire de l'Arabie Saoudite et le positionnement plus éloigné des populations d'Asie du Sud-Est.

	A	B	O	FYA	FYB	FYO	K	k	JKA	JKB	JKO	M	N	S	s	RZ	R1	ry	r'	R2	R0	r''	r
<b>Saoudiens</b>	0,134	0,123	0,741	0,194	0,225	0,591	0,114	0,896	0,537	0,463	0	0,755	0,245	0,435	0,565	0,008	0,459	0	0,014	0,149	0,099	0	0,271
<b>Somali</b>	0,155	0,077	0,74	0,041	0,04	0,92	0,005	0,995	0,687	0,312	0	0,51	0,49	0,152	0,848	0	0,155	0	0,006	0,013	0,654	0,007	0,163
<b>Béja</b>	0,193	0,083	0,723	0,127	0,086	0,786	0,005	0,995	0,546	0,454	0	0,48	0,52	0,272	0,728	0	0,27	0	0	0,085	0,428	0	0,217
<b>Lybiens</b>	0,216	0,117	0,667	0,39	0,3	0,311	0,051	0,949	0,533	0,468	0	0,529	0,471	0,304	0,598	0,013	0,361	0	0,034	0,138	0,11	0,022	0,322
<b>San</b>	0,234	0,037	0,72	0,216	0,37	0,39	0,053	0,957	0,904	0,096	0	0,547	0,453	0,177	0,823	0	0,049	0	0	0,02	0,905	0	0,046
<b>Sandawe</b>	0,17	0,129	0,7	0,007	0,002	0,991	0,003	0,997	0,717	0,282	0	0,6	0,4	0,372	0,627	0	0,035	0	0	0,052	0,708	0	0,206
<b>Hadza</b>	0,325	0,018	0,557	0	0	1	0	1	0,582	0,418	0	0,55	0,45	0,15	0,85	0	0,009	0	0	0,068	0,748	0	0,173
<b>Kavango</b>	0,174	0,183	0,635	0,029	0,05	0,92	0,002	0,993	0,883	0,091	0	0,647	0,276	0,191	0,809	0	0,03	0	0,02	0,05	0,737	0,03	0,13
<b>Bitonga</b>	0,161	0,121	0,717	0,003	0,003	0,994	0,002	0,998	0,762	0,237	0	0,529	0,47	0,28	0,72	0,002	0,027	0	0,03	0,093	0,691	0,016	0,134
<b>Zéruru</b>	0,17	0,118	0,71	0,009	0,017	0,973	0,002	0,998	0,81	0,187	0	0,51	0,49	0,2	0,8	0	0,044	0	0,016	0,098	0,675	0,001	0,164
<b>Kgalagadi</b>	0,154	0,103	0,743	0,076	0,089	0,835	0	1	0,883	0,091	0	0,585	0,415	0,243	0,757	0	0,009	0	0,015	0,125	0,63	0	0,221
<b>Bédik</b>	0,217	0,195	0,588	0,017	0	0,983	0	1	0,396	0,604	0	0,51	0,49	0,088	0,723	0,002	0,025	0	0	0,074	0,594	0	0,306
<b>Biaka</b>	0,185	0,148	0,667	0,001	0,001	0,998	0,004	0,996	0,808	0,192	0	0,545	0,454	0,137	0,706	0	0,007	0	0,033	0,014	0,816	0,002	0,128
<b>Sumatra</b>	0,139	0,213	0,702	0,885	0,115	0	0	1	0,589	0,364	0,124	0,4	0,6	0,093	0,907	0,007	0,846	0	0	0,106	0,04	0	0
<b>Indonésiens</b>	0,16	0,23	0,61	0,8	0,2	0	0	1	1	0	0	0,54	0,46	0,05	0,95	0,011	0,889	0	0	0,063	0,037	0	0
<b>Bétékés</b>	0,108	0,116	0,774	0	0,003	0,997	0,024	0,976	0,818	0,182	0	0,451	0,549	0,249	0,751	0	0,034	0	0,023	0,086	0,706	0	0,149
<b>Gadaba</b>	0,174	0,295	0,536	0,572	0,428	0	0,020	0,980	0,478	0,369	0,047	0,811	0,188	0,691	0,308	0,063	0,763	0	0,017	0	0	0,005	0,152

Table 16 : Fréquences géniques au sein des populations d'Afrique et du pourtour de l'Océan Indien (données de la littérature)

	Comoro	Saoudiens	Somali	Béja	Libyens	San	Sandawe	Hadza	Kavango	Bitonga	Zeruru	Kgalagadi	Bédik	Biaka	TobaBatak	Indonésie	Bétéké	Gadaba
<b>Comoro</b>	0.0000																	
<b>Saoudiens</b>	0.0722	0.0000																
<b>Somali</b>	0.0077	0.1123	0.0000															
<b>Béja</b>	0.0124	0.0579	0.0172	0.0000														
<b>Libyens</b>	0.1040	0.0396	0.1256	0.0624	0.0000													
<b>San</b>	0.0862	0.1791	0.0643	0.0953	0.1430	0.0000												
<b>Sandawe</b>	0.0139	0.1149	0.0161	0.0334	0.1586	0.0823	0.0000											
<b>Hadza</b>	0.0231	0.1453	0.0117	0.0313	0.1625	0.0865	0.0190	0.0000										
<b>Kavango</b>	0.0182	0.1304	0.0118	0.0420	0.1590	0.0559	0.0106	0.0173	0.0000									
<b>Bitonga</b>	0.0118	0.1299	0.0088	0.0318	0.1599	0.0712	0.0042	0.0153	0.0070	0.0000								
<b>Zéruru</b>	0.0126	0.1361	0.0067	0.0333	0.1557	0.0631	0.0103	0.0165	0.0063	0.0021	0.0000							
<b>Kgalagadi</b>	0.0177	0.1200	0.0177	0.0429	0.1361	0.0455	0.0161	0.0340	0.0100	0.0103	0.0070	0.0000						
<b>Bédik</b>	0.0402	0.1391	0.0300	0.0323	0.1447	0.1422	0.0464	0.0177	0.0456	0.0441	0.0450	0.0707	0.0000					
<b>Biaka</b>	0.0239	0.1679	0.0101	0.0484	0.1876	0.0605	0.0150	0.0146	0.0043	0.0067	0.0041	0.0142	0.0447	0.0000				
<b>TobaBatak</b>	0.2947	0.1977	0.3064	0.2264	0.0992	0.2813	0.4059	0.3922	0.3679	0.3823	0.3614	0.3396	0.3677	0.3947	0.0000			
<b>Indonésiens</b>	0.2894	0.2120	0.3048	0.2597	0.1328	0.2324	0.3913	0.4081	0.3323	0.3637	0.3334	0.2911	0.4278	0.3625	0.0355	0.0000		
<b>Bétéke</b>	0.0162	0.1470	0.0098	0.0374	0.1687	0.0686	0.0112	0.0233	0.0129	0.0028	0.0023	0.0108	0.0538	0.0080	0.3790	0.3558	0.0000	
<b>Gadaba</b>	0.3370	0.1219	0.3984	0.2787	0.1219	0.3423	0.3904	0.4735	0.4073	0.4280	0.4411	0.3809	0.4685	0.4867	0.1577	0.1717	0.4609	0.0000

Table 17 : Matrice de distances génétiques avec les populations d'Afrique et du pourtour de l'Océan Indien. Distances génétiques calculées avec les marqueurs érythrocytaires (23 allèles) et selon la formule de Reynolds.

Populations	Distances par ordre croissant	Localisation géographique régionale
Comores	0.0000	
Somali	0.0077	Afrique de l'Est
Bitonga	0.0118	Bantous Sud Est
Béja	0.0124	Afrique de l'Est
Zéruru	0.0126	Bantous Sud Est
Sandawe	0.0139	Tanzanie
Bétéké	0.0162	Bantous Nord West
Kgalagadi	0.0177	Bantous Sud Est
Kavango	0.0182	Bantous Sud Ouest
Hadza	0.0231	Tanzanie
Biaka	0.0239	Pygmées Centre Ouest
Bédik	0.0402	Afrique de l'Ouest
Saoudiens	0.0722	Péninsule arabique
San	0.0862	Afrique du Sud
Lybiens	0.1040	Afrique du Nord
Indonésiens	0.2894	Asie du Sud Est
Toba Batak	0.2947	Asie du Sud Est
Gadaba	0.3370	Inde

**Table 18** : Les populations d'Afrique et du pourtour de l'Océan Indien classées par ordre croissant des valeurs des distances génétiques calculées avec les marqueurs érythrocytaires (23 allèles) selon la formule de Reynolds.

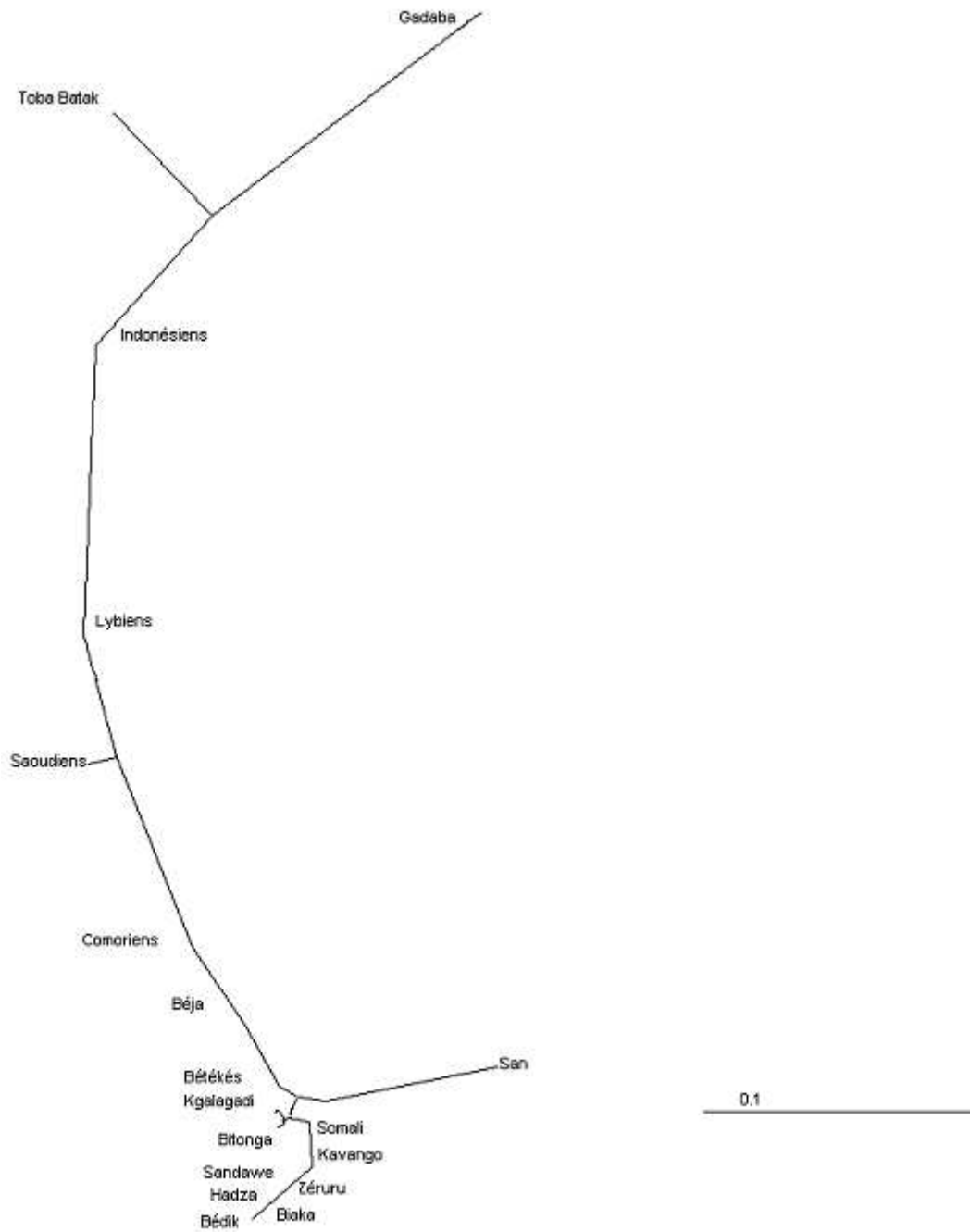


Figure 20.: Arbre phylogénétique positionnant la population des Comores par rapport aux autres populations d’Afrique et du pourtour de l’Océan Indien (calculs réalisés à partir des marqueurs érythrocytaires).

### **1.3 L'analyse multidimensionnelle**

L'analyse multidimensionnelle des 18 populations est représentée en Figure 21 p.142

Nous avons rapproché par un trait les populations dont la représentation ne rendait pas bien compte de la valeur de leur distance génétique. Nous constatons que la répartition des populations en fonction des différents systèmes érythrocytaires est corrélée à leur disposition géographique, bien que le schéma soit inversé par rapport aux dispositions cardinales. Dans la partie A, trois groupes de populations apparaissent plus différenciés par rapport à la population Comorienne. Un premier groupe qui comporte 3 populations apparaît très différencié : les Gadaba, les Toba Bataks et les Indonésiens. Les deux autres groupes occupent une position intermédiaire par rapport à la population des Comores. Il s'agit d'une part d'un groupe comprenant les deux populations de langue Afro-asiatique (Saoudiens et Libyens) et d'autre part d'une population, les San. La population comorienne se positionne au sein de l'ensemble des populations Africaines (exception faite des Sans). Dans la partie B, nous avons supprimé les groupes de populations apparaissant plus différenciés afin de mieux visualiser le positionnement de la population comorienne au sein des autres populations Africaines. Nous retrouvons la population comorienne à une position intermédiaire entre les populations d'Afrique de l'Est (Béja et Somalis) et les populations Bantoues principalement les Bitonga et les Zéruru.

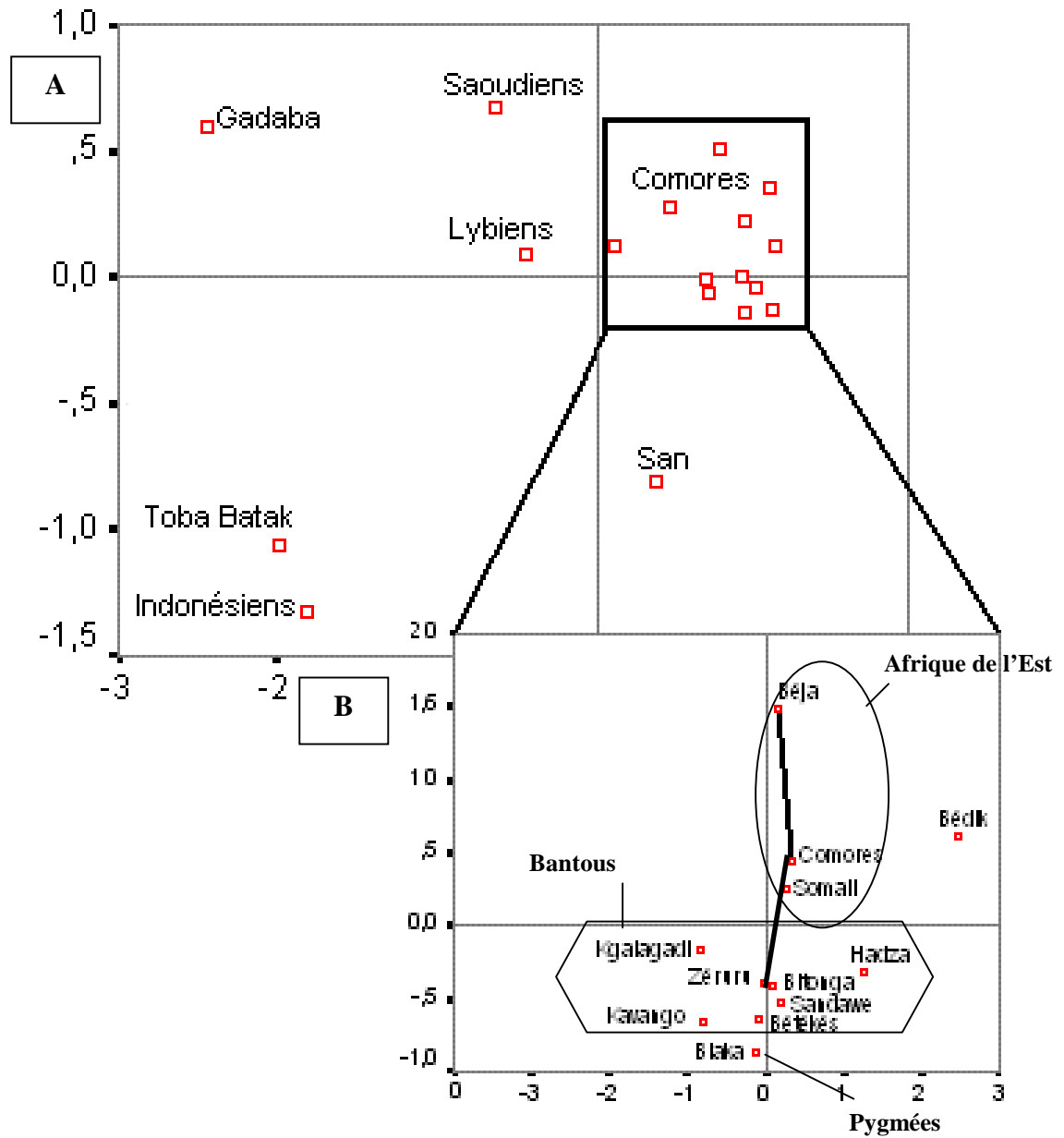


Figure 21: Analyse multidimensionnelle positionnant de la population des Comores par rapport aux autres populations Africaines et celles du pourtour de l'Océan Indien (calculs réalisés à partir des marqueurs érythrocytaires).



#### **1.4 La classification ascendante hiérarchique**

La lecture du dendrogramme (Figure 22 p.144) et le degré d'homogénéité des populations que nous avons fixé (distance = 8), nous permettent de dégager 5 groupes homogènes de populations :

1<sup>e</sup> groupe : Bitonga, Zéruru, Bétéké, Biaka, Sandawee, Kavango, Kgalagadi.

2<sup>e</sup> groupe : Comores, Somalis et Béja,

3<sup>e</sup> groupe : Hadza, Bédik,

4<sup>e</sup> groupe : Saoudiens, Libyens

5<sup>e</sup> groupe : Toba Bataks, Indonésiens.

Et deux populations isolées : les Sans et les Gadaba.

Ces groupes ont déjà été identifiés par l'analyse multidimensionnelle.

De plus, si on augmente le degré d'homogénéité des populations (distance = 12), nous pouvons alors dégager deux grands groupes homogènes:

1<sup>o</sup> groupe : l'ensemble des populations africaines et du Moyent Orient (Bitonga, Zéruru, Bétéke, Biaka, Sandawee, Kavango, Kgalagadi, Comores, Somalie, Béja, Hadza, Bédik, Saoudiens, Libyens et Sans).

2<sup>o</sup> groupe : l'ensemble des populations du Sud et Sud-Est asiatique (Toba Bataks, Indonésiens et Gadaba)

Au total, l'analyse multidimensionnelle (Figure 21 p.142) et la classification ascendante hiérarchique (Figure 22 p.144) confirment le regroupement de la population Comorienne avec l'ensemble des populations Africaines, la proximité avec les populations d'Afrique de l'Est (Somalis et Béja) ainsi qu'une proximité relative des Saoudiens.

Par contre, la population Comorienne apparaît éloignée des populations du Sud Est Asiatique.

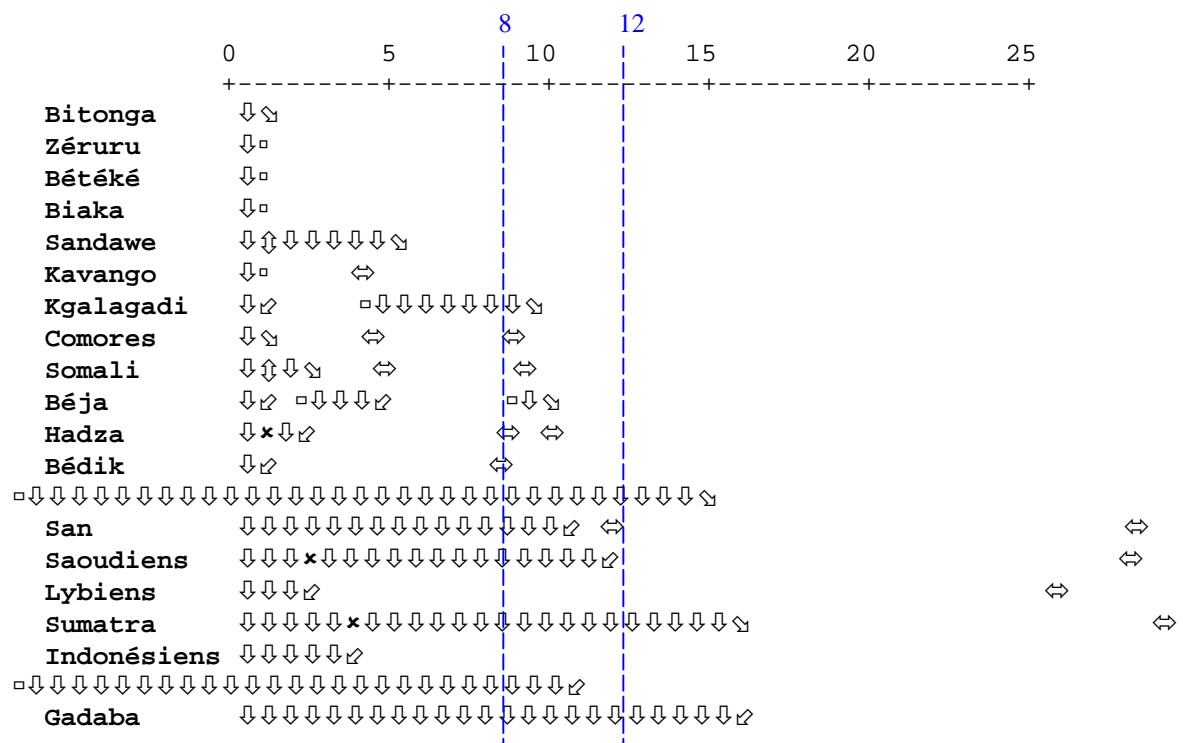


Figure 22 : Dendrogramme positionnant la population des Comores par rapport aux populations d’Afrique et du pourtour de l’Océan Indien selon la méthode UPGMA (calculs réalisés à partir des marqueurs érythrocytaires).

### ***1.5 L'admixture***

Nous avons réalisé les calculs de contribution génétique (admixture) sous le modèle tri-hybride. Compte tenu des données historiques, nous avons retenu la population Bantoue la plus proche (Bitonga) pour la contribution Africaine, les Saoudiens pour la composante Arabe et la population la plus proche d'Asie du Sud Est (Indonésiens) pour la composante Austronésienne. Les résultats montrent que les Comoriens sont une population métissée avec une forte composante Bantoue (Composante Bitonga  $69.5\% \pm 5.2\%$ ) suivie de la composante Arabe (Saoudiens  $29\% \pm 2.8\%$ ) et d'une très faible contribution Indonésienne ( $1.5\% \pm 3.7\%$ ). Ces calculs ont donné un "Model fitting value  $R^2 = 0.992$

## **II. Résultats des fréquences phénotypiques des donneurs de sang de l'EFS.**

Les fréquences phénotypiques des systèmes RH, KEL, Duffy, Kidd, MNS ont été calculées à partir de notre base de données. Les fréquences des principaux phénotypes susceptibles de présenter un écart significatif par rapport aux populations originaires d'Afrique subsaharienne sont données en table 19 p.147. La fréquence de la combinatoire antigénique Fya-, Jkb-, S- atteint 5.6% dans notre fichier. La prise en compte supplémentaire du phénotype Rh diminue le nombre d'unités compatibles. En effet, la fréquence d'une unité dd cc ee Fya-, Jkb-, S- tombe à 1% et si nous imposons qu'elle soit R0 (D cc ee) elle tombe à 0.14%.

<b>Phénotypes</b>	<b>Fréquences</b>
R0 (D cc ee)	2.6%
R (dd cc ee)	16%
S-	40%
Fya -	30%
Jkb-	20%
S-, Fya-, Jkb-	5.6%
r, S-, Fya-, Jkb-	0.98%
R0, S-, Fya-, Jkb-	0.14%
Fy(a-b-)	0% (4 donneurs)
S-s-	0% (1 donneur)
RH : -18	0%
RH : -34	0%
RH : -46	0%

Table 19 : Fréquences des phénotypes courants et de certains phénotypes publics négatifs (Duffy et RH) dans la population de donneurs de sang de l'EFS Alpes-Méditerranée.

### **III. Résultats de la détection des allèles « rares » du gène *RHCE*.**

#### **Bilan de l'étude en biologie moléculaire des allèles « rares » du gène *RHCE*.**

Les allèles associés aux phénotypes rares suivants ont été recherchés par PCR-ASP :

Phénotypes « publics négatifs » : RH :-34

Phénotypes RH5 partiels

Phénotypes RH10 et RH20 (antigènes de basse fréquence)

La recherche des allèles associés aux phénotypes RH :-18 et RH :-46 est toujours en cours de validation.

#### **Les allèles produisant les RH5 partiels (*ceMO*, *ce<sup>s</sup>(340)*)**

193 individus ont été analysés.

- Allèle *ceMO* :  
8 individus sont porteurs de cet allèle et tous sont hétérozygotes (PCR allèle sauvage positive). La fréquence est de 2%.
- Allèle *ce<sup>s</sup>(340)* :  
72 individus sont porteurs de cet allèle.  
3 individus sont homozygotes (PCR allèle sauvage négative pour 733 et 340) et 69 individus sont hétérozygotes. La fréquence est 19.4%.

**Haplotype produisant le phénotype RH :-34 ((C)ce<sup>s</sup>).**

24 individus ont été analysés à ce jour.

5 individus sont porteurs de l'allèle ((C)ce<sup>s</sup>). Ils sont tous probablement hétérozygotes du fait de la présence de l'antigène RhD et du gène *RHD*. L'analyse des phénotypes permet de définir les génotypes probables.

Phénotypes	Génotypes probables	
DCcee	(C)ce <sup>s</sup> /Dce	(C)ce <sup>s</sup> /DCe
DCEce	(C)ce <sup>s</sup> /DcE	(C)ce <sup>s</sup> /DCE
DCcee	(C)ce <sup>s</sup> /Dce	(C)ce <sup>s</sup> /DCe
DCEce	(C)ce <sup>s</sup> /DcE	(C)ce <sup>s</sup> /DCE
DCcee	(C)ce <sup>s</sup> /Dce	(C)ce <sup>s</sup> /Dce

Le génotype probable pourra être confirmé par sérologie en utilisant un anticorps composé anti-Ce ou par biologie moléculaire par séquence des gènes *RHCE*.

**Allèle produisant les antigènes RH10 et RH20 (ce<sup>s</sup>)**

193 individus ont été analysés.

120 individus sont porteurs de la mutation 733 et donc porteurs des antigènes RH10 et RH20.

10 individus sont homozygotes (PCR allèle sauvage négative) et 110 individus sont hétérozygotes (PCR allèle sauvage positive). La fréquence est de 70%.

## **IV. Résultats du polymorphisme *HLA-DRB1*.**

### ***IV.1 Les fréquences alléliques***

Les estimations des fréquences alléliques sont présentées dans la Table 20 p.152

L'analyse des fréquences alléliques montre que les trois allèles les plus courants (>15%) sont représentés par HLA-DRB1\***11** (26.5%), HLA-DRB1\***13** (16.3%) et HLA-DRB1\***15** (16.2%). Ces trois allèles apparaissent aussi comme les plus courants dans des populations Bantoues du Gabon et du Congo ainsi que dans la population Egyptienne. Par ailleurs, ces allèles (au moins deux d'entre eux) sont assez largement répandus dans de nombreuses populations d'origines géographiques diverses comme les populations originaires d'Afrique subsaharienne, de la partie Est du bassin Méditerranéen (Grèce, Moyen Orient...) et du Sud et Sud Est Asiatique.

L'allèle DRB1\*13 est caractéristique des populations d'Afrique subsaharienne où il peut atteindre plus de 30% en Afrique de l'Ouest (Bubi, Madenka et Fulani). L'allèle DRB1\*15 présente les valeurs les plus hautes dans les populations Bantoues du Gabon (31%) du Cameroun (28.9%) et du Congo (20.5%). Dans les populations non Bantoues d'Afrique de l'Ouest la fréquence de cet allèle est plus faible (Mossi {6.4%}, Bubi {6.3%}) ou totalement absent (Fulani, Madenka). En dehors de l'Afrique, la fréquence la plus haute est observée en Papouasie Nouvelle Guinée (40%). Les valeurs observées dans la population Comorienne sont comparables à celles observées en Egypte ou en Asie du sud (Inde et Pakistan).

L'allèle suivant, avec une valeur de 14.1%, est représenté par HLA-DRB1\***12** qui est caractéristique des populations d'Asie du Sud Est et de certaines populations Malgaches comme les Mérimina où sa fréquence atteint 36.4%.

Deux autres allèles sont retrouvés avec des fréquences presque comparables, l'allèle DRB1\***01** (10%), et l'allèle DRB1\***03** (9%). Avec une fréquence de près de 10% dans la



population comorienne l'allèle DRB1\*01 atteint la plus haute valeur observée en Afrique (Afrique du Nord incluse) juste derrière les Amhara où elle atteint 11.7%. Dans les autres populations Africaines, la fréquence reste inférieure à 7.5%. Ailleurs, les fréquences les plus élevées sont observées en Europe et au Pakistan mais elles restent toujours en deçà de 15%. En ce qui concerne l'allèle DRB1\*03, la fréquence observée dans la population comorienne est comparable à celle des populations d'Afrique subSaharienne, d'Afrique du Nord Est (Egypte), d'Afrique de l'Est (Ethiopie : Amhara) et du Moyen Orient.

L'allèle DRB1\*07 apparaît avec une fréquence relativement faible (5.6%) ce qui est comparable à ce qui est classiquement observé dans les populations d'Afrique subsaharienne à l'exception des Fulani. En dehors de l'Afrique de telles valeurs sont observées en Asie du Sud Est. Quatre allèles, enfin, présentent des fréquences inférieures à 3%. L'allèle DRB1\*09 (2.6%) dont les valeurs sont comparables aux autres populations Africaines. Cet allèle est essentiellement représenté en Asie du Nord et en Asie Centrale. L'allèle DRB1\*16 (2.1%) qui est essentiellement présents en Asie du Sud Est. L'allèle DRB1\*04 (1.7%) dont la fréquence dans la population Comorienne est comparable à celle observée dans la majorité des populations d'Afrique subsaharienne mais sans que cela apparaisse comme une exclusivité. En effet, de telles valeurs sont aussi retrouvées en Asie du Sud Est (Bornéo, Moluque), en Asie du Sud et à Madagascar. Ailleurs, cet allèle présente une fréquence plus élevée, dont les plus hautes valeurs sont retrouvées en Asie, en Europe et au Moyent Orient sans jamais dépasser 10%. L'allèle DRB1\*08 présente une valeur (1.7%) comparable à celle retrouvée au Moyen Orient. Par rapport aux populations d'Afrique subsaharienne, la fréquence observée dans la population comorienne représente la 2<sup>o</sup> valeur la plus faible après celle observée dans les populations bantoues du Gabon. Les plus hautes valeurs Africaines sont retrouvées en Afrique de l'Ouest non Bantoue (Mossi, Rimaibe, Fulani) avec 11%. En dehors de l'Afrique, cet allèle est essentiellement représenté dans les populations asiatiques. Enfin, l'allèle DRB1\*10 (1.7%) est essentiellement représenté dans les population de l'Est du bassin Méditerranéen et en Asie du sud (Inde et Pakistan).

<b>Allèles</b>	<b>Fréquences</b>
HLA-DR1*01	9.8%
HLA-DR1*03	9.0%
HLA-DR1*04	1.7%
HLA-DR1*07	5.6%
HLA-DR1*08	1.7%
HLA-DR1*09	2.6%
HLA-DR1*10	1.7%
HLA-DR1*11	26.5%
HLA-DR1*12	14.1%
HLA-DR1*13	16.3%
HLA-DR1*14	0.4%
HLA-DR1*15	16.2%
HLA-DR1*16	2.1%

Table 20 : Fréquence des spécificités génériques HLA-DRB1 dans la population Comorienne.

#### **IV.2 Les distances génétiques**

L'analyse des distances génétiques calculées à partir des fréquences alléliques HLA-DRB1 sont données dans la Table 21 p.154 et classées par ordre croissant de distance par rapport à la population comorienne.

Parmi les 10 populations apparaissant les plus proches ; 5 sont originaires d'Afrique subsaharienne [2 populations Bantoues du Congo et du Gabon. 2 populations d'Afrique de l'Ouest non Bantoue, les Rimaibe et les Mossi. 1 population d'Afrique de l'Est les Noba], 2 du Moyen Orient [Iran et Israël], 2 d'Afrique du Nord Est [Egypte et Siwa] et 1 d'Asie du Sud [Pathan]. Par ailleurs, 5 populations apparaissent plus différenciées. 4 sont originaires d'Asie du Sud Est (Bornéo, Moluque, Nusa, Java) et une d'Inde (Baloch).

#### **IV.3 L'analyse multidimensionnelle**

L'analyse multidimensionnelle de 37 populations d'origine africaine et du pourtour de l'Océan Indien est représentée en Figure 23 p.155. Nous avons rapproché par un trait les populations dont la représentation ne rendait pas bien compte de leur valeur de Fst. Nous constatons que la répartition des allèles HLA-DRB1 est corrélée à la disposition géographique des différentes populations, bien que l'Inde soit située à gauche sur le schéma. Dans la partie A, 10 populations apparaissent très différenciées : Java, Nusa, Moluque, Bornéo, Malay, Baloch, Brahui, Madenka, Fulani et Mérina. Dans la partie B, nous avons supprimé ces 10 populations afin de mieux visualiser le positionnement de la population comorienne. La population comorienne apparaît en position intermédiaire entre l'Afrique du Nord, l'Afrique de l'Est, le Moyen Orient et l'Afrique subsaharienne.

<b>Populations</b>	<b>Groupes</b>	<b>Distance</b>
CONGOLAIS	BCW	.806
IRANIENS	ASW	.1468
EGYPTIEN	AFE	.1638
GABONAIS	BCW	.1941
ISRAELIENS	ASW-MO	.2266
RIMAIBE	AFW	.2424
SIWA	AFE	.2499
PATHAN	ASS	.2900
MOSSI	AFW	.2931
NOBA	AFE	.3089
CAMEROUNAIS	BCW	.3206
SINDHI	ASS	.3553
KALASH	ASS	.3607
LIBANAIS	ASW-MO	.3759
INDE NORD	ASS	.4060
PYGMEES AKA	PYG	.4280
BANKOK	ASE	.4282
PALESTINIENS	ASW-MO	.4328
BURUSCHO	ASS	.5047
MERINA	MADA	.5095
JORDANIENS	ASW-MO	.5132
BUYI	ASE	.5134
BUBI	AFW	.5371
YEMENITE	ASW-PA	.5511
MALAY	ASE	.6568
MANDENKA	AFW	.6573
AMHARA	AFE	.6664
OROMO	AFE	.6727
DAI LUE	ASE	.7485
FULANI	AFW	.8305
BRAHUI	ASS	.8998
BALOCH	ASS	.10012
BORNEO	ASE	.10167
MOLUQUES	ASE	.10667
NUSA-TENGARRA	ASE	.10917
JAVA	ASE	.12624

**Table 21** : Les populations classées par ordre croissant de distances génétique par l'étude du système HLA

(Gibert M, 2003).

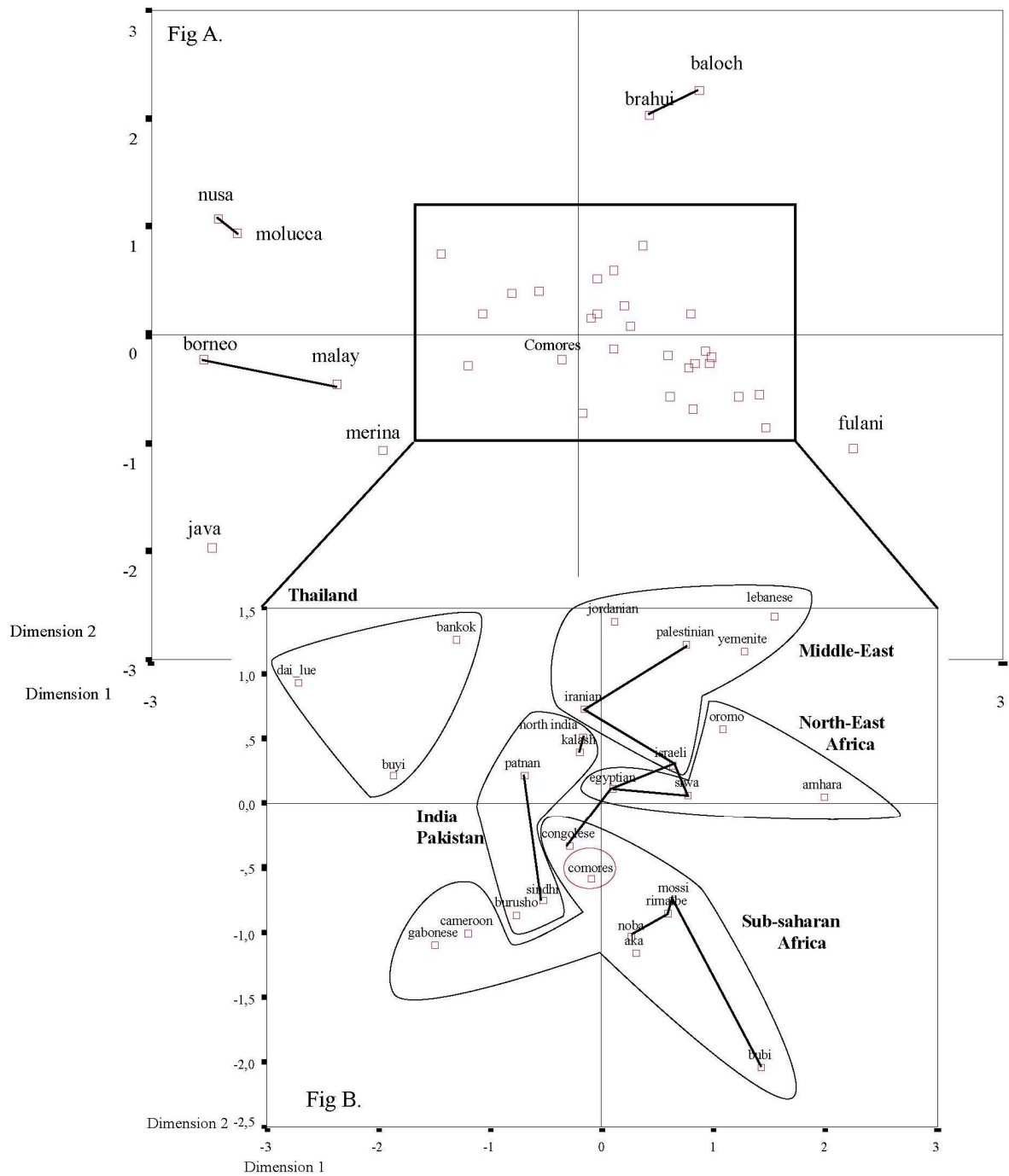


Figure 23 : Analyse multidimensionnelle de la population des Comores par rapport à 36 populations d’Afrique et du pourtour de l’Océan Indien (D’après le système HLA) (Gibert M, 2003)

## V. Résultats du polymorphisme des gènes *KIR*.

Nous présentons ici les résultats de l'étude du polymorphisme des gènes *KIR* dans la population comorienne de Marseille qui a fait l'objet d'un travail de DEA (C.Frassati, 2003).

### V.1 Les fréquences géniques

La fréquence des différents gènes est donnée en Table 22 p.157

La fréquence des gènes dits « d'ancrage », 2DL4, 3DL2 et 3DL3 est de 100% dans les deux populations testées, ce qui est en accord avec les données de la littérature.

Les autres gènes peuvent être divisés en deux groupes en fonction de leur fréquence dans la population comorienne. Les gènes dont la fréquence est > 60% représentent les gènes caractéristiques de l'haplotype A ; 2DL1 (100%), 2DL3 (95%), 3DL1 (98%) et 2DS4 (98%).

Les gènes dont la fréquence est < 60% représentent les gènes caractéristiques de l'haplotype B ; 2DL2 (29%), 2DL5 (55%) associés à un certain nombre de gènes activateurs comme 2DS1 (42%), 2DS2 (24%), 2DS3 (9%), 2DS5 (40%) et 3DS1 (31%).

En ce qui concerne les pseudogènes, ils sont retrouvés à une fréquence de 84% pour le 2DP1 et de 91% pour le 3DP1.

		KIR inhibiteurs					KIR activateurs					Pseudogènes					
		2DL1	2DL2	2DL3	2DL4	2DL5	3DL1	3DL2	3DL3	2DS1	2DS2	2DS3	2DS4	2DS5	3DS1	2DP1	3DP1
<b>France</b>	Testés positifs	41	28	44	45	22	40	45	45	35	27	11	44	10	7	27	41
<b>Sud-Est</b>	f(%)	91	62	98	100	49	89	100	100	78	60	24	98	22	15	60	91
(n=45)	F(%)	70	38	85	100	28	67	100	100	53	37	13	85	12	8	37	70
<b>Comores</b>	nombre	45	13	43	45	25	44	45	45	19	11	4	44	18	14	38	41
(n=45)	f(%)	100	29	95	100	55	98	100	100	42	24	9	98	40	31	84	91
	F(%)	100	16	79	100	33	85	100	100	24	13	4	85	22	17	60	70
$\chi^2$			10,08							11,85		11,66					
p corrigé		ns	<0,01	ns	ns	ns	ns	ns	ns	<0,001	<0,001	ns	ns	ns	ns	ns	ns

Table 22 : Fréquences observées (f %) et estimées (F%) des gènes KIR dans les deux populations testées. *ns* : non- significatif.

## **V.2 Les génotypes KIR.**

Nous avons analysé le nombre et le type (inhibiteurs KIR-L et activateurs KIR-S) de gènes KIR présents, pour chaque individu, et nous avons ainsi pu déterminer le génotype de chacun d'entre eux (Table 23 p.159). Nous avons calculé l'équilibre de Hardy Weinberg pour la distribution des génotypes et nos deux populations sont à l'équilibre.

La population comorienne et la population du Sud-Est de la France montrent une diversité dans les profils génétiques rencontrés, cette diversité porte sur le nombre et la combinaison des divers gènes KIR.

Dans la population Comorienne, nous avons trouvé 26 génotypes différents dont 7 n'avaient pas encore été décrits dans la littérature (ng : nouveau génotype dans la table 23 p.159).

Dans la population du Sud-Est de la France, nous avons trouvé 22 génotypes différents dont 4 nouveaux.

Le génotype AA1 est le plus fréquent dans les deux populations, il est présent à 20% dans la population comorienne et à 13.3% dans la population du Sud- Est de la France.

Ce génotype contient une combinaison de gènes caractéristiques de l'haplotype A (2DL1, 2DL3, 2DS4, 3DL1) et pas de gènes caractéristiques de l'haplotype B. Les individus qui possèdent ce génotype sont donc homozygotes pour l'haplotype A.

A l'inverse, le génotype BB ne présente pas les gènes caractéristiques de l'haplotype A mais il possède les gènes spécifiques de l'haplotype B ; 2DL2, 2DL5, et une combinaison variable de gènes activateurs à l'exception du 2DS4. Ce génotype n'a été trouvé que chez un individu de la population du Sud-Est de la France, il est donc homozygote pour l'haplotype B.

De plus, il existe des génotypes AB qui présentent des combinaisons entre les gènes caractéristiques des haplotypes A et B.



	Comores	France SE	2DL1	2DL2	2DL3	2DL4	2DL5	3DL1	3DL2	2DS1	2DS2	2DS3	2DS4	2DS5	3DS1
AA 1	9	6	1	0	1	1	0	1	1	0	0	0	1	0	0
ng	5	0	1	0	1	1	1	1	1	0	0	0	1	0	0
AA 1.1	2	0	1	0	1	1	1	1	1	0	0	0	1	1	0
AA 2	1	0	1	0	1	1	0	1	1	1	0	0	1	0	1
AA 2.1	1	0	1	0	1	1	1	1	1	1	0	0	1	1	1
AA 2.1*	1	2	1	0	1	1	1	1	1	1	0	0	1	1	1
AA 3	3	6	1	0	1	1	0	1	1	1	0	0	1	0	0
AA 3*	1	0	1	0	1	1	1	1	1	1	0	0	1	0	0
AA 103.1	1	1	1	0	1	1	0	1	1	0	0	0	1	1	1
AA103.1*	1	0	1	0	1	1	1	1	1	0	0	0	1	1	1
ng	2	0	1	0	1	1	1	1	1	1	0	0	1	1	0
ng	0	1	1	0	1	1	0	0	1	1	0	0	1	0	0
ng	0	1	1	0	1	1	0	1	1	0	0	1	1	0	0
ng	0	1	1	0	1	1	1	0	1	1	0	0	1	1	0
ng	1	0	1	0	1	1	0	1	1	1	1	0	1	0	0
ng	1	0	1	0	1	1	1	1	1	0	0	0	1	0	1
ng	2	0	1	0	1	1	1	1	1	1	1	0	0	1	0
ng	0	1	1	0	1	1	1	1	1	1	0	1	1	0	0
AB1	0	1	1	1	1	1	0	1	1	0	1	0	1	0	0
AB3.1	1	2	1	1	1	1	1	1	1	1	1	0	1	1	1
AB4.1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1
AB5	0	2	1	1	1	1	1	1	1	1	1	0	1	0	0
AB5.1*	1	2	1	1	1	1	1	1	1	1	1	0	1	1	0
AB8	1	0	1	1	1	1	0	1	1	0	0	0	1	0	0
AB8.1*	1	0	1	1	0	1	1	1	1	0	0	0	1	1	0
AB9	0	1	1	1	1	1	1	1	1	0	1	1	1	0	0
AB103.1*	0	1	1	1	1	1	1	0	1	1	1	0	0	1	1
ng	3	0	1	1	1	1	1	1	1	1	0	1	1	0	1
ng	0	5	1	1	1	1	1	1	1	1	1	1	1	0	0
ng	0	1	1	1	1	1	0	0	1	1	1	0	1	0	0
ng	0	1	1	1	1	1	1	0	1	1	1	1	1	0	0
ng	0	3	1	1	1	1	1	1	1	1	0	0	1	0	0
ng	0	1	1	1	1	1	1	1	1	1	1	0	1	0	1
ng	1	0	1	1	0	1	1	1	1	0	1	0	1	1	1
ng	1	0	1	1	1	1	0	1	1	1	0	0	1	0	0
ng	1	0	1	1	1	1	0	1	1	1	1	0	1	1	0
ng	1	0	1	1	1	1	1	0	1	1	1	0	1	1	1
ng	1	0	1	1	1	1	1	1	1	0	1	0	1	0	0
ng	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1
BB 5.1	0	1	0	1	0	1	1	1	1	1	1	0	1	1	0
C	0	3	0	1	1	1	0	1	1	1	1	0	1	0	0
n =	45	45													

**Table 23** : Génotypes KIR observés et nombre d'individus pour chaque génotype dans les deux populations. 1 : Présence du gène , 0 : absence du gène, ng : nouveau génotype non décrit dans la littérature, n: effectif.

### V.3 Comparaison du polymorphisme des gènes *KIR* dans les populations

- **Comparaison des génotypes.**

Nous avons comparé la fréquence des génotypes présents dans nos deux populations avec celle des populations décrites dans la littérature (Table 24 p.160). Nous constatons que le génotype AA1 est le génotype le plus fréquent dans toutes les populations à l'exception des Indiens et des Aborigènes Australiens. En dehors de ce génotype, on retrouve pour chaque population étudiée une grande diversité de génotypes dont les fréquences sont en moyenne < 6%.

Nomenclatures			Fréquences des génotypes (%)													
Witt et al., 1999; Norman et al., 2001,2002	Crum et al., 2000	Uhrberg et al., 1997	Comores	France SE	Inde	Angleterre	Palestine	Thaïlande	Vietnam	Australie (A)	Afrique	Trinidad (SA)	Karachi (SA)	Japon	Allemagne	
AA 1	1	1	<b>20</b>	<b>13,3</b>	5,6	<b>30,1</b>	<b>22,9</b>	<b>35,3</b>	<b>42,9</b>	1,7	<b>35,3</b>	<b>15,7</b>	<b>11,5</b>	<b>56</b>	<b>31,3</b>	
AA 1.1			4,4									0,9				
AA 2	2	4	2,2	4,4	2,8	8,8	6,7	1,7	8,5	4,5	6,5	2,7	2,6	19,5		
AA 2.1			2,2													
AA 2.1*			2,2	2,2			1,9	4,2				5,6	7,7	7,3		
AA 3			6,7	13,3	1,4	2,2	0,95				1,6					
AA 103.1			2,2	2,2							1,6		2,6			
AA 103.1*			2,2													
AB1	9	7		2,2	5,6	11	14,3	11,8			6,4	8,3	6,4	4,8	15,2	
AB3.1	10	9	2,2	4,4	1,4	4,4	1,9	3,4					2,6	4,8	6,1	
AB4.1	7+8		2,2	4,4	1,4	7,3	3,8	4,2	3,4	<b>16,9</b>		4,6	5,1		11,1	
AB5				4,4	2,8	2,9	1,9	0,8			1,6	5,6	1,3		5,1	
AB5.1*			2,2	4,4							1,6	3,7	3,8			
AB8		11	2,2		1,4						6,4					
AB8.1*			2,2								1,6	0,9		2,4		
AB9		3		2,2	<b>6,9</b>	5,1	11,4	3,4	5,1	4,5	8,1	8,3	<b>11,5</b>	2,4		
AB103.1*			2,2	2,2	2,8		2,9	1,7				1,8	2,6			
BB 5.1				2,2							1,6			2,4	1	
C				6,7							1,6					
n =			45	45	72	136	105	119	59	67	62	108	78			

Table 24 : Comparaison des génotypes avec les populations décrites dans la littérature. n: effectif

- **Comparaison des fréquences géniques.**

Les gènes 2DL4, 3DL2, 3DL3 ont été exclus de la comparaison car leurs fréquences étaient de 100 % dans toutes les populations. Le gène 2DL5 et les deux pseudogènes 2DP1 et 3DP1 ont aussi été exclus car ils n'étaient pas décrits dans la majorité des autres populations.

Gène KIR	2DL1	2DL2	2DL3	3DL1	2DS1	2DS2	2DS3	2DS4	2DS5	3DS1
<b>POPULATIONS</b>										
France SE	91	62	98	89	78	60	24	98	22	15
Comores	100	28	95	98	42	24	9	98	40	31
Afrique	79	52	85	98	23	45	19	97	24	13
Trinidad Asie S	82	64	83	88	55	69	27	81	37	44
Karachi Asie S	90	67	91	81	60	69	45	72	48	56
Angleterre	91	49	92	97	45	51	24	96	32	42
Palestine	83	62	85	88	44	64	37	88	27	39
Thaïlande	97	42	97	93	42	44	25	87	23	44
Inde	87	79	65	87	54	62	43	81	47	39
Vietnam	98	36	98	88	37	41	32	88	0	41
Australie	72	79	67	55	82	85	81	51	0	78
Allemagne	91	50	91	91	39	49	22	96	27	36
Japon	100	16	100	98	34	15	16	98	26	32

**Table 25 :** Comparaison des fréquences (en %) des différents gènes KIR dans les populations.

Afin de visualiser la répartition des populations en fonction du polymorphisme des gènes KIR, nous avons réalisé une analyse en composante principale (Figure 24 p.162)

L'analyse de cette ACP nous permet de dégager quatre groupes homogènes de populations ;

- France Sud-Est (Fse) / Angleterre (UK) / Allemagne (DE)
- Afrique (Afr) / Comores (Kmr)
- Karachi Asie du Sud (Kas) / Palestine (Pal) / Trinidad Asie du Sud (Tas)
- Thaïlande (Thai) / Japon / Vietnam (Viet)

L'Inde et l'Australie (Aus) sont isolées.

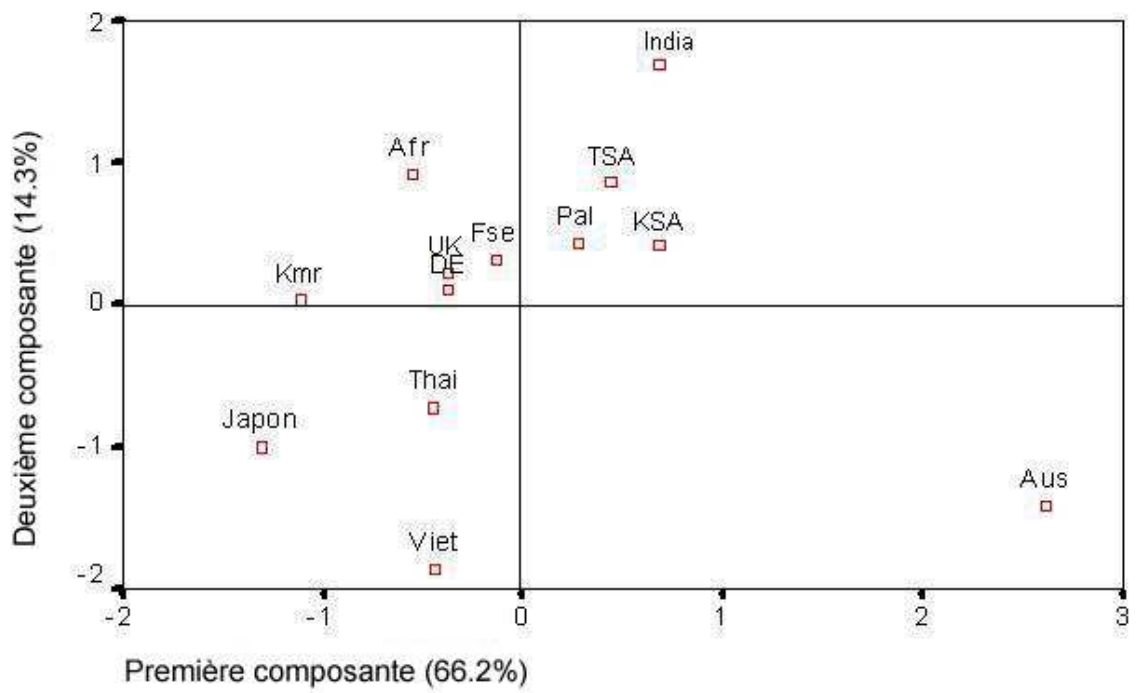


Figure24 : Diagramme de l'Analyse en Composante Principale (ACP).

## **DISCUSSION.**

L'étude génétique de la population Comorienne de Marseille était motivée par une double problématique : anthropologique et biomédicale. L'approche anthropologique avait pour objectif de décrire la structure génétique de cette population et de rechercher les traces génétiques des trois principales vagues migratoires impliquées historiquement dans le peuplement de l'archipel des Comores : Africaine, Arabe et Austronésienne. L'aspect biomédical devait évaluer l'impact transfusionnel des particularités immuno-hématologiques de cette communauté migrante résidant à Marseille et peu sensibilisée au don du sang.

## **I. Analyse de la contribution d'Afrique subsaharienne.**

### **I.1 Les données du polymorphisme érythrocytaire.**

L'analyse des fréquences phénotypiques et géniques des systèmes érythrocytaires évoque une grande similitude avec de nombreuses populations d'Afrique subsaharienne (Table 16 p.137). Cette similitude est détectable avec l'ensemble des systèmes de groupes sanguins étudiés.

#### **Les fréquences phénotypique et allélique.**

- **L'allèle *O* présente une haute fréquence (78.4%)** par rapport aux autres allèles *A* (14.1%) et *B* (6.9%).
- **L'allèle *K* (1.2%) est très rare et l'allèle *Kpa* est absent.** Ces deux allèles sont considérés comme les mutants respectifs des allèles *k* et *Kpb*. Ces derniers constituent l'haplotype originel (Redman CM, 1989 ; Blancher A, 2000) qui est le plus largement répandu en Afrique subsaharienne. En Europe, bien que les allèles *K* et *Kpa* soient considérés comme peu fréquents, ils atteignent tout de même des fréquences respectives de 4% et 2% (Daniels G, 2002).

- **L'haplotype *R0* est le plus courant des haplotypes du système RH (48%).** Cet haplotype représente un des traits les plus caractéristiques de l'Afrique subsaharienne puisqu'il dépasse fréquemment les 60% et il peut atteindre dans certaines populations 80 à 90% (Table 16 p.137). Il est très peu représenté en Europe (2%), en Extrême Orient et en Asie du Sud Est (3 à 4%) (Daniels G, 2002, Mourant AE, 1976). Cet haplotype, constitué d'un gène *RHD* et d'un allèle *RHce* du gène *RHCE*, représente probablement l'allèle ancestral (Carrit B, 1997) à l'origine des différents haplotypes du locus *RH*. Des événements génétiques (conversion, recombinaison, délétion) entre le gène *RHD* et l'allèle *Rhce* ont produit les autres allèles *RH* courants (*R1*, *R2*, *r*) plus fréquents dans d'autres populations probablement moins anciennes (Aspect évolutif du système RH Figure 12 p.73).
- **Le deuxième haplotype du système RH le plus fréquent est l'haplotype *r* (20.3%).** Sa fréquence est similaire à celle observée dans la majorité des populations d'Afrique subsahariennes (Table 16 p.137).
- **La fréquence du phénotype Fy (a-b-) (86.2%) atteint presque les plus hautes valeurs africaines observées chez les Bantous et les Pygmées (94%)** (Cavalli-Sforza L, 1994a). Les données de la biologie moléculaire démontrent que le mécanisme d'expression du phénotype Fy (a-b-) observé en Grande Comore est identique à celui observé en Afrique subsaharienne. Il est, en effet, caractérisé par la mutation T46C dans le promoteur du gène Duffy. Cette mutation prévient la fixation d'un facteur de transcription dans le tissu érythroïde (GATA-1) et empêche l'expression des molécules Duffy sur les hématies mais pas dans les autres tissus où un autre promoteur permet l'activation du gène (Tournamille C, 1995 ; Peipper SC 1995). La présence, à une telle fréquence, du phénotype Fy(a-b-) « Africain » dans ces régions peut être liée, comme l'évoquent les données historiques, à une contribution initiale et massive de peuples d'Afrique subsaharienne, notamment Bantous.
- **L'absence totale d'allèle *FYX*** caractéristique des populations Européennes est attestée par les explorations en biologie moléculaire (Chown B, 1965 ; Cedergren B, 1973)

- **L'allèle *Jka* présente une fréquence très élevée (70%)** Ceci est caractéristique des populations d'Afrique subsahariennes (Nurse, 1977) où ses valeurs sont fréquemment comprises entre 62 et 70% pour atteindre parfois plus de 80% chez les Pygmées (Table 16 p.137).
- **Les allèles dits "rares" du gène *RHCE* sont présents.** Les premiers résultats montrent la présence de certains de ces allèles (ceMO,ce<sup>s</sup>(340) et (C)ce<sup>s</sup>). Compte tenu de leur répartition exclusive dans les populations d'Afrique subsahariennes (Noizat Pirenne F, 2002), leur présence dans la population Comorienne représente un élément supplémentaire en faveur de la contribution Africaine.

### Les distances génétiques.

L'étude des distances génétiques confirme la proximité de la population Comorienne avec les populations d'Afrique subsaharienne et notamment Bantoues. En effet, parmi les 8 populations les plus proches, sur les 17 analysées, figurent 5 populations Bantouphones (Table 18 p.139). Cette similitude semble plus prononcée avec les Bantous de l'Est (Bitonga, Zéruru).

La population des Sandaweés pourrait en représenter une 6°. Bien que parlant une langue Khoi, les Sandaweés présentent plus de similitude génétique avec les Bantous qu'avec les Khoisans. Ils pourraient ainsi représenter une population Tanzanienne originelle qui a maintenu sa langue initiale (peut être imposée initialement par les Khoisan), mais qui a été génétiquement infiltrée lors de l'expansion Bantoue à partir du nucléus Congolais (Cavalli-Sforza L, 1994a). Ils illustrent un nouvel exemple de distorsion Gène – Langue par processus de remplacement de gènes.

La position des populations d'Afrique de l'Est (Somali et Béja) sera discutée lors de l'analyse de la contribution « arabe ».

Les autres populations Africaines (les Hadza, les Bédik, les Pygmées, les San et les Libyens) apparaissent plus éloignées.



La représentation des distances sous forme d'arbre phylogénétique (Figure 20 p.140) confirme le regroupement Bantous/Comores et l'éloignement relatif des San et des Bédik.

### **L'analyse multidimensionnelle et la classification ascendante hiérarchique**

L'analyse multidimensionnelle (Figure 21 p.142) et la classification ascendante hiérarchique (Figure 22 p.144) confirment le regroupement de la population Comorienne avec l'ensemble des populations Bantoues et sa proximité avec les populations d'Afrique de l'Est (Somalis et Béja).

#### **1.2 Les données du polymorphisme HLA-DRB1**

L'étude des distances génétiques confirme la proximité entre les populations des Comores et celles d'Afrique subsaharienne (Figure 23 p.155). En effet, parmi les 11 populations les plus proches 7 sont Africaines (Table 21 p.154). La population la plus proche est représentée par une population Bantoue du Nord Ouest (Congo) et d'une manière générale, ces populations Bantoues (Congo, Gabon) sont plus proches que celles d'Afrique de l'Ouest non Bantoues. Les populations des Comores, du Congo et du Gabon présentent d'ailleurs les allèles *HLA-DRB1- 11, 13, 15* à des fréquences similaires.

Enfin, on retrouve sur l'analyse multidimensionnelle un cluster Africain incluant les Comores

### **1.3 Les données du polymorphisme des gènes KIR.**

L'analyse de l'ACP (Figure 24 p.162) permet de dégager quatre groupes homogènes ( [*France Sud Est / Angleterre / Allemagne*] - [*Afrique / Comores*] - [*Karachi (Asie du sud) / Palestine / Trinidad (Asie du sud)*] - [*Thaïlande / Japon / Vietnam*] ) qui apparaissent corrélés avec la répartition géographique. La position des Comores proche de l'Afrique semble aller dans le sens de l'hypothèse d'une origine Africaine. Cependant, cet échantillon africain publié (Norman PJ, 2002) n'est pas clairement défini quant à son origine géographique. Il semblerait toutefois qu'il provienne d'Afrique subsaharienne de l'Ouest.

### **1.4 Conclusions relatives à la contribution Africaine.**

Ainsi, en accord avec les données linguistiques, historiques et culturelles la population des Comores semble bien appartenir au groupe populationnel Bantou et représenter la limite la plus orientale de cette expansion massive et récente. L'ensemble de ces éléments pourrait situer l'origine des premiers Bantous de la Grande Comore au Mozambique. Cette hypothèse, évoquée par la similitude génétique avec les Bitonga par exemple (Bantous du Mozambique), est renforcée par les données de la tradition orale de Said Ahmed (Robinau C, 1966) qui avançaient déjà cette origine et par les données linguistiques. Ces dernières rapprochent le langage du groupe Ouest des îles (Grande Comore et Mohéli) des langues usitées au Mozambique comme l'Esakaji (Vérin P, 1994). Ces résultats confirment la haute corrélation entre profil génétique et affiliation linguistique déjà rapportée en Afrique non seulement avec le système HLA (Sanchez-Mazas A, 2001) mais aussi avec d'autres marqueurs comme ceux des systèmes Gm et RH (Excoffier L, 1991), ceux du chromosome Y (Pereira L, 2002 ; Poloni ES, 1997 ; Scozzari, 1999) ou ceux de l'ADN mitochondrial (Pereira L, 2001).

## **II. Analyse de la contribution « Arabe ».**

### **II.1 Les données du polymorphisme érythrocytaire.**

L'étude des fréquences alléliques et des distances géniques semble montrer une similitude avec les populations d'Afrique de l'Est qui peut être interprétée comme un témoin indirect de la contribution « Arabe ». Par ailleurs, les distances génétiques positionnant les populations de la Péninsule Arabique parmi les populations les plus proches représentent, ici, un témoin direct de cette contribution.

#### **Les fréquences génétiques**

L'analyse des fréquences des haplotypes *RH* montre des valeurs proches de celles observées en Afrique de l'Est pour les haplotypes *Ro* et *RI*. En effet, bien que l'haplotype *Ro* soit le plus courant dans la population Comorienne sa fréquence représente un niveau intermédiaire entre les valeurs les plus basses d'Afrique, observées en Afrique de l'Est (43 à 65% (table 16 p.137)), et celles observées dans les autres populations parlant des langues afro-asiatiques comme celles d'Afrique du Nord (11 à 20% (table 16 p.137)) et de la Péninsule Arabique (10% (table 16 p.137) à 15% en moyenne (Cavalli-Sforza L 1994a)). Cette plus faible fréquence de l'haplotype *RO* est compensée en particulier par la fréquence de l'haplotype *RI*. La valeur de l'haplotype *RI* (19.5%) tend vers les plus hautes valeurs d'Afrique subsaharienne qui sont habituellement observées en Afrique de l'Est où elles atteignent en moyenne 16 à 27% (Table 16 p.137) alors elles restent comprises entre 7 et 10% (Table 16 p.137) dans les autres régions d'Afrique.

### **Les distances génétiques.**

Les valeurs des distances génétiques démontrent d'une part ;une proximité entre la population Comorienne et les populations d'Afrique de l'Est (Somali et Béja) qui figurent parmi les 8 populations les plus proches ; et d'autre part que la population des Saoudiens est plus proche que certaines populations Africaines (Sans et Libyens).

La représentation des distances sous forme d'arbre phylogénétique (Figure 20 p.140) situe la population Comorienne en position intermédiaire entre le "cluster" Africain, qui comprend les populations d'Afrique de l'Est (Somali et Béja), et l'Arabie Saoudite.

### **L'analyse multidimensionnelle et la classification ascendante hiérarchique.**

La proximité de la population Comorienne avec les populations d'Afrique de l'Est ainsi que sa position intermédiaire par rapport aux Saoudiens sont aussi révélées par l'analyse multidimensionnelle et le dendrogramme.

## **II.2 Les données du polymorphisme HLA-DRB1.**

L'analyse des données du polymorphisme HLA-DRB1 va dans le sens des résultats des marqueurs érythrocytaires en situant la population Comorienne proche des populations de l'Afrique de l'Est.

En effet, parmi les 10 populations apparaissant les plus proches (Table 21 p.154) nous retrouvons 1 population d'Afrique de l'Est (les Noba), 2 du Moyen Orient (Iran et Israël) et 2 d'Afrique du Nord Est (Egypte : le Caire et Siwa).

L'analyse multidimensionnelle (Figure 23 p.155) montre une proximité relative de la population comorienne avec l'Afrique du Nord, l'Afrique de l'Est et le Moyen Orient.

### **II.3 Les données du polymorphisme des gènes KIR.**

Malheureusement l'absence de données dans la littérature sur ce polymorphisme dans les populations « Arabes » ne nous a pas permis d'analyser leur contribution dans le pool génétique de la population Comorienne.

### **II.4 Conclusions relatives à la contribution « Arabe ».**

Le faible niveau de différenciation constaté entre les populations d'Afrique de l'Est et des Comores peut être considéré comme le témoin indirect d'une contribution "Arabe" sur le pool génétique comorien.

En effet, l'Afrique de l'Est représente un véritable corridor non seulement dans le sens Moyen Orient → Afrique du Nord (Manni F, 2002) mais aussi dans le sens Moyen Orient → Afrique subsaharienne (Richard, 2003). Cela est attesté par les faibles distances génétiques entre certaines populations d'Afrique subsaharienne et celles d'Afrique de l'Est (Egypte et Soudan [Noba, Siwa] pour le HLA et celles des Somalis et des Béja pour les groupes érythrocytaires). L'Afrique de l'Est est, en effet, un lieu de rencontre entre « Arabes » et Africains, tel qu'il est évoqué aux Comores où une population Bantoue initiale a vu des vagues migratoires Arabes provenant directement du Golfe Persique ou, surtout, de populations déjà établies sur la côte Est Africaine (Ottino P, 1974). Cette faible différenciation génétique entre les populations d'Afrique de l'Est et des Comores pourrait être liée au fait que ces deux régions représentent des lieux de rencontre de populations identiques, plus que d'une influence directe de l'Afrique de l'Est sur les Comores. Cette analyse permet aussi d'expliquer la proximité génétique, observée avec le HLA, entre les Comores et l'Egypte qui résulterait de ce flux génique et non d'une contribution directe. Ce témoin d'une contribution « Arabe » est par ailleurs plus directement évoqué par l'analyse des distances génétiques aussi bien avec les systèmes érythrocytaires qu'avec le système HLA. En effet, d'après les marqueurs érythrocytaires, les populations de la Péninsule

Arabique apparaissent non seulement plus proches que toutes autres populations non Africaines, mais aussi plus proches que certaines populations Africaines comme l'Afrique du Nord ou les Sans. En ce qui concerne le polymorphisme HLA, la place (2<sup>o</sup> position) de l'Iran est aussi très évocatrice de cette contribution.

### **III. Contribution de l'Asie du Sud (Inde et Pakistan).**

Au regard des distances génétiques (Table 18 p.139) et du positionnement sur l'arbre phylogénétique (Figure 20 p.140) réalisés à partir des marqueurs érythrocytaires, les populations d'Inde (Gadaba) apparaissent comme très différenciées par rapport à la population Comorienne. Cela est matérialisé aussi bien par les valeurs des distances génétiques que par le positionnement respectif de ces deux populations sur la représentation de l'analyse multidimensionnelle (Figure 21 p.142) et sur le dendrogramme (Figure 22 p.144).

En ce qui concerne le polymorphisme HLA, l'analyse multidimensionnelle (Figure 23 p.155) montre un positionnement des populations d'Asie du Sud (exception des Brahui et des Baloch) au sein d'un cluster constitué de l'Afrique Bantoue, de l'Afrique de l'Est, du Moyen Orient et des Comores. Cela est attesté par les faibles distances génétiques existant entre les populations d'Asie du Sud et celles d'Afrique de l'Est ou d'Afrique subsaharienne qui apparaissent parfois plus faibles qu'entre celle des Comores et celles d'Afrique subsaharienne (Gibert M, 2003). Cette proximité est probablement liée à un partage de gènes Africains dû à une contribution génétique commune plutôt qu'à une influence directe de l'Asie du Sud sur les Comores comme l'évoquent les hypothèses migratoires Afrique → Asie du Sud soutenues par des données génétiques. En effet, des études du polymorphisme du chromosome Y menées au Pakistan évoquent deux routes, l'une par le Nord du Golfe Persique au travers de l'Iran et une autre côtière au Sud (Quamar R, 1999) qui est le témoin

de l'expansion du "Out of Africa" (Kivisild T, 2003). De même, ce flux génique Afrique → Inde est aussi matérialisé par la détection, dans des populations Indiennes, de formes alléliques "Bantoues" de drépanocytose (Niranjan Y, 1999). Ainsi, sur la base de cette hypothèse, la proximité génétique existant entre les populations Comoriennes et celles d'Asie du Sud pourrait être surtout le fait de traces du « passage Africain » en Inde que d'un impact direct de cette région sur le pool génique comorien.

#### **IV. Contribution de l'Asie du Sud-Est.**

En ce qui concerne le polymorphisme érythrocytaire, l'étude des distances génétiques (Table 18 p.139) attribue aux populations d'Asie de Sud Est (Indonésiens et Toba batak) les valeurs les plus élevées. Cet éloignement est confirmé par leur position sur l'arbre phylogénétique (Figure 20 p.140), l'analyse multidimensionnelle (Figure 21 p.142) et sur le dendrogramme Figure 22 p.144°.

En ce qui concerne le polymorphisme HLA, les données disponibles montrent une forte différenciation entre les populations de Madagascar (Mérina) et celle des Comores. Seule la haute fréquence de l'allèle *HLA-DRI2* et le fait que cette population soit la plus proche parmi celles d'Asie du Sud Est pourraient évoquer un ancien contact austronésien. Celui semble être lié à une contribution indirecte des austronésiens à partir des raids provenant de Madagascar (Pirates Malgaches) ou à une contribution directe des Bantous sur la population Malgache.

Enfin, l'analyse du polymorphisme des gènes KIR révèle également une forte différenciation entre les populations d'Asie du Sud Est et les Comores.

## **V. Influence Européenne.**

Bien qu'il y ait eu des contacts Portugais (XVI<sup>e</sup> au XIX<sup>e</sup> siècle) et Français (1887 à 1975), nous ne pouvons considérer l'Europe comme une population parentale potentielle pour deux raisons. Tout d'abord parce que les migrations Portugaises en Afrique étant de faible effectif, elles ont eu à ce titre un faible impact sur les populations locales. Les études menées au Mozambique sur le chromosome Y (Pereira, 2002) et sur l'ADN mitochondrial (Pereira, 2001) confirment l'influence limitée des Portugais et des Européens. Les auteurs rapportent l'absence totale de lignage maternel et un impact génétique mâle très faible. La deuxième raison est directement liée aux critères de sélection des individus étudiés qui garantissent l'absence d'un apport génétique Français.

## **VI. Evaluation quantitative des diverses contributions.**

Cette évaluation est difficile. En effet, les régions de l'Océan Indien apparaissent soumises à des flux géniques continus rendant difficile la sélection des populations parentales putatives. De plus, la dérive génétique a pu agir et affecter en particulier les populations Comorienne et d'Asie du Sud Est qui sont très endogames (Cavalli-Sforza L, 1994a). Malgré tout une première estimation peut être réalisée. Cette estimation a été menée séparément avec le polymorphisme érythrocytaire d'une part et celui du système HLA-DRB1 d'autre part.

En ce qui concerne le polymorphisme érythrocytaire, les calculs "d'admixture" ont été réalisés en sélectionnant les populations putatives qui apparaissent historiquement impliquées dans l'histoire du peuplement des Comores à savoir : une population Bantoue (Bitonga) pour la contribution Africaine, une population de la Péninsule Arabique (Saoudiens) pour la contribution « Arabe » et enfin celle de l'Indonésie pour la contribution



Austronésienne. Les résultats font état d'une contribution Bantoue à hauteur de 69.5% +/- 5.2%, Arabe de 29% +/- 2.8% et une très faible contribution Indonésienne à hauteur de 1.5% +/- 3.7%. Ce dernier résultat est à interpréter avec une très grande prudence compte tenu de l'évaluation de l'erreur standard (SE) à 3.7% qui souligne à nouveau les difficultés que nous avons eu à définir avec précision les populations Indonésiennes qui ont réellement été en contact avec la population comorienne (Bertoni B, 2003).

En ce qui concerne le polymorphisme HLA, la contribution peut être estimée à partir des populations ayant des distances génétiques les plus faibles comme le Congo (Contribution Bantou Nord Ouest), l'Iran (Contribution "Arabe"), les Malay (Contribution directe d'Asie du Sud Est) ou les Mérina (Contribution indirecte d'Asie du Sud Est) et enfin les Pathan (Contribution d'Asie du Sud). L'Iran a été sélectionné plutôt que l'Egypte du fait de la théorie du "corridor" et sur la base de données historiques. Les résultats montrent une contribution à nouveau majoritaire en provenance d'Afrique de l'ordre de 50%, puis une contribution Arabe de l'ordre de 28%, d'Asie du Sud Est de l'ordre de 10 à 14% et enfin d'Asie du Sud de l'ordre de 6 à 8%.

Ces résultats sont en accord avec les rares études biologiques menées sur la population comorienne qui ont analysé les bases moléculaires des hémoglobinopathies et des déficits en G6PD des Comoriens de Marseille (Badens C, 2000). Cette étude montre que les allèles africains de ces pathologies sont prédominants. Par ailleurs, 17% des déficits en G6PD et toutes les mutations des bêta-thalassémies sont d'origine Méditerranéenne. Sur ces bases les auteurs concluent que la population Comorienne est née d'un mélange génique Africain et Arabe.

## **VII. Les premiers résultats du polymorphisme du chromosome Y.**

Les résultats préliminaires permettent une première estimation de la fréquence des haplogroupes D, E (YAP+) à hauteur de 72%, confirmant une forte influence Africaine.

## **VIII. L'exploration du polymorphisme KIR en Anthropologie.**

Le génotypage KIR se révèle utile pour mettre en évidence des différences entre les populations. Il semble donc intéressant, d'un point de vue anthropologique, d'étudier le polymorphisme des gènes *KIR*, et nous proposons de développer son utilisation en tant que nouveau marqueur dans l'analyse de la distribution de la variabilité biologique humaine et dans la reconstruction de l'histoire du peuplement humain.

## **IX. Analyse de l'effet de la sélection naturelle.**

La fréquence des différents phénotypes Duffy révèle des interactions "homme – milieu". En effet, la fréquence du phénotype silencieux à plus de 85% dans ces régions repose non seulement sur la nécessité d'un apport massif et initial par des populations subsahariennes mais aussi sur la nécessité du maintien d'une telle fréquence dans le temps. Comme les individus porteurs de ce phénotype sont résistants à la pénétration du *Plasmodium vivax* (Miller, 1975), l'allèle *FY0* est considéré comme un candidat sérieux pour être une cible de la sélection naturelle. L'expression des antigènes Duffy est par contre maintenue sur les cellules endothéliales, leur permettant de continuer à assurer leur fonction d'épurateur des interleukines. La persistance du *Plasmodium vivax* en Grande Comore (Blanchy S, 1999), pourrait être considérée comme le facteur de maintien, par pression sélective, de cette haute

fréquence dans le temps. De même, le maintien du *Plasmodium vivax* en Grande Comore pourrait être directement lié à la persistance, probablement suffisante, de phénotypes Duffy exprimant encore des antigènes érythrocytaires qui assurent ainsi le cycle vital de ce parasite (phénomène de co-évolution).

## **X. Communauté Comorienne migrante et impacts transfusionnels.**

### **X.1 Les antigènes et les phénotypes courants.**

Il s'agit tout d'abord du phénotype D cc ee qui est présent à près de 48% dans la population Comorienne. Ce phénotype est lié à la présence en double dose de l'allèle ancestral *Ro* caractéristique de nombreuses populations Africaines. Il est beaucoup plus rare dans notre population de donneurs de sang de l'EFS Alpes-Méditerranée, puisqu'il se situe à 2.6%. Ainsi, pour prévenir l'allo-immunisation chez ces sujets, il est souvent nécessaire d'avoir recours à des unités de sang de phénotype D négatif (dd cc ee) qui ne représentent que 16% de nos donneurs.

Les écarts de fréquences phénotypiques sont aussi observés avec d'autres antigènes. En effet, l'antigène MNS3 (S) est absent chez 60% des Comoriens alors qu'il est présent chez 60% de notre population de donneurs. L'antigène FY1 (Fya) est absent chez 88% des sujets Comoriens alors qu'il est présent chez 70% de notre population de donneurs. L'antigène JK2 (Jkb) est absent chez 50% des sujets Comoriens alors qu'il est présent à hauteur de 80% dans notre population de donneurs. Lorsqu'on considère l'absence simultanée de ces trois antigènes, la fréquence des donneurs de sang de l'EFS Alpes-Méditerranée ayant le phénotype compatible (Fya-, Jkb-, S-) pour transfuser un patient porteur d'un mélange d'anticorps anti-Fya-Jkb-S, est de 5.6% contre 26.4% dans la population Comorienne.

## **X.2 Les allèles « rares » et les phénotypes « rares ».**

Comme dans les populations originaires d'Afrique subsaharienne, la problématique transfusionnelle implique aussi des phénotypes qui apparaissent beaucoup plus rares dans d'autres populations comme celles originaires d'Europe et de France en particulier.

Le premier est représenté par le phénotype FY:-1,-2 (Fy (a-b-)). Il constitue un phénotype rare compte tenu qu'il est considéré comme exceptionnel dans les populations non africaines et donc dans la population de nos donneurs de sang. Il représente près de 90% des individus Comoriens. En général, la transfusion de ces sujets ne pose pas de problèmes majeurs compte tenu de la rareté de l'immunisation vis à vis de l'antigène de grande fréquence FY3. En effet, comme nous l'avons vu précédemment, ce phénotype retrouvé chez les Comoriens est identique à celui observé en Afrique subsaharienne. A ce titre, il est caractérisé par la présence en double dose d'un allèle silencieux *FY0* dont la mutation dans le promoteur du gène empêche l'expression de celui ci dans le tissu érythroïde mais pas dans les autres tissus et notamment sur les cellules endothéliales où Fyb et FY3 s'expriment. La présence de ces deux antigènes sur les cellules non érythroïdes explique la rareté d'immunisation correspondante et le fait qu'il faille, d'un point de vue transfusionnel, respecter la compatibilité Fya négatif. Le recours à des hématies phénotypés Fy (a-b-) chez les patients non immunisés n'est pas nécessaire. Cette indication peut se discuter dans un contexte de poly immunisation où le recours à des hématies Fy (a-b-), provenant toujours de donneurs d'origine Africaines, permet d'obtenir une meilleure phéno-compatibilité au niveau des antigènes courants (Noizat-Pirenne F, 2003).

Le second phénotype rare rencontré en Afrique subsaharienne est le phénotype S-, s-, U-. Ce phénotype n'a pas été mis en évidence dans notre échantillon d'individus Comoriens. Le fait que nous ayons eu à gérer, en contexte transfusionnel, deux patients d'origine Comorienne avec ce phénotype, nous laisse suspecter sa présence dans cette population.

D'un point de vue transfusionnel, il convient de recourir à des hématies MNS:-5 dès l'apparition de l'anti-S. La présence de ce phénotype chez de jeunes patients drépanocytaires, est un élément de mauvais pronostic compte tenu de la rareté des unités compatibles et doit faire discuter des alternatives thérapeutiques comme la greffe de cellules souches hématopoïétiques.

Les spécificités dites "rares" du système RH dans la population Africaine correspondent essentiellement à la perte d'expression d'un antigène de grande fréquence (RH:32,-46 / RH:-18 / RH:-34) et à la présence d'antigène RH5 partiel. Les résultats préliminaires concernent seulement la détection des allèles associés au phénotype RH5 partiel et de l'haplotype (C)*ce<sup>s</sup>* associé au phénotype RH:-34. L'allèle *ce<sup>s</sup>(340)* présente une fréquence élevée (19.4%). Les trois individus homozygotes sont porteurs d'un phénotype RH5 partiel et à ce titre en situation de blocage transfusionnel. Ces personnes seront contactées afin qu'elles puissent régulièrement donner du sang en vue d'alimenter la banque nationale de sangs rares. La fréquence de l'allèle *ce<sup>MO</sup>* présente une fréquence moins élevée 2%. Les valeurs des fréquences aboutissant au phénotype RH5 partiel sont difficilement comparables aux données de la littérature. En effet, la seule étude qui rapporte la fréquence de ces allèles repose sur l'analyse de 146 patients drépanocytaires d'origine afro-caribéenne provenant de 19 pays différents (Angola :1, Bénin :12, Centre- Afrique :2, Cameroun :18, Congo :19, Guinée :5, Côte d'Ivoire :7, Gabon :3, Ghana :1, Guadeloupe :29, Guyane :3, Madagascar :1, Mali :5, Martinique :12, Niger :1, Sénégal :6, Togo :13, Zaïre :4, origine ? : 4). Enfin, il existe une fréquence apparemment élevée de l'haplotype (C)*ce<sup>s</sup>*. En conséquence, le phénotype rare correspondant RH :-34 chez les sujets ddCcee doit être présent dans cette population et peut donc potentiellement être responsable de problèmes transfusionnels. De plus, les gènes *RHD* et *RHCE* anormaux de cet haplotype particulier doivent être séquencés pour vérifier s'ils sont identiques à ceux observés chez les sujets d'origine Africaine et Antillaise.

Sur le plan sérologique, les sujets RH :-34 peuvent être dépistés au sein des sujets présentant un phénotype ddCcee. Seule la biologie moléculaire peut permettre de mettre en évidence la présence de 2 haplotypes (C)ce<sup>s</sup> puisque il existe 3 génotypes probables pour ce phénotype : dCe/dce, dce/(C)ce<sup>s</sup>, (C)ce<sup>s</sup>/(C)ce<sup>s</sup>. La mutation 733 associée à l'expression des antigènes de basse fréquence RH10 et RH20 est largement représentée (70%) comme c'est le cas chez les sujets Antillais et Africains.

Enfin, certains antigènes, qui sont considérés comme des antigènes de basse fréquence dans notre population de donneurs de sang, sont plus répandus dans les populations originaires d'Afrique subsaharienne. On peut, citer, outre les antigènes RH10 et RH20, l'antigène Jsa qui est retrouvé chez près de 20% des sujets originaires d'Afrique subsaharienne et le RH23 associés à certains variants RHD. Ces antigènes de faible fréquence ne posent pas de problème transfusionnel dans un contexte classique. En revanche, ils doivent être pris en compte lorsque la transfusion nécessite du sang de phénotype rare, car les unités transfusées proviennent souvent des mêmes donneurs. L'exposition répétée aux mêmes antigènes RH20 ou Jsa peut aboutir à une immunisation et à terme à une impasse transfusionnelle. Tout comme on tient compte des antigènes courants, il est aussi nécessaire de tenir compte de ces marqueurs antigéniques spécifiques des populations Africaines lors des transfusions "intra-populationnelles" (Noizat-Pirenne F, 2003).

## **CONCLUSION.**

Les résultats concernant l'exploration de 6 systèmes de groupes sanguins érythrocytaires (ABO, RH, KEL, FY, JK, MNS), du polymorphisme *HLA-DRB1* et des gènes *KIR* ainsi que les résultats préliminaires du polymorphisme du chromosome Y, démontrent que le profil génétique de la population comorienne est proche de l'Afrique Bantoue dont la contribution peut être estimée entre 50 et 70%. Il apparaît que des populations Bantoues venues de la côte Est de l'Afrique ont atteint assez tôt les Comores et que leur origine pourrait se situer au Mozambique. Ainsi, cette influence Africaine est non seulement détectée dans la langue (Le Comorien appartient à la famille des langues Bantoues) et dans les traditions (pratique du grand mariage, position sociale centrale de la femme) mais aussi dans les gènes.

Ces données génétiques sont par ailleurs en accord avec la forte influence de la Péninsule Arabe, durant le Moyen Age, dans cette partie Ouest de l'Océan Indien (Ottino P, 1974). Cette influence est retrouvée aujourd'hui au niveau linguistique (alphabet arabe et emprunt de vocabulaire) et dans les pratiques religieuses (Islam). Les données historiques évoquent tout d'abord un véritable mythe Schirazien (Vérin P, 1994), modèle de construction identitaire des Comoriens. Elles évoquent aussi des mariages contractés entre les filles des chefs locaux Comoriens et les derniers migrants Arabes du moment, arrivés lors des différentes vagues migratoires (Verin P, 1994, Allibert C, 1993). On retrouve donc un processus de dominance élitiste complété par des « métissages génétiques » qui ont laissé une contribution de près de 30% dans le pool génique Comorien.



Enfin, et bien que ces données soient à prendre avec beaucoup de précaution compte tenu des difficultés précédemment exposées (sélection des populations impliquées), ces résultats évoquent une très faible, voir une absence d'influence génétique en provenance d'Asie du Sud Est. Ceci est en accord avec les données linguistiques qui ne détectent pas, contrairement à Madagascar, de contribution Austronésienne. L'influence Indonésienne semble donc limitée à l'adoption de certaines pratiques agricoles (riz, bétel, cocotier....) et de constructions navales (pirogue à balancier) (Vérin P, 1994).

En ce qui concerne l'aspect biomédical, nous avons constaté des écarts importants entre les fréquences phénotypiques des donneurs de sang de l'EFS Alpes-Méditerranée et de la communauté comorienne résidant à Marseille. Ces écarts concernent non seulement des allèles « rares » mais aussi certaines combinaisons d'allèles courants susceptibles de placer des individus en situation de blocage transfusionnel potentiel. Ainsi, l'approche anthropogénétique d'une population comme celle des Comores apporte non seulement des informations sur son histoire et sa diversité génétique actuelle mais peut s'avérer extrêmement utile pour appréhender les écarts phénotypiques inter-populationnels en contexte transfusionnel. Ceci est particulièrement utile dans une ville comme Marseille véritable carrefour culturel et lieu de rencontre historique de multiples communautés.

Au total, trois points forts ressortent de cette étude de la communauté Comorienne de Marseille :

La confirmation de la coexistence des influences Africaine Bantoue et « Arabe », dont les stigmates apparaissent non seulement au travers des pratiques culturelles (coexistence d'un Islam imprégné de traditions africaine) et linguistiques (langue bantoue avec nombreux emprunts arabes) mais aussi dans la structure génétique de cette population. L'influence Austronésienne, n'ayant pas eu d'impact linguistique, semble n'avoir eu que peu ou pas d'impact génétique.

Le risque potentiel de blocage transfusionnel par manque d'unités de sang compatibles lié à l'existence d'écarts phénotypiques constatés entre la communauté Comorienne de Marseille et la population des donneurs de sang répertoriée par l'EFS Alpes-Méditerranée. Ce point met en avant la nécessité d'actions spécifiques de sensibilisation de ces individus au don de sang.

Enfin, d'un point de vue plus général, cette étude confirme à nouveau, la similarité génétique existant entre les populations appartenant à des familles linguistiques identiques (Contribution Bantoue et Arabe). Elle est donc accord avec le fait que l'appartenance à une famille linguistique représente l'indice prédictif majeur de l'existence de relations génétiques entre les populations d'une même famille et qu'elle peut refléter les éventuels événements historiques migratoires survenus au cours de l'évolution d'une population (Excoffier L, 1987).

## **PERSPECTIVES.**

Afin de compléter la connaissance du polymorphisme génétique de cette population, cette étude peut être poursuivie par l'exploration :

- Du Chromosome Y (Thèse en cours A.Ducourneau),
- Des marqueurs Gm des Immunoglobulines (Collaboration Marseille-Toulouse avec JM.Dugoujon)
- Des autres allèles rares du gène *RHCE*. De plus, pour les individus porteur de l'allèle  $(C)ce^s$ , les gènes *RHD* et *RHCE* anormaux de cet haplotype particulier doivent être séquencés pour vérifier qu'ils sont identiques à ceux des haplotypes  $(C)ce^s$  des sujets d'origine africaine et antillaise. Il conviendra par ailleurs de confirmer leur génotype par sérologie en utilisant des anticorps composé anti-Ce ou par biologie moléculaire des gènes *RHCE*. Il sera intéressant d'analyser la réactivité de l'antigène partiel RhC (titrage, réactivité avec des monoclonaux)  
Pour les individus porteurs de l'allèle 733, il conviendra de tester l'expression de RH20 et / ou RH10 chez les sujets portant la mutation 733.
- Des bases moléculaires du phénotype D négatif. Celui-ci prend toute son importance dans la mise en place du génotypage *RHD* dans le cadre de la maladie hémolytique du nouveau né. Chez les sujets non africains, ce phénotype est caractérisé (Wagner, 2001) par la présence en double dose d'une délétion du gène *RHD* (Colin Y, 1991). Chez les sujets originaires d'Afrique subsaharienne, deux gènes supplémentaires sont retrouvés avec une fréquence non négligeable qui peuvent aboutir à ce phénotype : le pseudogène (Singleton, 2000) et le gène hybride *RHD-CE-D* associé à l'haplotype  $(C)ce^s$  produisant un antigène RH2 faible (Faas, 1997).

Le génotypage *RHD* étant fondé sur la présence ou l'absence de séquence génique spécifique du gène *RHD*, il en résulte des faux positifs lorsque ces gènes silencieux sont présents. Une stratégie de génotypage tenant compte de ces gènes doit être établie chez les sujets d'origine africaine.

Il conviendrait, par ailleurs, d'étudier selon les mêmes critères la population restée sur les îles de l'archipel afin de détecter une évolution spécifique (dérive génétique) par rapport à la population migrante analysée.

Enfin, il nous paraît intéressant d'appréhender, selon la même méthodologie, l'étude d'autres populations présentes à Marseille (Arméniens, Vietnamiens ...). Ce travail s'intégrerait dans un projet beaucoup plus large d'institutionnalisation d'un pôle de polymorphisme génétique humain au sein de notre établissement (EFS Alpes-Méditerranée) en collaboration avec les secteurs universitaires qui encadrent l'Anthropologie à Marseille (Ecole Doctorale et Unité Mixte de Recherche), à Toulouse voir à Turin.

## **BIBLIOGRAPHIE.**

1. Abdelaal MA, Anyaegber CC, Al Sobhi EM, Al Baz NM, Hodan K (1999) Blood group phenotype distribution in Saudi Arabia. *Afr Jour Med Med Sci* , 28, 133-5.
2. Adams EJ, Cooper S, Thomson G, Parham P (2000) Common Chimpanzees have greater diversity than humans at two of the three highly polymorphic MHC I genes. *Immunogenetics*, 51, 410-24.
3. Ahmed-Chamanga, Mohamed (1999). *La langue comorienne: histoire et évolution*, N°17.
4. Algarra I, Cabrera T, Garrido F (2000) The HLA crossroad in tumor immunology. *Hum Immunol*. Jan; 61(1):65-73.
5. Allibert.C et Vérin.P. (1993) Les Comores et Madagascar. Le premier peuplement. *Archéologia*, 290, 64-77.
6. Andre P, Biassoni R, Colonna M, Cosman D, Lanier LL, Long EO (2001) New nomenclature for MHC receptors. *Nat Immunol*. 2 (8):661.
7. Arce MA, Thompson ES, Wagner S (1993) Molecular cloning of Rh cDNA derived from a gene present in RhD positive, but not RhD negative individuals. *Blood*, 82, 651-5.
8. Aubin JT, Le Van Kim C, Mouro I, Colin Y, Bignozzi C, Brossard Y and Cartron JP. (1997) Specificity and sensitivity of RHD genotyping methods by PCR-based DNA amplification. *Br J of Haematol*, 98: 356-64.
9. Badens C, Martinez di Montemuros F, Thuret I, Michel G, Mattei J, Cappelini M, Lena-Russo D (2000) Molecular basis of haemoglobinopathies and G6PD deficiency in Comorian population. *Hamatol J*, 1, 264-8.
10. Barten R, Torkar M, Haude A, Trowsdale J, Wilson MJ (2001) Divergent and convergent evolution of NK-cell receptors. *Trends Immunol*. 22(1):52-7.
11. Becker S, Tonn T, Fussel T, Uhrberg M, Bogdanow M, Seifried E, Seidl C, 2003. Assessment of killer cell immunoglobulinlike receptor expression and corresponding

HLA class I phenotypes demonstrates heterogenous KIR expression independent of anticipated HLA class I ligands. *Hum Immunol*;64(2):183-93.

12. Bernard J, Ruffié J (1966) *hématologie géographique – écologie humaine, caractères héréditaires du sang*. Edts Masson, Cies.
13. Bertoni B, Budowle B, Sans M, Chakraborty R. (2003) Admixture in Hispanics : Distribution of ancestral population contribution in the continental united States. *Hum Biol* ,75/1, 1-11.
14. Blanchy S (1999) Stratification épidémiologique du paludisme dans l'archipel des Comores. *Bull. Soc. Pathol. Exot*, 92, 177-184.
15. Blancher A, Socher WW (1997) The Rhesus system. In A Blancher A, Klein J, Socha WW eds. *Molecular biology and evolution of blood group and MHC Antigens in Primates*. Berlin : Springer, 147-218.
16. Blancher A, Reid M, Socha WW (2000) Cross – reactivity of antibodies to human and primate red cell antigens. *Transf medicine rev*, 14, 161-79.
17. Bouloux C, Gomila J, Langaney A (1972) Hemotypology of the Bedik. *Hum Biol*, 44, 289-302.
18. Brunet G, Darlu P, Zei G (2001) *Le patronyme – Histoire, anthropologie, société* CNRS Editions.
19. Carritt B, Kemp TJ, Poulter M (1997) Evolution of the human RH (Rhesus) blood group genes : a 50 year old prediction (partially) fulfilled. *Hum Molec Genet*, 6, 843-50.
20. Cavalli-Sforza L (1986) *African Pygmies*. Orlando, Fla, Academic press.
21. Cavalli-Sforza L, Menozzi P, Piazza A (1994a) *The history and geography of human genes*. Princeton university Press. Princeton, New Jersey.
22. Cavalli-Sforza L (1994b) *Qui sommes nous ?* Edition Champs Flammarion.



23. Cavalli-Sforza L (1998) L'homme et la diversité de son génome. Une phase exceptionnelle dans l'histoire de la génétique des populations. *Pathol Biol*, 46 (2), 98-102.
24. Cavalli-Sforza L, Feldman MW (2003) The application of molecular genetic approaches to the study of human evolution. *Nature Genet* – Mar;33 suppl : 266-75.
25. Cedergren B, Giles CM (1973) An Fy<sup>x</sup>Fy<sup>x</sup> individual found in Northern Sweden. *Vox Sang.* 24 264-6.
26. Cerulli E. (1957) Somalia, Scritti vari editi ed Inediti : Storia della Somalia : l'Islâm in Somalia ; Il libro degli Zengi. A Cura dell'amministrazione. Fiducaria italiana della Somalia.
27. Chaabani H, Sanchez- Mazas A, Fadhel Sallami S (2000) Genetic differentiation of Yemeni people according to Rhesus and Gm polymorphism. *Ann de Genetique*, 43, 155-62.
28. Chakraborty R (1985) Gene identity in racial hybrid and estimation of admixture rates. *Genetic Microdifferentiation in Human and Other Animal Populations*. Y.Ahuja and J.V.Neel, eds Delhi India : India Anthropological Association, Delhi University Anthropology Department, 171-180.
29. Chagnoux.H et Haribou.A. Les Comores.(1980) Collection Que sais- je. PUF.
30. Chasis JA, Mohandas N (1992) The role of red cell glycoporphins in regulating membrane function. In : PC Agre, JP Cartron, eds. *Proteins Blood groups antigens of human red cells : structure function and clinical significance*. Baltimore John Hopkins, 152-69.
31. Charron D (1997) *Genetic Diversity of HLA. Functional and Medical implications* (eds). EDK Press, Paris.
32. Chaudhuri A, Polyakova J, Zbrezezna V (1993) Cloning of glycoprotein D cDNA, which encodes the major subunit of the Duffy group system and the receptor for the *Plasmodium vivax*. *Proc Natl Acad Sci USA*, 90, 10793-7.

33. Chaudhuri A, polyakova J, Zbrzezna V, Pogo AO (1995) The coding sequence of Duffy blood group gene in humans and simians : restriction fragments length polymorphism, antibody and malarial parasite specificities, and expression in nonerythroid tissues in Duffy-negative individuals. *Blood*, 85, 615-21.
34. Chittick N (1965) The Shirazi Colonization of East Africa . *Journal of Africa*.
35. Chown B, Lewis M, Kaita H (1965) The Duffy blood group system in Caucasians : evidence for a new allele. *Am J Hum Genet*, 17 384-9.
36. Clayton J, Lonjou C, Whittle D (1997) Allele and haplotype in various ethnic groups. In : Charron D (eds) : Genetic diversity of HLA : functional and Medical implication, vol I. Paris EDK.
37. Colin Y, Cherif-Zahar B, Le Van Kim C et al (1991) Genetic basis of the RhD-positive and RhD-negative blood group polymorphism as determined by Southern analysis. *Blood*, 78, 2747-52.
38. Cornevin M (1998) Chap 15 : *La côte orientale de l'Afrique avant le XVI<sup>e</sup> siècle. Dans : secrets du continent noir révélés par l'archéologie.* Paris, Maisonneuve et Larose.
39. Corrain C, Capitanio M, Gallio P (1968) Premiers résultats d'une recherche anthropologique parmi les populations du Fezzan (Libye). 8<sup>th</sup> Int Cong Anthro Ethn Sci. Tokyo and Kyoto, 195-9.
40. Cortajarena AL, Goni FM, Ostolaza H (2001) Glycophorin as a receptor for Echerichia coli alpha haemolysin in erythrocytes. *J Biol Chem*, 276, 12513-9.
41. Crow JF. Kimura M. (1970) Generation of gametic phase disequilibrium with epistasis. In an introduction to population genetics theory. Harper International Ed Harper and Row Publisher. New York.
42. Daniels GL, Faas BHW, Green CA (1998) The Rh VS and V blood group polymorphisms in Africans : a serological and molecular analysis. *Transfusion*, 38, 951-8.

43. Daniels GL (2002) *The human blood groups*. Second edition. Blackwell, Publishing Company.
44. Dichupa PJ, Anderson C, Chown B (1969) A further search for hypothetical *Kpb* of the Kell system. *Vox Sang*, 17, 1-4.
45. Direche-Slimani.K et Le Houerou.F (2002) Les Comoriens à Marseille. D'une mémoire à l'autre. Collection Autrement.
46. Ehret C (1982) Linguistic interferences about early Bantu history, in C.Ehret & M.Posnansky, *The Archeological and Linguistic Reconstruction of Africal History* Berkeley : University of California Press, pp. 57-65.
47. Ehret C (1999) Subclassifying Bantu : the evidence of stem morpheme innovation, in Hombert JM and Hyman LM (eds), *Bantu Historical Linguistics : Theoretical and Empirical Perspective*, Stanford : CSLI Publication, 43-147.
48. El Hassan AM, Godber MJ, Kopec AC, Mourant AE, Tills D, Lehmann H (1968) The hereditary blood factor of the Beja of Sudan. *Man*, 3, 272-83.
49. Excoffier L (1987) Genetic and history of Sub-Saharan Africa. *Yearbook of Physical Anthropol*, 30, 151-94.
50. Excoffier L, Harding RM, Sokal RR, Pellegrini B, Sanchez-Mazas A (1991) HLA-DR-DQ nucleotide sequence polymorphisms in five Melanesian populations. *Tissue Antigen*, 40, 31-7.
51. Faas BHW, Beckers EAM, Wildoer P (1997) Molecular background of VS and weak C expression in blacks. *Transfusion*, 37, 38-44.
52. Facer CA (1983) Erythrocytes sialoglycoproteins and *Plasmodium falciparum* invasion. *R Soc Trop Med Hyg*, 77, 524-30.
53. Fisher RA (1946a) The fitting of gene frequencies to data on Rhesus reactions. *Ann Eug*, 13; 150-55.
54. Fisher RA, Race RR (1946b) Rh gene frequencies in Britain. *Nature*, 157, 48-9.
55. Fisher RA (1947) Note on the calculation of the frequencies of Rhesus allelomorph. *Ann Eug*, 13, 223-4.

56. Felsenstein J (1989) PHYLIP - Phylogeny Inference Package (Version3.2). *Cladistics*, 5, 164-166.
57. Frassati C (2003) *Etude du polymorphisme des gènes KIR dans la population Comorienne de Marseille*. Mémoire de DEA d'Anthropologie biologique – Option Biologie humaine. Faculté de Médecine de Marseille, Université de la Méditerranée.
58. Freeman-Grenville G.S.P (1966) *The East African Coast. Select documents from the first to the earlier nineteenth century*. Oxford, Clarendon Press.
59. Furuholm U, Nevanlinna HR, Nurkka R (1968) The blood group antigen Ula (Karulha). *Vox Sang*, 15, 118-24.
60. Gessain R, Ruffié J, Kane Y, Kane O, Cabannes R, Gomila J (1965) Note sur la séro-anthropologie de trois populations de Guinée et du Sénégal : Coniagui, Bassari et Bedik (groupes ABO, MN, Rh, P, Kell, Gm et hémoglobines) *Bulletins et Mémoires de la Société Anthropologique de Paris*, 8, 5-18.
61. Gibert M (2001) *Etude anthropologique du polymorphisme HLA-DRB1 : associations HLA-Maladies à partir de l'exemple de la polyarthrite rhumatoïde*. Thèse d'anthropologie, Faculté de Médecine de Marseille, Université de la Méditerranée.
62. Godber M, Kopec AC, Mourant AE, Teesdale P, Tills D, Weiner JS, El-Nie LH, Wood CH, Barley S (1976). The blood group, serum groups, red-cell isoenzymes and haemoglobins of the Sandawe and Nyaturu of Tanzania. *Annals of Human Biology*. Volume 3, n°5, 463-473.
63. Greenberg JH (1963) *The languages in Africa*. Bloomington : Indiana University Press.
64. Groves JD, Tanner MJA (1992) Glycophorin A facilitates the expression of human Bande 3-mediated anion transport in *Xenopus* oocytes. *J Biol Chem*, 267, 22163-70.
65. Gueguen A, Salmon D, Gley S, Rouger P, Languillat G (1980) Les phénotypes érythrocytaires au Gabon. *Rev Fran Transf*, XXIII, 6, 683-98.
66. Haldane JBS (1955) The estimation and significance of the logarithm of the ratio of frequencies. *Ann Hum Genet* 20: 309-15.

67. Hamblin MT, Di Rienzo A (2000) Detection of the signature of natural selection in humans : evidence from the Duffy blood group locus. *Am.J.Hum.Genet*, 66,1669-79.
68. Hamblin MT, Thompson E, Di Rienzo A (2002) Complex signatures of natural selection at the Duffy blood group locus. *Am.J.Hum.Genet*, 70, 369-83.
69. Hamilton HB, Nakahara Y (1971) The rare Kell blood group phenotype K0 in Japanese family. *Vox Sang*, 20, 24-8.
70. Hammond -Tooke W.D (1974) The Bantu-speaking People of Southern Africa. London and Boston. Routledge and Kegan Paul.
71. Hawkins B, Elliot M, Kosasih EN, Simons MJ (1973) Red cell genetic studies of Toba Bataks of North Sumatra. *Hum biol Oceania*, 2 (2), 147-54.
72. Hedrick W, Thomson G (1983) Evidence for balancing selection at HLA. *Genetics*, 104, 449-56.
73. Heine B, Vossen R and Hoff H (1977) Neure Ergebnisse zur Territorialgeschichte der Bantu in Moehlig WJE (eds) , *Zur Sprachgeschichte und Ethnohistorie in Africa.Neure Beitrage afrikanistischer Forshungen* Berlin : D Reimer, 57-72.
74. Henry , Samuelson B (2000) ABO polymorphism and their putative biological relationship with diseases. In : King M J, ed *Human Blood Cells : consequences of genetic polymorphism and variations*. London : Imperial College Press, 1-103.
75. Hourani G.F (1963) Arab Seafaring In the Indian Ocean In Ancient and Early Medieval times. Beirut. Khayats oriental reprints. N°3.
76. Huang CH, (1998) The human Th50 glycoprotein gene. *Jbiol Chem*, 273, 2207-13.
77. Hughes AL, 2002. Evolution of the human killer cell inhibitory receptor family. *Mol Phylogenet Evol* .25(2):330-40
78. Irshaid NM, Henry SM, Olsson ML (2000) Genomic characterization of kidd blood group gene : different molecular basis of the Jk(a-b-) phenotype in Polynesians and Finns. *Transfusion*, 40, 69-74.

79. Irshaid NM, Eisher NI, Hustinx H, Pool J, Olsson ML (2002) Novel alleles at the JK blood group locus explain the absence of the erythrocyte urea transporter in European families. *Br J Haematol*, 116, 445-54.
80. Issitt PD, Obarski G, Hartnet PL, Wren MR, Prewitt PL (1990) Temporary suppression of Kidd system antigen expression accompanied by transient production of anti-Jk3. *Transfusion*, 30, 46-50.
81. Issitt PD (1994) Race-related red cell alloantibody problems. *Br J Biomed Sci*. 51, 158-67.
82. Iwamoto S, Omi T, Kajii E, Ikemoto S (1996) Genomic organization of the glycoprotein D gene : Duffy blood group Fya/Fyb alloantigen system is associated with a polymorphism at 44 amino acid residue. *Blood*, 85, 622-6.
83. Jacob T (1974) Studies on human variation in indonesia. *J Natl Med Assoc*, 66 (5), 389-99.
84. Jenkins T, Speirs J, Dunn DS, Nurse GT (1987) serogenetic and haematological studies onthe Kgalagadi of Bostwana. *Ann Hum Biol*, 14, (2), 143-53.
85. Jobling MA, Tyler-Smith C (2003) The human Y chromosome : an evolutionary marker comes of age. *Nature Genetics*, 4, 598-11.
86. Jokinen M, Ehnholm C, Vaisanen-Rhen V (1985) Identification of the major human sialoglycoprotein from red cells, glycophorin A<sup>M</sup>, as the receptor for *Escherichia coli* IH 11165 and characterization of recptor site. *Eur J Biochem*, 147, 47-52.
87. Kathan RH, Winzler RJ, Jonhson CA (1961) Preparation of an inhibitor of viral hemagglutination from human erythrocytes. *J Exp Med*, 113, 37-45.
88. Kitano T, Sumiyama K, Shiroishi T, Saitou N (1998) Conserved evolution of *the Rh50 gene compared to its homologous Rh blood group gene*. *Biochem Biophy Res Commun*, 249, 78-85.
89. Kitano T, Saitou N (2000) Evolutionary history of *Rh* blood group-related genes in vertebrates. *Immunogenetics*, 51, 856-62.

90. Kivisild T, Rootsi S, Metspalu A, Mastana S, Kaldma K, Parik J, Adojaan M, Tolk HV, Stepanov VA, Gölge M, Usanga EA, Papiha SS, Cinnioglu C, King R, Cavalli-Sforza L, Underhill PA, Villens R (2003) The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am J Hum Genet*, 72, 313-32.
91. Klein J, Sato Y, Nagl S, O'Huigin C (1998) Molecular trans-species polymorphism. *Annu Rev Ecol Syst*, 29, 1-21.
92. Kominato Y, Chester MA (1992) Heterogeneity of the blood group Ax allele : genetic recombination of common alleles can result in Ax phenotype. *Transfus Med*, 8, 154-64.
93. Kominato Y, Tsuchiya T, Hata N, Takizawa H, Yamamoto F (1997) Transcription of human ABO histo-blood group genes is dependent upon binding of transcription factor CBF/NF-Y to minisatellite sequence. *J Biol Chem*, 41, 25890-8.
94. Koscielak J (1986) A hypothesis on the biological role of ABH, Lewis and P blood group determinant structures in glycosphingolipids and glycoproteins. *Glycocon J*, 3, 95-108.
95. Laffon M (1974) Langues et contact de langues dans l'archipel des Comores. *Mémoire de fin d'année 3<sup>o</sup> cycle*. Université Aix Marseille.
96. Laffon M (1991) Travaux sur la syntaxe et le lexique franco-comorien. Collection L'Harmattan.
97. Languillat G, Cartron J, Gerbal R, Salmon C (1980) Les phénotypes érythrocytaires au Gabon. *Rev Fran Transf*, 6, 675-82.
98. Lawlor Da, Ward FE, Ennis PD, Jackson AP, Parham P (1988) HLA A and B polymorphisms predate the divergence of humans and chimpanzees. *Nature*, 335, 268-71.
99. Le Bourdieu F, Battistini R (1969) Atlas de Madagascar. Editeurs. Tananarive, Bureau pour le Développement de la Production Agricole et du Centre de l'institut Géographique National de Madagascar.

100. Lee S, Zambas ED, Marsh WL, Redman CM (1991) Molecular cloning and primary structure of Kell blood group protein. *Proc Natl Acad Sci USA*, 88, 6353-7.
101. Lee S, Russo DCW, Reiner AP (2001) Molecular defects underlying the Kell null phenotype. *J Biol Chem*, 276, 2728-9.
102. Lependu J, Clamagirant-Muelt C, Cartron JP, Gerard G, Vitrac D, Oriol R (1983) H-Deficient blood group of Reunion Island III. Alpha-2-L-fucosyl Transferase activity in sera. *Am J hum genet*, 35, (3), 497-507.
103. Le Pennec PY, Rouger P, Klein MT (1989) A serological study of red cells and sera from 18 Rh:32, -46 ( $R^N/R^N$ ) persons. *Transfusion*, 29, 798-802.
104. Le Van Kim C, Mouro I, Brossard Y, Chavinie J, Cartron JP (1994) PCR-based determination of Rhc and RhE status of fetuses at risk of Rhc and RhE haemolytic disease. *Br J of Haematol*, 88, 193-95.
105. Lewis M, Kaita H, Chown B (1972) The Duffy blood group system in Caucasians : a further population sample. *Vox Sang*, 23, 523-7.
106. Li J, Iwamoto S, Sugimoto N, Okuda H, Kajii E (1997) Dinucleotide repeat in the 3' flanking region provides a clue to the molecular evolution of the Duffy gene. *Hum Genet*, 99, 573-7.
107. Liu Z, Chen Y, Mo R (2000) Characterization of human RhCG and mouse Rhcg non erythroid Rh glycoprotein homologues predominantly expressed in kidney and testis. *J Biol Chem*, 275, 25641-55.
108. Lowe RF (1979) The distribution of blood group and HLA antigens of Zimbabwe African.
109. Lucien N, Sidoux Walter F, Olives B (1998) Characterization of the gene encoding the human Kidd blood group/urea transporter protein : evidence for splice site mutations in Jknull individuals. *J Biol Chem*, 273, 12973-80.
110. Lucien N, **Chiaroni J**, Cartron JP, Bailly P (2002) Partial deletion of the JK locus causing a Jknull phenotype. *Blood*, 99, 1079-81.



111. Mackay J, Wang JC, Wong KS (1969) The incidence of blood group in 4648 southern chinese from Toishan district and its vicinity. *J Hong Kong Med Tech Asso*, 1, 11-14.
112. Mallinson G, Soo KS, Schall TJ, Pisacka M, Anstee DJ (1995) Mutations in the erythrocyte chemokine receptor (Duffy) gene : the molecular basis of the Fy<sup>a</sup>/Fy<sup>b</sup>, antigens and identification of a deletion in the Duffy gene of an apparently healthy individual with the Fy(a-b-) phenotype. *Br J Haematol*, 90, 823-9.
113. Manni F, Leonardi P, Brakat A, Rouba H, Heyer E, Klintschar M, McElreaver K, Quintana-Muirci (2002) Y-chromosome analysis in Egypt suggests a genetic regional continuity in Northeastern Asia. *Human Biol*, 74, 645-58.
114. Mantran R (1969) L'expansion musulmane VII<sup>o</sup>-XI<sup>o</sup> siècles.Paris, Presses.
115. Martin AM, Freitas EM, Witt CS, Christians en FT, 2000. The genomic organization and evolution of the natural killer immunoglobulin-like receptor (KIR) gene cluster. *Immunogenetics*. 51(4-5):268-80
116. Martinko JM, Vincek V, Klein D, Klein J (1993) Primate ABO glycosyl transferases : evidence for trans-species evolution. *Immunogenetics*, 37, 274-8.
117. Matassi G, Cherif-Zahar B, Raynal V, Rouger P, Cartron JP (1998) Organisation of the human *RH50A* gene (RHAG) and evolution of base composition of the RH gene family. *Genomics*, 47, 286-93.
118. Matassi G, Cherif-Zahar B, Pesole G, Raynal V, Cartron JP (1999) The members of the *RH* genes family (*RH50* and *RH30*) followed different evolutionary pathways. *J mol Evol*, 48, 151-9.
119. Mattznetter T, Spielmann W (1969) Blutgruppen Moçambiquanisher Bantustame. *Zeitschrift fur Morphol. Anthropol*, 61, 57-61.
120. Merveilles de l'Inde (Les) (1957) Traduction de Sauvaget in Memorial. Jean Sauvaget tome 1 : 187-309. Damas, Institut Français de Damas.
121. Miller LH, Mason SJ, Dvorak JA, mc Ginniss MH, Rothman IK (1975) Erythrocyte receptor for malaria : Duffy blood group determinants. *Science*, 189, 561-3.

122. Miller LH, Haynes JD, McAuliffe FM (1977) Evidence for differences in erythrocyte surface receptors for malarial parasites *Plasmodium falciparum* and *Plasmodium knowlesi*. *J Exp Med*, 146, 277-81.
123. Mourant AE, Kopec AC, Domaniewska-Sobczak K (1976) Distribution of human blood group and other polymorphism . 2<sup>nd</sup> edn. London : Oxford University Press.
124. Mouro I, Le Van Kim C, Cherif-Zahar B (1994) Molecular characterization of the Rh-like locus and gene transcripts from the Rhesus monkey (*Macaca mulatta*). *J Mol Evol*, 38, 169-76.
125. Nepom GT, Erlich H. MHC class-II molecules and autoimmunity. (1991) *Annu Rev Immunol*, 9, 493-525.
126. Nichols ME, Rubinstein P, Barnwell J (1988) A new human Duffy blood group specificity defined by a murine monoclonal antibody. Immunogenetics and association with susceptibility to *Plasmodium vivax*. *J Exp Med*, 166, 776-85.
127. Niranjana Y, Chandak GR, Veerajug P, Singh L (1999) Some atypical and rare sickle cell gene haplotypes in populations of Andhra Pradesh, India. *Hum Biol*, 71, 333-40.
128. Noizat-Pirenne F, Mouro I, Le Pennec PY, Ansart-Pirenne H, Juszczyk G, Patereau C, Verdier M, Babinet J, Roussel M, Rouger P, Cartron JP (2001) Two new alleles of *RHCE* gene in black individuals : the *Rhce* allele *ceMO* and the *RhcE* allele *cEMI*. *Br J Haematol*, 113, 672-79.
129. Noizat-Pirenne F, Lee K, Le Pennec PY, Simon P, Kazup P, Bachir D, Rouzaud AM, Roussel M, Juszczyk G, Ménanteau C, Rouger P, Kotb R, Cartron JP, Ansart-Pirenne H (2002) Rare RHCE phenotypes in black individuals of afro-caribbean origin : identification and transfusion safety. *Blood*, 01, 2-29.
130. Noizat – Pirenne F (2003) Particularités immuno-hématologiques des populations Africaines et Antillaises. Implications transfusionnelles. *Transf Clin Biol*, 10, 185-91.
131. Norman PJ, Stephens HA, Verity DH, Chandanayingyong D, Vaughan RW, 2001. Distribution of natural killer cell immunoglobulin-like receptor sequences in three ethnic groups. *Immunogenetics*.52 (3-4):195-205.

132. Norman PJ, Carrington CV, Byng M, Maxwell LD, Curran MD, Stephens HA, Chandanayingyong D, Verity DH, Hameed K, Randath DD, Vaughan RW (2002) Natural Killer cell immunoglobulin-like receptor (KIR) locus profiles in Africa and South Asian populations. *Genes Immun*, 3(2), 86-95.
133. Nurse GT, Botha MC, Jenkins T (1977a). Sero-genetic studies on the San of south West Africa. *Hum Hered*, 27 (2), 81-98.
134. Nurse GT, Jenkins T (1977b). Serogenetic studies on the Kavango peoples of South Africa. *Ann Hum Biol*, 4 (5), 465-78.
135. Okubo Y, Yamaguchi H, Nagao N (1986) Heterogeneity of the phenotype Jk(a-b-) found in Japanese. *Transfusion*, 26, 237-9.
136. Okuda H, Kawano M, Iwamoto S (1997) The *RHD* gene is highly detectable in RhD negative Japanese donors. *J Clin Investigation*, 100, 373-9.
137. Olérup O, Zetterquist H (1992) HLA-DR typing by PCR amplification with sequence-specific primers (PCR-SSP) in 2 hours : an alternative to serological DR typing in clinical practice including donor-recipient matching in cadaveric transplantation. *Tissue Antigens*, 39, 225.
138. Olives B, Martiel S, Mattie MG (1996) Molecular characterization of a new kidney urea transporter in the human kidney. *FEBS Lett*, 386, 156-60.
139. Olives B, Merriman M, Bally P (1997) The molecular basis of the Kidd blood group polymorphism and its lack of association with type 1 diabetes susceptibility. *Hum Mol Genet*, 6, 1017-20.
140. Olson ML, Hansson C, Avent ND, Akesson IE, Green CA, Daniels GL (1998a) A clinically applicable method for determining the three major alleles at the Duffy (FY) blood locus using polymerase chain reaction with allele-specific primers. *Transfusion*, 38,168-73.
141. Olson ML, Smythe J, Hansson C, Poole J, Mallinson G, Jones J, Avent N, Daniels G (1998b) The Fy<sup>x</sup> phenotype is associated with a missense mutation in the Fy<sup>b</sup> allele predicting Arg89Cys in the Duffy glycoprotein. *Br J of haematol*, 103, 1184-91.

142. Ottino.P (1974) Le Moyen Age dans l'Océan Indien et le peuplement de Madagascar. Annuaire des pays de l'Océan Indien. Aix en Provence. Presse Universitaire d'Aix Marseille. Volume I.
143. Page, R. D. M. (1996). TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences*, **12**, 357-58.
144. Peiper SC, Wang Z, Neote K et al (1995) The Duffy antigens/receptor for chemokines (DARC) is expressed in endothelial cells of Duffy-negative individuals who lack erythrocyte receptor. *J Exp Med*, 181, 1311-17.
145. Pereira L, Macauley V, Torroni A, Scozzari R, Prata MJ, Amorim A (2001) Prehistoric and historic traces in mtDNA of Mozambique : insights into Bantu expansions and slave trade. *Ann Hum Genet*, 65, 439-58.
146. Pereira L, Cusmao L, Alves C, Amorim A and Prata MJ (2002) Bantu and European lineages in sud-Saharan Africa. *Ann Hum Genet*, 66, 369-78.
147. Poloni ES, Semino O, Passarino G, Santachiara-Benerecetti AS, Dupanloup I, Langaney A, Excoffier L (1997) Human genetic affinities for Y chromosome P49a,f/TaqI haplotypes show strong correspondence with linguistics. *Am J Hum Genet*, 61,1015-35.
148. Prins A.H.J (1961) The Swahili-Speaking peoples of Zanzibar and the East African Coast. (Arabes, Shirazi and Swahili). London, International African Institute.
149. Potts WK, Wakeland EK (1993) Evolution of MHC genetic diversity : a tale of incest, pestilence and sexual preference. *Trends Genet*, 9, 408-12.
150. Quamar R, Ayub Q, Khaliq S, Mansoor A, Karafet T, Mehdi SQ, Hammer MF (1999) African and Levantine origins of Pakistani YAP+ Y chromosomes. *Hum Biol*, 71, 745-55.
151. Race RR, Sanger R.(1975) Blood group in man. 6<sup>th</sup> edn Oxford : Blackwell Scientific Publications.
152. Radley E, Alderton RP, Kelly AP, Trowsdale J, Beck S. (1994) Genomic organization of HLA-DMA and HLA-DMB. Comparison of the gene organization of all

- six class II families in the human major histocompatibility complex. *J Biol Chem*, 269(29), 18834-8
153. Rajalingam R, Krausa P, Shilling HG, Stein JB, Balamurugan A, McGinnis MD, Cheng NW, Mehra NK, Parham P, (2002) Distinctive KIR and HLA diversity in a panel of north Indian Hindus. *Immunogenetics* .53(12):1009-19
  154. Rao PM, Kirk RL, Busi BR, Ramana GV, Reddi GG (1999) Genetic studies on Gadaba : a tribal population of Andhra Pradesh, India. *Anthropo.Anz* , 1, 41-49.
  155. Ramilison E (1951) Ny Loharanon'ny Andriana Nanjaka teto Imeria, 3-4 note page 27,29 Antananarivo.
  156. Redman CM, Lee S, Ten Bokkel Huinink D (1989) Comparison of human and Chimpanzee Kell blood group system. *Transfusion*, 29, 486- 90.
  157. Renfrew C (1987) L'énigme Indo-Européenne, Archéologie et langage. Champs-Flamarion.
  158. Reviron D (1999) *Polymorphisme du système HLA de classe II et prédisposition génétique dans la polyarthrite rhumatoïde et la pseudo polyarthrite rhizomélique*. Thèse d'immunologie, Aix Marseille II.
  159. Reynolds J, Weir BS, Cockerham CC (1983) Estimation of the coancestry coefficient: basis for a short term genetic distance. *Genetics*, 105, 767-79.
  160. Ruhlen M (1991) *A guide to the languages of the World* Stanford Univ. Press, Stanford, CA.
  161. Richards M, Rengo C, Cruciani F, Gratix F, Wilson JF, Scozzari R, Macaulay V, Torroni A (2003) Extensive female-mediated gene flow from Sub-Saharan Africa into Near East Arab populations. *Am J Hum Genet*, 72, 1058-64.
  162. Robinau C (1966) Islam aux Comores ; Une étude d'histoire culturelle de l'île d'Anjouan. *Revue de Madagascar*, 35 (3), 17-34.
  163. Robinson WP, Asmussen MA, Thomson G (1991) Three locus systems impose additional constraints on pairwise disequilibria. *Genetics*, 129, 925-30.

164. Rouger P, Salmon D, Ruffié J, Salmon C (1983) Studies of blood group antigens in Aka pygmies. *J Hum Evol*, 12, (4), 361-66.
165. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* , 4, 406-25
166. Saitou N, Yamamoto F (1997) Evolution of primate ABO blood group genes and their homologous genes. *Mol Biol Evol*, 14, 399-411.
167. Salamon H, Klitz W, Eastel S, Gao X, Erlich HA, Fernandez-Vina M, Trachtenberg EA, Mc Weeney SK, Nelson MP, Thomson G (1999) Evolution of HLA Class II molecules : allelic and amino acid site variability across populations. *Genetics*, 152, 393-400.
168. Salmon C (1995) Ne cherchez pas votre sosie. Les groupes sanguins , chroniques des années de recherche. Edts Frison Roche.
169. Salvignol I, Calvas P, Socha WW (1995) Structural analysis of the RH-like blood group gene products in non human primates. *Immunogenet*, 41, 271-81.
170. Sanchez-Mazas A (2001) African diversity from HLA point of view : influence of genetic drift, geography, linguistics and natural selection. *Hum Immunol*, 62 , 937-48.
171. Sandvig K and Van Deurs B (1994) Endocytosis without clathrin. *Trends cell Biol.*,4, 275-7.
172. Schneider, S.,Roessli, D., and Excoffier, L. (2000) Arlequin: A software for population genetics data analysis. Ver 2.000. Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva.
173. Scozzari R, Cruciani F, Santolamazza P, Malaspina P, Torroni A, Sellitto D, Arredi B, Destro-Bisol G, De Stefano G, Richards O, Martinez-Labarga C, Modiano D, Biondi G, Moral P, Olckers A, Wallace DC, Novelletto A (1999) Combined use of biallelic and microsatellite Y-chromosome polymorphisms to infer affinities among African populations. *Am J Hum Genet*, 65, 829-46.

174. Shilling HG, Guethlein LA, Cheng NW, Gardiner CM, Rodriguez R, Tyan D, Parham P (2002) Allelic polymorphism synergizes with variable gene content to individualize human KIR genotype. *J Immunol.* Mar 1; 168(5):2307-15.
175. Shimizu Y, Ao H, Soemantri A, Tiwawech D, Settheetham-Ishida W, Kayame OW, Kimura M, Nishioka T, Ishida T (2000) Sero- and molecular typing of Duffy blood group in Southeast Asians and Oceanians. *Hum Biol*, 72 (3), 511-8.
176. Sidoux Walter F, Lucien N, Nissinen R (2000) Molecular heterogeneity of the Jknull phenotype : expression analysis of the Jk(S291P) mutation found in Finns. *Blood*, 96, 1566-73.
177. Sim BKL, Chitnis CE, Wasniowska K, Hadley TJ, Miller LH (1994) Receptors and ligand domains for invasion of erythrocytes by *Plasmodium falciparum*. *Science*, 264, 1941-4.
178. Simonovic B, Muzziarelli A (1970) Pers. Comm.
179. Singleton BK, Green CA, Avent ND (2000) The presence of an *RHD* pseudogene containing a 37 base pair duplication and a non sense mutation in most Africans with the Rh-D negative blood group phenotype. *Blood*, 95, 12-18.
180. Sistonen J, Koistinen J, Aden Abdull O (1987) Distribution of blood groups in east African Somali population. *Hum Her*, 37 (5), 300-13.
181. Spielmann W, Teixidor D, Renninger W, Matznetter Th (1970). Blutgruppen and Lepra bei moçambiquanischen Völkerschaften. *Humangenetik*. Vol 10 (4): 304-17.
182. Stroup M, Mac Ilroy M, Walker R, Aydelotte JV (1965) Evidence that Sutter belongs to the Kell blood group system. *Transfusion*, 5, 309-14.
183. Takahata N, Satta Y, Klein J (1990) Polymorphism and balancing selection at major histocompatibility complex loci. *Genetics*, 130,025-938.
184. Templeton TJ, Keister DB, Muratova O, Procter JL, Kslow DC (1998) Adherence of erythrocytes during exflagellation of *Plasmodium falciparum* microgametes is dependent on erythrocyte surface sialic acid and glycophorin. *J Exp Med*, 187, 1599-609.

185. Tills D , Kopec AL, Walow A, Barnicot NA, Mourant AE, Marin A, Benett FJ, Woodburn JC (1982) Blood group, protein and red cell enzyme polymorphism of Hadza of Tanzania. *Hum Genet*, 61 (1), 52-9.
186. Tomita A, Radike EL, Parker CJ (1993) Isolation of erythrocyte membrane inhibitor of reactive lysis type II : identification as glycophorin A. *J Immunol*, 151, 3308-23.
187. Toneva M, Lepage V, Lafay G, Dulphy N, Busson M, Lester S, Vu-Trien A, Michaylova A, Naumova E, McCluskey J, Charron D (2001) Genomic diversity of natural killer cell receptor genes in three populations. *Tissue antigens*, 57(4), 358-62.
188. Trowsdale J, Barden R, Hande A, Stewart CA, Beck S, Wilson MJ (2001) The genomic context of natural killer receptor extended gene families. *Immunol Rev*, 181, 20-38.
189. Tournamille C, Colin Y, Cartron JP, Le Van Kim C (1995a) Disruption of a GATA motif in the *Duffy* gene promoter abolishes erythroid gene expression in Duffy-negative individuals. *Nature Genet*, 10 , 224-8.
190. Tournamille C, Le Van Kim C, Gane P, Cartron J-P, Colin Y (1995b) Molecular basis and PCR-DNA typing of Fy<sup>a</sup>/Fy<sup>b</sup> blood group polymorphism. *Hum Genet*, 95, 407-10.
191. Tournamille C, Le Van Kim C, Gane P, Le Pennec PY, Roubinet F, Babinet J, Cartron JP, Colin Y (1998) Arg89Cys substitution results in very low membrane expression of the Duffy antigen/receptor for chemokines in Fy(x) individuals. *Blood* , 92, 2147-56.
192. Trimingham J.Sp (1964) Islam in East Africa. Oxford, Clarendon Press.
193. Uhrberg M, Valiante NM, Shum BP, Shilling HG, Lienert-Weidenbach K, Corliss B, Tyran D, Lanier LL, Parham P (1997) Human diversity in killer cell inhibitory receptor genes. *Immunity* .7(6):753-63
194. Valdes AM, Mc Weeney SK, Meyer D, Nelson MP, Thomson G (1999) Locus and population specific evolution in HLA Class II genes. *Ann Hum Genet*, 63, 27-43.



195. Vansina J (1990) *Paths in the Rainforests : Towards a history of political traditions in equatorial Africa*. Madison, University of Wisconsin Press ; London, James Currey.
196. Vérin P (1971) Histoire ancienne du Nord-Ouest de Madagascar. Taloha. *Revue du Musée d'Art et d'Archéologie*. Numéro spécial. Tananarive.
197. Vérin.P (1994) Les Comores. Edition Karthala.
198. Viitala J, Jarnefelt J (1985) The red cell surface revised. *Trends Biol Sci*, 14, 392-5.
199. Vilches C, Rajalingam R, Uhrberg M, Gardiner CM, Young NT, Parham P (2000) KIR2DL5, a novel killer-cell receptor with a D0-D2 configuration of Ig-like domains *J Immunol.* 1; 164(11):5797-804
200. Wagtmann N, Biassoni R, Cantoni C, Verdiani S, Malnati MS, Vitale M, Bottino C, Moretta L, Moretta A, Long EO, 1995. Molecular clones of the p58 NK cell receptor reveal immunoglobulin-related molecules with diversity in both the extra- and intracellular domains. *Immunity*.2 (5):439-49
201. Welch SG, Mc Gregor IA, Williams K (1977) The Duffy blood group and malaria prevalence in Gambian West Africans. *Trans R Soc trop Med Hyg*, 71, 295-6.
202. Wells S (2003) *The journey of man : a genetic odyssey* – Edition Broché
203. Wende H, Colonna M, Ziegler A, Volz A (1999) Organization of the leukocyte receptor cluster (LRC) on human chromosome 19q13.4. *Mamm Genome* ; 10(2):154-60
204. Wiener AS, Unger LJ, Cohen L (1954) Distribution and heredity of blood factor-U. *Science*, 119, 734-5.
205. Wilson MJ, Torkar M, Haude A, Milne S, Jones T, Sheer D, Beck S, Trowsdale J (2000). Plasticity in the organization and sequences of human KIR/ILT gene families. *Proc Natl Acad Sci U S A*.97 (9):4778-83
206. Winddhof O, Walter H (1983). Blood group, serum protein and red cell enzyme polymorphism in Filipinos. *Hum Hered*, 33 (6), 357-64.
207. Witt CS, Dewing C, Sayer DC, Uhrberg M, Parham P, Christiansen FT (1999) Population frequencies and putative haplotypes of the killer cell immunoglobulin-like

- receptor sequences and evidence for recombination. *Transplantation*. Dec 15; 68(11):1784-9.
208. Woodfield G (1970) Pers. Comm.
209. Xie S-S, Hunag C-H, Reid ME, Blancher A, Blumenfeld OO (1997) The glycophorin A gene family in gorillas: structure, expression and comparison with human and chimpanzee homologues. *Biochem Genet*, 35, 59-76.
210. Y Chromosome Consortium (2002) A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res*.12,339-348.
211. Yamaguchi H, Okubo Y, Seno T, Matsushita K, Daniels GL (1979) A "new" allele, *Kpc*, at the *Kell* complex locus. *Vox Sang*, 36, 29-30.
212. Yamamoto F, Clausen H, White T, Marken J, Hakomori S (1990a) Molecular genetic basis of the histo-blood group ABO system. *Nature*, 345, 229-33.
213. Yamamoto F, Hakomori S (1990b) Sugar-nucleotide donor specificity of histo-blood group A and B transferases based on amino acid substitution. *J Biol Chem* , 265, 19257-62.
214. Yamamoto M, Lin XH, Kominato Y (2001) Murine equivalent of the human histo-blood group ABO gene is a cisAB gene and encodes a glycosyltransferase with both A and B transferase activity. *J Biol Chem*, 276, 13701-8.
215. Yasuda N, Kimura M (1968) A gene counting method of maximum likelihood for estimating gene frequencies in ABO and ABO-like systems. *Ann Hum Genet*, 31, 409.
216. Yawata M, Yawata N, McQueen KL, Cheng LW, Guethlein LA, Rajalingam R, Shilling HG, Parham P (2002) Predominance of group A KIR haplotypes in Japanese associated with diverse NK cell repertoires of KIR expression. *Immunogenetics*, 54(8), 543-50.
217. Zimmerman PA, Wooley I, Masinde GL (1999) Emergence of *FYAnull* in *Plasmodium vivax* endemic region of Papua New Guinea. *Proc Natl Acad Sci USA*, 96, 13973-7.

# **ANNEXES**

## Annexe 1

### **Etablissement Français du Sang “Alpes-Méditerranée”** 149 Bd Baille 13005 Marseille

#### **NOTICE D’INFORMATION ET DE CONSENTEMENT POUR UNE RECHERCHE SANS BENEFICE INDIVIDUEL DIRECT**

*Document établi en deux exemplaires*

*Un exemplaire cosigné est remis en main propre à chaque personne au cours de l’entretien médical préalable*

Nom / Nom de jeune fille	
Prénom	
Adresse	
Téléphone	

Le Docteur Jacques Chiaroni ou son représentant m’a proposé de participer à une recherche organisée par l’Etablissement Français du Sang Alpes Méditerranée sur la connaissance des groupes sanguins de la population d’origine comorienne de Marseille. Il m’a été précisé, au cours de cet entretien, que ce protocole avait fait l’objet d’une autorisation préalable du Comité Consultatif de Protection des Personnes dans la Recherche Biomédicale (CCPPRB) de Marseille (Loi Huriet) et que j’étais libre d’accepter ou de refuser de participer à cette étude. J’ai reçu et bien compris les informations suivantes :

- La population d’origine comorienne comme d’autres populations présente des particularités au niveau des groupes sanguins.
- Une des conséquences de cette particularité est la difficulté de trouver du sang compatible en cas de transfusion. Ceci est directement lié aux différences “sanguines” existant entre la population des donateurs de sang de Marseille et la population d’origine comorienne qui traditionnellement participe peu au don de sang.
- Cette étude, par l’analyse des groupes sanguins de la population d’origine comorienne, vise à atteindre trois objectifs principaux

Recenser les individus ayant les groupes sanguins les plus rares afin de les inciter à participer au don de sang bénévole. Cela permettra de disposer de réserves de sang congelé compatible pour eux mêmes et pour les individus possédant le même groupe sanguin.

Approfondir la connaissance scientifique de certains groupes sanguins du globule rouge comme les antigènes des systèmes Rh et Duffy pouvant présenter des variations spécifiques à cette population.

Définir la fréquence de plusieurs marqueurs sanguins (antigènes des globules rouges, antigènes du système HLA, protéines plasmatiques) dans la population d’origine comorienne afin de la comparer aux autres populations géographiquement voisines (continent africain, océan indien, continent indien).

J’accepte donc de répondre au questionnaire et de donner 3 tubes de 7 ml de sang afin de participer à cette recherche dans les conditions précisées ci dessus.

Mon consentement ne décharge pas les organisateurs de la recherche de leurs responsabilités. Je conserve tous mes droits garantis par la loi. Les données qui me concernent resteront strictement confidentielles. Je n’autorise leur consultation que par des personnes qui collaborent à la recherche, désignées par l’organisateur, **le Docteur Jacques Chiaroni** et éventuellement, un représentant des Autorités de Santé. Je pourrai, à tout moment, demander toute information complémentaire au **Docteur Jacques Chiaroni – Téléphone : 04.91.18.95.79 (83)**.

*Fait à Marseille, le.....*

*Signature de l’investigateur ou  
du médecin qui le représente*

*Signature du volontaire précédée de la mention “lu et  
approuvé”*

## Annexe 2

--

### **Individu**

Nom	
Prénom	
Sexe	
Date de Naissance	
Lieu de Naissance	
Langue Maternelle	

### **Conjoint**

Nom	
Prénom	
Sexe	
Lieu de Naissance	
Langue maternelle	
Lien de parenté	

### **Parents**

#### **Père**

Langue paternelle	
Langue maternelle	

#### **Mère**

Lieu de Naissance	
Langue maternelle	

### **Commentaire**

--

**Grands parents**

--

**Grand Père Paternel**

Lieu de Naissance	
Langue maternelle	

**Grand Mère Paternelle**

Lieu de Naissance	
Langue maternelle	

**Grand Père Maternel**

Lieu de Naissance	
Langue maternelle	

**Grand Mère Maternelle**

Lieu de Naissance	
Langue maternelle	

**Commentaire**

--

**Annexe 3.**

COMITE CONSULTATIF DE PROTECTION DES PERSONNES  
DANS LA RECHERCHE BIOMEDICALE - MARSEILLE 1

Marseille, le 21 Mai 2001

**Le Président**  
**Professeur J-C MANELLI**

Réf. du présent avis ou délibération: 10/04/01  
Projet de recherche enregistré *sous* le n°: 00/43

Le Comité a été saisi le 28 Novembre 2000

par Monsieur le Docteur J. CHIARONI

d'une demande d'avis concernant le projet de recherche sans bénéfice individuel direct, intitulé:

"ETUDE DU POLYMORPHISME GENETIQUE  
DE LA POPULATION COMORIENNE DE MARSEILLE",

dont le promoteur est:

L'ETABLISSEMENT FRANCAIS DU SANG ALPES MEDITERRANEE.

Le Comité a examiné les informations relatives à ce projet:

- Lors de ses séances du 12 Décembre 2000, 9 janvier, 13 Février, 13 Mars et 10 Avril 2001, au cours desquelles des renseignements complémentaires ont été demandés (protocole plus étayé, attestation d'assurance et protocole en concordance, CV des investigateurs),
- Le 18 Mai 2001, après réception des conclusions du rapporteur, portant sur les compléments reçus le 03 Mai 2001,
- Le 21 Mai 2001, après analyse des derniers compléments reçus ce jour.

Ont participé à la délibération:

M. J-C. MANELLI (I), M. B. AMBLARD (I), M. P. BENHAIM (II), Mme G. MACQUART-MOULIN (III),  
Mme A. FORESTIER-OLIVERO (III), Mme C. POTHIER (IV), M. P. AMBROSI (V), Mme M-I.  
MALAUZAT (VI), M. P. De ALCALA (VII), M R. BOU (VIII).

Le Comité a adopté la délibération suivante:

AVIS FAVORABLE



Professeur J-C. MANELLI

**HOPITAL D'ADULTES DE LA TIMONE - 2ème Etage - 13385 MARSEILLE - Cédex 5**  
**Tél. 04.91.47.71.00 / Poste interne: 84.844 - Fax: 04.91.47.71.94**

*Toutes les techniques de biologie moléculaire utilisées dans cette étude ont été mises au point et validées au laboratoire de biologie moléculaire de l'EFS Alpes-Méditerranée.*

*Nous exposerons brièvement le principe de l'amplification génique par PCR (Polymerase Chain Reaction), principale méthode utilisée, avant de présenter une fiche technique spécifique pour chacun des systèmes géniques étudiés.*



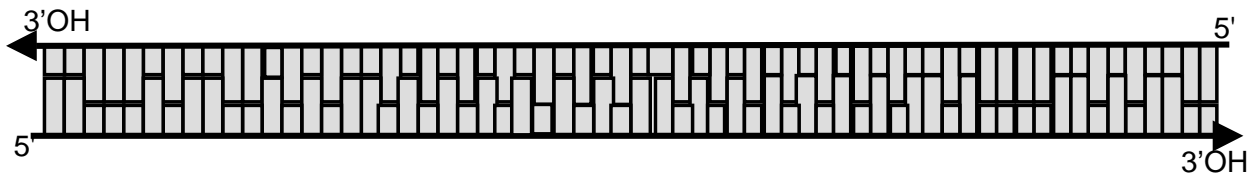
# LES ETAPES DE LA PCR

Les trois étapes de la PCR :

1. La dénaturation de l'ADN
2. L'appariement des deux amorces
3. L'élongation des amorces
4. Visualisation de l'amplification

## 1) La dénaturation de l'ADN :

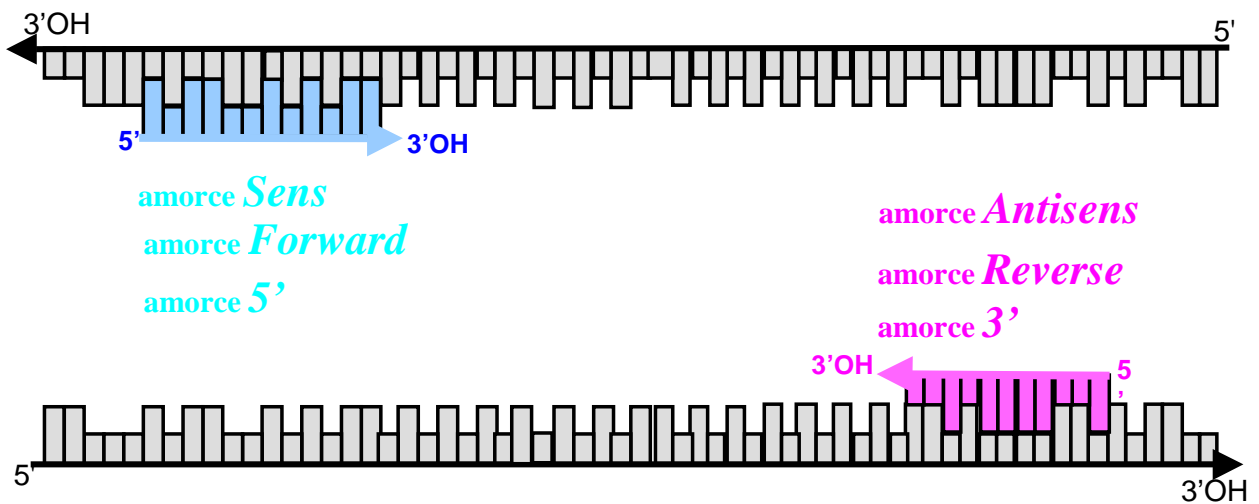
Elle se traduit par la rupture des liaisons hydrogènes à une température de 94-98°C



## 2) L'appariement des deux amorces

Une amorce est un oligonucléotide complémentaire et anti-parallèle à une séquence d'ADN à laquelle il pourra s'apparier.

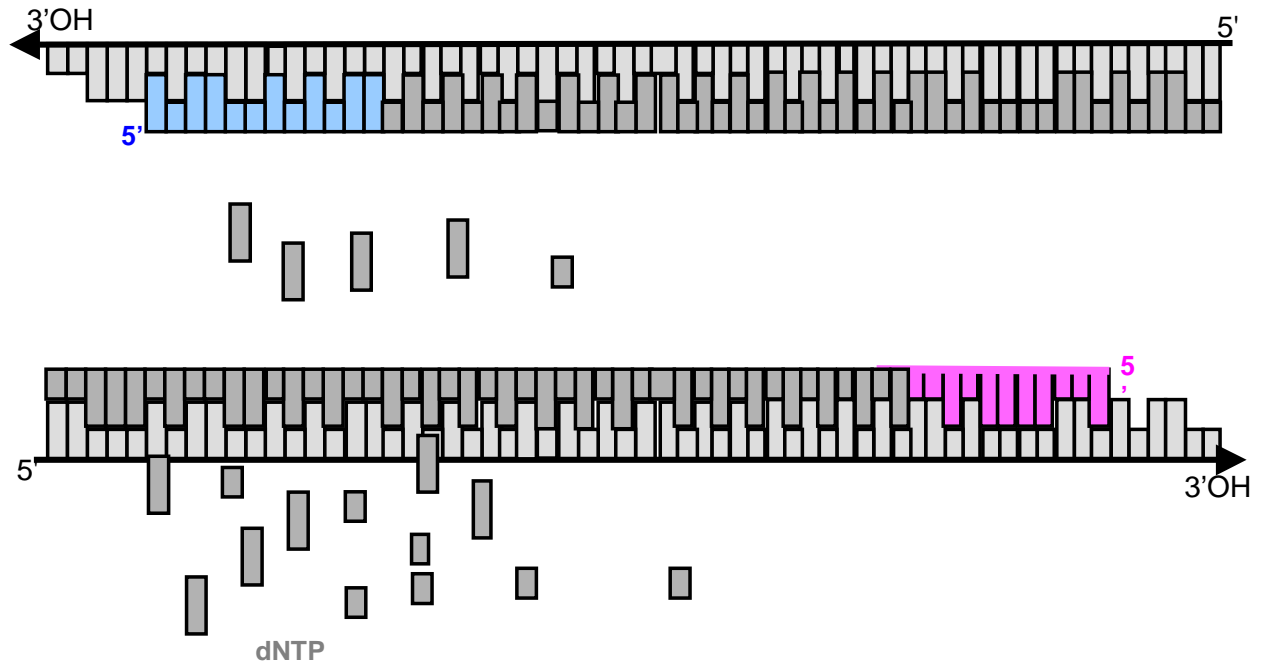
Cette étape est réalisée à la température de fusion des amorces : la  $T_m$ .

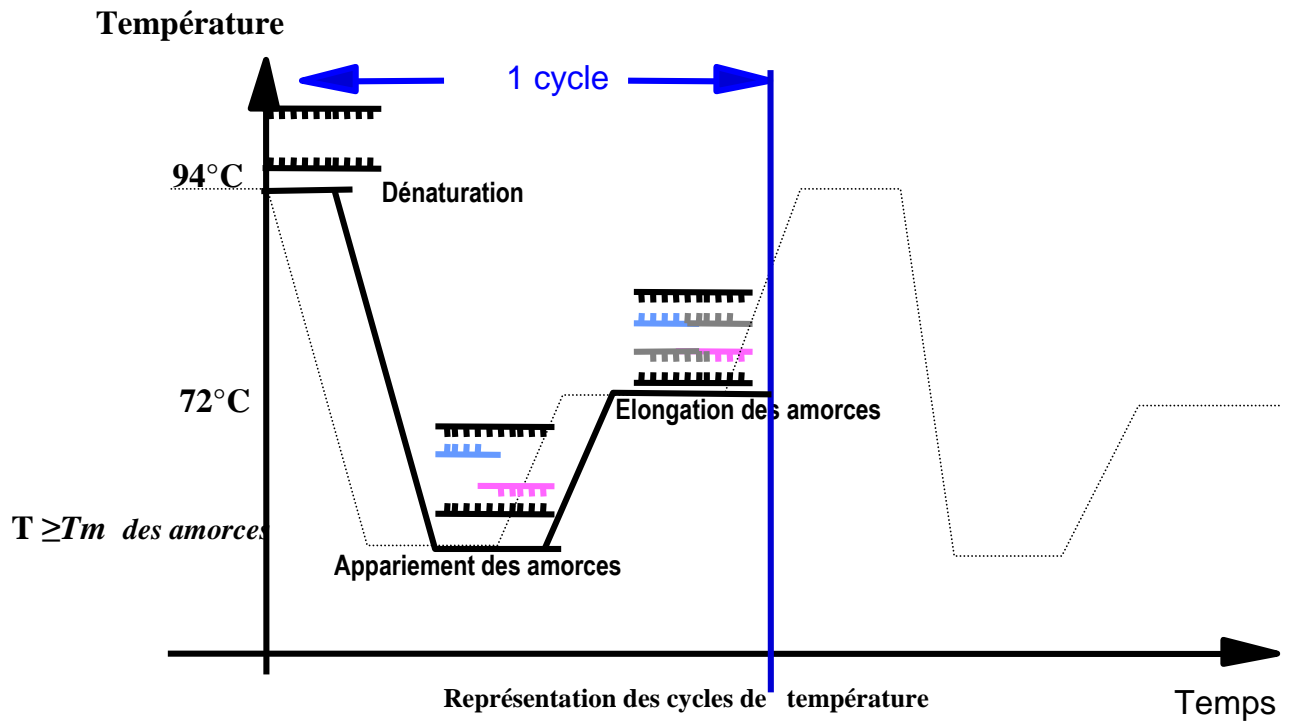


### 3) L'élongation des amorces

Cette étape est réalisée à la température optimale de 72°C.

Elle dépend de la durée de l'élongation et de la vitesse d'incorporation des nucléotides par la Taq polymérase. La **Taq Polymérase** est une enzyme thermo-résistante isolée de la bactérie *Thermus aquaticus*. Elle recopie une séquence d'ADN simple brin *dans le sens 5'-3'* à partir d'une amorce hybridée à celle-ci.

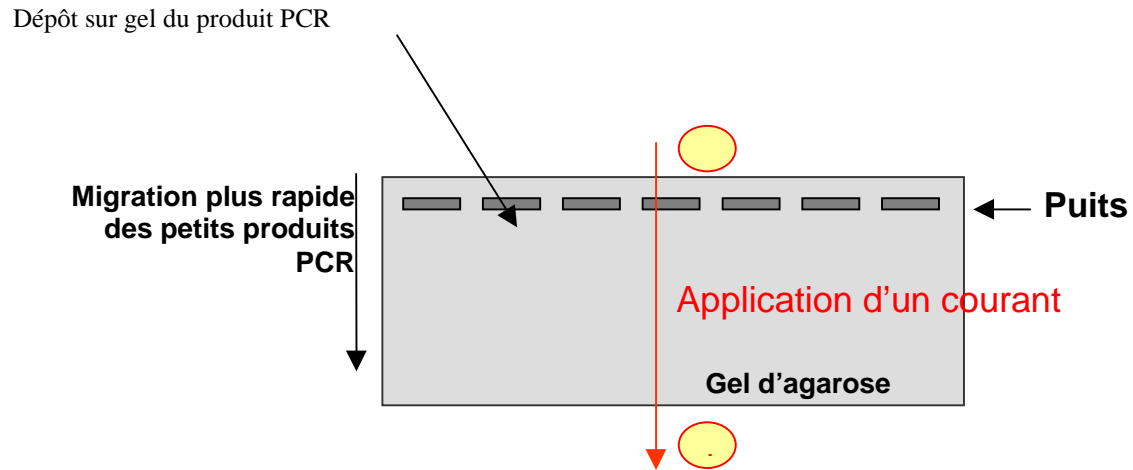




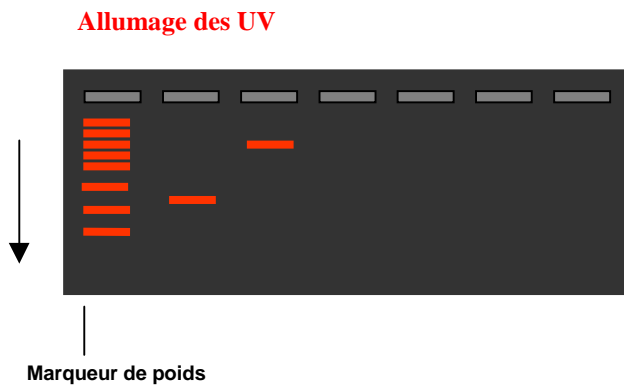
Il y a multiplication exponentielle des produits d'amplification délimités par les 2 amorces : si on considère ;  
 - n molécules d 'ADN génomique  
 - et m cycles

A la fin de la réaction d'amplification, on aura  $n \times 2^m$  molécules de PCR

4) Visualisation de l'amplification par migration électrophorétique sur gel horizontal d'agarose contenant du BET :

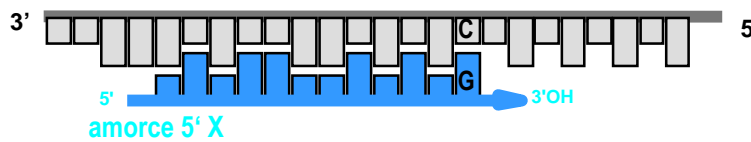
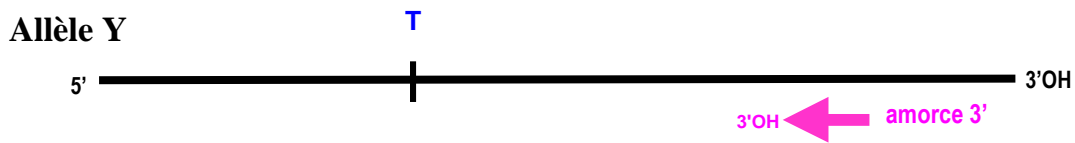


Photographie du gel sous UV :

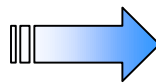


*Les fragments d'ADN apparaissent par fluorescence sous les UV par l'action du BET.*

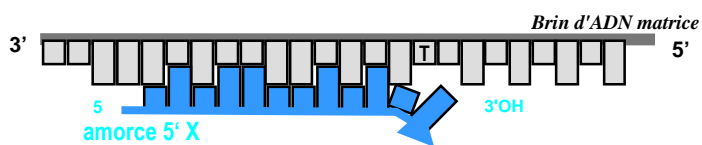
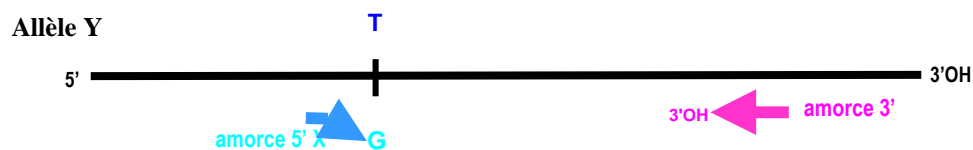
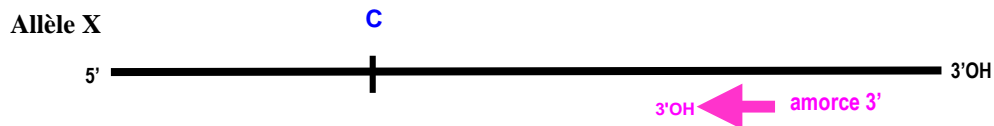
## LES ETAPES DE LA PCR-SSP (et -ASP)



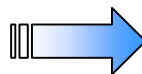
La **Taq Polymérase** peut catalyser la 1<sup>ère</sup> liaison phospho-diester entre l'extrémité 3'OH de l'amorce et le dCTP



*produit PCR attendu pour l'allèle X*

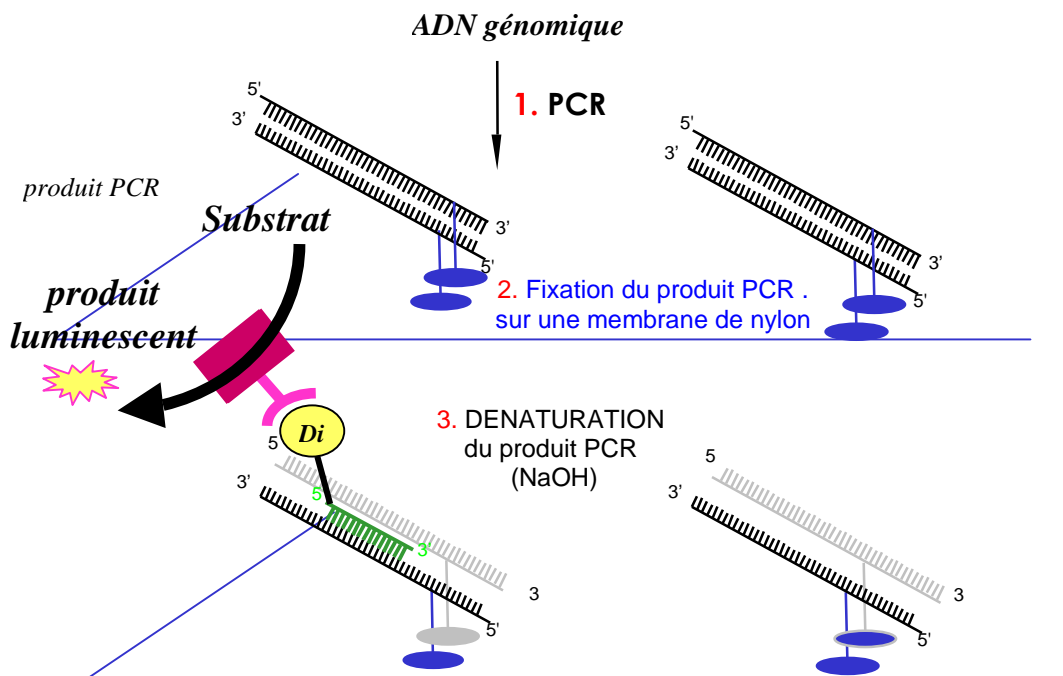


La **Taq Polymérase** ne peut pas catalyser la 1<sup>ère</sup> liaison phospho-diester entre l'extrémité 3'OH de l'amorce et le dCTP



*pas de produit PCR attendu pour l'allèle Y*

## LES ETAPES DE LA PCR-SSO



4. HYBRIDATION à  $T_m$  de la membrane avec une sonde spécifique marquée à la digoxigénine

5. LAVAGES STRINGENTS

6. REVELATION COLORIMETRIQUE

D'après A. Dormoy, 2003, Communication personnelle

**Annexe 4**

**Génotypage DUFFY par PCR-ASP**

## Détection du polymorphisme FY\*A/FY\*B par PCR-ASP

**1 - Extraction de l'ADN génomique à partir de sang total (Kit commercial Qiagen)**

**2 - Amplification génique par PCR-ASP (Polymerase Chain Reaction using Allele Specific Primers)**

Amorces:

FYAB2	: 5' CTCATTAGTCCTTGGCTCTTAT 3'	Gène DARC
FYA_REV	: 5' AGCTGCTTCCAGGTTGGCAC 3'	Allèle FY*A
FYB_REV	: 5' AGCTGCTTCCAGGTTGGCAT 3'	Allèle FY*B

Contrôle interne:

Ces amorces reconnaissent et amplifient une séquence du gène ABO

FYABO	: 5' GCTGGAGGTGCGCGCCTAC 3'
FYABO_REV	: 5' ACGAATTCTACTTGTTCAGGTGGCTCTCGTC 3'

**3 - Migration sur gel d'agarose**

**4- Interprétation des résultats**

Une bande contrôle		419 pb
Avec: FYAB2 + FYA_REV:	Homozygote ou Hétérozygote FY*A	711pb
Avec: FYAB2 + FYB_REV:	Homozygote ou Hétérozygote FY*B ou FY*X	711 pb

N.B:

Le couple d'amorces utilisé ne peut pas donner d'amplification, pour l'allèle FY\*Fy, puisque l'amorce FYAB2 est spécifique des allèles contenant un site GATA-1 fonctionnel.



## Détection du polymorphisme FY\*B/FY\*X par PCR-ASP

**1 - Extraction de l'ADN génomique à partir de sang total (Kit commercial Qiagen)**

**2 - Amplification génique par PCR-ASP (Polymerase Chain Reaction using Allele Specific Primers)**

Amorces:

FYAB2	:	5' CTCATTAGTCCTTGGCTCTTAT 3'	Gène DARC
ASPB_REV	:	5' CAGGGCAGAGCTGCCAGCG 3'	Allèle FY*B
ASPX_REV	:	5' CAGGGCAGAGCTGCCAGCA 3'	Allèle FY*X

Contrôle interne:

Ces amorces reconnaissent et amplifient une séquence du gène ABO

FYABO	:	5' GCTGGAGGTGCGCGCCTAC 3'
FYABO_REV	:	5' ACGAATTCTACTTGTTTCAGGTGGCTCTCGTC 3'

**3 - Migration sur gel d'agarose**

**4 - Interprétation des résultats**

Une bande contrôle	419 pb
Avec: FYAB2 + ASPX_REV: Homozygote ou Hét. FY*A	850 pb
Avec: FYAB2 + ASPB_REV: Homozygote ou Hét. FY*B ou FY*X	850 pb

N.B:

Le couple d'oligonucléotides utilisé ne peut pas donner d'amplification, pour l'allèle FY\*Fy, puisque l'amorce FYAB2 est spécifique des allèles contenant un site GATA-1 fonctionnel.

**Annexe 5**

**Génotypage RH par PCR-ASP**

## Les bases du génotypage *RHD*

Compte tenu des bases moléculaires, la détection de la présence ou de l'absence de *RHD* peut être basée sur :

- La recherche de la séquence de la région 3' de l'exon 10 spécifique du gène *RHD* où l'obtention d'un amplifiat signe la présence du gène alors que l'absence du produit atteste de son absence.
  - L'amplification d'une région de l'intron 4 qui donne un produit de 600 pb pour le gène *RHD* et 1200 pb pour le gène *RHCE*. L'absence du produit de 600 pb signant l'absence du gène *RHD*.
  - Enfin, l'amplification d'une région de l'exon 7 commune aux gènes *RHD* et *RHCE* sert de témoin.
- Chacune de ces méthodes détectant une seule région du gène *RHD*, elles peuvent donner des résultats faussement négatifs et parfois faussement positifs en cas de variant :
- Les hématies DVI et DVa, sont des D+, alors que manquants de l'intron 4 de *RHD*, elles donneront un résultat négatif.
  - Le gène hybride, *RHCE-D-CE*, responsable de DBT manque de l'exon 7, aboutira à une réaction négative avec les techniques utilisant cet exon.
  - Le gène hybride *DCE-D-CE* associé au complexe d(C)ce<sup>s</sup> ne produit pas de D sur l'hématie mais possède l'exon 10 dérivé de *RHD*.
  - Le pseudogène *RHDψ* des populations Africaines possède un gène *RHD* mais il est inactif.

Ainsi, la technique décrite par Singleton (Singleton, 2000) permet de détecter lors d'une PCR multiplex l'exon 7 et l'intron 4 du gène *RHD*, ainsi que le pseudogène *RHDψ*. Toutes ces méthodes ne permettent pas de distinguer un individu homo- ou hétérozygote pour le gène *RHD*.

## Les bases du génotypage *RHCE*

### Bases moléculaires du polymorphisme C/c.

D'un point de vue moléculaire, le polymorphisme C/c est habituellement associé à la substitution de 6 nucléotides entre les deux allèles aboutissant à 4 substitutions d'acides aminés (table). Seule la substitution en 103 (T307C) est extra membranaire (2<sup>e</sup> boucle). La détection de l'allèle *c* repose sur une PCR spécifique de C307 dans l'exon 2 (143,315). La détection de l'allèle *C* ne peut prendre en compte ce polymorphisme de l'exon 2 qui est identique à celui du gène *RHD*. De plus, compte tenu de l'absence de corrélation stricte entre l'antigène C et le polymorphisme C48 de l'exon 1 (Cys16 absent alors que C présent dans l'haplotype *d(C)e<sup>s</sup>* [Exon 1, 2 et fin 5' de l'exon 3 dérivent de *RHD*] et Cys16 présent chez 74% des individus C-c+ d'origine africaine), la détection de cet allèle repose sur une PCR spécifique d'une séquence de 109 pb dans l'intron 2 (314).

### Bases moléculaires du polymorphisme E/e.

Le polymorphisme E/e est associé à la substitution C676G dans l'exon 5 aboutissant à Pro226Ala dans la quatrième boucle extra cellulaire de la protéine RhCcEe. La détection de l'allèle *E* repose sur une PCR spécifique de l'allèle au niveau de l'exon 5. La détection de l'allèle *e* ne peut prendre en compte l'unique substitution nucléotidique G676 qui est aussi présente dans le gène *RHD*. Sa détection impose, non seulement la prise en compte de G676 mais aussi d'une partie spécifique de l'exon 5 du gène *RHCE* comme A787 par exemple.

Polypeptides	Exon 1		Exon 2		Exon 5
	aa16	aa60	aa68	aa103	Aa226
ce	Trp	Leu	Asn	Pro	Ala
cE	Trp	Leu	Asn	Pro	Pro
Ce	Cys	Ile	Ser	Ser	Ala
CE	Cys	Ile	Ser	Ser	Pro
D	Trp	Ile	Ser	Ser	Ala
<b>Génotypage c (RH4)</b> • PCR allèle spécifique <i>c</i> T307C Exon 2 <b>Génotypage C (RH2)</b> • Primer spécifique Exon 2 du gène <i>RHCE</i> • PCR allèle spécifique <i>C</i> T307C Exon 2 <b>Nb</b> d(C)ce <sup>s</sup> exon 2 absent alors que C présent.			<b>Génotypage E (RH3)</b> • PCR allèle spécifique <i>E</i> C676G Exon 5 <b>Génotypage e (RH5)</b> • Primer spécifique Exon 5 du gène <i>RHCE</i> • PCR allèle spécifique <i>e</i> C676G Exon 5		

## Génotypage RH par PCR-ASP

### 1- Extraction de l'ADN génomique à partir de sang total (Kit commercial Qiagen)

### 2 - Amplification génique par PCR-ASP (Polymerase Chain Reaction using Allele Specific Primers)

#### 2.1 Génotypage classique et détection des allèles RHD, RHc, RHC, RHE et RHe

Voir les séquences des amorces dans le tableau plus bas.

- Amplification de l'allèle RHD

La détection du gène RHD, présent exclusivement chez les sujets de phénotype RHD positif (RH1), est réalisée à l'aide d'un couple d'amorces **D4/D5** dans la région 3' non traduite de l'exon 10. Un témoin interne de PCR est utilisé; il s'agit de l'amplification d'une séquence commune à au gène RHD et au gène RHCE de l'exon 4 à partir du couple d'amorces **P4/P5**.

- Amplification des séquences RHc par PCR

Elle est basée sur l'existence d'une substitution dans l'exon 2 du gène RHCE à la position 307 (RHC=T/RHc=C) qui permet de discriminer les allèles RHC et RHc. La technique utilisée est une PCR multiplex (PCR RHc spécifique à partir d'un couple d'amorces **P2/P3** avec pour témoin d'amplification interne, l'amplification de l'exon 4 à partir du couple d'amorces **P4/P5**)

- Amplification des séquences RHE par PCR

Elle est basée sur l'existence d'une substitution dans l'exon 5 du gène RHCE à la position 676 (RHe=G/RHE=C) qui permet de discriminer les allèles RHe et RHE. La technique utilisée est une PCR multiplex (PCR RHE spécifique à partir d'un couple d'amorces **E1/E2** avec pour témoin d'amplification interne, l'amplification de l'exon 4 à partir du couple d'amorces **P4/P5**)

Allèles	Amorces	Séquences Nucléotidiques	Taille (pb)
RHD	D4	5'-GGATTTTAAGCAAAAGCATCCAAGAA-3'	291
	D5	5'-ACTGGATGACCACCATCATAT-3'	
RHc	P2	5'-TCGGCCAAGATCTGACCG-3'	177
	P3	5'-TGATGACCACCTTCCCAGG-3'	
RHC	P2	5'-TCGGCCAAGATCTGACCG-3'	177
	P4	5'-TGATGACCACCTTCCCAGA-3'	
RHE	E1	5'-TGGCCACGTGTCAACTCTC-3'	149
	E2	5'-ctcacCATGCTGATCTTCT-3'	
RHe	e1	5'-TGGCCACGTGTCAACTCTG-3	149
	E2	5'-ctcacCATGCTGATCTTCT-3'	
Contrôle interne	P4	5'-CGCAGCCTATTTTGGGCTG-3'	110
	P5	5'-CCAGCATGGCAGACAAACT-3'	

## 2.2 Détection des allèles RHCE "rares"

Plusieurs étapes correspondant à différentes amplifications sont indispensables pour la détection de ces allèles RHCE rares.

- L'antigène RH5 partiel peut être produit par deux allèles.

### Allèle **ce<sup>s</sup>** (340) (MAYET)

Cet allèle est caractérisé par deux mutations. Une, dans l'exon 5 (733C>G) et une, dans l'exon 3 (340C>T). Il est détecté par deux PCR-ASP utilisant respectivement les couples d'amorces **P18/P19** et **P26/P27** chez les sujets RH5 (e). La définition du génotype est basée sur la détection des allèles sauvages correspondants détectés avec les couples d'amorces P18/P19 et P26/P27

### Allèle **ceMO**

Cet allèle est caractérisé par une mutation dans l'exon 5 (667G>T). Cet allèle est détecté par une PCR-ASP avec le couple d'amorces **P24/P25** chez les sujets RH5 (e). La définition du génotype est basée sur la détection de l'allèle sauvage correspondant détecté avec le couple d'amorce **P24/P25**. Les sujets RH5 partiels présentent un phénotype RH5. Ils doivent être homozygotes pour cet allèle c'est-à-dire porter la mutation en double dose.

- Un phénotype RH :-34 peut être produit par l'allèle (C)ce<sup>s</sup>.

Cet allèle est caractérisé par la présence de deux mutations dans le gène RHCE [Une dans l'exon 5 (733C>G) et une dans l'exon 7 (1006G>T)] et par un gène RHD hybride [D-CE(3-8)-D] qui produit un antigène C partiel. Cet allèle est détecté par trois PCR ASP utilisant les couples d'amorces **P18/P19** pour la mutation dans l'exon 5, **P20/P21** pour la mutation dans l'exon 7 et **P22/P23** pour le gène hybride.

La définition du génotype est basée sur la détection des allèles sauvages correspondants détectés avec les couples d'amorces **P18/P19**, **P20/P21** et **P22/P23**. Les sujets RH :-34 présentent un phénotype ddCcee. Ils doivent être homozygotes pour cet allèle c'est à dire porter les deux mutations du gène RHCE et le gène RHD hybride en double dose

Les couples d'amorces spécifiques des différentes mutations recherchées leurs séquences amorces, leurs spécificités ainsi que la taille des produits de PCR résultant des amplifications sont récapitulés dans le tableau plus bas.

**3 - Détection des produits d'amplification par migration sur gel d'agarose**

**4 - Purification des produits de PCR à séquencer**

**5 - Séquençage**

*Les produits sont séquencés en utilisant un kit de séquençage d'ADN en D-Rhodamine et d'un séquenceur ABI Prism 377 (Perkin-Elmer), ou sont envoyés directement à Genom-Express à (Grenoble, France).*

**Séquences des amorces et conditions PCR-ASP utilisées pour la détection des allèles *RHCE* "rares"**

<u>Amorces</u>	<u>Spécificité</u>	<u>Localisation</u>	<u>Séquences 5'- 3'</u>	<u>Produit (bp)</u>
P18-s	CE	In 4	gcaacagagcaagagtcca	
P19/P19*-a	D/CE*	Ex 5	TGTCACCACACTGACTGCTAC(G*)	428
P20/P20*-s	1006T/G*	Ex 7	ACTCCATCTTCAGCTTGCTGT(G*)	
P21-a	CE	In 7-Ex 7	accacATGCCATTGCCGTTT	94
P22-s	D	Ex 3	TCGGTGCTGATCTCAGTGGA	
P23/P23*-a	CE/D*	Ex 3	ACTGATGACCATCCTCAGGG(T*)	110
P24/P24*-s	D/CE*	Ex 5	TGGATGTTCTGGCCAAGTT(G*)	
P25-a	CE	In 5-Ex 5	tcacCATGCTGATCTTCCT	150
P26/P26*-s	340 T/C*	In 2-Ex 3	ccttctacccccagTATTT(C*)	
P27-a	CE	Ex 3	CTGATGACCATCCTCAGGG	152

**Annexe 6**

**Génotypage Complet HLA DRB par PCR-SSO**



## Génotypage Complet HLA DRB par PCR-SSO

### 1 - Extraction de l'ADN par *Salting-out* (Miller, 1988) ou utilisation des noyaux cellulaires

### 2 – Amplification génique par PCR-SSO (Polymerase Chain Reaction using Sequences Specifics Oligoprobes )

- Amplification générique par PCR du locus DRB1 à l'aide des amorces consensus (Voir tableau ci-dessous)
  - Vérification de l'amplification génique par migration électrophorétique
  - Dépôt de l'ADN amplifié et dénaturé sur une membrane de nylon
- Hybridation de l'ADN amplifié à l'aide des amorces spécifiques marquées en 3' à la digoxigénine (Voir tableau ci-dessous)

Nomenclature		
XIIth Workshop	Identification	SEQUENCE 5'-3'
HLA SSO	Labo HLA	
Amorces consensus HLA-DRB1		
DRBGH46	P1DRB1	CCGGATCCTTCGTGTCCCCACAGCACG
DRBTIERCY	P2DRB1	TCG CCG CTG CAC TGT GAA G
Amorces spécifiques des allèles HLA-DRB1		
	2801 DR1	CGG TTG CTG GAA AGA TGC
1002W	DR2	AG CCT AAG AGG GAG TGT C
A85 Tiercy	DR16	CTA CGG GGC TGT GGA GAG C
	7004 DR3	GGC CGG GTG GAC AAC TAC
	1004 DR 4	GAG CAG GTT AAA CAT GAG
	5703 DR11	G CCT GAT GAG GAG TAC TG
RB1 Y16	DR12	GGT GAG TGT TAT TTC TTC
	7007 DR13	AC ATC CTG GAA GAC GAG C
	5704 DR14	G CCT GCT GCG GAG CAC TG
	1006 DR7	TGG CAG GGT AAG TAT AAG
FG 7400	DR8	AGG CGG GCC CTG GTG GAC
	1007 DR9	G AAG CAG GAT AAG TTT GA
	1008 DR10	GAG GAG GTT AAG TTT GAG
K 71-1	DR1303	TG GAA GAC AAG CGG GCC G
	3703 DR3+6	AC CAG GAG GAG AAC GTG C
	7012 DR14.2	AC CGC GGC CCG CCT CTG C *
	1003 DR n°17	G TAC TCT ACG TCT GAG TG
RB1A74	DR n°18	GAA GAC AGG CGG GCC GCG G
	2807 DR n°19	G CGG TAC CTG GAC AGA TA
RB1F47A	DRN15	GAC GTG GGG GAG TTC CGG
	7003 DR1602	GAC CTC CTG GAA GAC AGG

### 3 - Révélation par photoluminescence

### 4 - Interprétation des résultats

**Annexe 7**

**Génotypage Complet du système KIR par PCR-SSP**

# Génotypage Complet du système KIR par PCR-SSP

1- Extraction de l'ADN génomique à partir de sang total par *Salting-out* (Miller, 1988)

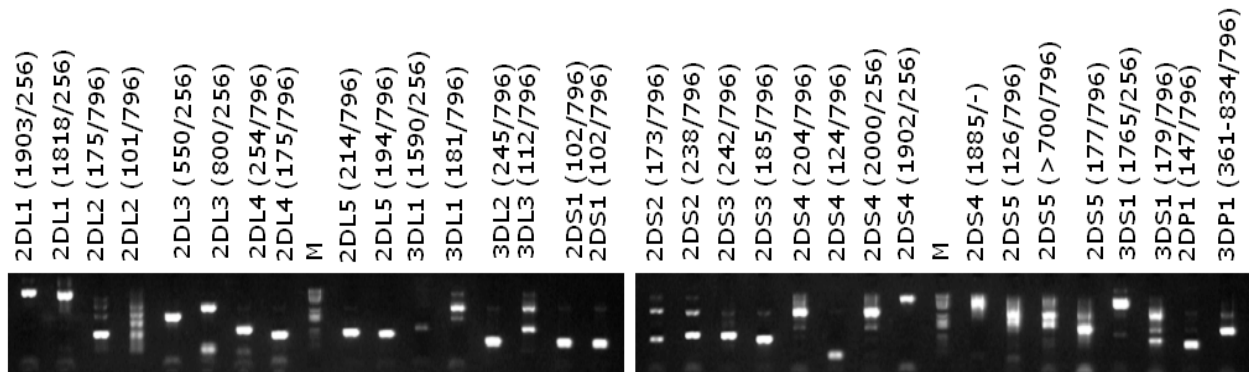
2- Amplification génique par PCR-SSP (Polymerase Chain Reaction using Sequences Specifics Primers)

Voir liste d'amorces ci-dessous

3- Migration sur gel d'agarose

4- Résultats:

↳ Gel d'électrophorèse :



**Exemple d'un gel représentant le génotypage KIR complet d'un individu réalisé à l'aide du kit de 32 couples d'amorces**

Le premier numéro à la suite du nom du gène correspond à la taille du produit spécifique amplifié, le second à celle du contrôle d'amplification.

**M** : marqueur de poids moléculaire qui permet de vérifier la taille des bandes.

↳ **Interprétation des résultats :**

La majorité des gènes KIR est testé par deux couples d'amorces. Pour le rendu des résultats, la positivité de l'un des deux couples ou des deux donne un résultat positif, la double négativité et la présence de la bande du contrôle d'amplification donne un résultat négatif.

*Liste des amorces retenues pour le géotypage KIR*

	Couples d'amorces	Amorce sens (5'-3')	Amorce anti-sens (3'-5')
1	M2DL1b	tgg acc aag agt ctg cag ga	act cag cat ttg gaa gtt ccg
2	G2DL1a	cca tca gtc gca tga cg	cca ctc gta tgg aga gtc at
3	G2DL1b	cca tca gtc gca tga cg	aat gtt ccg ttg acc ttg gt
4	M2DL2a	gtt ggt cag atg tca gtt tga a	ggt cac tgg gag ctg aca c
5	N2DL2b	cag cac ttc ctt ctg cac ac	gcc ctg cag aga acc tac a
6	M2DL3a	ctt cat cgc tgg tgc tg	agg ctc ttg gtc cat tac aa
7	M2DL3b	tcc ttc atc gct ggt gct g	ggc agg aga caa ctt tgg atc a
8	M2DL4a	cag gac aag ccc ttc tgc	ctg ggt gcc gac cac t
9	N2DL4NOR	gcc agt ggg gaa gcc cat gaa ct	acc aac ctg tga cag aaa cag
10	C2DL5a	gcg ctg tgg tgc ctc g	gac cac tca atg ggg gag c
11	C2DL5b	tgc agc tcc agg agc tca	ggg tct gac cac tca tag ggt
12	G3DL1a	cca tcg gtc cca tga tgc t	aga gag aag gtt tct cat atg
13	M3DL1b	ggt gaa atc agg aga gag	gta ggt ccc tgc aag ggc aa
14	M3DL2a	caa acc ctt cct gtc tgc cc	gtg ccg acc acc cag tga
15	M3DL3a	gtc aga tgt cag gtt tga gcg	cat gga ata gtt gac ctg gga ac
16	C 2DS1a	ctt ctc cat cag tcg cat gaa	aga ggg tca ctg gga gct gac
17	C 2DS1b	ctt ctc cat cag tcg cat gag	aga ggg tca ctg gga gct gac
18	M2DS2a	ttc tgc aca gag agg gga agt a	agg tca ctg gga gct gac aa
19	M2DS2b	cgg gcc cca cgg ttt	ggt cac tcg agt ttg accact ca
20	M2DS3a	tgg ccc acc cag gtc g	tga aaa ctg ata ggg gga gtg agg
21	M2DS3b	cta tga cat gta cca tct atc cac	aag cag tgg gtc act tga c
22	M2DS4a	ctg gcc ctc cca ggt ca	tct gta ggt tcc tgc aag gac ag
23	M2DS4b	tag gct ccc tgc agt gcg	gag ttt gac cac tcg tag gga gc
24	D2DS4VE	atc ctg caa tgt tgg tcg	ctg gat aga tgg tac atg tc
25	D2DS4AN (KIR1D)	atc ctg caa tgt tgg tcg	ctg gat aga tgg agc tgc ag
26	M2DS5a	tga tgg ggt ctc caa ggg	tcc aga ggg tca ctg ggc
27	N2DS5a	aga gag ggg acg ttt aac c	gga aag agc cga agc atc
28	N2DS5b	aga gag ggg acg ttt aac c	tcc aga ggg tca ctg gga
29	G3DS1a	ggc aga ata ttc cag gag g	agg ggt cct tag aga tcc a
30	M3DS1b	ggt gaa atc agg aga gag	gtc cct gca agg gca c
31	M2DP1a	tgt tgg tca gat gtc agg ttt c	ggt ccc tgc cag gtc ttc
32	N3DP1a	cca tgt cgc tca tgg tcg	tga cca ccc agt gag ga