HAL
open science

Variabilité génétique et flux de gènes chez les moustiques anophèles arabiensis et anophèles gambiae, vecteurs du paludisme en Afrique. Apport des loci microsatellites

Frédéric Simard

To cite this version:

Frédéric Simard. Variabilité génétique et flux de gènes chez les moustiques anophèles arabiensis et anophèles gambiae, vecteurs du paludisme en Afrique. Apport des loci microsatellites. Génétique. Université Henri Poincaré - Nancy 1, 1999. Français. NNT : 1999NAN10147 . tel-01747577

HAL Id: tel-01747577

https://hal.univ-lorraine.fr/tel-01747577v1

Submitted on 29 Mar 2018

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 LORRAINE

AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4
Code de la Propriété Intellectuelle. articles L 335.2- L 335.10
http://www.cfcopies.com/V2/leg/leg_droi.php
http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm

Université Henri Poincaré – NANCY I
Faculté des Sciences
U.F.R. Sciences et Techniques Biologiques
Ecole Doctorale Biologie et Santé

UNIVERSITE NANCY 1 BIBLIOTHEQUE SCIENCES

Thèse

présentée pour l'obtention du titre de

Docteur de l'Université Henri Poincaré, Nancy-I

en Génétique

par **Frédéric SIMARD**

Variabilité génétique et flux de gènes chez les moustiques
Anopheles arabiensis et *Anopheles gambiae*, vecteurs du paludisme en Afrique.
Apport des loci microsatellites.

Soutenue publiquement le 22 Juillet 1999

Membres du jury:

Président :	B. DECARIS	Professeur, Université Henri Poincaré, Nancy I
Rapporteurs:	M. RAYMOND	Directeur de Recherche, Université de Montpellier II
	F. RODHAIN	Professeur, Institut Pasteur, Paris
Examinateurs:	C. DOURNON	Professeur, Université Henri Poincaré, Nancy I
	V. ROBERT	Directeur de Recherche, IRD, Dakar, Sénégal
	D. FONTENILLE	Chargé de Recherche, IRD, Dakar, Sénégal

Structure d'accueil: **UPRES EA 2401 «Génétique et Interactions cellulaires en Reproduction», Laboratoire de Biologie Expérimentale et Immunologie, Faculté des Sciences, 54506 Vandœuvre les Nancy.**

Laboratoire d'accueil: **Laboratoire IRD de Zoologie Médicale à l'Institut Pasteur, BP 220, Dakar, Sénégal.**

Remerciements

Les travaux présentés dans ce mémoire sont le résultat d'un long parcours qui m'aura enrichi à plusieurs titres. Je tiens ici à remercier tous ceux qui, de près ou de loin ont contribué à la réalisation de ce projet et m'ont soutenu tout au long de ces trois années de recherche.

L'intégralité de ces travaux a été réalisée au laboratoire IRD de Zoologie Médicale à l'Institut Pasteur de Dakar, Sénégal. Pour m'avoir permis de résider à Dakar, d'abord en tant que Coopérant du Service National, puis comme thésard, je souhaite remercier Messieurs ***B. Philippon*** *et* ***C. Bellec*** *du Département Santé de l'IRD en France, Messieurs* ***P. Mathieu*** *et* ***J.-R. Durand****, représentants successifs de l'ORSTOM, maintenant devenu IRD, au Sénégal ainsi que Messieurs* ***J.-P. Digoutte*** *et* ***J.-P. Moreau****, ancien et actuel directeurs de l'Institut Pasteur de Dakar.*

Cette thèse doit son existence à ma rencontre avec ***Didier Fontenille*** *lorsque j'étais en stage de DEA à l'Institut Pasteur de Paris. Didier, je ne saurais assez te remercier pour avoir tout mis en œuvre pour concrétiser ma venue à Dakar. Ton dynamisme à toute épreuve, ton insatiable soif de connaissances et ta motivation communicative m'ont constamment soutenu et aidé dans mes choix. Tu resteras un modèle d'organisation et de rigueur scientifique pour moi et pour les nombreux étudiants qui ont eu et qui auront la chance de bénéficier de tes conseils et de travailler sous ta direction. Merci de m'avoir fait profiter de ton expérience et pour avoir dirigé ce travail avec tant d'enthousiasme.*

Je tiens tout particulièrement à exprimer ici ma plus profonde gratitude à M. ***François Lacoste*** *pour l'inestimable aide qu'il m'a apportée en acceptant de me fournir une bourse d'étude privée. Dès notre premier entretien, vous m'avez accordé votre confiance et je vous en serai éternellement reconnaissant. C'est grâce à vous que j'ai pu revenir au Sénégal et y vivre pendant ces trois années. Je vous dois également un voyage à Atlanta qui restera une étape capitale de ma formation. L'intérêt permanent que vous avez porté à mes travaux et le soutien sans faille que vous m'avez accordé m'ont permis de mener à bien ce projet et de me consacrer entièrement à mes recherches. Votre générosité et votre gentillesse exceptionnelles vous honorent et sont unanimement saluées par toutes les personnes qui ont eu connaissance de vos actions à mon égard. Merci encore pour tout ce que vous avez fait pour moi.*

C'est sans aucun doute à ***Tovi Lehmann*** *que je dois la plus importante part de mes connaissances en génétique des populations. Merci Tovi pour ton extrême gentillesse, tes encouragements et ta disponibilité de chaque instant et surtout pour ton aide si précieuse pour le traitement et l'interprétation de mes résultats. Je garde un excellent souvenir de mes deux voyages à Atlanta et je suis persuadé que notre collaboration future sera riche d'enseignements et très fructueuse. A très bientôt!*

Je souhaite également remercier Messieurs ***Bernard Decaris*** *et* ***Christian Dournon****, professeurs à l'Université de Nancy I pour avoir autorisé mon inscription en thèse dans cette Université qui a vu se dérouler tout mon cursus. Je suis particulièrement honoré par votre présence dans mon jury et vous en remercie très sincèrement.*

*Je tiens aussi à remercier Messieurs **François Rhodain** et **Michel Raymond** pour avoir accepté de juger ce travail et me faire bénéficier de leur incomparable expérience en entomologie médicale et en génétique des populations.*

*J'exprime mes plus sincères remerciements à **Vincent Robert** pour avoir accepté de siéger dans mon jury et pour le soutien et l'intérêt qu'il m'a manifestés à maintes reprises. A **Laurence Lochouarn** également, directrice du laboratoire après le départ de Didier, en souvenir de ces années passées ensemble: bonne chance et bon courage pour la suite!*

*Je ne peux résumer en quelques lignes tout ce je dois à **Jean-Jacques Lemasson**. Ton aide m'a été primordiale et souvent même salutaire au labo comme lors de nos nombreuses missions de collecte de moustiques où ton expérience de l'entomologie de terrain force le respect. Merci pour m'avoir fait bénéficier de tes innombrables talents tout au long de ce travail et particulièrement pour ton soutien et tes encouragements lors de la rédaction. Il existe aussi une vie à côté du boulot, et là encore ton rôle a été multiple: tour à tour conseiller, confident, ami et même père parfois (pardon!), infatigable saï-saï lors de nos virées nocturnes, tu as été incontestablement le plus grand complice de ma vie dakaroise. Je pars vers l'Ouest et toi vers l'Est (à moins que...) mais notre amitié née en Afrique ne saurait en souffrir et le souvenir de tout ce que nous avons vécu ensemble restera à jamais gravé dans ma mémoire. Sois assuré de ma plus profonde gratitude et de mon amitié la plus sincère.*

*Un grand merci également à **Frank H. Collins** et **Nora J. Besansky** pour l'accueil qu'ils m'ont réservé lors de mes visites au CDC d'Atlanta puis à l'Université de Notre Dame et pour les discussions très pertinentes que nous avons eues en ces occasions.*

*Pour leur accueil au laboratoire de Dakar, leur complicité et leur soutien, je souhaite remercier tous les djambars du laboratoire de Zoologie Médicale: **Mamoudou Diallo, Ibrahima «Stid» Dia, Mathurin Diatta, Mawlouth Diallo, Khalilou Ba, Yamar Ba, Ibrahima Baber, Baba, Ousmane Faye, Massamba Sylla, Teddy Diop, Ibrahima «goor gui» N'Diaye.** Bon courage et bonne chance à tous et encore merci pour tout.*

*Pour leur contribution à la collecte des moustiques et leur hospitalité, je tiens à rendre hommage à tous les habitants de Barkedji avec une mention particulière pour le «boss» **Sadjo Sow** et les deux «tueurs de serpents» **Batch** et **Gaousso**, palmes d'or de la capture d'anophèles. Je remercie également **Abdoulaye Diop** pour m'avoir accompagné lors de la première année de captures et avoir contribué de manière significative à ma formation aux techniques de collecte et d'investigations sur le terrain.*

*Plusieurs personnes m'ont apporté une aide à divers niveaux au cours de ce travail, qu'ils en soient remerciés ici: les docteurs **A. Sall, O. Faye, J.-F. Molez, L. Brutus, R. Gopaul** et **B. Delay**, Messieurs **S. Aris-Brosou** (merci pour les papiers), **R. Girod, O. Niang, M. Diagne** et **A. Thiaw**, Mesdames **A. Traore, K. Briot** et **H. Diallo-Agne** et tous les collègues de l'Institut Pasteur et de l'IRD de Dakar.*

Dans un tout autre registre, je souhaite exprimer ma gratitude à tous ceux qui ont partagé ma vie ici au Sénégal et qui ont contribué à faire de mon séjour un moment inoubliable. Pour leur soutien et leur hospitalité, je remercie particulièrement tous les membres de ma désormais belle-famille: Dieuredieuf ***Sana Touré, Boury, Cheikh, N'Diémé, Mara, Binetou, Lamine, Ousmane, Hassane, Omar, Khadim, Mame-Awa,*** *akk nyi ma waxul.... Merci à* ***Didi Sène*** *en particulier; à* ***Marie-Jo, Olymp, Dior....*** *et toutes celles et ceux régulièrement «de passage». A tous merci pour m'avoir prouvé que la Téranga n'est pas qu'un terme du guide du routard et pour l'accueil que vous avez réservé aux toubabs venus nous rendre visite.*

Un dernier mot pour remercier tous mes supporters de derrière l'écran qui grâce à leur E-mails m'ont permis de garder le contact avec la métropole et m'ont encouragé lors des moments difficiles: mes parents ***Liliane et Jean-Marie*** *pour leur soutien sans relâche et leur compréhension, le second thésard de la famille,* ***Arnaud*** *et* ***Céline*** *pour leur humour décapant, la famille* ***Bramardi*** *au grand complet, la famille* ***Clerc*** *exilée à Lyon,* ***Barb, Baud-Baud, Jojo, Pat, James*** *et tous les autres.....*

A tous, mille mercis.

A Moune,

TABLE DES MATIERES

INTRODUCTION **11**

1. La génétique des populations **14**

1. 1. - Principes 14

1. 2. - Les modèles mathématiques 15

1. 2. 1. - Le modèle d'Hardy-Weinberg 15

1. 2. 2. - Les modèles démographiques 18

1. 3. - Approches expérimentales et traitement des données 21

1. 3. 1. - Méthode directe 21

1. 3. 2. - Méthode indirecte 22

1. 4. - Outils moléculaires: le cas des microsatellites 24

1. 4. 1. - Caractéristiques générales 26

1. 4. 2. - Avantages et propriétés 28

1. 4. 3. - Les contraintes 30

1. 4. 4. - Applications 35

2. Le complexe *Anopheles gambiae* **36**

2. 1. - Identification des membres du complexe *An. gambiae* 38

2. 2. - Biologie et répartition géographique 41

2. 3. - Stratégies de lutte 45

2. 3. 1. - La lutte antivectorielle aujourd'hui 46

2. 3. 2. - Axes de recherche et perspectives 47

3. Structure génétique et spéciation au sein du complexe *An. gambiae* **49**

3. 1. - Les inversions chromosomiques comme marqueurs de populations 49

3. 2. - Polymorphisme des marqueurs isoenzymes 51

3. 3. - Un schéma contrasté fourni par les marqueurs moléculaires 52

3. 4. - Application des marqueurs microsatellites 53

3. 4. 1. - Structure géographique des populations d'*An. gambiae* Savane 53

3. 4. 2. - Présentation de l'étude 55

METHODOLOGIE GENERALE 56

1. Collecte et identification des moustiques sur le terrain 57

1. 1. - Capture par pulvérisation intra-domiciliaire d'insecticide 57

1. 2. - Capture sur sujet humain volontaire 59

1. 3. - Capture à l'aspirateur à bouche 61

1. 4. - Identification morphologique 61

2. Détermination de l'espèce et du génotype 62

2. 1. - Identification spécifique 62

2. 2. - Détermination du génotype aux loci microsatellites 64

3. Analyse des données 67

3. 1. - Test de conformité à l'équilibre d'Hardy-Weinberg 67

3. 2. - Test du déséquilibre de liaison entre paires de loci 68

3. 3. - Test d'homogénéité des fréquences alléliques ou génotypiques 68

3. 4. - Les statistiques *F* 68

3. 5. - Les statistiques *R* 69

PREMIERE PARTIE
Polymorphisme des marqueurs microsatellites chez *Anopheles arabiensis*: étude de la répartition spatiale de la variabilité génétique 70

1. Description des sites d'étude 73

1. 1. - Sénégal 73

1. 2. - Madagascar 73

1. 3. - L'archipel des Mascareignes 75

2. Analyse génétique **78**

2. 1. - Polymorphisme des marqueurs microsatellites chez *An. arabiensis* 78

2. 2. - Variabilité génétique au sein des populations 80

2. 3. - Différenciation génétique entre populations et flux de gènes 84

3. Structure génétique des populations d'*An. arabiensis*: le rôle des facteurs écologiques et historiques **86**

3. 1. - Perte de polymorphisme et contraintes sur la taille des populations: influence de la dérive génétique sur les populations locales 86

3. 2. - Différenciation génétique entre sous-populations: dérive génétique et flux de gènes 90

4. Retour sur les conditions d'application des modèles **93**

5. Conclusion **97**

DEUXIEME PARTIE
Suivi temporel de la variabilité génétique dans une population d'*Anopheles arabiensis* en zone sahélienne **99**

1. Dynamique des populations d'anophèles à Barkedji **101**

1. 1. - Observations de terrain 101

1. 2. - La problématique 106

2. L'approche temporelle : estimation de la taille effective des populations grâce aux variations de fréquences alléliques **108**

2. 1. - Principes 108

2. 2. - Formulation mathématique 110

2. 3. - Application aux données expérimentales et commentaires 112

3. Application de l'approche temporelle à la population d'*An. arabiensis* de Barkedji **113**

3. 1. - Evolution du polymorphisme des loci microsatellites au cours du temps 113

3. 2. - Taille effective de la population 120

3. 3. - Implications sur la biologie d'*An. arabiensis* en zone sahélienne 124

3. 3. 1. - L'estivation 125

3. 3. 2. - La dispersion des individus 127

4. Structure génétique au Sénégal 128

5. Conclusion 132

5. 1. - Fiabilité de l'estimation de Ne 132

5. 2. - Conclusion générale 135

TROISIEME PARTIE
Etude de la différenciation génétique entre les formes Mopti et Savane chez *Anopheles gambiae* s.s. 137

1. Introduction 138

1. 1. - Compléments sur les études cytogénétiques 138

1. 2. - Nouvelles données et présentation de l'étude 140

2. Résultats 142

2. 1. - Distribution des fréquences alléliques 142

2. 2. - Test de la panmixie 145

2. 3. - Différenciation génétique entre formes chromosomiques 147

3. Discussion 151

3. 1. - Le temps de divergence entre populations 153

3. 2. - Différents régimes de sélection 155

3. 3. - Des populations en expansion 158

4. Conclusion 160

CONCLUSIONS & PERSPECTIVES 163

1. Les marqueurs microsatellites et leur utilisation pour l'étude de la génétique des populations d'anophèles **164**

2. Structure des populations et flux de gènes au sein du complexe *An. gambiae* : implications sur les moyens de lutte antivectorielle **171**

REFERENCES BIBLIOGRAPHIQUES **177**

ANNEXES **198**

Annexe 1 :

Lehmann T, Hawley WA, Kamau L, Fontenille D, Simard F et Collins FH. 1996. Genetic differentiation of *Anopheles gambiae* populations from East and West Africa: a comparison of microsatellite and allozyme loci. *Heredity* **77**: 192-208.

Annexe 2 :

Simard F, Fontenille D, Lehmann T, Girod R, Brutus L, Gopaul R, Dournon C et Collins FH. 1999. High amounts of genetic differentiation between populations of the malaria vector *Anopheles arabiensis* from West Africa and Eastern Outer islands. *Am. J. Trop. Med. Hyg.* **60(6)**: 1000-1009.

Annexe 3 :

Lemasson JJ, Fontenille D, Lochouarn L, Dia I, Simard F, Ba K, Diop A, Diatta M et Molez J-F. 1997. Comparison of behavior and vector efficiency of *Anopheles gambiae* and *An. arabiensis* (Diptera: Culicidae) in Barkedji, a sahelian area of Senegal. *J. Med. Entomol.* **34(4)**: 396-403.

Annexe 4 :

Simard F, Lehmann T, Lemasson JJ, Diatta M et Fontenille D. Persistence of *Anopheles arabiensis* during severe dry-season conditions in West Africa : an indirect approach using microsatellite loci. In preparation.

INTRODUCTION

En dépit des campagnes de lutte instaurées depuis plus d'un demi siècle, le paludisme reste aujourd'hui la maladie parasitaire à laquelle l'humanité paye le plus lourd tribut. Actuellement en pleine recrudescence, il sévit à l'état endémique dans 102 pays et plus de deux milliards d'individus vivent en zone de transmission, principalement en Afrique sub-saharienne (Danis et Mouchet, 1991). Malgré l'absence de statistiques fiables, on estime que 300 à 500 millions de personnes sont atteintes par an, 1,7 à 3 millions des cas ayant une issue fatale en particulier chez les enfants de moins de 5 ans (Gysin et Doury, 1997; Declan, 1997; OMS, 1998). C'est une parasitose à transmission vectorielle. Les moustiques du genre *Anopheles* sont les seuls à pouvoir transmettre les 4 espèces de *Plasmodium* responsables des paludismes humains.

L'apparition chez le parasite de phénomènes de résistance aux antipaludiques courants et la diminution de la sensibilité des vecteurs aux insecticides ont mis un terme aux espoirs d'éradication de la maladie suscités dans les années 50 suite à la découverte des propriétés insecticides du DDT. La biologie des parasites comme celle des vecteurs et la complexité de leurs interrelations restent encore en bien des points obscures. L'extrême variabilité des profils épidémiologiques de la maladie ainsi que le manque cruel de moyens alloués à la recherche dans ce domaine constituent également des obstacles majeurs à la mise au point de moyens de lutte fiables et efficaces. En outre, l'accessibilité aux soins et le dépistage précoce des cas de paludisme est encore aléatoire dans la plupart des Etats les plus touchés et le manque de ressources dû à une diminution des investissements dans la plupart des pays concernés (Collins et Paskewitz, 1995) entrave la mise en œuvre convenable des moyens de prévention et de traitement dont nous disposons. Le vaccin en est toujours au stade de la recherche et de nouvelles stratégies de lutte moins coûteuses sont aujourd'hui nécessaires.

La communauté scientifique internationale s'est récemment mobilisée afin de coordonner les recherches et d'identifier les priorités en matière de lutte contre le paludisme (Declan, 1997). Désormais, l'objectif n'est plus l'éradication pure et simple de la maladie mais les efforts se concentrent pour limiter la mortalité et la morbidité palustre et parvenir à l'endiguement du fléau. Les nouvelles stratégies visent à diminuer l'intensité de la transmission: elles requièrent donc une connaissance approfondie de la biologie et de l'écologie des populations vectorielles afin d'en optimiser les moyens de contrôle.

Depuis la fin des années 50, de très nombreuses études entomologiques, souvent particulièrement détaillées, résultant du suivi des sites de transmission du paludisme ont mis en évidence la complexité du système vectoriel impliqué dans la zone afrotropicale (Hamon *et al.*, 1956; Coluzzi, 1984; Mouchet *et al.*, 1993). Dans la grande majorité des cas, plusieurs espèces transmettent le parasite en même temps ou de façon alternée au cours de l'année. Cinq espèces sont considérées comme vecteurs d'importance majeure sur le continent africain: *Anopheles gambiae*, *Anopheles arabiensis*, *Anopheles funestus*, *Anopheles nili* et *Anopheles moucheti*. A elles seules, *An. gambiae* et *An. arabiensis*, deux espèces jumelles appartenant au complexe *An. gambiae* s.l. (sens large) sont globalement responsables de 90% de la transmission dans la zone intertropicale (Robert et Carnevale, 1984). Ces deux vecteurs sont donc la cible essentielle des actions de lutte contre le paludisme en Afrique. Bien connaître ces populations vectorielles et définir leurs interrelations est primordial pour mieux comprendre les mécanismes de la transmission du paludisme et pouvoir agir de manière efficace.

Dans cette optique, nous avons entrepris une étude de la génétique des populations des anophèles du complexe *An. gambiae* grâce aux marqueurs ADN microsatellites. Au cours de cette introduction, nous allons dans un premier temps

présenter les principes, modèles et outils utilisés en génétique des populations en décrivant plus particulièrement les propriétés des marqueurs que nous avons choisi (Chapitre 1). Nous exposerons ensuite quelques généralités sur le complexe *An. gambiae*, les différentes espèces qui le composent et les stratégies utilisées ou préconisées pour les contrôler (Chapitre 2), pour terminer par une présentation des données disponibles sur la structure génétique des populations au sein du complexe *An. gambiae* qui nous permettra de préciser les objectifs de nos investigations (Chapitre 3).

1. LA GENÉTIQUE DES POPULATIONS

1. 1. - Principes

La génétique des populations est l'étude de la répartition de la variabilité génétique entre et au sein de populations naturelles et des forces génétiques, écologiques ou environnementales qui la gouvernent. Ces forces, nombreuses et interactives, sont notamment la mutation, la recombinaison, la sélection, la migration, la distance géographique entre individus et le système de reproduction. Reconnaître et quantifier ces facteurs est essentiel pour déterminer ou prédire l'évolution des populations et des espèces.

En entomologie médicale plus particulièrement, une description précise et intelligible de la variabilité génétique des populations vectorielles et de sa répartition dans l'espace et/ou dans le temps peut fournir des indications précieuses permettant de mieux comprendre l'épidémiologie de la maladie transmise. En effet, les forces régissant la structure génétique des populations de vecteurs et son évolution vont

également jouer un rôle primordial dans la répartition de la capacité à transmettre l'agent pathogène ou à résister à un moyen de contrôle.

La subdivision des populations naturelles en sous-unités locales apparaît d'un point de vue empirique comme un élément majeur à prendre en compte. En effet, chez la plupart des espèces, il existe des variations géographiques qui se manifestent tant au niveau morphologique qu'au niveau des fréquences de certains gènes. L'amplitude de ces différences résulte de la balance entre les forces qui tendent à individualiser les populations locales d'une part, et les forces qui tendent à homogénéiser la variabilité génétique au sein de l'espèce d'autre part. La complexité de ces systèmes nécessite le recours à des modèles mathématiques reposant sur un certain nombre de postulats simples qui permettront l'analyse des structures génétiques existantes.

1. 2. - Modèles mathématiques

La description des modèles utilisés en génétique des populations nécessite au préalable la définition de certaines notions spécifiques qui figurent dans l'Encadré 1. Dans cette discipline, la population est vue comme un "pool de gènes" dont la composition est susceptible d'évoluer, les individus n'étant que les porteurs transitoires de ces gènes.

1. 2. 1. - Le modèle d'Hardy-Weinberg

C'est le modèle le plus simple, qui sert de base à la génétique des populations. Il a été publié simultanément et indépendamment par G.H. Hardy et par W. Weinberg en 1908. Ce modèle décrit les relations entre fréquences alléliques et génotypiques dans une population idéale.

Encadré 1

Définition de quelques notions importantes en génétique des populations

- **Gène:** Un gène est un enchaînement de bases sur l'ADN destiné à être transcrit puis traduit en une succession d'acides aminés qui formeront une protéine. En génétique des populations, ce terme désigne, plus généralement, un fragment d'ADN non recombinant qui peut être codant ou non.

- **Allèle**: ce terme représente l'état d'un gène sur un chromosome. Chez les organismes diploïdes, un gène possède deux allèles, chacun sur l'un des deux chromosomes homologues. Si les deux copies sont identiques, l'individu est dit **homozygote**; si les deux copies sont différentes, l'individu est **hétérozygote**. La proportion d'individus hétérozygotes définit l'**hétérozygotie** de la population pour le marqueur considéré. C'est un des paramètres permettant de quantifier le degré de diversité génétique au sein d'une population (Nei, 1978).

- **Locus:** c'est l'emplacement d'un gène sur le chromosome (= localisation cytologique).

- **Population**: ensemble d'individus de la même espèce entre lesquels les croisements sont libres (c'est la **panmixie**).

- **Sous-population** (synonymes: dème, population locale): subdivision de la population.

- **Taille effective** d'une population (Wright 1931,1938): taille (en nombre d'individus) d'une population idéale, non soumise à la migration, ayant les mêmes propriétés que la population naturelle étudiée. Le sex-ratio est 1:1, les croisements sont libres et tous les individus ont la même capacité de reproduction. De ce fait, la taille effective de la population est très souvent inférieure à la taille réelle de la population naturelle car elle correspond au nombre d'individus qui vont effectivement avoir une descendance.
Symbole conventionnel: *Ne*.

- **Dérive génétique**: c'est une force évolutive qui traduit l'échantillonnage aléatoire des allèles de la population parentale qui seront représentés dans la descendance. Elle est directement liée à la taille finie des populations naturelles: plus les populations seront de taille restreinte, plus l'influence de la dérive génétique sera importante. Elle agit en faveur de l'augmentation de la différenciation entre populations.

- **Flux de gènes**: cette appellation regroupe toutes les formes de migration des gamètes, des individus ou de groupes d'individus d'une sous-population à une autre. Cette force s'oppose à la dérive génétique et favorise l'homogénéisation de la population.

Les hypothèses permettant de simplifier les calculs sont:

- les gamètes s'associent au hasard par rapport aux gènes considérés (panmixie),
- la population est de taille infinie,
- la fréquence des gènes n'est pas modifiée d'une génération à l'autre par la mutation, la sélection ou l'apport extérieur de nouveaux gènes (pas de migration).

Dans ces conditions, les fréquences alléliques restent stables au cours des générations et les fréquences génotypiques se déduisent des fréquences alléliques par la relation:

homozygotes $A_i A_i = p_i^2$

hétérozygotes $A_i A_j = 2 p_i p_j$

où p_i et p_j sont les fréquences des allèles A_i et A_j respectivement dans la population.
La population est dite **en équilibre d'Hardy-Weinberg**.

Cette relation entre fréquences alléliques et génotypiques est vraie dès qu'il y a panmixie, l'influence de la taille de la population, de la mutation, de la sélection et de la migration n'intervenant qu'au niveau de la constance des fréquences alléliques au cours des générations.

Ainsi, il est possible de tester si, à l'intérieur d'une population donnée, les croisements se font de manière aléatoire ou non en comparant les fréquences génotypiques observées à celles attendues sous l'hypothèse de l'équilibre d'Hardy-Weinberg par une statistique de type χ^2 ayant comme degré de liberté le nombre d'allèles-1 (nombre d'allèles indépendants) observés à un locus.

Une déviation significative par rapport à cet équilibre peut être due (i) au système de reproduction (autogamie, consanguinité, homogamie ...) ou (ii) au mélange de deux ou plusieurs sous-populations qui peuvent être elles-mêmes panmictiques mais

qui ne sont pas interfécondes: c'est l'**effet Wahlund** qui se manifeste par un déficit en hétérozygotes. Ce modèle servira constamment de base dans tout ce qui va suivre et permet de définir une population (voir Encadré 1).

1. 2. 2. – Les modèles démographiques

Les modèles décrivant la structure de populations dispersées dans l'espace appartiennent à deux grandes catégories: ceux décrivant une structure continue dans l'espace et ceux pour lesquelles les populations sont subdivisées en sous-unités discrètes. Tous ces modèles reposent sur l'étude de l'équilibre entre dérive génétique et flux de gènes (voir Encadré 1). Les populations sont supposées persister suffisamment longtemps pour que cet équilibre soit atteint. Les croisements sont libres à l'intérieur de chaque sous-population. Parmi ces modèles, on distingue:

- les modèles en îles (Wright, 1931; Latter, 1973; Slatkin, 1985a)

En 1931, Wright propose un modèle dans lequel une population d'individus diploïdes à générations discrètes est subdivisée en un nombre infini d'îles ou sous-populations de tailles finies N individus. Chaque île échange des migrants avec l'ensemble des autres îles. Les fréquences alléliques sur la population globale sont stables mais chaque île est soumise à la dérive génétique et à la migration. Plus tard, Latter (1973) puis Slatkin (1985a) vont affiner ce modèle en considérant un nombre fini d'îles d'effectifs égaux N. La population globale est alors elle aussi soumise à la dérive génétique et les fréquences alléliques peuvent changer au cours des générations.

Ces modèles ne supposent aucune géographie particulière des îles entre elles, chacune contribuant de manière égale à la formation d'un groupe de migrants équitablement redistribué à chaque génération. Wright (1931) a montré que si une

fraction m des N individus d'une île est remplacée par des immigrants provenant d'une autre île, les deux sous-populations ne pourront diverger génétiquement l'une de l'autre qu'à la condition où $Nm<1$. En d'autres termes, l'échange d'un seul individu par génération entre deux sous-populations est suffisant pour maintenir l'homogénéité de la variabilité génétique au sein de la population et contrecarrer l'effet de la dérive génétique. Le paramètre Nm est très souvent utilisé en génétique des populations pour quantifier les flux de gènes entre sous-populations.

- les modèles de "stepping stone" (Kimura, 1953)

Ces modèles reposent sur les mêmes caractéristiques que les modèles en îles mais font référence à un système géographique structuré à l'intérieur duquel les sous-populations ne peuvent échanger des migrants qu'avec leurs plus proches voisins, de façon réciproque. Différentes variantes existent en fonction de la position dans l'espace des voisins pouvant échanger des migrants: ce sont les modèles de "stepping stone" à une, deux ou trois dimensions. Un modèle à une dimension pourra décrire la répartition de populations le long d'un rivage ou d'un axe de communication, par exemple. Un modèle à deux dimensions sera plus adapté à l'étude de populations insulaires dans un archipel.

- le modèle de populations subdivisées et transitoires ou métapopulations

C'est Levins (1969, 1970) qui utilise pour la première fois le terme de métapopulation qu'il définit comme "une population de populations locales qui sont établies par des individus colonisateurs, survivent pendant un certain temps, envoient des migrants et, éventuellement, disparaissent. La persistance d'une espèce dans une

région donnée dépend de la balance entre le taux de colonisation et le taux d'extinction".

Dans un environnement constitué par plusieurs sites semblables capables de permettre le développement des individus (similaire au modèle en îles), la taille des populations locales occupant ces différents sites est soit 0 (site inoccupé), soit K (effectif en nombre d'individus de la population locale). La dynamique interne est ignorée, seuls les évènements de colonisation ou d'extinction sont pris en compte. Le taux d'extinction est fixe et identique pour chacune des populations locales; le taux de colonisation est variable en fonction du temps et dépend de p, la fraction de sites occupés à un temps donné (source de migrants colonisateurs) et de $(1-p)$, la fraction de sites inoccupés (cibles de la migration). La valeur de p à l'équilibre est $p=1-e/m$ où m et e sont respectivement les taux de colonisation et d'extinction. Si le taux d'extinction surpasse le taux de colonisation, l'espèce s'éteint. Ce modèle permet d'étudier l'impact des extinctions et des colonisations sur la répartition de la variabilité génétique entre sous-populations au sein d'une même espèce.

- le modèle de voisinage ou d'isolement par la distance (Wright, 1943)

A l'inverse des modèles précédents, aucune subdivision de la population n'est supposée: les individus sont librement dispersés dans l'espace, de manière continue. Ce modèle suppose que la dispersion des individus, ou plus généralement des gènes est une fonction décroissante de la distance géographique. Le paramètre caractérisant la migration n'est plus la fraction d'individus échangés entre deux sous-unités mais la taille du voisinage, fonction de la densité de la population et de la capacité de dispersion des individus (distance entre le lieu de naissance d'un individu et celui de ses parents).

Ces modèles ont fait l'objet d'études mathématiques et statistiques très poussées visant à définir et à estimer un certain nombre de paramètres généraux permettant de rendre compte de la force de la structure génétique et de l'intensité des relations qui existent entre et au sein de populations naturelles. Ils sont à l'origine d'outils permettant l'analyse des données expérimentales.

1. 3. - Approches expérimentales et traitement des données

Deux méthodes peuvent être utilisées dans le but d'étudier la structure génétique des populations et plus particulièrement les flux de gènes:

1. 3. 1. - Méthode directe

Elle consiste en l'observation directe sur le terrain des flux migratoires et de la dispersion des individus et fait notamment appel aux techniques de capture-marquage-recapture. Cette méthode a été appliquée à l'étude de la dispersion des moustiques, particulièrement chez *Ae. aegypti* mais peu d'exemples ont été recensés chez *An. gambiae* (Service, 1993). Les moustiques sont marqués par des poudres fluorescentes. Des modèles spécifiques permettent d'estimer la capacité de dispersion des individus, leur espérance de vie (taux de mortalité journalier), la densité de la population et d'en déduire une estimation de la taille effective des populations (Fisher et Ford, 1947; Service, 1993). Elle apporte donc des informations très importantes sur la capacité vectorielle des populations, leur biologie et leur importance épidémiologique.

Cependant, la méthode directe souffre de différents biais pour l'estimation des flux de gènes (Slatkin, 1987). En effet, la capacité de dispersion peut ne pas refléter les flux de gènes parce qu'elle n'est pas représentative du succès d'établissement et de reproduction des individus ayant migré. De plus, elle est nécessairement limitée dans le

temps (l'impact des flux de gènes en termes évolutifs devrait se mesurer sur un grand nombre de générations) et dans l'espace (répartition des points de recapture). Chez les moustiques, le taux de recapture est généralement très faible (1 à 5%) et cette méthode est inapplicable en saison sèche, la densité des populations étant extrêmement faible dans ces conditions (Service, 1993; Costantini *et al.*, 1996; Touré *et al.*, 1998).

1. 3. 2. - Méthode indirecte

Cette méthode s'appuie sur l'observation de la répartition géographique des fréquences alléliques, le plus souvent obtenues en utilisant des marqueurs allozymiques, mais aussi, depuis l'avènement des techniques de biologie moléculaire et de la PCR (Polymerase Chain Reaction) en particulier, des marqueurs permettant d'avoir un accès direct au polymorphisme de l'ADN (Nei, 1987; Avise, 1994; Raymond, 1996). Cette méthode diffère foncièrement de la précédente dans le sens où l'on se place en amont des flux de gènes, le but étant alors d'estimer le taux des flux géniques qui ont dû se produire pour expliquer le profil de répartition de la variabilité génétique observé. Cela justifie l'appellation "indirecte".

La méthode d'analyse des données la plus utilisée est celle développée par Wright afin d'estimer la variance standardisée des fréquences alléliques entre les sous-populations (Wright, 1951, 1965). Wright a construit un ensemble d'outils mathématiques basé sur les rapports d'hétérozygotie observée et calculée sous l'hypothèse de l'équilibre d'Hardy-Weinberg dans chaque sous-population et sur la population totale, les **statistiques *F***. Elles permettent de décrire la répartition de la variabilité génétique entre et au sein des populations. Trois coefficients sont utilisés:

- ***Fis*** mesure la déviation par rapport aux proportions attendues sous l'hypothèse de l'équilibre d'Hardy-Weinberg au sein de chacune des sous-populations

(*Fis* est également appelé coefficient de consanguinité des individus dans la sous-population à laquelle ils appartiennent),

- ***Fst*** mesure la différenciation génétique entre sous-populations (*Fst* est appelé index de fixation, quantifiant la consanguinité dans les sous-populations par rapport à la population totale),

- ***Fit*** mesure la déviation par rapport aux proportions attendues sous l'hypothèse de l'équilibre d'Hardy-Weinberg sur l'ensemble des sous-populations (*Fit* est le coefficient de consanguinité des individus dans la population totale).

Ces trois paramètres sont reliés par l'équation: $(1\text{-}Fit) = (1\text{-}Fis)\,(1\text{-}Fst)$

Weir et Cockerham (1984) ont montré la correspondance entre les paramètres *F* de Wright et les composantes de la variance des fréquences alléliques dans des populations structurées: la variance entre individus au sein d'une sous-population estime *Fis*, la variance entre sous-populations au sein de la population totale estime *Fst*, et la variance entre individus au sein de la population totale estime *Fit*. L'analyse par les statistiques *F* est donc analogue à une analyse hiérarchique de variance (nested ANOVA) ayant comme variables les fréquences alléliques et comme facteurs l'individu et la sous-population.

Ces estimateurs sans biais des paramètres *F* ont été généralisés à l'étude de plusieurs allèles et plusieurs loci et sont très utilisés dans les logiciels informatiques de génétique des populations. Ils sont sensibles à des différences d'effectifs trop importantes entre les échantillons testés.

Sous l'hypothèse d'un équilibre entre mutation et dérive, d'un modèle en nombre infini d'îles, d'un taux de mutation négligeable par rapport à la taille effective des

populations et si le taux de migration $m<<1$, le paramètre *Fst* est relié simplement à une mesure du flux de gènes entre sous-populations par la formule:

$$Fst = 1/(1+ 4\,Nm) \qquad \text{(Wright, 1951)}$$

Une autre méthode d'estimation de *Nm* a été proposée par Slatkin (1985b) sur la base de la fréquence des allèles privés, allèles qui n'existent que dans une seule sous-population (Neel, 1973). Pour les modèles en îles et de stepping-stone, il existe une relation linéaire entre le logarithme de la fréquence allélique moyenne de ces allèles et le logarithme de *Nm*. Cette méthode permet en outre un ajustement par rapport à la taille des échantillons.

Il existe encore d'autres méthodes pour étudier la structure génétique des populations et les flux de gènes telles que la méthode du maximum de vraisemblance (Rannala et Hartigan, 1996) ou les analyses de corrélation spatiales (Epperson, 1993) mais elles ne seront pas développées ici.

Les approches indirectes sont généralement les approches les plus souples pour analyser la structuration des populations. Les données nécessaires à ces analyses sont des effectifs (ou fréquences), des génotypes et des allèles. Différents outils permettent d'avoir accès à ces informations.

1. 4. - Outils moléculaires: le cas des microsatellites

L'approche indirecte pour estimer les flux de gènes s'intéresse à la distribution des gènes entre et au sein des populations. Afin d'avoir accès à ces informations, l'outil privilégié est longtemps resté l'électrophorèse enzymatique (Shaw et Prasad, 1970;

Richardson *et al.*, 1986). La migration sur gel d'amidon d'extraits enzymatiques puis leur révélation grâce à un substrat spécifique de l'enzyme étudié permet la mise en évidence d'isomorphes considérés comme autant d'allèles dont la fréquence peut dès lors être estimée dans les populations étudiées. La séparation des isomorphes sur gel d'amidon est due à une différence de la charge globale de la protéine qui révèle une différence dans la séquence des acides aminés qui la constituent. Ces systèmes ont l'avantage d'être codominants mais sont généralement peu polymorphes et peuvent s'avérer peu résolutifs en termes de génétique des populations.

La mise au point de la technique PCR en 1985 a permis le développement de nombreux outils moléculaires permettant d'avoir accès directement au polymorphisme de l'ADN, à partir de peu de matériel de départ. Parmi ces nouveaux marqueurs, les plus utilisés sont:

- les RAPD (Random Amplified Polymorphic DNA),
- les RFLP (Restriction Fragment Length Polymorphism),
- les SSCP (Single Strand Conformation Polymorphism),
- les micro- et minisatellites multilocus,
- les microsatellites,
- les séquences d'ADN (ADN mitochondrial ou nucléaire).

Nous allons maintenant décrire plus en détail les propriétés des marqueurs utilisés lors de notre étude: les **loci microsatellites**.

1. 4. 1. - Caractéristiques générales

Ce sont de petites séquences d'ADN non codant de deux à cinq paires de bases répétées en tandem présentant un polymorphisme de longueur (Encadré 2). Ils sont très répandus sur le génome des Eucaryotes (Hamada *et al.*, 1982; Tautz et Renz, 1984). Leurs allèles sont codominants et, étant situés dans des régions non codantes du génome, ils sont généralement considérés comme des gènes neutres. Leur fonction biologique reste peu connue (Epplen *et al.*, 1993). Cependant, les séquences de type $(CA)_n$ pourraient constituer des points chauds de recombinaison lors de la réplication ou intervenir dans la conformation de l'ADN (organisation de la chromatine). Les séquences de type $(CTG)_n$ quelque fois associées à des maladies génétiques humaines pourraient jouer un rôle dans la régulation de l'expression de certains gènes (Nordheim et Rich, 1983; Pardue *et al.*, 1987; Brahmachari *et al.*, 1995).

Bien que l'appellation la plus courante pour ce type de marqueurs soit "microsatellites", la littérature abonde de termes variés pour les décrire dont les plus courants sont SSR pour "Simple Sequence Repeats", STR pour "Simple Tandem Repeats", "short VNTR" où VNTR signifie "Variable Number of Tandem Repeats" ou encore SSLP pour "Single Sequence Length Polymorphism".

L'utilisation de ces marqueurs pour générer des empreintes génétiques apparaît vers 1985 mais la première véritable étude de structure des populations utilisant les microsatellites est celle de Edwards *et al.* (1992) sur des populations humaines.

Encadré 2

Les marqueurs microsatellites

(d'après Viard, 1996; Jarne et Lagoda, 1996)

Définition:

Séquences d'ADN répétées en tandem d'un motif de deux à cinq paires de bases.

5' GCCACGATGGTTTCG.....**GTGTGTGTGTGT**.....CGTCCAAGTACGATC 3'

séquence flanquante **6 répétitions** séquence flanquante

Les amorces de PCR correspondant à des séquences uniques (soulignées) sont déterminées dans les régions adjacentes aux répétitions (séquences flanquantes).

Structure:

Les marqueurs microsatellites peuvent être classés en trois familles selon leur structure interne.

Purs (parfaits)	GTGTGTGTGTGTGTGTGTGT
Composés	GTGTGTGTGT**CACACACACA**
Interrompus	GTGTGT**TT**GTGT**TT**GT**TT**GT

Base de la variabilité:

Les séquences microsatellites présentent un polymorphisme de longueur, du au nombre variable de répétitions qui les composent. Chaque allèle correspond à un nombre de répétitions (en théorie) et à une taille d'amplifiat (en pratique).

Dans l'exemple ci-dessous, les séquences flanquantes font 80 paires de bases et le motif répété est un dinucléotide (2 pb):

Nombre de répétitions	Fragment amplifié	Taille du fragment
12	❒❒❒❒®®®®®®®®®®®®❒❒❒❒	12 x 2 + 80 = **104** pb
11	❒❒❒❒®®®®®®®®®®®®❒❒❒❒	11 x 2 + 80 = **102** pb
10	❒❒❒❒®®®®®®®®®®❒❒❒❒	10 x 2 + 80 = **100** pb
9	❒❒❒❒®®®®®®®®®❒❒❒❒	9 x 2 + 80 = **98** pb

❒ = séquences flanquantes ® = motif répété (2pb)

1. 4. 2. – Avantages et propriétés

Quelques caractéristiques importantes pour les études de génétique des populations sont à retenir:

- une mise en évidence aisée des allèles

Grâce à des amorces spécifiques désignées dans les séquences adjacentes aux répétitions (séquences flanquantes, voir Encadré 2), l'amplification par PCR puis l'électrophorèse sur gel d'agarose ou d'acrylamide permettent d'avoir directement accès au génotype des individus (Queller *et al.*, 1993).

- un taux de mutation élevé générant un important polymorphisme

Les mécanismes moléculaires à l'origine du polymorphisme important des microsatellites ne sont pas encore aujourd'hui parfaitement élucidés. Cependant, le glissement de la polymérase ("slippage") lors de la réplication de l'ADN et les crossing-over inégaux semblent être la source majeure de la variation (Levinson et Gutman, 1987; Tautz et Schlötterer, 1994; Dover, 1995). Ces mécanismes particuliers confèrent aux microsatellites un taux de mutation très élevé. Les estimations chez l'homme (Weber et Wong, 1993), la souris (Dallas, 1992) ou *An. gambiae* (Lehmann *et al.*, 1998) sont de l'ordre de 10^{-3} à 10^{-5}, soit de 10 à 1000 fois supérieures aux valeurs avancées pour les isoenzymes (Ayala, 1976). Ces forts taux de mutations rendent les microsatellites très polymorphes et donc potentiellement très résolutifs pour les études de génétique des populations.

- la conservation de loci microsatellites d'une espèce à une autre a été démontrée dans des groupes aussi variés que les cétacés (Schlötterer *et al.*, 1991), les oiseaux (Primmer *et al.*, 1996a), les primates (Deka *et al.*, 1994; Garza *et al.*, 1995; Blanquer-Maumont et Crouau-Roy, 1995), les tortues marines (FritzSimmons *et al.*, 1995) ou les mammifères en général (Stallings, 1995) et peut permettre l'utilisation

d'amorces hétérologues pour des études comparatives de la variabilité génétique entre espèces proches.

- des modèles de mutation et d'interprétation des données spécifiques

Le modèle de mutation généralement utilisé pour les isoenzymes est l'IAM (Infinite Allele Model) décrit par Kimura et Crow, 1964. Dans ce modèle adapté à des marqueurs ayant un faible taux de mutation, chaque nouvel allèle apparaissant par mutation est différent de ceux qui existaient déjà dans la population. Par conséquent, les allèles sont indépendants les uns des autres. Shriver *et al.* (1993) et Valdes *et al.* (1993) ont montré par simulation sur ordinateur, que l'évolution des microsatellites humains à motif dinucléotide serait mieux décrite par un modèle de type SMM ("Stewise Mutation Model", Kimura et Otha, 1978). Le SMM suppose qu'à chaque mutation est associée une suppression ou une addition d'une répétition du motif. Néanmoins, d'autres études n'ont pas pu rejeter ni l'un ni l'autre de ces modèles pour expliquer le profil de variation obtenu (Estoup *et al.*, 1995a). Une évolution du SMM, le TPM (Two-Phase Model) a également été proposé par Di Rienzo *et al.* (1994): dans ce modèle, une proportion p de mutations suit un SMM et une proportion $1-p$ se traduit par des ajouts ou des délétions de plusieurs motifs. Au contraire de l'IAM, les modèles SMM et TPM tiennent donc compte de la relation entre allèles en fonction de leur taille: la taille d'un allèle mutant est liée à la taille de l'allèle qui a muté, et deux allèles de même taille n (répétitions) peuvent avoir été générés par des processus différents ($n+1 \rightarrow n$ ou $n-1 \rightarrow n$ sous un SMM).

Ces différents modèles ont permis le développement de nouvelles méthodes d'analyse des données spécifiques aux loci microsatellites qui tiennent compte de leur fort taux de mutation et de l'interdépendance de la taille des allèles. Ce sont les statistiques *R* (Slatkin, 1995). Les estimateurs obtenus (*Ris* et *Rst*) sont comparables aux

estimateurs définis par Weir et Cockerham (1984) pour les statistiques *F*, l'analyse de variance étant effectuée sur la taille des allèles (en nombre de répétitions) et non plus sur la simple identité ou non-identité des allèles (Goldstein *et al.*, 1995a; Michalakis et Excoffier, 1996; Kimmel *et al.*, 1996). L'estimation de *Nm* est reliée à *Rst* par une formule analogue à celle de Wright (1951) corrigée pour les effets d'échantillonnage (Slatkin, 1995).

1. 4. 3. - les contraintes

L'utilisation des marqueurs microsatellites en biologie et en génétique des populations est allée grandissante depuis 1992 (voir par exemple les revues de Buford et Wayne, 1993; Ashley et Dow, 1994; Jarne et Lagoda, 1996; Goldstein et Pollock, 1997). Cette profusion d'études originales a permis de mettre en évidence l'existence de contraintes inhérentes aux microsatellites qui peuvent se manifester à plusieurs niveaux et conduire à des interprétations erronées en termes de flux de gènes et de structure des populations. Parmi ces contraintes, on peut retenir notamment:

- des contraintes agissant sur la taille des allèles

A l'évidence, le nombre de répétitions que comprend un allèle microsatellite ne peut pas s'étendre indéfiniment: la taille des fragments est donc obligatoirement limitée et le nombre d'allèles différents à un locus n'est pas infini. Cette observation a été confirmée aussi bien par des données expérimentales (Bowcock *et al.*, 1994; Deka *et al.*, 1994; Garza *et al.*, 1995) que par des simulations sur ordinateur (Nauta et Weissing, 1996; Lehmann *et al.*, 1996). Combinées à de forts taux de mutation tels que ceux observés au niveau des loci microsatellites, ces contraintes pourraient introduire un biais dans l'estimation des flux de gènes en surestimant l'identité des allèles entre différentes populations. Dans ces conditions, les marqueurs microsatellites pourraient s'avérer

relativement peu résolutifs en ce qui concerne l'étude de la structure des populations de grande taille ou la reconstruction de phylogénies d'espèces ayant divergé depuis longtemps.

- le phénomène d'homoplasie (voir Encadré 3)

Le même raisonnement a été à l'origine de la mise en évidence du phénomène d'homoplasie: deux allèles identiques par état (dans le cas des microsatellites, deux allèles de même taille) peuvent résulter de processus évolutifs différents (et donc ne pas être identiques par ascendance). Ces excès de similarité entre allèles peuvent se manifester par des différences (substitutions, insertions ou délétions) dans les séquences flanquantes ou dans la place des interruptions dans le cas des microsatellites "interrompus" qui ne seront révélées que par séquençage des différents allèles. Néanmoins, ce phénomène semble plus commun entre espèces éloignées qu'entre sous-espèces phylogénétiquement proches (Estoup *et al.*, 1995b). Les microsatellites peuvent donc s'avérer peu recommandables pour l'étude de phylogénies interspécifiques (Garza *et al.*, 1995; Doyle *et al.*, 1998).

- un taux de mutation hétérogène

Bien que le taux de mutation des microsatellites soit généralement très élevé comme nous l'avons vu précédemment, certaines observations suggèrent qu'il ne soit pas uniforme pour tous les loci. Ainsi, on a pu constater des différences de taux de mutation entre espèces pour le même locus (Rubinsztein *et al.*, 1995); en fonction de la nature du locus (pur, interrompu ou composé) et du motif répété (di-, tri- ou tetranucléotide) au sein de la même espèce (Chung *et al.*, 1993; Pépin *et al.*, 1995; Brinkmann *et al.*, 1998) ou en fonction de la taille des allèles, les allèles les plus longs étant généralement plus exposés à la mutation (Wierdl *et al.*, 1997; Schlötterer *et al.*, 1998; Brinkmann *et al.*, 1998). D'autres indications plaident en faveur d'un taux de

Encadré 3

Exemples de mise en évidence de phénomènes d'homoplasie

Variations dans la nature et la place des interruptions (Estoup *et al.*, 1995b)

Allèles observés au locus A113 chez différentes sous-espèces d'*Apis mellifera* (*A. m.*).

Espèce	Séquence de l'allèle (5'→ 3')	Taille
A. m. mellifera	$(TC)_2C(TC)_2\mathbf{TT}(TC)_4\mathbf{TT}(TC)_2\mathbf{TT}(TC)_9GTTTCG(TC)_2$	214 pb
A. m. ligustica	$(TC)_2C(TC)_2\mathbf{TT}(TC)_4\mathbf{TT}(TC)_2\mathbf{TT}(TC)_9GTTTCG(TC)_2$	214 pb
A. m. scutellata	$(TC)_2C(TC)_2\mathbf{TT}(TC)_5\mathbf{TT}(TC)_{11}GTTTCG(TC)_2$	214 pb
A. m. capensis	$(TC)_2C(TC)_2\mathbf{TT}(TC)_5\mathbf{TTTCTT}(TC)_3\mathbf{TT}(TC)_{10}GTTTCG(TC)_2$	214 pb

Ces allèles sont tous identiques par état (même taille du fragment amplifié) mais l'analyse de leur composition montre des différences substantielles révélant qu'ils ne sont pas identiques par ascendance. La position et la nature des interruptions peuvent être très informatives dans ce cas.

Nombre variable de répétitions au niveau d'un locus composé (Garza et Freimer, 1996)

Allèles observés au locus Mfd79 chez le chimpanzé *Pan troglodytes*:

	Séquence de l'allèle (5'→ 3')	Taille
Allèle 1	$(CA)_1(TA)_{20}(CA)_{12}$	172 pb
Allèle 2	$(CA)_1(TA)_{21}(CA)_{11}$	172 pb

Le nombre de répétitions des différents motifs constituant ce locus sont variables et leur différences peuvent s'annuler, induisant de l'homoplasie.

Variations dans les séquences flanquantes (Grimaldi et Crouau-Roy, 1997)

Allèles observés au locus MIB chez l'homme:

	Séquence de l'allèle (5'→ 3')	Taille
Allèle 1	...GTTATCC$\mathbf{(CA)_{23}}$TAAA - - - - - - CATAAG...	350 pb
Allèle 2	...GTTATCC$\mathbf{(CA)_{20}}$TAAACATAAACATAAG...	350 pb

Des délétions ou insertions dans les séquences flanquantes peuvent également permettre la mise en évidence de phénomènes d'homoplasie, y compris dans le cas de loci purs.

mutation biaisé vers l'augmentation de taille des allèles pour un locus donné (Garza *et al.*, 1995; Rubinsztein *et al.*, 1995; Primmer *et al.*, 1996b; Amos *et al.*, 1996). Toutes ces données soulignent que le mécanisme moléculaire d'évolution des loci microsatellites est encore mal compris et ceci doit être pris en compte lors des études entreprises avec ces marqueurs (Kruglyak *et al.*, 1998; Colson et Goldstein, 1999).

- l'existence d'allèles nuls (ou non détectables sur gel)

L'une des contraintes les plus importantes et certainement la plus répandue concernant les microsatellites est l'existence d'allèles nuls. Leur présence se manifeste généralement par un excès d'homozygotes observé au locus affecté, associé à l'impossibilité d'amplifier l'ADN de certains individus (Callen *et al.*, 1993). Ils sont dus à des mutations ponctuelles qui rendent impossible l'appariement des amorces utilisées pour la PCR (Callen *et al.*, 1993; Lehmann *et al.*, 1996). Dès lors, les individus hétérozygotes pour cet allèle apparaissent comme homozygotes pour l'allèle amplifié, et les individus homozygotes pour l'allèle nul ne donnent aucun amplifiat. Leur mise en évidence indubitable se fait par séquençage des zones d'appariement des amorces (Callen *et al.*, 1993; Koorey *et al.*, 1993), par des confrontations de génotypes parents-descendance (Paetkau et Strobeck, 1995; Pemberton *et al.*, 1995; Van Treuren, 1998) ou par la désignation de nouvelles amorces ("alternative primers") qui évitent la mutation (Paetkau et Strobeck, 1995; Lehmann *et al.*, 1996). Lorsque la séquence des zones flanquantes n'est pas connue, différentes méthodes existent pour évaluer la fréquence de tels allèles en fonction du déficit en hétérozygotes qu'ils génèrent (Chakraborty *et al.*, 1992; Brookfield, 1996). Selon Jarne et Lagoda (1996), la fréquence de ces allèles peut atteindre 15% et ils pourraient toucher jusqu'à 25% des loci microsatellites. Ce sont donc des particularités très importantes qu'il convient d'avoir en mémoire. L'existence

de tels allèles a également été rapportée chez les isoenzymes (Foltz, 1986). Dans ce cas, l'allèle nul ne possède pas l'activité enzymatique qui permet sa révélation.

Toutes les contraintes agissant sur les loci microsatellites décrites dans ce paragraphe doivent être sérieusement prises en compte lors des investigations visant à déterminer le degré de flux géniques entre sous-populations afin de minimiser les biais qu'elles introduisent. Néanmoins, ces contraintes s'exercent toutes spécifiquement sur un locus (voire sur un allèle) particulier et ne sauraient toucher tous les loci avec la même influence. Par contre, les forces régissant la structure des populations (dérive génétique et flux de gènes en particulier) s'exercent de façon identique sur tout le génome. Par conséquent, lors de l'analyse des données, il apparaît judicieux de prêter une attention particulière aux informations fournies par chacun des loci utilisés, indépendamment des autres loci, et de vérifier l'homogénéité des tendances révélées par l'ensemble des loci.

De telles analyses détaillées fournissent un moyen pertinent d'estimer l'influence relative des deux catégories de forces évolutives s'exerçant sur les loci microsatellites. Dans la mesure du possible, l'utilisation de différents marqueurs, soumis à des mécanismes moléculaires d'évolution différents, permettra de valider les interprétations. Au cours de ce travail, nous aurons l'occasion de revenir sur certains de ces points.

1. 4. 4. - Applications

En résumé, grâce à leur ubiquité, leur abondance sur le génome et le fort taux de mutation qui leur confère un important polymorphisme, les microsatellites apparaissent aujourd'hui comme des marqueurs de choix dans de nombreux domaines:

- la biologie et la génétique des populations (Bossart et Pashley-Prowell, 1998),

- la cartographie du génome (Bell et Ecker, 1994; Dib *et al.*, 1996; Zheng *et al.*, 1996) et son application dans l'analyse de traits quantitatifs (QTL, voir Gorman *et al.*, 1997),

- la génétique humaine, en particulier comme marqueurs des maladies génétiques (Sutherlands et Richards, 1995), et de plus en plus pour l'étude des phylogénies et de l'histoire de l'Humanité (Bowcock *et al.*, 1994; Goldstein *et al.*, 1995b; Reich et Goldstein, 1998).

- la médecine légale: autopsie, recherche de paternité, police scientifique (Holland, 1993; Jin et Chakraborty, 1995).

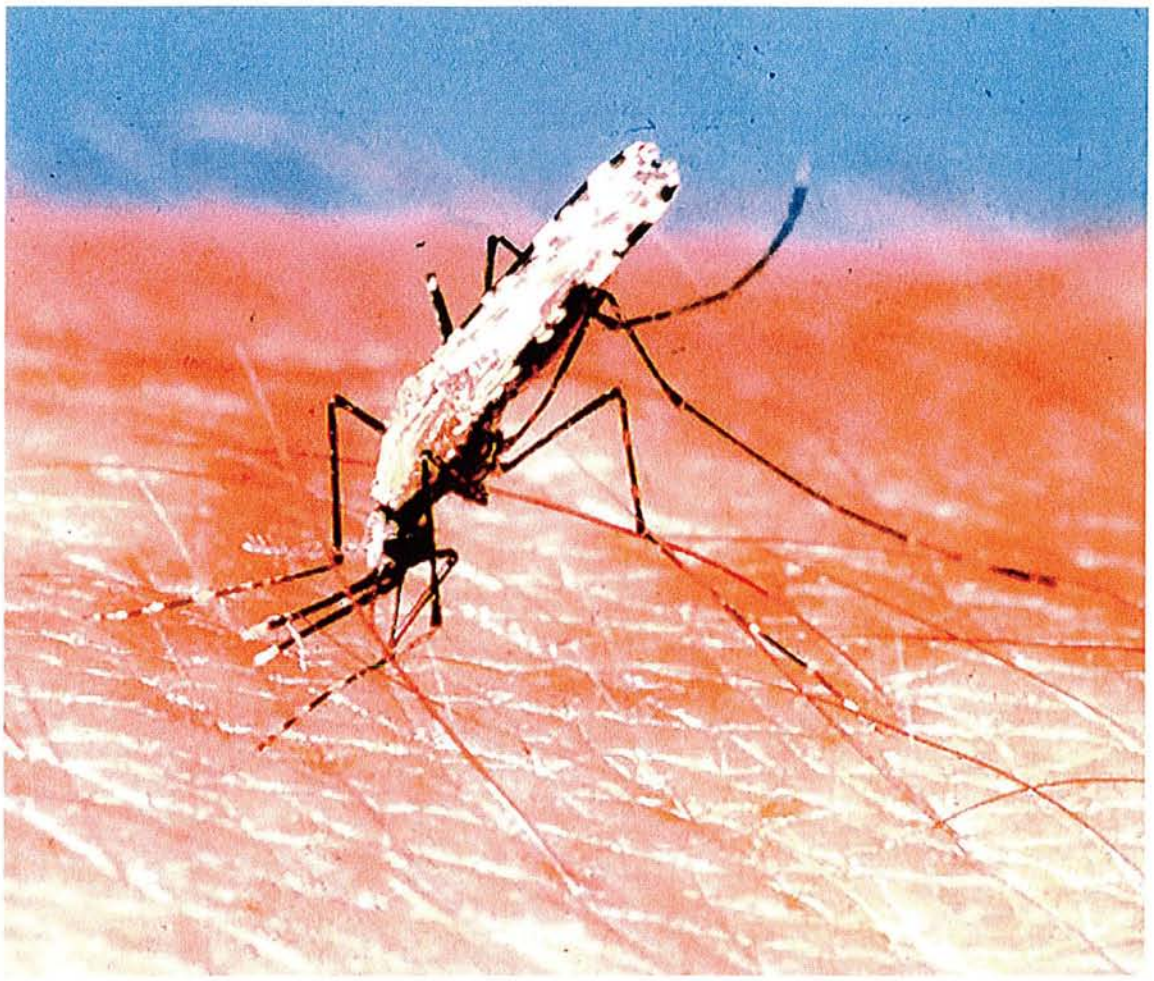
Nous avons utilisé ce type de marqueurs afin d'étudier la génétique des populations de moustiques appartenant au complexe *An. gambiae*, un système vectoriel d'une importance primordiale en entomologie médicale, dont nous allons maintenant décrire les caractéristiques biologiques, écologiques et génétiques.

2. Le complexe *Anopheles gambiae*

L'intérêt que portent les scientifiques à l'étude des anophèles et plus particulièrement aux membres du complexe *An. gambiae* (Illustration 1) est très étroitement lié à l'importance de la menace qu'ils représentent pour la santé publique. Ce sont incontestablement les vecteurs les plus efficaces du paludisme en Afrique sub-Saharienne mais ils sont également impliqués dans la transmission de *Wuchereria bancrofti*, agent de la filariose lymphatique. Ils ont en outre joué un rôle prépondérant lors des flambées épidémiques des fièvres dues au virus O'Nyong-Nyong en Afrique de l'Est et ont été trouvés porteurs de plusieurs autres souches d'arbovirus (Gillies et Coetzee, 1987; Digoutte, 1993). A ce titre, ce sont certainement les plus importants vecteurs d'agents pathogènes humains au monde (Gwadz et Collins, 1996). L'impact de leur macabre contribution à la transmission de pathologies humaines est énorme en termes de santé publique et d'économie (Nàjera et Hempel, 1996).

Les anophèles sont des Diptères nématocères appartenant la famille des *Culicidae* qui regroupe l'ensemble des insectes connus sous le nom de moustiques. Parmi plus de 500 espèces d'anophèles recensées dans le monde, seule une vingtaine sont considérées comme vecteurs majeurs du paludisme, d'autres espèces pouvant localement jouer un rôle secondaire (Rhodain et Perez, 1985). Cette aptitude à transmettre le parasite découle de particularités biologiques, écologiques et génétiques très spécifiques qui déterminent la capacité vectorielle d'une espèce. Elle sera notamment fonction:

- de la préférence trophique de l'espèce c'est à dire de la propension à piquer l'homme (anthropophilie) plutôt que d'autres vertébrés (zoophilie). C'est en effet lors



Source : OMS.

Illustration 1: Femelle *Anopheles gambiae* s.l. prenant un repas de sang.

du prélèvement d'un repas de sang sur un individu porteur du parasite que l'anophèle va s'infecter.

- de la longévité du moustique car le parasite doit effectuer un cycle de multiplication et de différenciation avant que l'anophèle ne puisse le transmettre lors du prélèvement d'un nouveau repas de sang.

- de facteurs génétiques, pour le moment en grande partie inconnus, qui autorisent le développement spécifique des parasites chez certaines espèces d'anophèles seulement, déterminant ainsi la compétence vectorielle de l'espèce.

Bien que semblables morphologiquement, tous les membres du complexe *An. gambiae* ne possèdent pas la même aptitude à transmettre le paludisme ce qui rend particulièrement nécessaire l'identification et l'individualisation des différentes espèces qui le constituent.

2. 1. - Identification des membres du complexe *An. gambiae*

La notion d'espèces et de complexe d'espèces est définie dans l'Encadré 4.

La spéciation au sein des membres d'un complexe d'espèces morphologiquement identiques se traduit généralement par des différences au niveau comportemental et génomique. La capacité à démembrer un complexe d'espèces est donc fonction des outils disponibles pour mettre en évidence ces différences et identifier les espèces jumelles.

Les premières indications d'une hétérogénéité au sein des populations d'*An. gambiae* furent en effet basées sur des considérations purement écologiques avec la

Encadré 4

Notion d'espèce et de complexe d'espèces

La notion d'espèce a varié avec les progrès de la biologie.

La définition originale, donnée par Cuvier était « une collection de tous les corps organiques nés les uns des autres ou de parents communs et de ceux qui leur ressemblent ». La base de l'espèce était donc essentiellement morphologique bien que l'idée d'interfécondité soit déjà présente.

En 1963, Mayr définit une espèce comme « un groupe de populations naturelles capables d'intercroisement et reproductivement isolées d'autres groupes semblables ». La discrimination entre espèces se fait donc dès lors sur la base d'une capacité à donner des descendants viables.

Cette notion fait encore référence aujourd'hui mais certains cas particuliers sont ambigus. En effet, des spécimens morphologiquement indifférenciables peuvent être reproductivement isolés et appartenir à des espèces distinctes que l'on a appelées « **espèces jumelles** » (ou « sibling species »). Chez les moustiques par exemple, des degrés différents d'isolement sont observés lors des croisements: pas d'œufs viables, adultes F1 non viables ou stériles (généralement les mâles). Ces différentes barrières de reproduction peuvent être dues à des incompatibilités d'ordre génétique, écologique ou comportemental. Cette incompatibilité peut se manifester avant ou après l'accouplement. Il est fréquemment advenu qu'après avoir séparé les espèces jumelles selon ces critères, on observe des différences morphologiques qui jusque là étaient passées inaperçues.

L'ensemble des espèces jumelles résultant du démembrement d'une espèce morphologique est appelé **complexe d'espèces**.

description de «variants» capables de coloniser des gîtes larvaires saumâtres (Muirhead-Thomson, 1951). Sur la base de croisements effectués en laboratoire, ces variants se sont avérés partiellement isolés des souches inféodées aux gîtes d'eau douce, elles-mêmes formant trois groupes de croisement distincts (Davidson 1962, 1964 a et b; Davidson et Jackson, 1962; Paterson, 1964). Les mâles issus de ces croisements interspécifiques sont stériles mais les femelles sont fertiles. Les travaux de Coluzzi en cytogénétique (Coluzzi et Sabatini, 1967, 1968 a et b, 1969) ont permis une avancée considérable des moyens d'identification des différentes espèces constitutives du complexe *An. gambiae*, basée sur l'existence d'inversions chromosomiques polymorphes mises en évidence par coloration des chromosomes géants isolés des glandes salivaires des larves au stade IV (voir ci-dessous) ou des ovaires des femelles semi-gravides (Green, 1972). D'autres techniques, basées sur des différences de profil électrophorétique des hydrocarbones cuticulaires ou sur les variations de fréquences de certains isoenzymes ont également été développées mais ne sont pas aussi discriminantes (Miles, 1978, 1979; Carlson et Service, 1980; Hamilton et Service, 1983; Bullini, 1984; Marchand et Mnzava, 1985). L'identification par cytogénétique, malgré sa mise en œuvre difficile et le niveau d'expertise qu'elle requiert, est longtemps restée la technique de choix dans ce domaine. Les progrès de la biologie moléculaire ont alors permis le développement de sondes ADN spécifiques (Collins *et al.*, 1987; Crampton et Hill, 1993) qui ont débouché sur la mise au point d'un test diagnostic basé sur l'amplification par PCR des gènes codant pour les ARN ribosomaux (Scott *et al.*, 1993). Ce test permet l'identification rapide des spécimens collectés sur le terrain, quels que soient le sexe et le stade de leur développement. Il nécessite très peu de matériel de départ. Son efficacité a été grandement démontrée (Fontenille *et al.*, 1993) et il est maintenant pratiqué en routine dans de nombreux laboratoires.

Le complexe *An. gambiae* comprend aujourd'hui sept espèces bien différenciées:

- *An. gambiae* s.s. (sens strict) Giles, 1902.
- *An. arabiensis* Patton, 1904.
- *An. bwambae* White, 1985.
- *An. melas* Theobald, 1903.
- *An. merus* Doenitz, 1902.
- *An. quadriannulatus* Theobald, 1911.
- *An. quadriannulatus* B, dénomination provisoire de la septième espèce du complexe identifiée très récemment en Ethiopie (Hunt *et al.*, 1998).

2. 2. – Biologie et répartition géographique

Tous les membres du complexe ont un cycle biologique typique de la plupart des anophèles (voir Encadré 5). Néanmoins, certaines particularités comportementales et écologiques distinguent les différentes espèces qui n'ont pas toutes la même importance en entomologie médicale.

An. gambiae s.s. et *An. arabiensis* sont les plus importants vecteurs du paludisme et les plus répandus sur le continent Africain ainsi qu'à Madagascar. Le premier occupe les régions de forêt mais remonte loin en savane; le second, plus adapté aux climats secs, est présent au Sahel et sur les plateaux d'Afrique méridionale et de Madagascar, mais s'enfonce assez profondément dans les savanes humides concentriquement au premier. Leur zone de recouvrement est très large (Figure 1). Les larves se développent dans les collections d'eau claire peu profondes, ensoleillées et dépourvues de végétation (empreintes de pas, de roues, fossés, flaques....). En saison des pluies, de nombreux gîtes de ce type sont disponibles et les densités de

Encadré 5

Cycle Biologique des anophèles

Comme tous les Diptères, les anophèles sont holométaboles, c'est-à-dire qu'ils présentent des métamorphoses complètes. Ils passent successivement par **quatre stades de développement**: œuf, larve, nymphe et adulte ou imago. Les trois premiers stades sont aquatiques alors que les adultes mènent une vie aérienne. Les mâles ne piquent pas et seules les femelles sont hématophages. Elles trouvent dans le sang des vertébrés les protéines nécessaires à la maturation des œufs. Par conséquent, seules les femelles anophèles sont susceptibles de transmettre le paludisme.

- Biologie des stades larvaires

Une seule femelle peut pondre de 200 à 300 œufs de 0,6 à 0,8 mm de long à chaque ponte. Les œufs, entourés de flotteurs, sont déposés isolément à la surface de l'eau et éclosent généralement au bout de deux à trois jours. Ils sont très sensibles à la dessiccation (Beier *et al.*, 1990).
On distingue ensuite quatre stades larvaires successifs, séparés par des mues. Les larves sont détritiphages et se nourrissent près de la surface de l'eau où elles doivent remonter pour respirer. La durée du développement larvaire varie de 8 à 30 jours selon l'espèce et les conditions éco-climatiques.
Après la dernière mue, la larve au quatrième stade se transforme en nymphe très mobile qui ne se nourrit plus. Au cours de ce stade qui dure généralement moins de 48 heures, l'insecte subit de profondes transformations morphologiques qui lui permettent de passer du stade aquatique au stade aérien lors de l'émergence.

- Biologie des adultes

Après émergence, les moustiques doivent se reposer 2 ou 3 jours pour que leur exosquelette durcisse et pour atteindre leur maturité sexuelle. Il leur faut également prendre un repas sucré pour satisfaire leurs besoins énergétiques. La fécondation a alors lieu au crépuscule lors d'essaimages (« swarming »). La femelle n'est probablement fécondée qu'une seule fois et les spermatozoïdes, stockés dans une spermathèque, seront relargués au cours de chaque ponte. Les ovocytes sont fécondés lors de leur passage dans l'oviducte. La prise d'un repas de sang sur un hôte vertébré est primordiale pour la maturation des ovocytes et conditionne la ponte.
L'ensemble des processus qui débute par la recherche de l'hôte, se poursuit par la prise d'un repas de sang puis sa digestion concomitante au développement des ovocytes et se termine par la ponte est connu sous le nom de **cycle gonotrophique**. Après la ponte, la femelle se met en quête d'un nouvel hôte et un autre cycle commence.
Pour la plupart des espèces tropicales, la durée d'un cycle gonotrophique est de 2 à 3 jours. L'accomplissement du premier cycle après l'émergence nécessite souvent la prise de deux repas de sang.

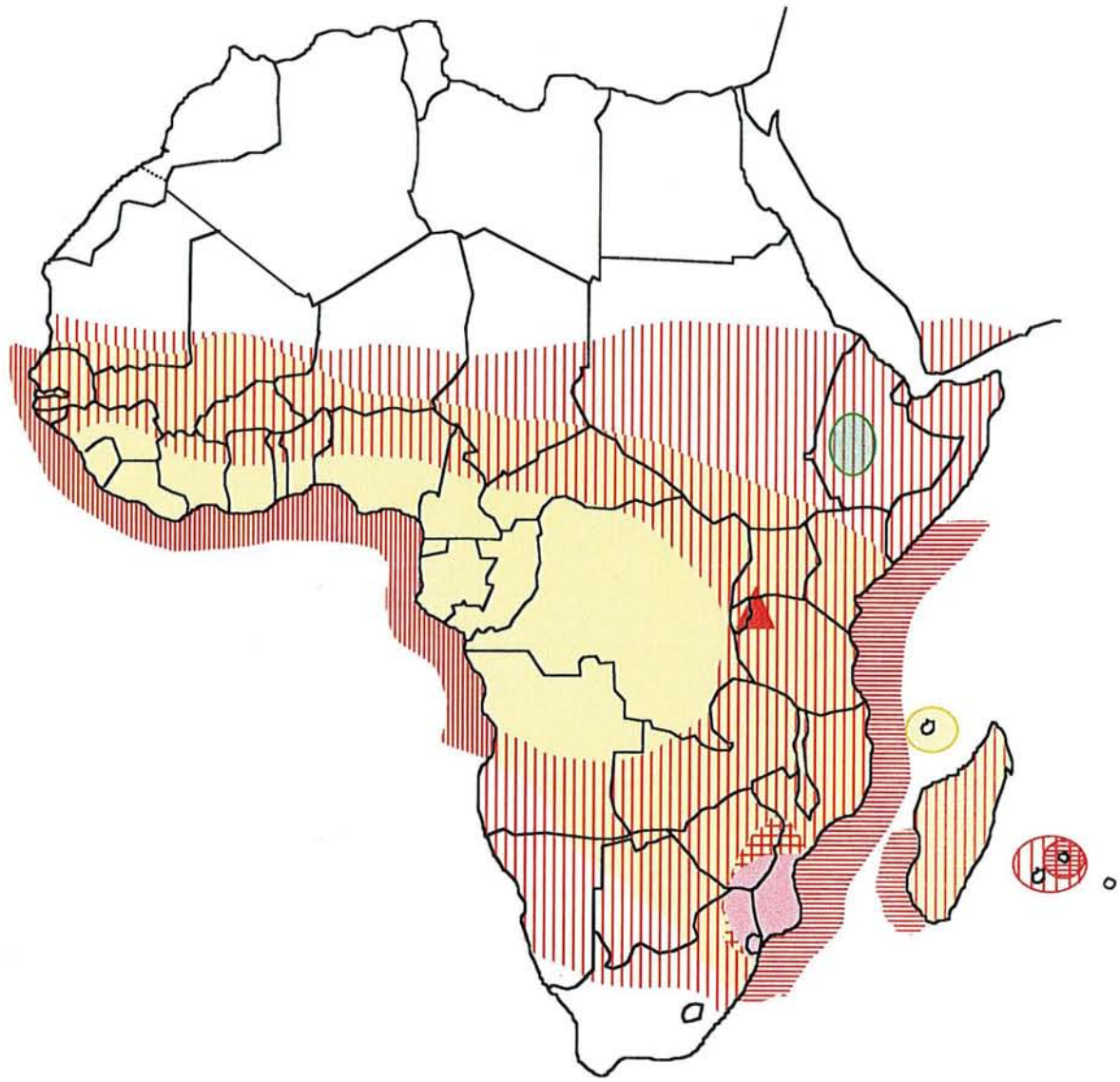
populations sont maximales à cette période de l'année. Les adultes sont très anthropophiles et très étroitement associés à l'homme à tel point qu'*An. gambiae* est souvent qualifié d'espèce domestique. Cette association se manifeste également par un comportement endophage (les femelles entrent à l'intérieur des maisons pour y piquer l'homme) et endophile (elles restent à l'intérieur des maisons pour digérer le repas de sang). A la différence d'*An. gambiae*, *An. arabiensis* est capable de se développer dans des zones totalement sauvages, adoptant alors un comportement exclusivement zoophile. C'est le seul vecteur potentiel toujours présent à Maurice et sur l'île de La Réunion, dans l'Océan Indien.

An. melas et *An. merus* sont deux espèces inféodées aux gîtes d'eau saumâtre. Leur répartition géographique est de ce fait limitée aux zones côtières bien qu'ils puissent pénétrer loin à l'intérieur des terres au niveau des estuaires. Ils occupent respectivement les mangroves des côtes occidentales et orientales d'Afrique (Figure 1) où ils peuvent jouer un rôle local dans la transmission du paludisme. Ils sont toutefois préférentiellement zoophiles. *An. merus* a été identifié sur la côte Ouest de Madagascar, il est également présent sur l'île Maurice mais semble absent de La Réunion.

Les larves d'*An. bwambae* n'ont été trouvées que dans les sources géothermales de la vallée du Rift et sont très dépendantes de la composition minérale de leur milieu (Davidson et White, 1972). En conséquence, cette espèce est circonscrite à cette région géographique très limitée.

An. quadriannulatus est zoophile et n'est donc pas vecteur de pathologies affectant l'homme. Ses larves se développent en eau douce et sa présence n'a été signalée que dans quelques pays d'Afrique orientale méridionale (Mozambique, Zimbabwe, Afrique du Sud). Il est probable que son aire de répartition reste à complèter

Figure 1: Répartition des espèces du complexe *Anopheles gambiae* en Afrique.



An. gambiae

An. arabiensis

An. quadriannulatus

An. merus

An. bwambae

An. quadriannulatus B

An. melas

Zone de sympatrie de ces quatre espèces

D 'après Danis et Mouchet (1991), actualisé.

mais cette espèce n'a que peu d'intérêt en entomologie médicale. *An. quadriannulatus* B a été récemment isolée en Ethiopie et sa caractérisation est en cours. Son élévation au rang d'espèce est consécutive à l'obtention d'individus mâles stériles dans la descendance issue de croisements avec une souche de référence d'*An. quadriannulatus* d'Afrique du Sud, bien que les chromosomes polytènes soient en tous points semblables entre ces deux taxa (Hunt *et al.*, 1998).

Les travaux de cytogénétique en particulier ont permis de mettre en évidence une très forte hétérogénéité génétique à l'intérieur même de l'espèce *An. gambiae* suggérant des phénomènes de spéciation très récents ou en cours. De très fortes structures génétiques corrélées aux conditions environnementales ont été mises en évidence par ce biais, comme nous le verrons dans le chapitre 3. Le complexe *An. gambiae* dans son ensemble apparaît ainsi constitué de nombreuses sous-unités plus ou moins isolées génétiquement et géographiquement qui reflètent la diversité des profils épidémiologiques de la transmission du paludisme en Afrique.

2. 3. – Stratégies de lutte

Compte tenu de l'apparition de résistances aux antimalariques chez *Plasmodium falciparum*, agent infectieux responsable des accès pernicieux de la maladie, la chimioprophylaxie antipaludique n'est envisageable que dans des cas bien particuliers (voyageurs, femmes enceintes). Les médicaments étant réservés aux traitements curatifs, la lutte antivectorielle est donc actuellement la seule méthode de prévention de masse applicable en zone d'endémie palustre. Le but est d'empêcher la prolifération des vecteurs et de réduire le contact homme-vecteur afin de limiter la transmission.

2. 3. 1. – La lutte antivectorielle aujourd'hui

Les moyens sont multiples et regroupent:

- les actions mécaniques: l'assainissement de l'environnement et le drainage des marais comme moyens de protection collectifs par limitation des gîtes larvaires disponibles. Les anophèles ayant une agressivité surtout nocturne, l'utilisation de moustiquaires est une bonne protection individuelle.

- les insecticides chimiques: les composés organochlorés, le DDT, la dieldrine et le lindane, premiers insecticides synthétiques, ont provoqué une véritable révolution dans la lutte antivectorielle mais les espoirs furent rapidement déçus suite à l'apparition de résistances chez les anophèles. La recherche de nouveaux composés a abouti à la mise sur le marché d'insecticides organophosphorés (malathion et fénitrothion contre les imagos, téméphos contre les larves), de carbamates (propoxur et bendiocarb utilisés en aérosol comme imagocides) et surtout des pyréthrinoïdes synthétiques tels que la perméthrine, la deltaméthrine et la lambdacyhalothrine. Ces derniers sont très utilisés actuellement pour les traitements intra-domiciliaires mais le problème des résistances et de leur rémanence dans l'environnement devient très préoccupant.

- les agents de lutte biologique: ils concernent exclusivement la lutte anti-larvaire et regroupent des parasites (champignons, nématodes ou microsporidies), des prédateurs (poissons larvivores) et des pathogènes (Baculovirus ou bactéries entomopathogènes). Ces méthodes sont peu probantes concernant la lutte contre les anophèles, à cause de la nature même des gîtes larvaires qu'ils colonisent (voir Danis et Mouchet, 1991).

La stratégie actuelle de lutte contre le paludisme vise essentiellement au contrôle de la maladie (OMS, 1996). Plusieurs études ont montré que l'utilisation de moustiquaires et de rideaux imprégnés d'insecticide réduisait le taux de morbidité et de mortalité palustre (voir Lengeler, 1998, pour une analyse combinée des résultats de plus de 60 études réparties dans 38 pays). Cependant, dans les zones de paludisme stable où les taux d'infection sont élevés, ces méthodes n'ont pas permis de faire chuter le taux de transmission au-dessous d'un seuil permettant une diminution réelle de la morbidité palustre et leur efficacité sur du long terme reste sujet à controverse (Trape et Rogier, 1996; Snow *et al.*, 1997; D'Allessandro et Coosemans, 1997; Lengeler *et al.*, 1997; Robert et Trape, 1997).

2. 3. 2. - Axes de recherche et perspectives

Le premier objectif en matière de lutte contre le paludisme reste, en dépit des réticences qui s'expriment parfois et des déceptions entraînées par certains essais cliniques qui s'avéraient prometteurs (Patarroyo *et al.*, 1988; Herrington *et al.*, 1987; Ballou *et al.*, 1987; Valero *et al.*, 1993; Alonso *et al.*, 1994; D'Allessandro *et al.*, 1995; Ferreira, 1996; Valero *et al.*, 1996; Nosten *et al.*, 1996), le développement d'un vaccin qui doit prendre en compte les différents stades du parasite (Mons, 1997; Miller *et al.*, 1997; Engers et Godal, 1998). Des vaccins altruistes dirigés contre les antigènes des stades sexués du parasite qui ne sont pas pathogènes pour l'homme sont également à l'étude. Ils pourraient permettre de bloquer la transmission chez le moustique (Shahabuddin *et al.*, 1996, 1999; Kaslow, 1997). Parallèlement, la recherche de nouveaux anti-parasitaires continue par l'investigation de pharmacopées traditionnelles.

D'autre part, les recherches sur le génome des moustiques et le génie génétique («entomologie moléculaire» selon Mulla, 1994) sont en plein essor et la manipulation

génétique des vecteurs pour en altérer la capacité vectorielle ouvre des perspectives énormes (Crampton *et al.*, 1990; Kidwell et Ribeiro, 1992; Aldous, 1993; Crampton, 1994; Mulla, 1994; Collins, 1994; Curtis, 1994; Collins *et al.*, 1994; Collins et James, 1996). L'intérêt de cette nouvelle orientation en termes de lutte antivectorielle réside dans l'auto-propagation du principe actif (diffusion d'un ou plusieurs gènes au sein de populations naturelles) et le faible coût de l'application sur le terrain (lâcher d'individus génétiquement transformés).

La lutte antivectorielle par moustiques transgéniques nécessite en premier lieu l'identification des gènes cibles sur lesquels agir afin de rendre les populations vectorielles réfractaires aux *Plasmodium*. Les recherches dans ce domaine sont nombreuses et progressent rapidement. Des souches d'*An. gambiae*, d'*An. stephensi* et *d'Aedes aegypti* réfractaires à différentes espèces de *Plasmodium* humains et animaux ont été sélectionnées en laboratoire (Collins *et al.*, 1986; Feldmann et Ponnudurai, 1989; Thathy *et al.*, 1994). L'étude des croisements entre souches réfractaires et susceptibles a permis de préciser les mécanismes responsables de ces phénotypes et d'en étudier le déterminisme génétique (Vernick *et al.*, 1989; Vernick et Collins, 1989; Crews-Oyen *et al.*, 1993; Gorman *et al.*, 1996, 1997; Gorman et Paskewitz, 1997; Yan *et al.*, 1997; Zheng *et al.*, 1997; Collins *et al.*, 1997). Parallèlement, une carte génomique d'*An. gambiae* a également été dressée (Zheng *et al.*, 1996).

En second lieu, la mise au point d'une technique de transgénèse et la recherche de véhicules permettant d'introduire les gènes d'intérêt chez l'anophèle est primordiale. Dans ce domaine également, les recherches sont très actives (Collins et James, 1996; Curtis et Townson, 1998; Ashburner *et al.*, 1998).

Enfin, une condition *sine qua non* à la mise en pratique de cette nouvelle stratégie de lutte est une connaissance approfondie de la structure génétique des

populations naturelles de vecteurs et des flux de gènes qui la gouvernent. Ces recherches sont également primordiales pour aider à la gestion des moyens de lutte existant, apportant des informations particulièrement pertinentes sur la capacité de diffusion des gènes de résistance. C'est à ce niveau que se situe l'étude présentée dans ce mémoire. Nous allons maintenant approfondir ce sujet et faire le point sur l'état actuel des connaissances accumulées dans ce domaine chez *An. gambiae* s.l..

3. Structure génétique et spéciation au sein du complexe *An. gambiae*

Les études indirectes de génétique des populations sur les moustiques du complexe *An. gambiae* ont commencé par le démembrement du complexe et l'identification des différentes espèces qui le composent (voir chapitre 2). La plupart des études ont ensuite concerné le taxon *An. gambiae* s.s., le plus efficace vecteur du paludisme; *An. arabiensis* restant peu étudié en dépit de sa contribution non négligeable à la transmission de la maladie. Les autres espèces du complexe, beaucoup plus localisées géographiquement et dont l'importance vis-à-vis de l'épidémiologie du paludisme est nettement moindre sont peu ou pas étudiées en génétique des populations.

3. 1. – Les inversions chromosomiques comme marqueurs de populations

Les connaissances accumulées sur *An. gambiae* reposent en grande partie sur la cytogénétique, l'étude des inversions chromosomiques paracentriques polymorphes (Frizzi et Holstein, 1956; Coluzzi *et al.*, 1979). Certaines de ces inversions sont fixées et servent de marqueurs pour l'identification spécifique des membres du complexe (Green, 1972; Coluzzi et Kitzmiller, 1975), d'autres sont très polymorphes et révèlent une

organisation particulière des populations au sein de l'espèce *An. gambiae* (Coluzzi *et al.*, 1979; Coluzzi, 1984). En effet, l'étude de la structure des populations d'*An. gambiae* dans différentes conditions environnementales (climatiques et écologiques) a révélé l'existence de plusieurs "formes" chromosomiques caractérisées par la présence de certaines inversions. Ces différentes formes sont également désignées sous le terme cytotypes. Cinq cytotypes ont été identifiés en Afrique de l'Ouest: les cytotypes Forêt, Savane, Bissau, Mopti et Bamako (voir Touré *et al.*, 1983 et Coluzzi *et al.*, 1985 pour une définition complète des caractéristiques de ces cytotypes). Seul le cytotype Savane est représenté en Afrique de l'Est. Nous reviendrons de manière plus précise sur les particularités de ces cytotypes au cours de la Troisième Partie de ce mémoire. Brièvement, chaque cytotype est associé à des conditions environnementales particulières (notamment climatiques, selon un gradient hygrométrique) et, en zone de sympatrie, des variations saisonnières dans la composition de la population ont été mises en évidence, les formes adaptées à un climat plus sec étant majoritaires en saison sèche et les formes rencontrées en zone de forêt étant majoritaires en saison des pluies (Bryan *et al.*, 1982; Coluzzi *et al.*, 1985; Petrarca *et al.*, 1983, 1986, 1987; Touré *et al.*, 1983, 1987, 1994; Coosemans *et al.*, 1989). Ces formes apparaissent partiellement ou totalement reproductivement isolées dans la nature mais leur descendance est parfaitement viable en laboratoire ce qui laisse supposer un mécanisme de barrière pré-copulatoire à la reproduction (voir références citées précédemment) et une spéciation récente, voire en cours, au sein du taxon *An. gambiae*.

Chez *An. arabiensis*, ces inversions sont également très polymorphes mais les différents arrangements sont généralement en équilibre d'Hardy-Weinberg suggérant la panmixie à l'intérieur de cette espèce (Mekuria *et al.*, 1982; Coluzzi *et al.*, 1985). Toutefois, une corrélation entre la fréquence de certaines inversions et le comportement

trophique de ce vecteur (zoo ou anthropophilie, endo ou exophilie) ou l'environnement dans lequel il évolue (ville ou village, climat plus ou moins sec, saison des pluies ou saison sèche) a parfois été rapportée (Petrarca et Beier, 1992; Petrarca *et al.*, 1986, 1987; Mnzava *et al.*, 1995) mais ces associations sont beaucoup moins fortes que chez *An. gambiae*.

Les inversions chromosomiques apparaîtraient alors comme de larges unités de recombinaison impliquant un ensemble de gènes co-adaptés, conférant aux individus des capacités différentielles d'adaptation à tel ou tel milieu ou influençant certaines composantes comportementales, voire la susceptibilité des individus aux *Plasmodium* (Petrarca et Beier, 1992), en fonction du caryotype (Carson, 1982; Coluzzi, 1982; Coluzzi *et al.*, 1985).

3. 2. – Polymorphisme des marqueurs isoenzymes

Les études réalisées sur ces deux espèces avec les isoenzymes sont peu nombreuses et ont eu essentiellement pour but la recherche d'allèles spécifiques permettant d'identifier les différents membres du complexe; elles se sont révélées peu résolutives à ce niveau, seuls quelques loci permettant de différencier les deux espèces (Miles, 1979; Bullini, 1984; Mahon *et al.*, 1976), mais ont permis de trouver des taux de différenciation, certes faibles, mais significatifs entre des populations d'*An. gambiae* d'Afrique de l'Est et d'Afrique de l'Ouest (Miles, 1978, 1979; voir annexe 1). Par contre, aucune différentiation n'a été révélée par ces marqueurs entre les différents cytotypes, qu'ils soient sympatriques ou non (Bullini et Coluzzi, 1978; Cianchi *et al.*, 1983; Coluzzi *et al.*, 1985). Certains loci particuliers ont pourtant suggéré une possible structuration géographique et/ou comportementale à l'intérieur des populations d'*An.*

gambiae et d'*An. arabiensis* en Afrique de l'Est mais la panmixie n'a jamais pu être mise en doute par ce biais (Mukiama, 1987; Smits *et al.*, 1996; Coosemans *et al.*, 1998).

3. 3. – Un schéma contrasté fourni par les marqueurs moléculaires

L'accès aux techniques de biologie moléculaire a permis des investigations plus poussées de la phylogénie et de la génétique des populations au sein du complexe *An. gambiae*, notamment grâce aux travaux de Zheng *et al.* (1991, 1993, 1996) qui ont établi une carte génétique d'*An. gambiae* recensant 131 loci microsatellites dont 47 ont été localisés précisément sur la carte physique réalisée par coloration des chromosomes (Coluzzi *et al.*, 1979). Des loci RAPD et RFLP ont rapidement complété cette carte (Dimopoulos *et al.*, 1996; Della Torre *et al.*, 1996). La correspondance entre ces deux cartes, génétique et physique, s'est révélée d'une utilité primordiale pour les études qui ont suivi en fournissant la possibilité d'étudier comparativement le comportement de marqueurs situés à l'intérieur ou à l'extérieur des inversions chromosomiques.

Les RAPD, dans une certaine mesure, puis les RFLP ont permis récemment de conforter l'hypothèse de spéciation récente entre les différents cytotypes au sein du taxon *An. gambiae* ainsi que le rôle des inversions dans l'évolution des phénomènes adaptatifs (Favia *et al.*, 1994, 1997; Mathiopoulos et Lanzaro, 1995). L'analyse de la séquence de certains gènes nucléaires ou localisés sur l'ADN mitochondrial a par contre révélé des phylogénies conflictuelles au sein du complexe, remettant parfois en cause l'origine monophylétique des inversions, l'un des principes de base permettant la reconstruction d'arbres phylogénétiques à partir des observations de cytogénétique, et suggérant des phénomènes d'introgression entre *An. arabiensis* et *An. gambiae* via des femelles hybrides viables et fertiles en zone de sympatrie (Besansky *et al.*, 1994, 1997; Garcia *et al.*, 1996; Della Torre *et al.*, 1997; Caccone *et al.*, 1998). De telles femelles

hybrides sont effectivement observées *in natura* dans les zones de sympatrie à des fréquences de 1 à 2 ‰ (White, 1970; Coluzzi *et al.*, 1979).

3. 4. – Application des marqueurs microsatellites

La première étude utilisant les loci microsatellites fut publiée par Lanzaro *et al.* en 1995 et visait à évaluer le potentiel résolutif des microsatellites par rapport aux marqueurs isoenzymes chez *An. gambiae*. Les 11 loci microsatellites testés se révélèrent hautement polymorphes (nombre moyen d'allèles par locus = 7,45; hétérozygotie moyenne calculée = 0,732) par rapport aux 23 systèmes enzymatiques dont seulement 40% présentèrent plus d'un allèle par locus (nombre moyen d'allèle par locus = 1,55; hétérozygotie moyenne calculée = 0,097) sur les 24 spécimens utilisés suggérant que ces loci puissent se révéler particulièrement adaptés pour des études de génétique des populations chez *An. gambiae*.

3. 4. 1. – Structure géographique des populations d'*An. gambiae* Savane

La comparaison des populations d'*An. gambiae* cytotype Savane entre le Sénégal et le Kenya grâce à 5 loci microsatellites et 6 loci isoenzymes est présentée en annexe 1:

LEHMANN T, HAWLEY WA, KAMAU L, FONTENILLE D, SIMARD F et COLLINS FH. 1996. Genetic differentiation of *Anopheles gambiae* populations from East and West Africa: a comparison of microsatellite and allozyme loci. *Heredity* 77: 192-208.

Les travaux de Lehmann *et al.* (article présenté en annexe 1) représentent la première application des marqueurs microsatellites à l'étude la structuration génétique des populations naturelles d'*An. gambiae*. Deux populations d'*An. gambiae* cytotype Savane séparées par plus de 6000 km sur le continent Africain, entre le Sénégal et le Kenya (Ouest de la vallée du Rift) ont été comparées. Bien que les 5 loci microsatellites utilisés lors de cette étude se soient révélés très polymorphes, un très faible degré de

différenciation génétique a été estimé entre ces populations (Fst = 0,016; Rst = 0,036), comparable à celui observé par Miles (1978) grâce aux isoenzymes (Fst = 0,036, annexe 1). Ces faibles taux de différenciation génétique, similaires bien qu'estimés à partir de marqueurs dont le mécanisme d'évolution est très différent, suggèrent des flux de gènes importants chez *An. gambiae* cytotype Savane sur tout le continent Africain (Nm>3), les populations du Sénégal et du Kenya étant reliées par un habitat continu.

Par la suite, d'autres études comparant des populations séparées par 700 km de part et d'autre de la vallée du Rift au Kenya, un système de dépressions géographiques arides et très faiblement peuplées ne permettant pas l'établissement de populations d'*An. gambiae*, ont révélé des taux de différenciation supérieurs à ceux observés précédemment (microsatellites: Fst =0,08; Nm=1,54). Ces résultats suggèrent un rôle important de la vallée du Rift comme barrière à la migration des individus, restreignant ainsi les flux de gènes entre populations (McLain *et al.*, 1989 utilisant les RFLP localisés dans les segments intergéniques codant pour les sous-unités ribosomiques; Kamau *et al.*, 1998a, utilisant les microsatellites). L'homogénéité des populations à l'échelle microgéographique a été vérifiée dans la région Ouest du Kenya par une étude hiérarchique de la répartition de la variabilité génétique en fonction de l'unité d'échantillonnage (maison, village et région); aucune indication de déviation par rapport à la panmixie n'étant révélée à l'intérieur d'une même maison, ni entre maisons à l'intérieur d'un village, ni entre villages (Lehmann *et al.*, 1997; Besansky *et al.*, 1997).

Ainsi, la structure génétique des populations d'*An. gambiae* Savane semble essentiellement générée par l'existence de zones géographiques particulières ne permettant pas l'établissement de populations. En l'absence de fortes barrières à la

migration des individus, les populations restent faiblement différenciées, suggérant des flux de gènes importants même sur de très grandes distances géographiques.

3. 4. 2 – Présentation de l'étude

Sur la base de ces données, nous avons entrepris d'estimer si le même phénomène était observable au sein de l'espèce *An. arabiensis*, par la comparaison de 4 populations très isolées géographiquement (Première Partie). Cette étude nous a permis de vérifier que les marqueurs microsatellites initialement utilisés chez *An. gambiae* pouvaient se révéler de très bons outils pour l'étude d'autres membres du complexe. Il est particulièrement important de s'intéresser à cette espèce dans la mesure où *An. arabiensis* est souvent le seul vecteur présent dans les zones de paludisme instable, là où la lutte antivectorielle est réputée efficace en termes de santé publique, permettant une réduction sensible de la morbidité et de la mortalité palustre.

Nous avons ensuite testé la stabilité au cours du temps de la variabilité génétique au sein d'une population d'*An. arabiensis* endémique d'une région de savane sahélienne au Sénégal (Deuxième Partie). La méthode d'analyse appliquée à ces données nous a permis d'estimer la taille effective de cette population soumise à d'importantes fluctuations saisonnières en densité.

Nous nous sommes finalement intéressés à l'évaluation du degré de divergence génétique révélé par les marqueurs microsatellites entre les cytotypes Mopti et Savane chez *An. gambiae* (Troisième Partie). Ce phénomène de "spéciation en cours" (Coluzzi *et al.*, 1985) au sein même de l'espèce *An. gambiae* représente un modèle d'étude particulièrement intéressant pour tester le potentiel résolutif des loci microsatellites et comparer les indications qu'ils fournissent à celles obtenues à partir d'autres marqueurs génétiques.

METHODOLOGIE GENERALE

1. Collecte et identification des moustiques sur le terrain

Les moustiques étudiés au cours de ce travail ont été collectés selon trois méthodologies différentes, classiquement utilisées par les entomologistes.

1. 1. - Capture par pulvérisation intra-domiciliaire d'insecticide

C'est la méthode la plus rentable en termes d'effectif de l'échantillon collecté. Elle consiste en la pulvérisation intra-domiciliaire d'insecticides dérivés du pyréthre. Les chambres sont visitées au petit matin pour éviter au maximum la fuite des moustiques dérangés par les activités journalières des habitants. Après avoir évacué les vivres et ustensiles de vaisselle qui sont parfois stockés dans la pièce, des draps blancs sont étendus sur toute la surface du sol et toutes les fenêtres et portes sont fermées (Illustration 2). De l'insecticide à base de pyréthre en aérosol est alors pulvérisé à l'intérieur de la chambre. Pendant ce temps, une seconde personne fait le tour de la chambre par l'extérieur et pulvérise de l'insecticide au niveau des ouvertures fréquemment observées dans les murs ou à la base du toit pour éviter la fuite des insectes (Illustration 3). La chambre est maintenue fermée pendant une dizaine de minutes, temps nécessaire à l'action de l'insecticide. Les insectes, tués par le pyréthre tombent au sol, sur les draps. Ils sont alors aisément repérables et collectés (Illustration 4). Les anophèles sont facilement identifiables par rapport aux autres culicidés et le tri peut être réalisé à ce moment. Une fois la collecte effectuée, la chambre est aérée pour éliminer les dernières effluves de l'aérosol.

Les anophèles collectés par cette méthode représentent la fraction endophile de la population, au repos dans les habitations. Service (1993) estime que 80 à 95% des moustiques présents dans la chambre sont effectivement collectés, certains spécimens

Collecte des moustiques par pulvérisation intra-domiciliaire d'insecticide

Illustration 2: Des draps blancs sont étendus sur toute la surface du sol de la pièce visitée.

Illustration 3: L'insecticide à base de pyréthre est pulvérisé à l'intérieur et autour de la pièce.

Illustration 4: Les moustiques sont collectés après action de l'insecticide.

Photos: Jean-Jacques Lemasson IRD, Dakar.

parvenant à s'échapper (en fonction de l'état général de la construction), d'autres s'enfonçant profondément dans le treillis des pailles du toit ou les lézardes des murs avant de mourir.

Le taux d'anthropophilie des différentes espèces peut être déterminé par la suite par analyse du sang contenu dans l'estomac des moustiques collectés (Beier *et al.*, 1988).

1. 2. - Capture sur sujet humain volontaire

Cette méthode consiste en la capture de la faune culicidienne agressive vis-à-vis de l'homme. Les anophèles ayant une activité nocturne, les séances de capture en zone tropicale sont organisées à partir de 18 heures (heure solaire) jusqu'au lendemain matin 6 heures. Des collecteurs volontaires munis de tubes à hémolyse et d'une lampe torche occupent un poste déterminé, à l'intérieur ou à l'extérieur des habitations (Illustration 5). Une jambe de pantalon relevée, ils capturent tous les moustiques qui viennent les piquer grâce aux tubes à hémolyse qu'ils stockent ensuite dans de petits sacs en tissu portant mention de la tranche horaire durant laquelle le moustique a été capturé (18-19h, 19-20h, ..., 5-6h). Ces sacs sont collectés toutes les deux heures et apportés à la station de recherche où ils sont entreposés au réfrigérateur pour éviter le dessèchement des spécimens. En général, deux collecteurs se relaient sur le même poste de capture si la séance de capture dure toute la nuit. Les tenants et les aboutissants de ces séances sont explicitement présentés aux villageois dont l'accord préalable est requis. Cette méthode de capture permet de déterminer le taux de piqûre, le taux d'endophagie des espèces (préférence à piquer à l'intérieur des habitations) ainsi que leur cycle d'agressivité nocturne. C'est une méthode très pertinente, et pour le moment irremplaçable, pour quantifier le taux de transmission du paludisme. Le taux de parturité des femelles peut

Méthodes permettant la capture de spécimens vivants

Illustration 5: Volontaire capturant les femelles anthropophiles.

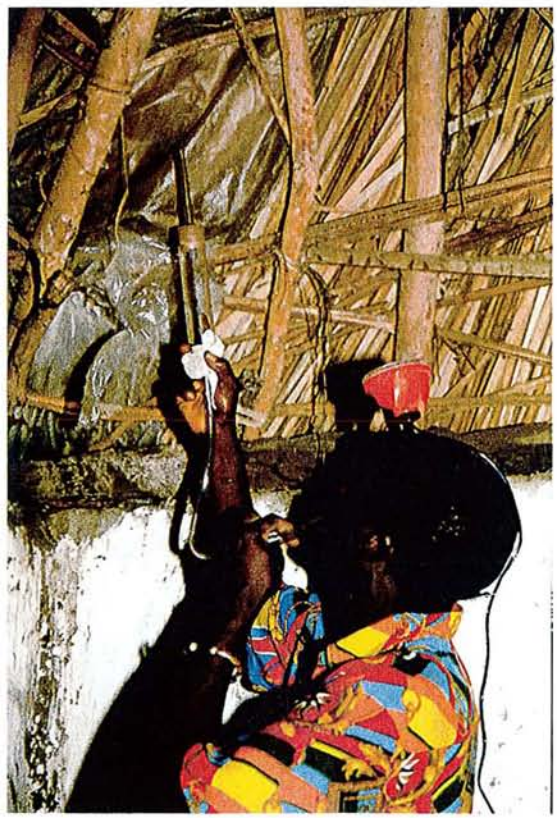


Illustration 6: Capture de la faune endophile grâce à un aspirateur.

Photos: Jean-Jacques Lemasson, IRD, Dakar.

également être déterminé par examen des ovaires selon la technique de Detinova (1962).

1. 3. - Capture à l'aspirateur à bouche

Cette technique a pour cible les moustiques au repos dans les habitations, les étables ou les gîtes extérieurs. Les moustiques sont repérés à la lampe torche et collectés grâce à un aspirateur à bouche muni d'un gobelet de collecte (Service, 1993; Illustration 6). L'avantage de cette méthode est que les moustiques sont collectés vivants et peuvent ensuite être utilisés pour initier un élevage en insectarium, pour des études de descendance ou comme matériel de départ pour les expériences de marquage-recapture. Le rendement est nettement plus faible que lors de collectes par pulvérisation intra-domiciliaire d'insecticide car les anophèles sont difficiles à repérer sur les murs ou les toits en paille.

Il existe de nombreuses autres méthodes de captures mais les techniques citées dans ce paragraphe sont les plus utilisées pour collecter les anophèles lors des études de la transmission du paludisme. Ce sont en outre les trois seules méthodes que nous ayons utilisées.

1. 4. - Identification morphologique

L'identification individuelle des spécimens collectés se fait sous la loupe binoculaire à l'aide des clés de détermination taxonomique générales basées sur des critères morphologiques (Gillies et de Meillon, 1968; Gillies et Coetzee, 1987; Hervy *et al.*, 1998). D'autres clés spécifiques à certaines zones géographiques sont également disponibles (voir par exemple Diagne *et al.*, 1994 pour les anophèles du Sénégal).

Les spécimens appartenant au complexe *An. gambiae* sont ensuite stockés individuellement dans des tubes de 1,5 ml contenant un dessiccateur (silicagel) et conservés au laboratoire à -20°C.

2. Détermination de l'espèce et du génotype

2. 1. - Identification spécifique

Comme nous l'avons déjà précisé dans l'introduction, il est impossible de déterminer l'espèce à laquelle appartient un individu du complexe *An. gambiae* sur des critères morphologiques. Le test basé sur la technique PCR décrit par Scott *et al.* (1993) est alors appliqué. Un cocktail d'amorces est utilisé pour l'amplification de l'ADN localisé dans les fragments intergéniques codant pour les sous-unités ribosomiques: une amorce correspondant à une séquence conservée chez tous les individus du complexe *An. gambiae* (amorce s'hybridant du côté 5' du fragment à amplifier) et plusieurs amorces dérivées de séquences uniques et spécifiques de chaque membre du complexe (amorces s'hybridant du côté 3' du fragment à amplifier) permettent l'amplification de fragments de taille différente selon l'espèce. La séquence des amorces utilisées ainsi que la taille des fragments diagnostics obtenus après amplification sont présentées dans l'Encadré 6.

La PCR est réalisée dans un volume total de 12,5 µl. L'ADN matrice amplifié provient d'une patte ou d'une aile du spécimen soit directement plongée dans le milieu réactionnel, soit après extraction selon le protocole décrit par Collins *et al.* (1987). Un exemple du profil de migration obtenu après électrophorèse sur gel d'agarose (1,5%) est également donné dans l'Encadré 6.

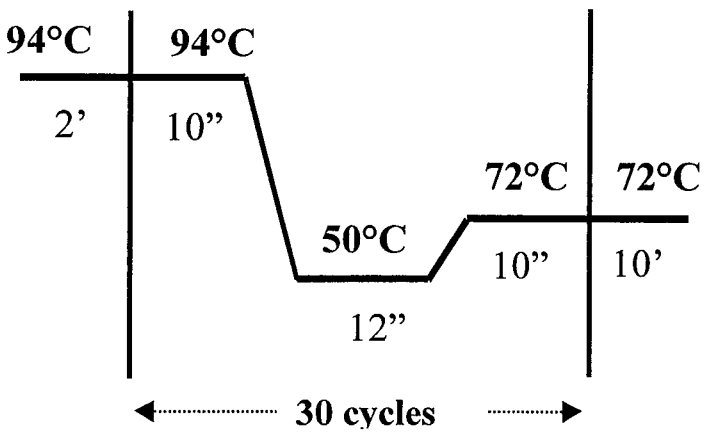
Encadré 6
Identification spécifique des membres du complexe *An. gambiae* par PCR

- Amorces utilisées:

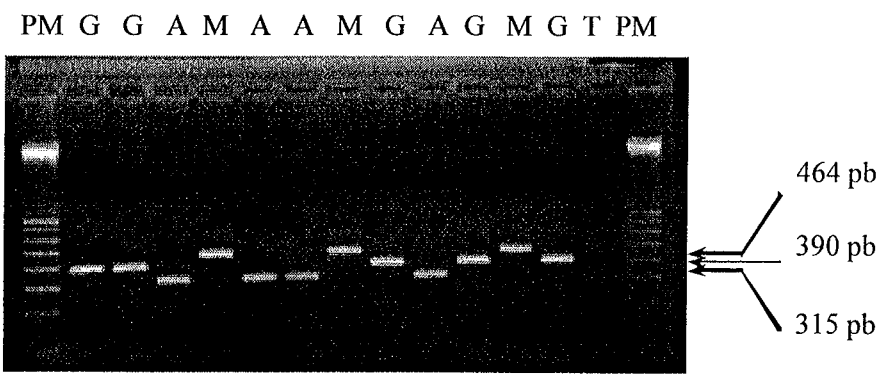
Spécificité	séquence	taille
Universelle	5' GTGTGCCCCTTCCTCGATGT 3'	-
An. arabiensis	5' AAGTGTCCTTCTCCATCCTA 3'	315 pb
An. gambiae	5' CTGGTTTGGTCGGCACGTTT 3'	390 pb
An. melas	5' TGACCAACCCACTCCCTTGA 3'	464 pb

d'après Scott *et al.*, 1993. pb=paires de bases

- Conditions de PCR:



- Exemple de profil de migration obtenu sur gel d'Agarose à 1,5%:



PM: Echelle de poids moléculaire (100 pb).
G: *An. gambiae*; A: *An. arabiensis*; M: *An. melas*; T: Témoin négatif.

2. 2. - Détermination du génotype aux loci microsatellites

Neuf loci microsatellites répartis sur les trois chromosomes ont été sélectionnés parmi les 131 loci isolés chez *An. gambiae* par Zheng *et al.* (1996) sur la base:

- d'une amplification effective chez *An. arabiensis*,
- d'un polymorphisme important détecté chez *An. gambiae* et *An. arabiensis* lors d'expériences préliminaires sur 10 individus issus de notre élevage,
- d'une taille des allèles observés comprise entre 100 et 200 paires de bases (pb) pour optimiser la précision de la lecture des bandes sur gel d'acrylamide à 10%.

Les caractéristiques de ces 9 loci sont présentées dans le Tableau 1. La PCR a été réalisée dans un volume total de 25 µl, ramené par la suite à 12,5 µl sans qu'une baisse de sensibilité ni de rendement d'amplification ne soit décelée. Dix microlitres de chaque amplifiat sont ensuite déposés sur gel d'acrylamide/bisacrylamide (19:1) à 10% et l'électrophorèse est effectuée en conditions non dénaturantes pendant 3 heures, à une intensité constante de 50 mA/gel. Les bandes correspondant aux allèles sont révélées par le protocole de coloration rapide au nitrate d'argent décrit par Cairns et Murray (1994). Cette méthodologie permet de ne pas utiliser de radioactivité, ce qui est un avantage dans les Pays en Voie de Développement qui ne possèdent pas toujours les structures de décontamination nécessaires. La taille exacte des allèles est déterminée grâce au dépôt sur le même gel d'une échelle de poids moléculaire réalisée par pool de plusieurs amplifiats représentant un large spectre de variation au locus étudié, dont la séquence d'un des allèles au moins a été déterminée.

Un exemple du profil de migration obtenu et de son interprétation pour 25 spécimens *An. arabiensis* au locus 803 est donné en Encadré 7.

Tableau 1: Caractéristiques des loci utilisés pour l'analyse de la variabilité génétique chez *An. arabiensis*

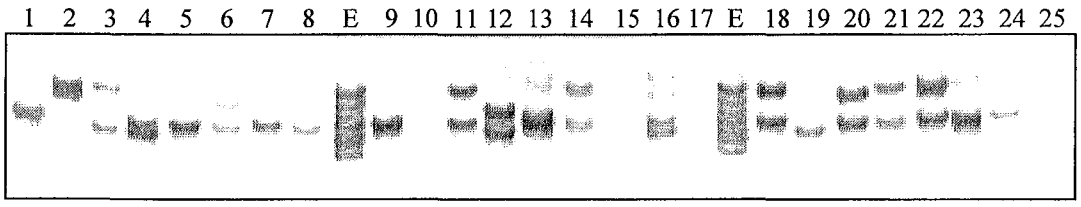
Locus	Localisation cytologique*	Motif répété	Amorces	Taille des allèles (pb) §
7	X : 1C	GT	5'CACGATGGTTTTCGGTGTGG 3' 5'ATTTGAGCTCTCCCGGGTG 3'	90-100
49	X : 1D	GT	5'CAGCGCCTCCATATAGAACG 3' 5'GATCATTCAGCTGAACCTGC 3'	87
24D	2 La : 24D	CT	5'GGCGAGCAGTTCATTCAAGT 3' 5'CGTCTGGAAGTTTCGTTGAG 3'	105 - 121
147	2 R : 19	GT	5'CTGCTGTTGCTGCCAAAATG 3' 5'AGCTTCACGGAAAGCAAAGG 3'	171-177
141	2 Rd	GT	5'CGGAGCAAATCTGAACCGTG 3' 5'CCTTGGCCACAACAACATCG 3'	102-112
26	2 Rb : 12	GT	5' GGTTCCTGTTACTTCCTGCC 3' 5'CCGGCAACACAAACAATCGG 3'	86-100
803	2 R : 7B	TG	5'CTCGATAAATCCCGTCGGTG 3' 5'GTCGGTTTGAGGTTGTAAAGC 3'	120-128
45C	3 L : 45C	GA	5'AAAAGTGGTGACCGAGTGAC 3' 5'ATCTTCAACACTTCAGCACG 3'	146-156
29C	3 R : 29C	TGA	5'ATGTTCCAGAGACGACCCAT 3' 5'TGTTGCCGGTTTGTTGCTGA 3'	145-148

D'après Zheng *et al.* (1993, 1996).
* Division et subdivision correspondant aux différents fragments chromosomiques selon la nomenclature de Zheng *et al.* (1996).
§ Allèles observés chez *An. arabiensis* lors d'essais préliminaires réalisés sur 10 moustiques d'élevage.

Encadré 7

Collecte des données pour l'analyse génétique

- Exemple de profil de migration obtenu après amplification de l'ADN au locus 803 et migration en conditions non dénaturantes sur gel vertical de polyacrylamide à 10%. Les bandes sont colorées au nitrate d'argent.



- Interprétation du gel:

E = Echelle de taille des différents allèles présents au locus 803.
(120-122-124-126-128-130-132 pb)

Spécimen:

N°1 : Homozygote 124-124 ;
N°2 : Hétérozygote 130-132 ;
N°3 : Hétérozygote 124-132 ;
N°4 : Hétérozygote 124-126 ;
N°5 : Homozygote 126-126 ;
N°6 : Hétérozygote 126-130 ;
N°7 : Homozygote 126-126 ;
N°8 : Homozygote 126-126 ;
N°9 : Homozygote 126-126 ;
N°10 : Négatif ;
N°11 : Hétérozygote 126-132 ;
N°12 : Hétérozygote 124-128 ;
N°13 : Homozygote 126-126 ;
N°14 : Hétérozygote 126-132 ;
N°15 : Négatif ;
N°16 : Hétérozygote 124-126 ;
N°17 : Négatif ;
N°18 : Hétérozygote 126-132 ;
N°19 : Homozygote 124-124 ;
N°20 : Hétérozygote 126-130 ;
N°21 : Hétérozygote 126-132 ;
N°22 : Hétérozygote 126-132 ;
N°23 : Homozygote 126-126 ;
N°24 : Homozygote 126-126 ;
N°25 : Témoin Négatif.

Les fréquences alléliques et génotypiques sont déterminées par simple «comptage des gènes».

3. Analyse des données

Les données ainsi collectées ont été analysées grâce aux logiciels informatiques Biosys1 (Swofford et Selander, 1989); GENEPOP v3.1 (Raymond et Rousset, 1995a) et ARLEQUIN v1.1 (Schneider *et al.*, 1997). Dans certains cas, la détermination de paramètres spécifiques non évalués par ces logiciels a été réalisée grâce à l'adaptation de programmes écrits en langage SAS (SAS Institute Inc., 1990) aimablement fournis par T. Lehmann.

Les principaux tests auxquels nous ferons référence tout au long de ce travail sont brièvement présentés ci-dessous.

3. 1. - Test de la conformité à l'équilibre d'Hardy-Weinberg

Ce test a été généralement effectué grâce au logiciel GENEPOP v3.1 (Raymond et Rousset, 1995a). Les fréquences génotypiques observées sont comparées à celles estimées à partir des fréquences alléliques sous l'hypothèse de l'équilibre d'Hardy-Weinberg par un test exact de Fisher (Haldane, 1954). Si le nombre d'allèles par locus est supérieur à 4, la méthode des chaînes de Markov (Guo et Thompson, 1992) est utilisée pour définir la probabilité exacte associée à H_0 : «union aléatoire des gamètes». Ces tests exacts sont plus appropriés que le test classique du χ^2 pour des loci hypervariables tels que les microsatellites présentant généralement un grand nombre d'allèles dont certains représentés en très faibles fréquences (Guo et Thompson, 1992; Chakraborty et Zhong, 1994; Rousset et Raymond, 1995). Un test global est également effectué sur l'ensemble des loci selon la méthode de Fisher, assumant l'indépendance des loci. Une valeur significative signifie que H_0 n'est pas respectée au moins une fois sur l'ensemble des données.

3. 2. - Test du déséquilibre de liaison entre paires de loci

L'hypothèse nulle testée est «le génotype à un locus est indépendant du génotype à un autre locus». Pour chaque paires de loci à l'intérieur de chaque population, un test exact de Fisher est effectué sur des tables de contingences génotypiques selon la méthode des chaînes de Markov (Guo et Thompson, 1992).

3. 3. - Test d'homogénéité des fréquences alléliques ou génotypiques

L'hypothèse nulle testée est ici «la distribution des allèles (ou des génotypes) est identique entre les populations». Le test utilisé compare des tables de contingence dans lesquelles les lignes représentent les populations et les colonnes les allèles (ou les génotypes). L'homogénéité de la distribution des fréquences alléliques est testée par un test exact de Fisher utilisant les chaînes de Markov (Raymond et Rousset, 1995a et b), et un test G (Goudet *et al.*, 1996) est utilisé pour tester l'homogénéité de la distribution des génotypes entre populations.

3. 4. - Les statistiques *F*

Les indices de fixation de Wright (*Fis*, *Fst* et *Fit*) sont estimés selon Weir et Cockerhamm (1984), soit pour chaque paires de populations, soit sur l'ensemble des populations, pour chaque locus indépendamment ainsi que sur l'ensemble des loci (Raymond et Rousset, 1995a). *Fis* est testé par le test exact de conformité à l'équilibre d'Hardy-Weinberg à l'intérieur de la population considérée et *Fst* est testé par le test exact d'homogénéité de la répartition des génotypes entre populations précédemment décrit (Goudet *et al.*, 1996). Alternativement, *Fst* a également été déterminé par une analyse de variance moléculaire (AMOVA, Excoffier *et al.*, 1992) et testé par permutation des génotypes entre populations (Schneider *et al.*, 1997). Une estimation de

l'index de migration (Nm) est dérivée de *Fst* selon la formule de Wright (1951) présentée en introduction.

3. 5. - Les statistiques *R*

Les indices de fixation basés sur l'analyse de variance de la taille des allèles (statistiques *R* de Slatkin, 1995) ont été estimés selon Kimmel *et al.* (1996). *Rst* a été testé dans ce cas par un test F réalisé sur l'analyse hiérarchique de variance du nombre de répétitions (nested ANOVA) incluant la population et les individus comme facteurs (voir annexe 1). Alternativement, *Rst* a également été estimé en utilisant la même méthode que pour l'estimation de *Fst* (AMOVA), l'analyse de variance étant conduite sur la taille des allèles (en nombre de répétitions du motif de base) au lieu des fréquences des différents allèles (Schneider *et al.*, 1997). Cette estimation est testée par permutation des génotypes entre populations et Nm déterminé à partir de *Rst* selon la formule de Slatkin (1995), analogue à celle de Wright (1951) concernant *Fst*.

D'autres tests ont utilisés au cours de ce travail et seront décrits au moment opportun.

PREMIERE PARTIE

Polymorphisme des marqueurs microsatellites chez *Anopheles arabiensis*: étude de la répartition spatiale de la variabilité génétique

Les modèles mathématiques décrits dans le chapitre introductif permettent d'étudier les facteurs agissant sur la structure génétique des populations en se basant sur la distribution géographique des gènes. Les travaux présentés en annexe 1 ont suggéré un flux de gènes important au sein du taxon *An. gambiae* cytotype Savane entre des populations séparées par plus de 6000 km, mais reliées par un habitat continu où le trafic lié aux activités humaines est intense. L'aire de répartition géographique d'*An. arabiensis* est plus vaste que celle d'*An. gambiae* (voir Figure 1) et les populations apparaissent beaucoup plus homogènes génétiquement sur tout le continent Africain (Coluzzi *et al.*, 1979; Besansky *et al.*, 1997). De plus, *An. gambiae* est généralement présenté comme une espèce hautement domestique, très anthropophile et très adaptée à l'environnement humain qui ne peut se développer qu'en très étroite association avec l'homme ce qui laisse supposer un habitat plus ou moins fragmenté pour cette espèce, en fonction des densités humaines. De telles contraintes semblent beaucoup moins affecter *An. arabiensis*, anthropophile préférentiel mais qui peut se nourrir alternativement sur d'autres vertébrés suivant leur disponibilité et qui a été trouvé dans des zones géographiques totalement sauvages telles que le parc national Kruger en Afrique du Sud (Braack *et al.*, 1994). Ces observations suggèrent une plasticité comportementale plus importante chez *An. arabiensis* qui pourrait se traduire par une plus forte homogénéité des populations sur tout le continent Africain.

Afin de tester le potentiel des marqueurs microsatellites isolés chez *An. gambiae* pour l'étude de la structure génétique des populations naturelles d'*An. arabiensis*, nous avons comparé quatre populations qui, *a priori*, représentent le cas le plus extrême d'isolement que l'on puisse trouver au sein de l'espèce. En effet, comme l'ont encore récemment souligné Goldstein *et al.* (1999), il apparaît particulièrement judicieux, pour

estimer la capacité des marqueurs moléculaires à décrire correctement la structure génétique de populations naturelles, d'étudier en premier lieu des populations dont les connaissances écologiques et historiques laissent *a priori* supposer une différenciation conséquente.

Dans cet esprit, des échantillons d'*An. arabiensis* prélevés au Sénégal, sur les hauts plateaux malgaches et dans deux îles de l'archipel des Mascareignes dans l'Océan Indien, La Réunion et l'île Maurice, ont été comparées. Ces quatre sites sont effectivement séparés par une distance géographique considérable qui inclut une barrière évidente à la migration des individus qu'est l'océan. De plus, ces populations d'*An. arabiensis* ont un historique bien connu et très différent d'un site à l'autre (voir ci-dessous).

Quatre questions seront finalement posées : (i) L'utilisation d'amorces hétérologues, décrites pour des loci isolés chez *An. gambiae* est-elle adaptée à l'étude de la structure génétique des populations d'*An. arabiensis*? (ii) Quel est le degré de variabilité génétique chez *An. arabiensis*? (iii) La répartition de la variabilité génétique entre populations est-elle homogène ou liée aux caractéristiques spécifiques des populations étudiées? (iv) Quelles en sont les conséquences en termes de différenciation génétique entre populations et de flux de gènes et quels facteurs peuvent être impliqués?

Les résultats de cette étude sont présentés en annexe 2:

SIMARD F, FONTENILLE D, LEHMANN T, GIROD R, BRUTUS L, GOPAUL R, DOURNON C et COLLINS FH. 1999. High amounts of genetic differentiation between populations of the malaria vector *Anopheles arabiensis* from West Africa and Eastern Outer Islands. *Am. J. Trop. Med. Hyg.* **60(6)**: 1000-1009.

1. Description des sites d'étude

La localisation géographique des différents sites d'étude où les moustiques ont été collectés est présentée sur la Figure 2.

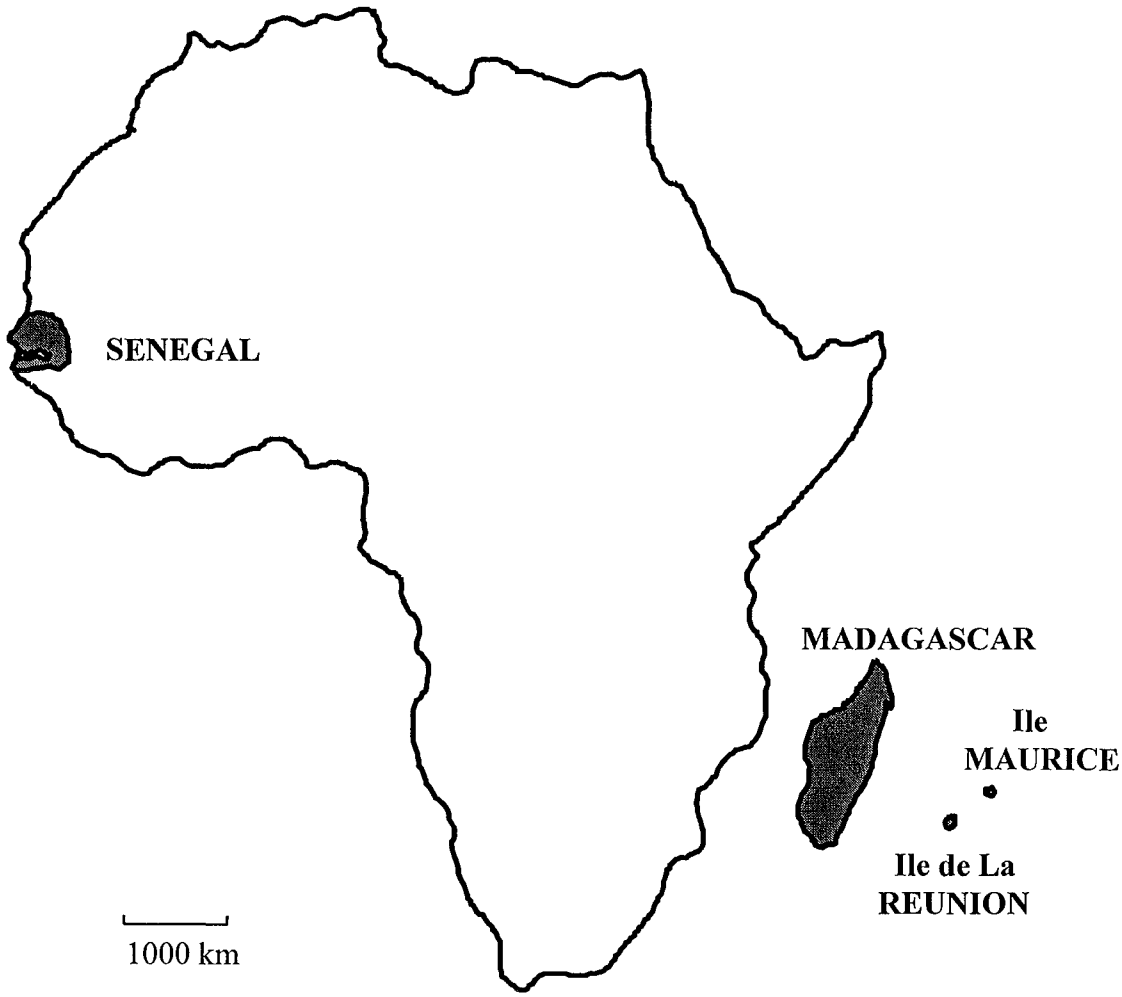
1. 1. - Sénégal

La seule population continentale étudiée lors de ce travail a été collectée durant le mois d'Août 1994 dans le village de BARKEDJI (15°17'N, 14°53'O), localisé dans la vallée du Ferlo en zone de savane sahélienne en plein centre du Sénégal. La saison des pluies est courte, s'étendant de Juillet à Octobre avec un total annuel de précipitations enregistré au village même de Barkedji variant entre 300 et 400 mm. *An. gambiae* et *An. arabiensis*, sympatriques dans cette région, assurent une transmission saisonnière du paludisme. Ils occupent des gîtes larvaires temporaires, uniquement présents en saison des pluies et qui s'assèchent généralement vers la fin du mois de Décembre. La densité de la population vectorielle est très faible voire nulle en saison sèche et augmente très rapidement après les premières pluies, dès que les gîtes sont en eau. Les insecticides sont très peu utilisés dans cette région et aucun programme de lutte anti-vectorielle n'a jamais été organisé. Une description détaillée du site et de l'écologie des vecteurs sera donnée dans la deuxième partie de ce rapport.

1. 2. - Madagascar

La population d'*An. arabiensis* des hauts plateaux malgaches a été échantillonnée dans le village de FENOARIVO, lors d'enquêtes entomologiques qui se sont déroulées en Novembre-Décembre 1996. Madagascar est l'île la plus vaste de l'Océan Indien (587 000 km^2), située à 350 km de la côte Est du continent Africain. Le

Figure 2: Localisation géographique des différentes populations d'*Anopheles arabiensis.*



village de FENOARIVO (19°47'S, 46°34'E), à une altitude de 1235 m, est entouré de rizières qui fournissent pléthore de gîtes aux deux vecteurs *An. arabiensis* et *An. funestus* (Fontenille, 1992; Laventure *et al.*, 1996). *An. gambiae*, abondant sur la côte, n'est pas présent sur les plateaux (Fontenille *et al.*, 1990). La saison des pluies dure de Novembre à Avril avec un total annuel de précipitations autour de 1400 mm.

La colonisation des zones d'altitude par les vecteurs est consécutive aux modifications anthropiques du milieu (terrassement, riziculture) réalisées à partir de la seconde moitié du XIXème siècle (Mouchet *et al.*, 1997). Les plateaux, contrairement aux zones côtières, sont restés indemne de paludisme jusqu'en 1878 lorsqu'éclata une épidémie meurtrière suivie de l'endémisation de la maladie jusqu'en 1949 (Mouchet *et al.*, 1997). Des aspersions domiciliaires de DTT ainsi que la distribution de chloroquine ont ensuite permis d'endiguer la maladie qui fut déclarée éradiquée en 1962 (Zahar, 1985; Mouchet *et al.*, 1997). Seuls quelques foyers sont restés sous surveillance jusqu'en 1975 avant l'abandon complet des programmes de lutte. La situation s'est alors progressivement dégradée jusqu'à l'apparition d'une nouvelle épidémie en 1986-1988, marquant le retour à un profil épidémiologique et entomologique identique à celui d'avant l'éradication. Cette épidémie a été en grande partie contrôlée grâce à l'accès à une prophylaxie de masse. Parallèlement, les aspersions de DTT ont repris et furent maintenues jusqu'en 1998, de manière localisée et discontinue. Aujourd'hui, *An. arabiensis* et *An. funestus* sont toujours présents sur les plateaux.

1. 3. - L'archipel des Mascareignes

L'île de La Réunion (2500 km^2) et l'île Maurice (2000 km^2) sont localisées dans l'Océan Indien, entre 20° et 25° de latitude Sud et 55° de longitude Est. Elles sont distantes de 240 km entre elles, et situées respectivement à 800 et 1000 km de la côte

Est de Madagascar. Le total annuel de précipitations est très variable sur chacune des îles, passant de 5000 mm dans les régions sous le vent à 500 mm dans les zones protégées. Ces deux îles sont souvent soumises au passage de cyclones durant la saison des pluies qui s'étend de Décembre à Avril. A La Réunion, les moustiques ont été collectés dans une seule localité située au nord-ouest de l'île, «Etang de Saint-Paul», en Février 1995. A Maurice, le peu de spécimens dont nous disposions nous a forcé à regrouper au sein du même échantillon des anophèles collectés sur toute la surface de l'île en Mars-Avril 1997.

A l'inverse de Madagascar, véritable sous-continent qui s'est détaché du continent Africain depuis quelques 100 millions d'années, emportant avec lui sa flore et sa faune qui se sont endémisées; les îles de l'archipel des Mascareignes sont «océaniques», d'origine volcanique, elles n'ont jamais eu de liaison terrestre avec le continent Africain, ni avec Madagascar, ni avec l'Asie. Par conséquent, toutes les espèces animales ou végétales peuplant ces îles résultent d'apports exogènes. Depuis un millénaire au moins, l'homme est à l'origine de l'importation, volontaire ou non, d'espèces qui ont bouleversé les milieux insulaires. L'action de l'homme se manifeste également par la création de biotopes favorables au développement et à la reproduction des espèces importées. C'est particulièrement le cas des Diptères hématophages dont les anophèles vecteurs du paludisme qui furent introduits à Maurice à la fin du XIX éme siècle, consécutivement à l'instauration d'une ligne de navigation à vapeur entre Madagascar et Maurice qui réduisit la durée du voyage à 3 jours (Julvez *et al.*, 1990; Julvez et Mouchet 1994a). La courte durée du voyage a permis aux anophèles, incapables de se reproduire sur les bateaux (à la différence des *Aedes* par exemple) de survivre du point d'embarquement jusqu'à l'arrivée. Il apparaît également plausible que le peuplement de La Réunion par *An. gambiae* s.l. ait eu lieu à partir de Maurice, par

migration assistée par les vents d'un cyclone (Barat, 1869; Julvez et Mouchet, 1994a et b). La datation de ces événements de colonisation est possible grâce aux archives de la Médecine Navale de la fin du siècle dernier qui reportent avec une grande précision l'apparition de nuisances dues aux Culicidae sur des îles jusqu'alors réputées pour leur salubrité, les dates précises des premières épidémies de paludisme et dressent un premier bilan des espèces d'arthropodes rencontrées (voir par exemple Bory de Saint-Vincent, 1804; Barat, 1869; Daruty de Grandpre et Emmerrez de Charmoy, 1900). La première épidémie eut lieu en 1866 à Maurice et en 1868 à La Réunion et le paludisme s'y est maintenu à l'état endémique pendant près d'un siècle (Julvez *et al.*, 1990). La maladie fut déclarée éradiquée à l'île Maurice en 1973 et à La Réunion en 1979, à la suite du programme mondial d'éradication du paludisme instauré au début des années 50 par l'OMS, basé sur la lutte antivectorielle grâce au DDT et sur une prophylaxie de masse à la chloroquine (Zahar, 1985). Mais, si très peu de cas de transmission autochtone ont été reportés depuis l'éradication officielle de la maladie, *An. arabiensis* est toujours présent sur ces deux îles et représente une menace pour le maintien de l'éradication. Depuis une vingtaine d'année, La Réunion et Maurice sont donc dans un état «d'anophélisme sans paludisme» qui impose une surveillance entomologique et épidémiologique particulière. La stratégie actuelle s'articule autour:

- de la lutte antilarvaire permanente qui tient compte du cycle biologique des anophèles et de la rémanence des produits larvicides utilisés (Téméphos en particulier), ciblée autour des zones prioritaires identifiées sur la base de prospections entomologiques régulières;

- du dépistage actif des cas de paludisme importé par prélèvement sanguin systématique des voyageurs en provenance de zones impaludées, avec intervention prophylactique et entomologique autour des cas positifs,

- de la démoustication également systématique par imagocides (pyréthrinoïdes de synthèse, fénitrothion) des moyens de communication (avions ou bateaux) et une surveillance étroite des aéroports et ports internationaux.

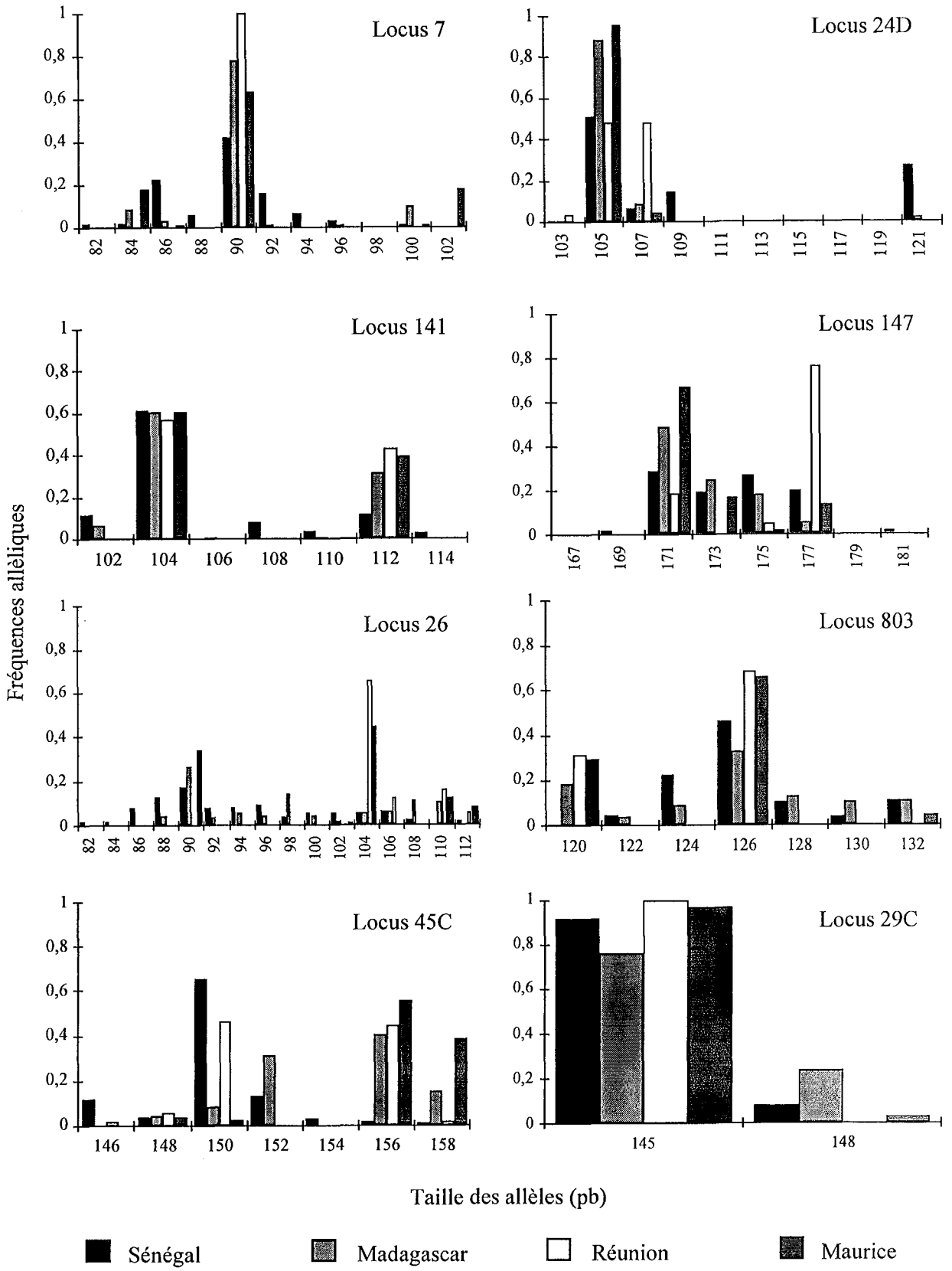
Les populations d'*An. arabiensis* peuplant ces sites ont donc des caractéristiques écologiques et une histoire bien spécifiques qu'il convient d'avoir en mémoire pour interpréter les résultats de l'analyse génétique. D'un point de vue historique, toutes les populations insulaires que nous avons citées sont logiquement issues d'une population souche ancestrale africaine.

2. Analyse génétique

2. 1. - Polymorphisme des marqueurs microsatellites chez *An. arabiensis*

Pour chaque locus, 54, 53, 44 et 52 individus respectivement collectés au Sénégal, à Madagascar, à Maurice et à La Réunion ont été testés. L'utilisation d'amorces hétérologues, déterminées chez *An. gambiae* pour étudier le polymorphisme génétique chez l'espèce jumelle *An. arabiensis* s'est révélée particulièrement adaptée. En effet, sur les neuf loci testés, huit ont présenté un polymorphisme important chez *An. arabiensis*, avec un nombre d'allèles distincts observés sur les 203 individus testés variant de 2 pour le locus 29C à 16 pour le locus 26. Ce résultat confirme que les deux espèces sont génétiquement très proches l'une de l'autre (Besansky *et al.*, 1994, 1997). La Figure 3 représente la composition allélique et la distribution des fréquences observées pour chaque locus polymorphe, au sein de chacune des 4 populations étudiées.

Figure 3: Variabilité génétique intra-population chez *An. arabiensis*



Le locus 49, localisé sur le chromosome X dans la division 1D n'a présenté qu'un seul allèle de 87 pb chez tous les individus testés. Il n'a donc pas été utilisé dans les analyses qui vont suivre. Ce locus n'est pas informatif pour étudier les populations d'*An. arabiensis* bien qu'il soit très polymorphe chez *An. gambiae* (voir troisième partie). Une baisse du nombre de loci polymorphes ainsi que du taux de polymorphisme a souvent été rapportée lorsque des amorces isolées chez une espèce sont utilisées chez une espèce voisine (Irwin *et al.*, 1991; Primmer *et al.*, 1996a). Chez *An. gambiae* s.l., le chromosome X est le siège d'inversions qui permettent le diagnostic entre les différentes espèces du complexe: l'inversion X^{ag}, qui englobe la presque totalité du chromosome (75% du bras chromosomique, selon Green, 1980) est fixée chez *An. gambiae* et l'inversion X^{bcd} est fixée chez *An. arabiensis* (Coluzzi *et al.*, 1979). La perte de polymorphisme observée chez *An. arabiensis* pourrait donc résulter d'une conformation différente de l'ADN à ce niveau, consécutive au phénomène de spéciation. Par contre, le locus 7 situé dans la division 1C, à moins d'1 centiMorgan du locus 49 s'est révélé très polymorphe chez les populations étudiées sauf à La Réunion (voir ci-dessous) indiquant que le monomorphisme général observé au locus 49 résulte de contraintes certainement structurales spécifiques à ce locus chez *An. arabiensis* et non de phénomènes d'auto-stop liés à la sélection d'allèles avantageux ou à l'élimination d'allèles délétères à un gène adjacent soumis à sélection (voir Hedrick, 1980).

2. 2. - Variabilité génétique au sein des populations

Les paramètres classiquement utilisés pour synthétiser l'information fournie par les marqueurs génétiques sont donnés dans le Tableau 2 et la Figure 3 pour les huit loci polymorphes. Deux résultats majeurs apparaissent.

Tableau 2: Variabilité génétique intra-population.

Population	Locus (n)	Nall	Ho	He	Fis	P
Sénégal						
	7 (53)	9	0,604	0,749	0,195	0,022
	24D (54)	5	0,481	0,648	**0,259**	**0,0056**
	147 (54)	7	0,667	0,773	0,138	0,276
	141 (54)	6	0,519	0,594	0,128	0,299
	26 (54)	15	0,833	0,915	0,090	0,261
	803 (53)	7	0,660	0,714	0,076	0,625
	45C (54)	7	0,537	0,539	0,004	0,903
	29C (54)	2	0,167	0,154	-0,082	1
	Tous loci (53,8)†	7,3	0,558	0,636	0,122	0,011
Madagascar						
	7 (53)	6	0,396	0,388	-0,021	0,361
	24D (53)	4	0,226	0,208	-0,09	1
	147 (52)	6	0,596	0,667	**0,107**	**0,0025**
	141 (53)	5	0,509	0,539	0,055	0,330
	26 (53)	12	0,887	0,876	-0,013	0,051
	803 (53)	7	0,736	0,812	0,094	0,043
	45C (53)	5	0,755	0,713	-0,059	0,067
	29C (53)	2	0,283	0,364	0,224	0,129
	Tous loci (52,9)†	5,9	0,549	0,571	**0,039**	**0,0015**
Maurice						
	7 (43)	5	0,512	0,551	0,072	0,585
	24D (44)	2	0,091	0,088	-0,036	1
	147 (44)	4	0,455	0,508	0,106	0,556
	141 (43)	2	0,558	0,484	-0,156	0,352
	26 (44)	5	0,591	0,673	0,123	0,354
	803 (44)	3	0,545	0,482	-0,134	0,420
	45C (44)	4	0,432	0,545	0,210	0,153
	29C (44)	2	0,068	0,067	-0,024	1
	Tous loci (43,8)†	3,4	0,406	0,425	0,106	0,556
La Réunion						
	7 (52)	1	-	-	-	-
	24D (52)	3	0,462	0,541	0,149	0,038
	147 (51)	3	0,373	0,382	0,025	1
	141 (52)	2	0,596	0,496	-0,205	0,165
	26 (52)	4	0,577	0,532	-0,085	0,689
	803 (52)	2	0,519	0,593	-0,189	0,211
	45C (52)	5	0,538	0,593	0,093	0,076
	29C (52)	1	-	-	-	-
	Tous loci (51,9)†	2,6	0,383	0,373	-0,028	0,0848
Population Totale	7 (201)	10	0,373	0,488	**0,236**	**$<10^{-4}$**
	24D (203)	5	0,325	0,478	**0,320**	**$<10^{-4}$**
	147 (201)	8	0,527	0,715	**0,263**	**$<10^{-4}$**
	141 (202)	7	0,545	0,547	0,004	0,0156
	26 (203)	16	0,729	0,854	**0,147**	**$<10^{-4}$**
	803 (202)	7	0,619	0,667	**0,072**	**$<10^{-4}$**
	45C (203)	7	0,571	0,747	**0,235**	**$<10^{-4}$**
	29C (203)	2	0,133	0,166	**0,199**	**0,0024**
	Tous loci (202,3)†	7,8	0,478	0,583	**0,214**	**$<10^{-4}$**

(n) : Effectif de l'échantillon (nombre d'individus diploïdes dont le génotype a été déterminé).
Ho: Hétérozygotie observée ; He: hétérozygotie calculée (Nei, 1978).
P = Probabilité de conformité à l'équilibre d'Hardy-Weinberg (test exact de Fisher).
- = non calculé (locus monomorphe).
† = moyenne sur tous les loci. P est la probabilité du test combiné de Fisher.
Valeurs soulignées = P<0,05 ; Caractères en **gras souligné**: P<0,05 après application de la procédure de Bonferroni.

(i) - Des populations panmictiques et individualisées

L'analyse de conformité par rapport aux proportions attendues sous l'hypothèse de l'équilibre d'Hardy-Weinberg (Tableau 2) a montré un très fort déficit en hétérozygotes (Fis positive et très significative) à tous les loci lorsque les différentes populations prélevées sont considérées comme appartenant au même pool de gènes, suggérant un important effet Wahlund (Tableau 2, «Population Totale»). En effet, lorsque les tests sont réalisés sur chaque population indépendamment, l'équilibre d'Hardy-Weinberg est respecté dans la plus grande majorité des cas. Seules deux déviations sur 32 tests réalisés sont néanmoins restées significatives après que la procédure de Bonferroni (Holm, 1979; Rice, 1989) ait été appliquée pour prendre en compte les différents tests effectués: au locus 24 D dans la population sénégalaise et au locus 147 dans la population malgache. Ces deux déviations décelées au niveau d'un locus particulier dans une seule population représentent probablement des variations stochastiques car elles ne sont pas soutenues par le reste des données, tant au niveau de la population que du locus considérés. Elles pourraient néanmoins indiquer la présence d'allèles nuls à ces loci particuliers.

L'effet Wahlund détecté lors du test effectué sur les données groupées a été confirmé par des analyses du déséquilibre de liaison entre paires de loci. Un fort taux de déséquilibre de liaison a été révélé lorsque le test exact de Fisher a été effectué sur les données groupées, indiquant une répartition non uniforme des génotypes multi-loci: sur 28 combinaisons possibles, 24 tests ont été significatifs avec des valeurs de probabilité associée variant de $P<10^{-6}$ à $P=0,0045$. Par contre, aucune indication de ce type n'a été trouvée lorsque les tests ont été effectués sur les populations individualisées: seuls 5 tests sur les 112 comparaisons possibles se sont révélés significatifs (avec $0,0046 \leq P \leq 0,0388$), impliquant des paires de loci différentes suivant les populations.

Aucune liaison entre loci n'a donc été mise en évidence vérifiant ainsi l'indépendance génétique des marqueurs utilisés.

Ces analyses de la répartition des fréquences alléliques et génotypiques ont permis de démontrer que la panmixie est respectée au sein de chacune de ces 4 populations mais qu'elles représentent des entités distinctes entre lesquelles les gènes ne circulent pas librement.

(ii) - Les différentes populations testées présentent une variabilité génétique hétérogène (Tableau 2, Figure 3)

Les loci 29C et 7 n'ont présenté qu'un seul allèle à La Réunion alors qu'ils sont polymorphes dans les trois autres populations. Le nombre moyen d'allèles par locus, observé sur ces échantillons de taille comparable ($46<n<54$), varie significativement entre populations (Test de Kruskal-Wallis, $\chi^2=14,45$; ddl=3; P=0,0024) et peut être classé dans un ordre décroissant selon la hiérarchie: Sénégal ≥ Madagascar ≥ Maurice ≥ La Réunion. Une tendance similaire a été observée sur l'hétérozygotie moyenne calculée à partir des fréquences alléliques sous l'hypothèse vérifiée de panmixie à l'intérieur de chaque population, bien que les différences ne soient pas significatives ($\chi^2=7,63$; ddl=3; P=0,0543). La distribution des allèles entre les différentes populations présentée sur la Figure 3 fait également ressortir un profil particulier: sur un total de 62 allèles distincts observés sur les huit loci, aucun allèle privé n'a été trouvé à La Réunion, 1 allèle est spécifique de la population mauricienne (allèle 102 au locus 7), 3 sont uniques à Madagascar et 11 n'ont été observés que dans la population sénégalaise. La population sénégalaise en particulier présente un profil de variabilité génétique plus large que les autres (Figure 3). L'allèle majoritaire est le même pour les 4 populations sur 5 des 8 loci

testés. Ce profil de répartition de la variabilité génétique est compatible avec l'historique de la colonisation des îles à partir d'une population souche africaine.

2. 3. - Différenciation génétique entre populations et flux de gènes

L'analyse de la structure génétique des populations a été effectuée par l'estimation des paramètres Fst (Wright, 1951; estimé selon Weir et Cockerhamm, 1984) et Rst (Slatkin 1995; estimé selon Kimmel *et al.*, 1996). Les résultats obtenus sont présentés dans le Tableau 3.

Un très fort taux de différenciation génétique a été observé entre les quatre populations testées pour chacun des loci microsatellites considérés ainsi que sur l'ensemble des loci avec une valeur de Fst moyenne de 0,169 très significative ($P<10^{-4}$) et une valeur de Rst légèrement supérieure, de 0,213. Les populations ont également été comparées deux à deux: dans la majorité des cas, les estimations de Rst ainsi obtenues ont été plus faibles que les estimations de Fst, exception faite de la comparaison entre le Sénégal et l'île de La Réunion (Tableau 3, moyenne sur tous les loci). Les taux de différenciation les plus élevés ont été observés entre La Réunion et chacune des autres populations ($0,206 \leq Fst \leq 0,215$; $0,132 \leq Rst \leq 0,302$) alors que les populations de Madagascar et Maurice apparaissent comme les plus proches génétiquement (Fst=0,08; Rst=0,022). Ces forts taux de différenciation se traduisent par des estimations de l'index de migration (Nm) très faibles et supérieures à 1 uniquement entre le Sénégal et Madagascar ($1,19 \leq Nm \leq 2,04$, selon la méthode de calcul) et entre Madagascar et l'île Maurice ($0,87 \leq Nm \leq 5,61$). Les trois méthodes utilisées pour estimer ce paramètre, à partir de l'estimation de Fst (Wright, 1951), de Rst (Slatkin, 1995) ou selon la méthode des allèles privés de Slatkin (1985b) ont donné des résultats homogènes sur l'ensemble des comparaisons effectuées.

Tableau 3: Différenciation génétique entre populations.

Locus		Sénégal x Madagascar	Sénégal x Réunion	Sénégal x Maurice	Madagascar x Réunion	Madagascar x Maurice	Réunion x Maurice	Global sur toutes les populations
7	Fst (Nm)	**0,148** (1,91)	**0,357** (0,60)	**0,111** (2,7)	**0,142** (2,0)	**0,058** (5,4)	**0,280** (0,86)	**0,183** (1,49)
	Rst (Nm)	0,010 (16,5)	0,005 (33,2)	0,025 (6,5)	0,001 (166)	0,0007 (238)	0,016 (10,3)	0,016 (15,2)
24D	Fst (Nm)	**0,202** (0,99)	**0,176** (1,17)	**0,262** (0,70)	**0,297** (0,59)	0,012 (21,1)	**0,379** (0,41)	**0,243** (0,78)
	Rst (Nm)	**0,162** (0,65)	**0,144** (0,74)	**0,184** (0,55)	0,001 (12,8)	0,006 (20,0)	**0,162** (0,65)	**0,218** (0,90)
147	Fst (Nm)	0,039 (6,1)	**0,255** (0,73)	**0,131** (1,65)	**0,386** (0,40)	0,046 (5,2)	**0,424** (0,34)	**0,222** (0,88)
	Rst (Nm)	**0,053** (2,2)	**0,136** (0,79)	**0,090** (1,26)	**0,340** (0,24)	0,004 (28,3)	**0,382** (0,40)	**0,258** (0,72)
141	Fst (Nm)	**0,031** (7,8)	**0,094** (2,4)	**0,074** (3,1)	0,011 (22,7)	0,001 (227)	0,0000 (>>1)	**0,038** (6,23)
	Rst (Nm)	0,010 (12,4)	**0,059** (2,0)	**0,042** (2,9)	0,012 (10,0)	0,003 (35,6)	0,0000 (>>1)	**0,031** (7,90)
26	Fst (Nm)	**0,022** (11,1)	**0,241** (0,79)	**0,128** (1,70)	**0,249** (0,75)	**0,110** (2,0)	**0,122** (1,80)	**0,150** (1,42)
	Rst (Nm)	**0,060** (1,96)	**0,517** (0,12)	**0,137** (0,79)	**0,295** (0,30)	0,016 (7,6)	**0,157** (0,67)	**0,275** (0,66)
803	Fst (Nm)	**0,036** (6,8)	**0,155** (1,37)	**0,128** (1,71)	**0,126** (1,74)	**0,101** (2,2)	0,0000 (>>1)	**0,095** (2,37)
	Rst (Nm)	0,0001 (>>1)	**0,141** (0,76)	**0,086** (1,33)	**0,064** (1,83)	**0,032** (3,7)	0,0000 (>>1)	**0,076** (3,00)
45C	Fst (Nm)	**0,300** (0,58)	**0,172** (1,2)	**0,441** (0,32)	**0,158** (1,33)	**0,115** (1,92)	**0,223** (0,87)	**0,246** (1,83)
	Rst (Nm)	**0,406** (0,18)	**0,181** (0,57)	**0,696** (0,05)	**0,055** (2,2)	**0,150** (0,71)	**0,302** (0,29)	**0,421** (0,34)
29C	Fst (Nm)	0,074 (3,1)	0,074 (3,1)	0,011 (21,6)	**0,225** (0,86)	0,141 (1,52)	0,027 (9,0)	**0,120** (1,83)
	Rst (Nm)	**0,039** (3,1)	**0,038** (3,2)	0,005 (24,9)	**0,128** (0,85)	**0,077** (1,50)	0,014 (9,1)	**0,094** (2,42)
Tous loci	Fst †(Nm)	**0,109** (2,04)	**0,206** (0,96)	**0,188** (1,08)	**0,214** (0,92)	**0,080** (2,87)	**0,215** (0,91)	**0,169** (1,23)
	Rst † (Nm)	**0,095** (1,19)	**0,302** (0,29)	**0,169** (0,61)	**0,174** (0,59)	0,022 (5,61)	**0,132** (0,82)	**0,213** (0,69)
	Nm (PA)	1,35	0,56	0,69	0,59	0,87	0,50	1,33

Valeurs soulignées: P< 0,05; caractères **gras soulignés**: P< 0,05 après application de la procédure de Bonferroni.

† Estimations testées par permutation des génotypes entre populations.

Nm (PA): estimations de Nm selon la méthode des allèles privés de Slatkin (1985).

Ces résultats mettent donc en évidence une structure génétique forte entre ces différentes populations très isolées géographiquement et soulignent la capacité de ce type de marqueurs à détecter des différences dans la répartition de la variabilité génétique entre sous-populations au sein de la même espèce.

3. Structure génétique des populations d'*An. arabiensis*: le rôle des facteurs écologiques et historiques

Les résultats obtenus peuvent être analysés à la lumière des connaissances que nous avons sur l'écologie et l'historique des populations dans ces différents sites géographiques.

3. 1. - Perte de polymorphisme et contraintes sur la taille des populations: influence de la dérive génétique sur les populations locales

L'un des paramètres principaux influençant le taux de polymorphisme neutre dans une population en panmixie est la taille effective de la population. En effet, comme nous l'avons vu, dans des populations de grande taille, de nombreux allèles pourront être générés par mutation et l'influence restreinte de la dérive génétique permettra un polymorphisme important, ce qui est observé pour la population sénégalaise et, dans une moindre mesure, pour la population malgache. Par contre, dans des populations de petite taille, une très forte dérive génétique va empêcher le développement de forts taux de polymorphisme en éliminant les allèles de faible fréquence conduisant à une variabilité génétique réduite telle que celle caractérisant les populations des petites îles de l'archipel des Mascareignes. De plus, les populations insulaires ont, au cours de leur histoire, subi des goulots d'étranglement démographiques importants consécutifs à

l'établissement de la population originale dans ces sites (effets fondateurs évidents sur les îles des Mascareignes) et à la lutte intensive menée dans les années 50 qui a considérablement réduit les densités de population à Madagascar, Maurice et La Réunion. Les goulots d'étranglement sont à l'origine d'une réduction de l'hétérozygotie et du nombre d'allèles sous différents modèles de mutation (IAM: Nei *et al.*, 1975; Maruyama et Fuerst, 1985a; SMM: Chakraborty et Nei, 1977; Cornuet et Luikart, 1996). La perte de polymorphisme associée à ces phénomènes est fonction de l'amplitude du goulot d'étranglement, du taux de croissance de la population résiduelle, du taux de mutation des marqueurs considérés et du temps (nombre de générations) écoulé depuis la réduction de taille. La diminution du nombre d'allèles est rapide et surtout fonction de la taille du goulot d'étranglement car la dérive génétique va éliminer un grand nombre d'allèles dans ces conditions alors que la chute de l'hétérozygotie est plus lente et fortement influencée par le taux de croissance de la population après l'évènement. Un fort taux de croissance, associé à un fort taux de mutation des marqueurs, va rapidement restaurer une variabilité génétique importante ce qui ne sera pas le cas si la population est maintenue à une taille réduite. Lorsque l'équilibre entre mutation et dérive génétique est de nouveau atteint, l'hétérozygotie et le nombre d'allèles sont stabilisés à une valeur qui sera fonction de la taille effective de la population et du taux de mutation des loci.

Or, le taux de croissance des populations d'anophèles peut raisonnablement être estimé comme fort si l'on se base sur le taux de fécondité des femelles (voir cycle biologique des anophèles en introduction) et les exemples de dynamique des populations naturelles en conditions favorables: explosions démographiques en saison des pluies en Afrique continentale (Gillies et De Meillon, 1968; Service, 1993), la colonisation rapide du Brésil par *An. gambiae* dans les années 1930-40 (Soper et Wilson, 1943) ou la célérité avec laquelle les sites des hauts plateaux malgaches ont été

recolonisés après l'arrêt des pulvérisations de DDT (Mouchet *et al.*, 1997). Le taux de mutation des loci microsatellites étant par ailleurs très élevé, il est probable que le retour aux conditions d'équilibre soit relativement rapide dans un tel système.

Le degré variable de polymorphisme observé sur les mêmes marqueurs suggère donc des différences de taille effective des populations étudiées:

- une variabilité génétique importante au Sénégal implique des populations de grande taille sur le continent Africain, conformément aux résultats obtenus avec d'autres marqueurs (inversions chromosomiques, Coluzzi *et al.*, 1985; isoenzymes, Smits *et al.*, 1996; séquences d'ADN, Besansky *et al.*, 1997),
- des caractéristiques similaires s'appliquent à la population des hauts plateaux malgaches soumise à une faible pression insecticide due à l'irrégularité des traitements focalisés sur des zones particulières,
- une variabilité génétique réduite sur les petites îles des Mascareignes dont la population est constamment soumise à un très fort contrôle suggère une taille effective réduite.

La perte complète de polymorphisme aux loci 7 et 29C dans la population de La Réunion soutient cette hypothèse de taille réduite des populations insulaires. En effet, le locus 29C est peu polymorphe dans les autres populations (seulement 2 allèles) et l'allèle de plus faible fréquence a pu être perdu sous l'influence unique de la dérive génétique dans cette population sans cesse contenue. Le même raisonnement peut être tenu à propos du monomorphisme observé au locus 7. Ce locus est situé sur le chromosome X, hétérochromosome qui n'est représenté qu'en une seule copie chez les individus mâles. La taille effective de la population pour ce chromosome particulier est

donc diminuée de 25% par rapport à celle estimée pour les chromosomes homologues et la recombinaison est impossible chez les mâles. La localisation cytologique du locus 7 l'expose ainsi à une plus forte influence de la dérive génétique et la fixation de l'allèle majoritaire sera donc plus rapide. Il est impossible de séparer, à partir des données que nous possédons, l'effet relatif des goulots d'étranglement dus aux effets fondateurs initiaux et à la limitation de la densité des populations consécutive à l'utilisation d'insecticides mais ces évènements mettent en jeu la même force évolutive: la dérive génétique. Néanmoins, il apparaît plausible au vu du fort taux de croissance des populations d'anophèles et du fort taux de mutation des loci microsatellites que les différences observées en termes de variabilité génétique (nombre d'allèles par locus et hétérozygotie) soient essentiellement imputables au contrôle assidu et durable de la densité des populations sur les petites îles. Le fait que l'allèle majoritaire soit le même au sein de chacune des populations pour 5 des 8 loci utilisés est également en faveur d'une influence prépondérante de la dérive génétique par rapport aux mutations d'une part car ces populations ne possèdent pas ou très peu d'allèles privés, ou à des phénomènes d'auto-stop moléculaire d'autre part. En effet, malgré des conditions écologiques et un historique similaires, les populations de La Réunion et de l'île Maurice présentent un profil de variabilité génétique différent compatible avec une évolution essentiellement stochastique plutôt que dirigée par des forces adaptatives. De plus, aucune résistance aux insecticides utilisés n'a été signalée chez *An. arabiensis* à Maurice comme à La Réunion suggérant une bonne gestion de la pression exercée qui n'entraîne pas la sélection de génotypes particuliers tout en limitant la densité des populations. Cette absence de résistance sur les îles soutient l'hypothèse de flux de gènes limités avec les populations continentales.

3. 2. - Différenciation entre sous-populations: dérive génétique et flux de gènes

Comme nous l'avons déjà précédemment remarqué, le taux de différenciation génétique entre populations sera fonction de l'influence relative de la dérive génétique comme force visant à individualiser les populations locales et de la migration (ou flux de gènes) qui tend à homogénéiser les populations. Dans le cas étudié ici, de très faibles valeurs de Nm ont été observées suggérant des flux de gènes limités entre les différents isolats géographiques. Le fait que les valeurs Rst soient généralement plus faibles que les valeurs de Fst estimées au cours de ce travail plaide en faveur d'une différenciation essentiellement générée par dérive génétique plutôt que par mutation (Slatkin, 1995). La forte différenciation génétique observée est le résultat de la combinaison de plusieurs facteurs, favorisant la dérive génétique au détriment les flux de gènes qui sont notamment :

(i) - Des effets fondateurs initiaux suivis de l'expansion des populations sur les îles

L'établissement des populations d'*An. arabiensis* sur les îles implique nécessairement un effet fondateur initial dont l'amplitude peut être plus ou moins importante selon le mécanisme de colonisation impliqué (vicariance ou colonisation de sites vierges par migration d'un nombre restreint d'individus). Si, suite à cet effet fondateur les populations n'échangent plus de gènes, les différents pools de gènes ainsi créés vont évoluer indépendamment les uns des autres et diverger par le seul fait des forces stochastiques. Or, si la diversité génétique à l'intérieur de chacune de ces nouvelles populations peut être restaurée rapidement comme nous l'avons dit précédemment, estompant ainsi l'influence de ces effets fondateurs sur la diversité génétique intra-population, il n'en est pas de même lorsque ces populations sont comparées les unes aux autres. L'évolution indépendante de ces dèmes par mutation et

dérive génétique va créer des différences dans la nature et les fréquences des allèles représentés au sein de chaque population, permettant ainsi la mise en évidence d'une différenciation génétique significative. En l'absence de sélection ou d'autres types de contraintes (sur la taille des allèles par exemple) qui pourraient diriger l'évolution des différents pools de gènes de façon uniforme, l'influence des effets fondateurs initiaux sera donc primordiale à ce niveau. Seuls des flux de gènes importants (Nm>1) pourront alors favoriser l'homogénéisation de la composition allélique entre populations.

(ii) - Une distance géographique importante incluant une barrière physique restreignant la migration par dispersion

L'Océan Indien apparaît comme une barrière évidente à la migration active des individus donc au flux de gènes entre populations. L'isolement génétique particulier de la population de La Réunion par rapport aux trois autres sites et le fait que les populations de Madagascar et de l'île Maurice apparaissent comme les moins différenciées alors qu'elles sont séparées par la plus grande distance géographique au dessus de l'océan (1000km) suggèrent néanmoins que cette barrière physique n'explique pas complètement le profil révélé. Une telle absence de corrélation entre le taux de différenciation génétique et la distance géographique entre les populations a été observée lors d'une étude récente conduite sur *An. polynesiensis* (Failloux *et al.*, 1997). Les auteurs ont par contre trouvé une corrélation très significative entre les flux de gènes (Nm) et le niveau de trafic commercial entre les différents îlots étudiés suggérant que le transport passif des moustiques puisse représenter une forme de migration non négligeable. De nombreux exemples de colonisation impliquant le transport passif des moustiques ont été cités dans un revue récente de Service (1997) et une preuve évidente de l'importance de ce moyen de dispersion à grande échelle des vecteurs du paludisme

réside en particulier dans la description de plusieurs cas de «paludisme d'aéroport» en Europe ou en Israël (Holvoet *et al.*, 1983; Blazquez, 1987; Rosci *et al.*, 1987; Shpilberg *et al.*, 1988; Service, 1997). En effet, c'est par ce biais que les îles étudiées ici ont été colonisées et le trafic commercial et touristique est aujourd'hui intense à travers l'Océan Indien. Les index de migration supérieurs à 1 observés entre le Sénégal et Madagascar, et entre Madagascar et Maurice confirment qu'une large zone géographique pourtant très inhospitalière ne soit pas un obstacle insurmontable à la migration d'*An. arabiensis*, bien que son influence soit incontestable. De plus, la population de Madagascar ne présente pas les caractéristiques d'endémisation auxquelles on aurait pu s'attendre si la fragmentation des populations avait été due exclusivement à un évènement de vicariance lors de la séparation de «la Grande Ile» et du continent Africain ce qui confirme l'existence de flux de gènes contemporains. En outre, comme nous l'avons cité plus haut, l'hypothèse d'une migration anémochore exceptionnelle ne peut pas être complètement écartée (Julvez *et al.*, 1998; Julvez et Mouchet, 1994a et b; Service, 1997). De tels évènements exceptionnels de migration en masse peuvent en effet avoir une influence prépondérante sur la structure génétique des populations (Slatkin, 1987). L'application du modèle d'isolement par la distance n'apparaît donc pas adéquate pour l'étude d'un tel système car il ne rend pas bien compte de ces phénomènes de migration assistée des individus, qui sont par ailleurs difficilement généralisables et paramétrables.

(iii) - Les traitements insecticides

L'influence des traitements insecticides appliqués de façon différentielle sur les populations étudiées peut s'exercer à plusieurs niveaux. Nous avons vu précédemment que l'emploi intensif des insecticides aboutit à la limitation de la densité donc de la taille effective des populations exposées au contrôle. Cette hétérogénéité des tailles des

populations comparées aura une influence très importante sur l'estimation du taux de flux géniques, inhérente aux modèles mathématiques utilisés (voir ci-dessous). A un autre niveau, sur l'île de La Réunion par exemple, une très forte pression insecticide est constamment appliquée aux points d'entrée potentiels des vecteurs sur l'île (ports et aéroports) résultant en un contrôle effectif de ces moyens de dispersion passive des moustiques qui, indépendamment de la limitation sur la taille des populations locales, va jouer en faveur de la différenciation des populations sous l'action de la dérive génétique en empêchant l'apport de nouveaux allèles par immigration.

Tous ces facteurs agissent simultanément et touchent tous les loci dans les mêmes proportions. Il est impossible d'estimer précisément dans quelle mesure chacun de ces mécanismes contribue au profil observé avec les modèles mathématiques dont nous disposons actuellement (voir Slatkin, 1987). Mais le fait est clair que tous ces facteurs augmentent l'influence de la dérive génétique et s'opposent au libre échange de gènes entre populations, favorisant ainsi l'évolution indépendante des différents dèmes.

4. Retour sur les conditions d'application des modèles

La dérive génétique apparaît donc comme la force majeure à l'origine de la structure génétique observée entre les populations d'*An. arabiensis* du Sénégal, de Madagascar, de l'île Maurice et de l'île de La Réunion. Mais, de fortes indications suggèrent que la taille effective des populations diffère en fonction de la population considérée. Or, dans un modèle en îles ou de stepping stone, toutes les sous-populations sont supposées avoir la même taille constante N (Wright, 1931; Slatkin, 1985a). Cette

condition a de grandes chances de ne pas être respectée ici et des différences de taille entre populations non prises en compte par les modèles utilisés peuvent considérablement biaiser les estimations de l'index de migration, Nm, et conduire à des conclusions erronées en termes de flux de gènes et de temps de divergence entre populations (Nei et Chesser, 1983; Whitlock, 1992; Slatkin, 1995).

Selon le modèle de mutation considéré, à l'équilibre mutation-dérive, lorsque la taille de la population est stable, la diversité génique et la taille effective des populations sont liées selon l'équation:

$$He = \frac{4Ne\mu}{1+4Ne\mu}$$ sous un modèle IAM (Nei, 1987),

$$He = 1 - \frac{1}{\sqrt{1+8Ne\mu}}$$ sous un modèle SMM (Ohta et Kimura, 1973; Nei, 1987),

où Ne est la taille effective des population et μ le taux de mutation du locus considéré.

En inversant ces formules, il est alors possible de déterminer Neμ à partir de He. Mais, il a été démontré que l'estimation du paramètre Neμ à partir de l'hétérozygotie observée est biaisée (Zouros, 1979) et qu'il est beaucoup plus correct de se baser sur le nombre d'allèles (Ewens et Gillespie, 1974). Cornuet et Luikart (1996) ont écrit un programme (Bottleneck) permettant, à partir du nombre d'allèles observés, d'estimer l'hétérozygotie attendue dans les conditions d'équilibre entre mutation et dérive génétique (Heq) sous différents modèles de mutation. Il est alors possible connaissant Heq et en inversant les formules présentées ci-dessus de déterminer une valeur relative de Neμ dans chacune des populations testées. Sous l'hypothèse d'un taux de mutation moyen sur tous les loci (μ) identique dans chacune des populations testées (les mêmes loci ayant été utilisés), le rapport entre les estimations de Neμ pour chaque population va refléter le rapport entre les tailles effectives des différentes populations.

Nous avons donc estimé les valeurs de l'hétérozygotie moyenne attendue à l'équilibre mutation-dérive (Heq) sous un modèle de mutation de type IAM et de type SMM (Cornuet et Luikart, 1996). Les valeurs obtenues pour Heq ont été comparées aux valeurs de l'hétérozygotie moyenne calculée sous l'hypothèse d'Hardy-Weinberg (He) pour chaque population par un test du signe de Wilcoxon (Luikart, 1997; Luikart et Cornuet, 1997) afin de déceler un éventuel écart significatif par rapport aux conditions d'équilibre. Les résultats obtenus sont présentés dans le tableau suivant:

		IAM		SMM	
Population	He	Heq	P	Heq	P
Sénégal	0,636	0,579	0,055	0,705	0,04
Madagascar	0,571	0,518	0,195	0,655	0,31
Maurice	0,425	0,350	0,195	0,461	0,31
La Réunion	0,373	0,246	0,016	0,330	0,69

P=probabilité du test de Wilcoxon Valeurs soulignées P<0,05

Comme nous l'avons précisé précédemment, les loci microsatellites ne suivent pas exclusivement l'un ou l'autre de ces deux modèles de mutation (voir Di Rienzo *et al.*, 1994; Estoup *et al.*, 1995a) ce qui explique les deux déviations significatives observées par rapport à la valeur de l'hétérozygotie moyenne attendue à l'équilibre mutation-dérive (à La Réunion sous un modèle IAM et au Sénégal sous un modèle SMM). Néanmoins, ces deux modèles représentent les extrêmes généralement considérées pour étudier le polymorphisme des loci microsatellites (le modèle TPM étant un intermédiaire entre l'IAM et le SMM). Dans ces conditions, les valeurs obtenues par simulation pour Heq en fonction du nombre d'allèles observés encadrent les valeurs réelles. Ces simulations ont également été réalisées en supposant un modèle

de mutation de type TPM, accordant une proportion p=0,05 de mutations provoquant une variation de la taille des allèles supérieure à une unité du motif répété (Di Rienzo *et al.*, 1994). Dans ces conditions, aucune déviation significative par rapport aux valeurs d'équilibre n'a été mise en évidence (P>0,25) suggérant que les populations considérées sont bien à l'équilibre entre mutation et dérive génétique et qu'aucun goulot d'étranglement démographique de forte amplitude n'a eu lieu récemment (Cornuet et Luikart, 1996).

Les valeurs de Heq ainsi estimées ont été utilisées pour estimer Neμ dans chaque population sous l'hypothèse d'un IAM et d'un SMM et le ratio de ces estimations a été calculé. Les valeurs obtenues sont données dans le tableau suivant, les ratios étant calculés par rapport à la population de La Réunion:

Modèle de mutation		Sénégal	Madagascar	Maurice	Réunion
IAM	Neμ	0,344	0,269	0,135	0,082
	ratio	4,2:1	3,3:1	1,6:1	1
SMM	Neμ	1,31	0,925	0,305	0,153
	ratio	8,6:1	6:1	2:1	1

Ces rapports n'ont qu'une valeur indicative mais l'amplitude de la différence entre les tailles effectives des populations est flagrante et l'estimation des flux géniques entre populations pourrait alors être biaisée. Néanmoins, ces différences ne remettent pas en cause l'amplitude de la différenciation génétique observée entre les sous-populations étudiées mais doivent attirer l'attention sur le fait que la structure génétique mise en évidence parait liée à une hétérogénéité *a priori* importante de la taille des populations considérées.

5. Conclusion

Les loci microsatellites utilisés se sont révélés très polymorphes chez *An. arabiensis* permettant la mise en œuvre d'études visant à déterminer le degré de structuration des populations naturelles au sein de cette espèce.

Les résultats obtenus lors de notre étude confirment que l'utilisation d'amorces hétérologues s'avère un très bon outil pour l'analyse de la génétique des populations chez des espèces voisines et que les amorces décrites chez *An. gambiae* pourront permettre de préciser la biologie et l'écologie d'autres espèces appartenant à ce complexe, notamment *An. arabiensis*. L'image révélée par les variations de polymorphisme entre et au sein des populations étudiées est en parfait accord avec ce qui pouvait être prédit *a priori*, à partir des caractéristiques historiques et écologiques des différentes populations étudiées, confirmant ainsi la pertinence et la valeur des informations fournies par ces marqueurs moléculaires.

L'analyse du polymorphisme observé au niveau de ces loci et de sa distribution entre les différentes sous-populations a mis en exergue l'influence prépondérante de la taille des populations et de ses fluctuations au cours du temps sur le degré de diversité génétique existant au sein des populations locales et, par conséquent, sur le taux de différenciation génétique entre populations. L'équilibre entre la dérive génétique qui érode la variabilité et la migration qui la restaure fait donc intervenir des notions de dynamique qui sont étroitement liées aux paramètres du milieu dans lequel évoluent les populations naturelles. Les méthodes d'analyse spatiale ponctuelles doivent donc être validées par des investigations menées de façon récurrente afin de tester la stabilité au cours du temps de la structure des populations.

De nombreuses incertitudes demeurent quant à l'influence relative des différents facteurs impliqués dans la structuration géographique des populations étudiées et leur stabilité dans le temps. Des études menées sur différentes échelles géographiques et de manière hiérarchique sur le continent Africain comme sur chacune des îles étudiées devraient permettre de se passer de l'interprétation d'effets locaux et de préciser l'influence des facteurs touchant certains groupes de populations et pas d'autres (l'utilisation d'insecticides par exemple).

L'étude que nous avons menée avait essentiellement pour but d'estimer le potentiel résolutif des marqueurs microsatellites chez *An. arabiensis.* Sur la base des résultats obtenus lors de la comparaison des ces quatre populations, l'évolution au cours du temps de la variabilité génétique au sein de la population d'*An. arabiensis* du village de Barkedji au Sénégal a été suivie sur plusieurs années afin d'étudier d'une manière indirecte l'écologie de cette espèce en zone de savane sahélienne.

DEUXIEME PARTIE

Suivi temporel de la variabilité génétique dans une population d'*Anopheles arabiensis* en zone sahélienne

Nous avons vu lors de la Première Partie que les loci microsatellites isolés chez *An. gambiae* pouvaient s'avérer de très bons marqueurs moléculaires pour les études de génétique des populations d'*An. arabiensis*. Le fort taux de polymorphisme de ces marqueurs peut également être utilisé afin de rendre compte de l'évolution au cours du temps de la variabilité génétique à l'intérieur d'une population donnée. Cette approche dynamique est particulièrement pertinente pour la mise en évidence de goulots d'étranglements démographiques dont les conséquences sur le taux de variabilité génétique sont immédiates (diminution du nombre d'allèles et de l'hétérozygotie, voir Nei *et al.*, 1975; Maruyama et Fuerst, 1985a et b; Cornuet et Luikart, 1996).

La dynamique des populations d'*An. arabiensis* dans les zones de savane sahélienne en bordure du Sahara présente des particularités singulières qui en font un excellent modèle d'étude pour ce type d'investigations. En effet, les densités de populations fluctuent énormément au cours de l'année: présentant un pic d'abondance en saison des pluies, les populations de vecteurs sont terriblement réduites, voire introuvables en saison sèche où les conditions environnementales sont très défavorables à la survie des anophèles (Holstein, 1954; Gillies et de Meillon, 1968; Faye *et al.*, 1993; Taylor *et al.*, 1993; Charlwood *et al.*, 1995; voir ci-dessous). Cette très faible densité de population interdit l'utilisation des méthodes directes d'étude de la dispersion des individus par marquage et recapture (Taylor *et al.*, 1993; Touré *et al.*, 1998) et la biologie des populations en saison sèche reste aujourd'hui très mal connue. Une approche indirecte, utilisant à la fois les connaissances écologiques accumulées au cours du suivi entomologique des sites de transmission du paludisme et des connaissances génétiques issues de l'analyse du polymorphisme des loci microsatellites chez les populations vectorielles devrait permettre de faire le lien entre les paramètres

«dynamique de l'habitat», «dynamique des populations» et «composition génétique», et de préciser le devenir des moustiques en saison sèche.

Cette méthodologie a été appliquée à la population d'*An. arabiensis* du village de Barkedji situé en zone de savane sahélienne, au Sénégal. L'objectif de ce travail a été d'étudier les variations du polymorphisme neutre au sein de cette population sur quatre années consécutives, ce qui nous a permis d'obtenir des informations sur la taille effective de la population et son évolution au cours du temps.

Dans un premier temps, nous allons présenter les données écologiques relatives aux fluctuations saisonnières de la densité des populations vectorielles à Barkedji afin de définir clairement la problématique invoquée.

1. Dynamique des populations d'anophèles à Barkedji

L'écologie et la dynamique des populations du complexe *An. gambiae* à Barkedji ont été décrites pour la période 1994-1996 dans l'article présenté en annexe 3:

LEMASSON JJ, FONTENILLE D, LOCHOUARN L, DIA I, SIMARD F, BA K, DIOP A, DIATTA M et MOLEZ JF. 1997. Comparison of behavior and vector efficiency of *Anopheles gambiae* and *An. arabiensis* (Diptera: Culicidae) in Barkedji, a sahelian area of Senegal. *J. Med. Entomol.* **34(4)**: 396-403.

1. 1. - Observations de terrain

La localisation géographique du village de Barkedji est donnée sur la Figure 4. Comme nous l'avons dit dans la Première Partie, ce village est localisé dans la zone de savane sahélienne du Sénégal, exposée à de très faibles précipitations annuelles (de 300 à 400 mm) concentrées sur une période restreinte, de Juillet à Octobre. Les seuls gîtes larvaires permettant le développement des anophèles dans cette région sont les mares

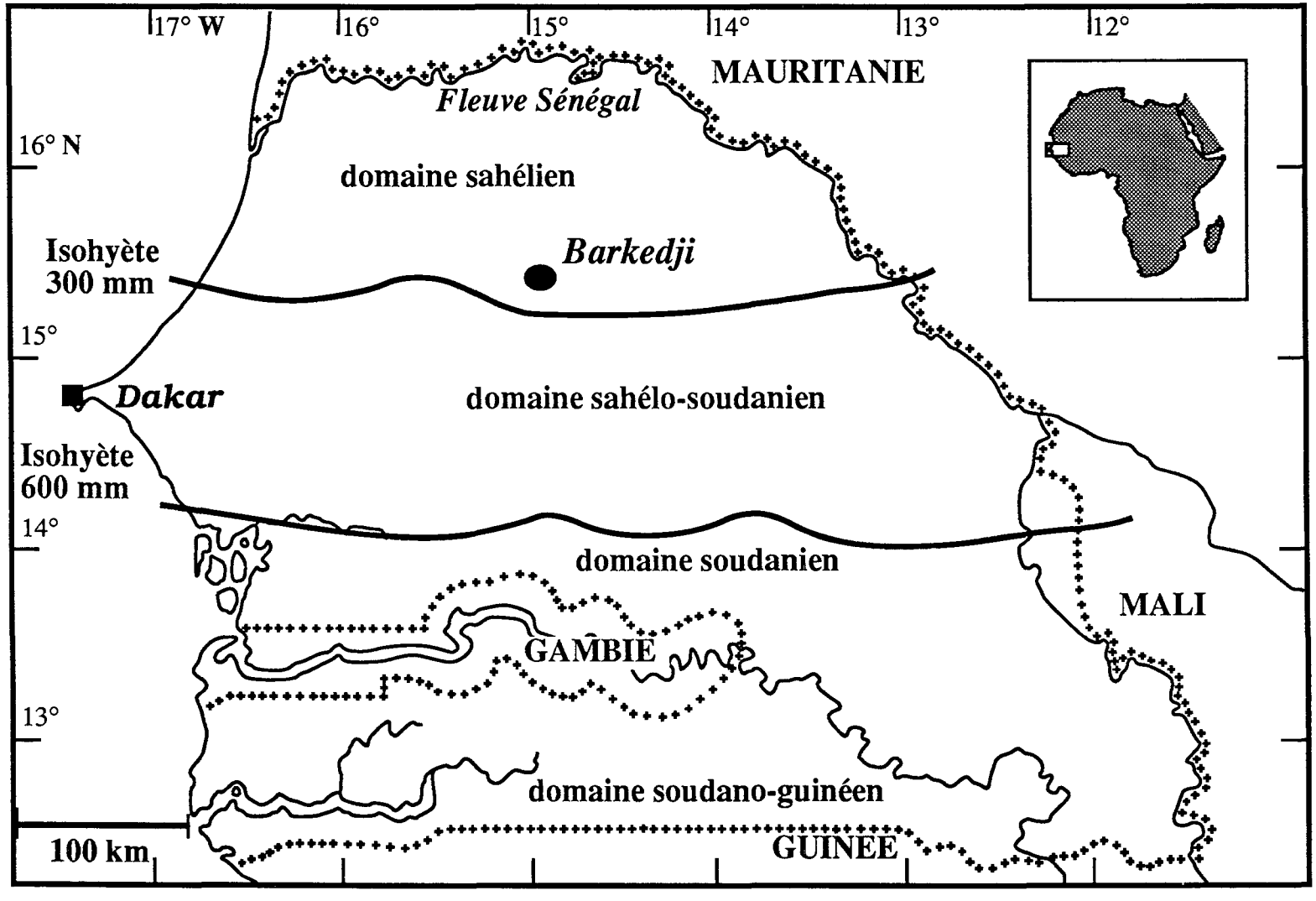


Figure 4: Localisation géographique du village de Barkedji au Sénégal.

temporaires qui se forment en début de saison des pluies par collecte des eaux de ruissellement dans les cuvettes argileuses qui entourent le village (Illustration 7). Ces mares s'assèchent progressivement pour disparaître complètement à la fin du mois de Décembre (Illustration 8).

La densité des populations d'*An. gambiae* et d'*An. arabiensis*, seuls vecteurs du paludisme et seuls membres du complexe rencontrés dans cette zone est étroitement dépendante de la présence de ces gîtes larvaires. La courbe présentée sur la Figure 5, basée sur les résultats de captures sur sujets humains volontaires reflète cette fluctuation de la densité des populations à Barkedji pour la période 1994-1996. A noter, en Décembre 1995, une chute de pluie inhabituelle a permis aux mares temporaires de persister à un faible niveau jusqu'en Février 1996. Le suivi entomologique du site a été maintenu jusqu'en Décembre 1997 et des profils de répartition similaires ont été observés.

Dès que les gîtes larvaires sont en eau, peu de temps après les premières pluies, les vecteurs apparaissent et les densités de populations croissent très vite pour atteindre leur optimum en Octobre-Novembre. Elles décroissent rapidement ensuite lorsque les alizés soufflant du Nord font chuter le degré d'hygrométrie. *An. arabiensis*, plus adapté aux environnements arides est plus largement représenté et reste présent plus longtemps qu'*An. gambiae*. De ce fait, c'est le vecteur majoritaire du paludisme à Barkedji, bien qu'*An. gambiae* possède un taux d'infestation plus élevé durant la période où les deux espèces sont simultanément présentes.

Fluctuation saisonnière de la disponibilité en eau dans les mares temporaires, seuls gîtes larvaires permettant le développement des anophèles à Barkedji.

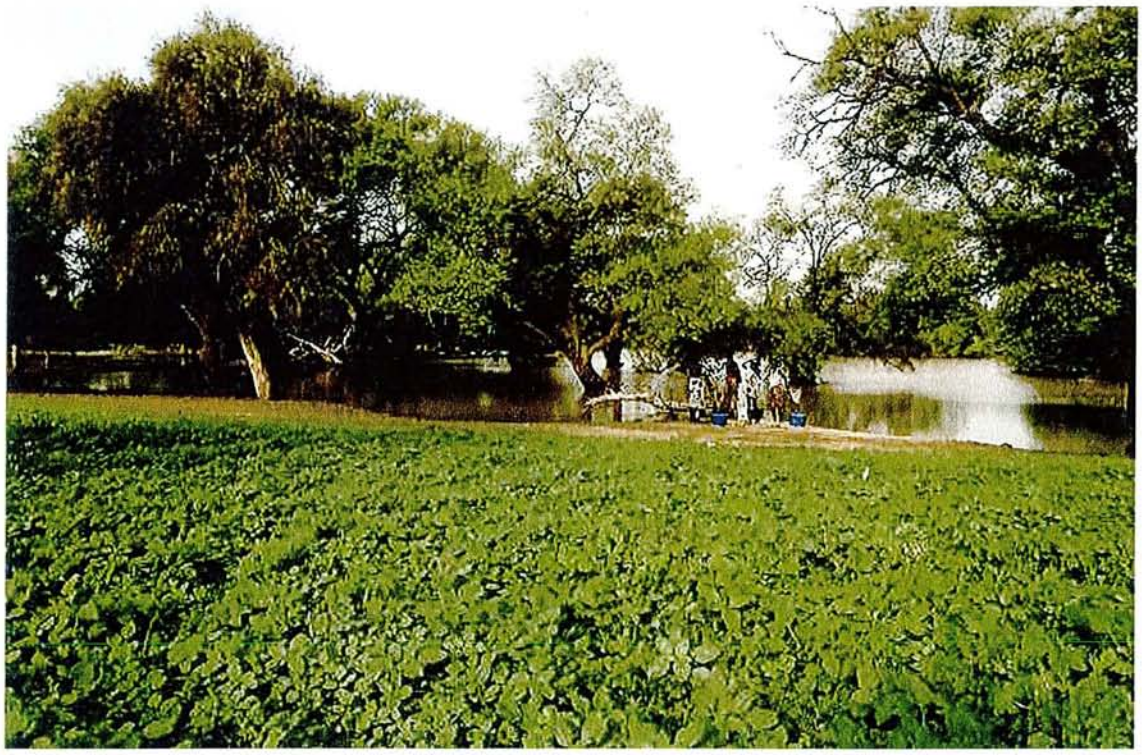
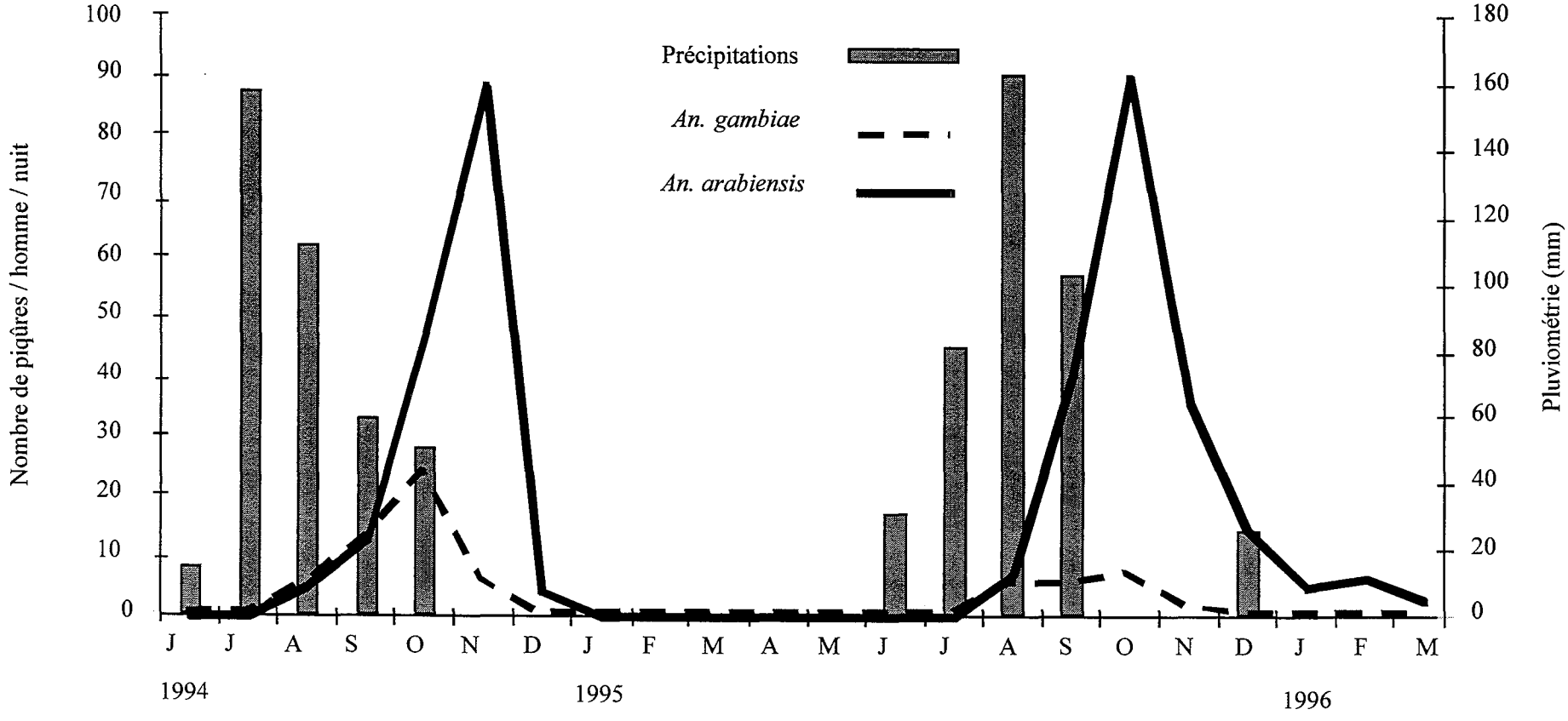


Illustration 7: Durant la saison des pluies, l'eau de ruissellement s'accumule dans les cuvettes argileuses qui entourent le village fournissant d'excellents gîtes larvaires pour les anophèles (de Juillet à Décembre).

Illustration 8: Aucun gîte larvaire n'est plus disponible pendant la saison sèche (de Décembre à Juillet).

Photos: Jean-Jacques Lemasson, IRD, Dakar.

Figure 5: Fluctuations saisonnières de l'agressivité des membres du complexe *An. gambiae* à Barkedji (Sénégal), en zone de savane sahélienne Ouest-Africaine pour la période Juin 1994 - Mars 1996. La présence des vecteurs est étroitement liée à la disponibilité en eau dans les gîtes temporaires.



1. 2. - La problématique

La disparition apparente des vecteurs en saison sèche pousse à se demander quels mécanismes entrent en jeu pour permettre aux populations de recoloniser rapidement les sites en saison des pluies. Plusieurs hypothèses, non exclusives, sont communément avancées.

D'une part, les populations pourraient effectivement disparaître localement lorsque les conditions sont trop restrictives pour permettre la survie et la reproduction des individus. Les gîtes larvaires temporaires formés en début de saison des pluies seraient alors recolonisés chaque année par de nouvelles femelles fondatrices provenant de zones voisines où la disponibilité en eau permet le développement des anophèles tout au long de l'année. Dans ce cas, les populations d'*An. arabiensis* obéissent à une dynamique de métapopulation et de fortes différences de la composition génétique d'une année sur l'autre devraient être observées, suite aux effets fondateurs du début de saison des pluies.

Alternativement, de larges populations pourraient être maintenues localement par l'intermédiaire d'individus en estivation. Beier *et al.* (1990) ont démontré que les œufs d'anophèles sont très sensibles à la dessiccation. Le maintien des populations locales par cet intermédiaire est donc inconcevable sur un intervalle de temps aussi long. Par contre, Omer et Cloudsley-Thompson (1968, 1970) ont décrit l'existence de femelles appartenant au complexe *An. gambiae* (très probablement *An. arabiensis*) capables de survivre pendant les neuf mois que dure la saison sèche au Soudan. Ces femelles en état de «dissociation gonotrophique» sortent périodiquement de leur léthargie pour prendre un repas de sang sur un hôte vertébré mais le développement des

ovaires est extrêmement ralenti et prend toute la saison sèche pour s'accomplir. La ponte n'a donc pas lieu, d'où le terme de «dissociation gonotrophique» utilisé précédemment. Dans ce cas, les dernières femelles fécondées seront les fondatrices de la population de l'année suivante. Si un nombre suffisant de femelles persiste, on ne devrait pas déceler de différences dans la composition génétique des populations d'une année sur l'autre malgré une baisse considérable du nombre de moustiques actifs (et donc accessibles aux techniques de collecte utilisées) en saison sèche.

Nous avons entrepris de suivre au cours du temps le polymorphisme génétique présenté par la population d'*An. arabiensis* grâce aux marqueurs microsatellite. Des échantillons de 50 individus ont été prélevés chaque année, de 1994 à 1997, au mois d'Août, peu de temps après la mise en eau des gîtes temporaires et au mois de Décembre, juste avant qu'ils ne s'assèchent afin d'encadrer respectivement le pic d'abondance de la population (intervalles Août-Décembre) et la saison sèche au cours de laquelle la population semble disparaître (intervalles Décembre-Août). En effet, comme nous allons le voir maintenant, il est possible d'estimer la taille effective des populations pendant un intervalle de temps donné à partir des fluctuations des fréquences alléliques observées aux bornes de cet intervalle (Waples 1989a, 1991). Le protocole d'échantillonnage que nous avons choisi devrait nous permettre d'estimer la taille effective de la population d'*An. arabiensis* de Barkedji en saison des pluies et en saison sèche, respectivement, et de comparer les différentes estimations obtenues sur plusieurs années.

2. L'approche temporelle: estimation de la taille effective des populations grâce aux variations de fréquences alléliques

2. 1. - Principes

La théorie de "l'approche temporelle" repose sur le simple principe que, sous l'action unique de la dérive génétique, de fortes variations de fréquences alléliques neutres sont attendues au cours du temps dans des populations de petite taille alors que, pour des populations de grande taille, ces variations seront moindres (Krimbas et Tsakas, 1971; Waples, 1991). Comme nous l'avons déjà expliqué, ce principe de base de la génétique des populations souligne que l'influence de la dérive génétique sera une fonction inversement proportionnelle de la taille effective des populations, Ne (Nei, 1987; Hartl et Clark, 1989; Waples, 1989a; Caballero, 1994). Dès lors, la différence entre les fréquences alléliques observées dans la même population à deux points successifs dans le temps doit permettre d'estimer la taille effective de la population durant cet intervalle de temps. Si la taille de la population fluctue au cours du temps, la taille effective estimée pour cet intervalle sera très proche de la taille minimale qu'aura atteint la population. En effet, la taille effective de la population peut être vue comme la moyenne harmonique des différentes tailles qu'aura pris la population au cours du temps (Nei et Tajima, 1981; Nei, 1987): si la taille de la population à chaque génération *i*, N_i, varie entre la génération *1* et la génération *t*, la taille effective de la population pour l'intervalle de temps $1 \rightarrow t$ est donnée par la formule (Nei et Tajima, 1981; Nei, 1987; Hartl et Clark, 1989):

$$\frac{1}{Ne} = \frac{1}{t}\left[\frac{1}{N_1} + \frac{1}{N_2} + \frac{1}{N_3} + \ldots\ldots + \frac{1}{N_t}\right]$$

Cette estimation est très sensible aux goulots d'étranglement démographiques. Par exemple, dans notre cas, si la population d'*An. arabiensis* passe par un minimum de

100 individus pendant 6 générations en saison sèche puis augmente à 1 000 000 d'individus pendant six générations en saison des pluies, la taille effective de la population sur l'année, estimée selon la formule ci-dessus donnera Ne=200, proche du minima atteint en saison sèche. C'est sur ce principe que repose l'étude conduite sur la population d'*An. arabiensis* de Barkedji: en estimant la taille effective de la population sur des intervalles de temps réguliers et sur l'intervalle de temps global durant lequel cette population a été suivie, on peut obtenir une évaluation de la taille de la population résiduelle en saison sèche.

Différentes méthodes ont été élaborées pour estimer la taille effective des populations à partir des fluctuations de fréquences alléliques selon le type de population auquel on s'intéresse et le plan d'échantillonnage effectué (voir les revues publiées sur ce sujet par Waples, 1989a ou Caballero, 1994). La migration, la mutation et la sélection sont négligées et la population est considérée homogène (pas de subdivision du pool de gènes) et panmictique (Nei, 1987; Waples, 1989a, 1991). Sous ces hypothèses, la variation de fréquence allélique est attribuée à deux processus:

- l'erreur introduite lors du prélèvement de l'échantillon par l'expérimentateur (les fréquences alléliques sont estimées à partir d'un échantillon de la population et non à partir de données obtenues sur la totalité des individus),
- les variations dues à la dérive génétique, en d'autres termes à l'échantillonnage des allèles de la génération parentale qui seront représentés dans la descendance.

Nei et Tajima (1981) puis Waples (1989a) ont suggéré deux manières différentes d'estimer Ne selon la méthode d'échantillonnage utilisée: dans le premier cas (sampling plan I), les individus sont collectés après la reproduction (ou replacés avant que la

reproduction n'ait lieu) de telle sorte que leurs propres allèles pourront être transmis à la génération suivante. Dans le second cas (sampling plan II) par contre, les individus sont prélevés avant la reproduction et ne sont pas replacés; ils ne peuvent alors contribuer à la formation de la génération suivante. En ce qui concerne les populations d'*An. arabiensis* qui nous intéressent, les spécimens étudiés (50 par date) ont été prélevés au sein d'une population *a priori* relativement grande, possédant un taux de fécondité élevé et n'ont pas été relâchés après capture. Les équations relatives à la seconde méthode d'échantillonnage (sampling plan II: individus collectés avant la reproduction) ont alors été utilisées.

2. 2. - Formulation mathématique

Plusieurs formules ont été proposées pour déterminer la variance standardisée des différences de fréquences alléliques, F, successivement par Krimbas et Tsakas (1971), Nei et Tajima (1981) puis finalement Pollak (1983). Au cours de ce travail, nous avons considéré les équations suivantes:

$$F_c = \frac{1}{K}\sum_{i=1}^{K}\frac{(x_i - y_i)^2}{(x_i + y_i)/2 - x_i y_i}$$ selon Nei et Tajima, 1981.

et

$$F_k = \frac{1}{K-1}\sum_{i=1}^{K}\frac{(x_i - y_i)^2}{(x_i + y_i)/2}$$ selon Pollak, 1983.

où K est le nombre d'allèles au locus considéré, x_i et y_i représentent les fréquences de l'allèle *i* à la génération *0* et *t* respectivement.

L'estimation de la taille effective de la population pour un intervalle de temps *t* (en générations) sera alors donnée par (Krimbas et Tsakas, 1971; Nei et Tajima, 1981; Pollak, 1983; Waples, 1989a):

$$Ne = \frac{t}{2\left(F - \frac{1}{2S_0} - \frac{1}{2S_t}\right)}$$

où *t* est le nombre de générations (discrètes) séparant les deux événements d'échantillonnage, F correspond à F_c ou F_k selon la méthode utilisée, S_0 et S_t sont respectivement l'effectif (en nombre d'individus) de l'échantillon observé à la génération *0* et à la génération *t*. La prise en compte de ces éléments permet de corriger l'estimation de Ne pour l'erreur d'échantillonnage due à l'expérimentateur et de ne s'intéresser qu'à la déviation provenant de l'action de la dérive génétique (Krimbas et Tsakas, 1971).

Pour combiner l'information générée par plusieurs loci, la moyenne pondérée de F est utilisée (Tajima et Nei, 1984):

$$F = \frac{\sum (K_j - 1) F_j}{\sum (K_j - 1)}$$

où les indices *j* font référence à chaque locus considéré. Parallèlement, la moyenne harmonique des effectifs des échantillons testés aux deux dates est utilisée pour remplacer S_0 et S_t dans l'équation précédente (Tajima et Nei, 1984; Waples, 1989a).

Waples (1989a) a étudié les propriétés statistiques de ce paramètre et a démontré que la méthode la plus appropriée pour définir l'intervalle de confiance de cette estimation est de calculer tout d'abord l'intervalle de confiance sur F selon la formule:

Intervalle de confiance à 95% de F = $\left[\frac{nF}{\chi^2_{0,025[n]}}, \frac{nF}{\chi^2_{0,975[n]}}\right]$ (Waples, 1989a)

où n = K-1 (nombre d'allèles indépendants observés au locus considéré).

Pour l'estimation moyenne sur tous les loci, $n = \sum_j K - 1$.

Les valeurs de F ainsi calculées sont ensuite reportées dans la formule d'estimation de Ne pour en déterminer les bornes supérieures et inférieures.

2. 3. - Application aux données expérimentales et commentaires

L'approche temporelle peut générer des estimations de Ne négatives, si F est inférieure à $(1/2S_0+1/2S_t)$ ce qui signifie que la variance entre fréquences alléliques observées entre les deux bornes considérées est inférieure à celle imputable uniquement à l'erreur d'échantillonnage. Dans ce cas, aucune indication de taille effective finie n'est donnée et l'estimation légitime est une taille infinie (Waples, 1989a, 1991). Ce résultat est commun pour des populations de grande taille.

Les allèles représentés en très faibles fréquences peuvent introduire des biais dans l'estimation de Ne (Waples, 1989a). En conséquence, les allèles dont la fréquence ne dépassait pas 2% à chaque borne de l'intervalle considéré ont été regroupés en une seule classe (voir Lehmann *et al.*, 1998).

Les calculs effectués prennent le nombre de générations comme unité de temps. Nous avons considéré une valeur moyenne de 12 générations par an pour *An. arabiensis*, conformément à ce qui est généralement estimé pour cette espèce (Taylor *et al.*, 1993; Lehmann *et al.*, 1998). Une seconde estimation sera également donnée en ne comptant que deux générations entre Décembre et Août, afin de prendre en compte une possible diapause des femelles en saison sèche.

Les résultats de cette étude sont présentés en annexe 4:

SIMARD F, LEHMANN T, LEMASSON JJ, DIATTA M et FONTENILLE D. Persistence of *An. arabiensis* during the severe dry-season conditions in West-Africa: an indirect approach using microsatellite loci. En préparation.

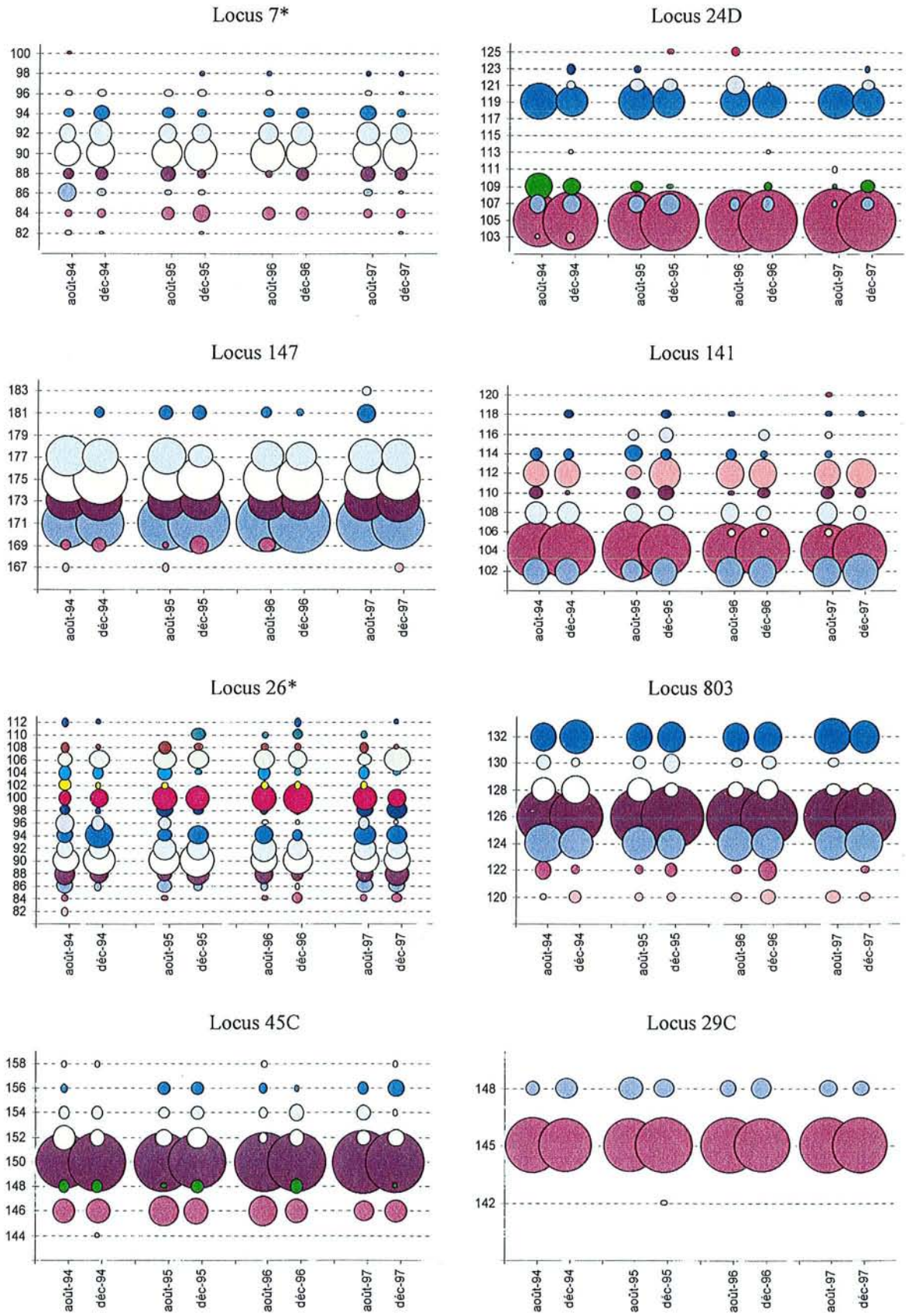
3. Application de l'approche temporelle à la population d'*An. arabiensis* de Barkedji

3. 1. - Evolution du polymorphisme des loci microsatellites au cours du temps

Les fréquences alléliques ont été déterminées pour les neuf loci microsatellite précédemment décrits sur des échantillons de 50 à 54 individus prélevés à Barkedji, en Août et en Décembre chaque année, de 1994 à 1997. Comme lors de l'étude précédente, le locus 49 est resté monomorphe et n'a donc pas été pris en considération pour la suite des analyses. L'évolution des fréquences alléliques au cours du temps est représentée sur la Figure 6, pour chaque locus indépendamment. Les paramètres caractérisant le taux de variabilité génétique détecté sont présentés dans le Tableau 4, pour chaque date d'échantillonnage: nombre d'allèles par locus, hétérozygotie observée et calculée ainsi que leurs moyennes respectives sur tous les loci pour chaque date étudiée et par locus sur l'ensemble des dates testées.

Quelle que soit la date, aucune indication de déviation par rapport à la panmixie n'a été observée à l'intérieur de la population d'*An. arabiensis* de Barkedji suggérant un pool de gènes homogène, sans subdivision apparente. Néanmoins, suivant le locus considéré, quelques déviations ponctuelles significatives par rapport aux proportions d'Hardy-Weinberg ont été observées (voir Tableau 4). C'est le cas au locus 147 en Août 1995 (Fis=0,315, P<0,01), au locus 45C en Août 1996 (Fis=-0,098, P<0,05), au locus 26 en Décembre 1997 (Fis=0,176, P<0,05) et au locus 24D dans 3 cas (en Août 1994: Fis=0,259, P<0,05; en Décembre 1994: Fis=0,343, P<0,01; et en Décembre 1997: Fis=0,384, P<0,01). Alors que les déviations ponctuelles observées aux loci 147, 45C et 26 peuvent être vraisemblablement imputées aux seules fluctuations stochastiques dues à l'échantillonnage d'un petit nombre d'individus sur la population totale, le déficit

Figure 6: Evolution des fréquences alléliques au sein de la population d'*An. arabiensis* de Barkedji pour la période Août 1994 – Décembre 1997.



La surface des disques est proportionnelle à la fréquence de l'allèle dans la population à la date considérée.

* La surface des disques a été réduite de moitié pour éviter un encombrement stérique trop important dû au grand nombre d'allèles observés.

Pour tous les graphiques: Axe des abscisses = date ; axe des ordonnées = taille des allèles (pb).

Tableau 4: Suivi de la variabilité génétique de la population d'*An. arabiensis* à Barkedji.

Date		Locus								Moyenne
		7	24D	147	141	26	803	45C	29C	tous loci#
Août 1994										
	Nall	9	5	7	6	15	7	7	2	7,25
	H obs	0,604	0,481	0,667	0,519	0,833	0,660	0,537	0,167	0,558
	H exp	0,749	0,648	0,773	0,594	0,915	0,714	0,539	0,154	0,636
	Fis	+0,195	+0,259	+0,138	+0,128	+0,090	+0,076	+0,004	-0,082	+0,122
Décembre 1994										
	Nall	8	8	6	7	13	7	7	2	7,25
	H obs	0,600	0,380	0,680	0,460	0,826	0,680	0,460	0,260	0,543
	H exp	0,722	0,577	0,768	0,583	0,861	0,736	0,435	0,228	0,614
	Fis	+0,170	**+0,343**	+0,115	+0,213	+0,041	+0,077	-0,057	-0,140	+0,117
Août 1995										
	Nall	7	6	7	7	13	7	6	2	6,87
	H obs	0,680	0,560	0,520	0,420	0,841	0,660	0,520	0,300	0,563
	H exp	0,685	0,561	0,757	0,535	0,876	0,674	0,477	0,258	0,603
	Fis	+0,007	+0,001	**+0,315**	+0,216	+0,041	+0,021	-0,090	-0,167	+0,068
Décembre 1995										
	Nall	9	6	6	8	11	7	6	3	7
	H obs	0,640	0,460	0,580	0,540	0,694	0,600	0,520	0,200	0,529
	H exp	0,654	0,531	0,712	0,687	0,846	0,625	0,511	0,216	0,598
	Fis	+0,021	+0,134	+0,186	+0,216	+0,181	+0,041	-0,018	+0,073	+0,115
Août 1996										
	Nall	8	5	6	8	14	7	6	2	7
	H obs	0,620	0,360	0,660	0,520	0,813	0,600	0,460	0,140	0,522
	H exp	0,587	0,489	0,719	0,636	0,884	0,621	0,419	0,165	0,565
	Fis	-0,057	+0,265	+0,083	+0,184	+0,081	+0,035	-0,098	+0,155	+0,078
Décembre 1996										
	Nall	5	6	5	8	15	7	6	2	6,75
	H obs	0,640	0,360	0,680	0,500	0,800	0,640	0,460	0,280	0,545
	H exp	0,668	0,464	0,705	0,619	0,874	0,709	0,451	0,243	0,592
	Fis	+0,042	+0,226	+0,036	+0,193	+0,085	+0,098	-0,021	-0,153	+0,079
Août 1997										
	Nall	8	5	5	10	14	6	5	2	6,87
	H obs	0,776	0,400	0,540	0,551	0,837	0,640	0,340	0,180	0,533
	H exp	0,738	0,413	0,724	0,689	0,899	0,660	0,337	0,165	0,578
	Fis	-0,051	+0,033	+0,256	+0,202	+0,070	+0,031	-0,010	-0,089	+0,079
Décembre 1997										
	Nall	9	6	6	6	13	6	7	2	6,87
	H obs	0,500	0,300	0,700	0,520	0,735	0,600	0,420	0,160	0,492
	H exp	0,573	0,485	0,754	0,630	0,890	0,624	0,423	0,149	0,566
	Fis	+0,128	**+0,384**	+0,072	+0,176	+0,176	+0,039	+0,008	-0,077	+0,132
Moyenne toutes dates	Nall	7,87	5,87	6	7,5	13,5	6,75	6,25	1,12	6,9
	H obs	0,645	0,423	0,636	0,490	0,803	0,624	0,438	0,210	0,533
	H exp	0,669	0,545	0,731	0,631	0,887	0,661	0,426	0,196	0,593
	Fis*	+0,061	+0,211	+0,155	+0,191	+0,097	+0,054	-0,035	-0,069	+0,099

Caractères soulignés: P<0,05; **gras soulignés**: P<0,01.
valeur moyenne de Fis (estimée selon Weir et Cockerham, 1984) testée par le test combiné de Fisher (Raymond et Rousset, 1995a).
* valeur de Fis donnée à titre indicatif et non testée statistiquement.

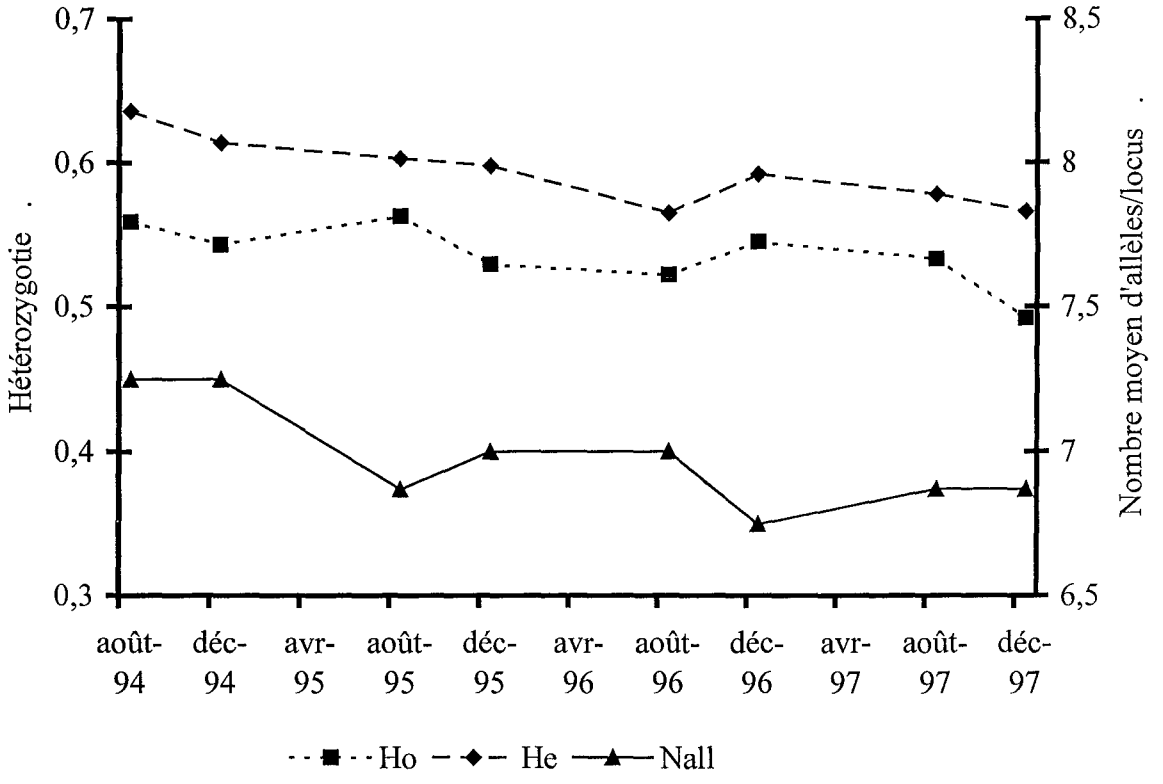
significatif en hétérozygotes observé de manière récurrente au locus 24D soutient l'hypothèse formulée précédemment concernant la présence possible d'allèles nuls à ce locus.

L'étude du déséquilibre de liaison entre paires de loci reflète également cette homogénéité génétique. Sur 28 tests effectués par échantillon (8x7/2), seuls les loci 26 et 147 sont apparus significativement liés dans l'échantillon prélevé en Décembre 1994 ($P<0,001$). Aucun autre déséquilibre de liaison n'a été mis en évidence ($P>0,05$ dans tous les autres cas lorsque la procédure de Bonferroni est appliquée), confirmant ainsi l'absence de structure génétique interne et l'indépendance génétique des loci microsatellites utilisés.

La Figure 7 représente l'évolution au cours temps du nombre moyen d'allèles par locus (Nall) et de l'hétérozygotie moyenne observée (Ho) et calculée (He) sur toute la durée de l'étude. Ces trois paramètres apparaissent relativement stables. Le test du signe de Wilcoxon appliqué pour chaque locus entre paires de dates consécutives n'a pas permis de mettre en évidence de différence significative entre les échantillons (ddl=7, $P>0,07$ pour l'hétérozygotie observée ou calculée, $P>0,27$ pour le nombre d'allèles). Néanmoins, la diminution de l'hétérozygotie moyenne observée et calculée entre Août 1994 (Hobs=0,558; Hexp=0,636) et Décembre 1997 (Hobs=0,492; Hexp=0,566) s'est révélée marginalement significative (ddl=7; P=0,05 dans les deux cas). Cette légère diminution significative de l'hétérozygotie pourrait suggérer que la population d'*An. arabiensis* de Barkedji soit effectivement passée par un ou plusieurs épisodes de taille réduite entre 1994 et 1997.

Les fréquences alléliques observées à tous les loci sont très fortement corrélées d'une date à l'autre (coefficient de corrélation de Pearson entre les fréquences de chaque allèle à chaque locus pour deux dates consécutives compris entre 0,962 et 0,987; $61<n<64$) ainsi qu'entre les deux dates bornant l'intervalle de temps total

Figure 7: Evolution au cours du temps de trois paramètres classiques de la génétique des populations chez *An. arabiensis* à Barkedji.



(coefficient de corrélation de Pearson=0,955; n=62). En utilisant le test exact d'homogénéité (Raymond et Rousset, 1995b), aucune variation significative dans la répartition des fréquences alléliques sur l'ensemble des dates testées n'a pu être décelée pour les loci 147, 141, 803, 45C et 29C (P>0,16), suggérant également une stabilité temporelle de la variabilité génétique au sein de cette population. Des variations très significatives ont été observées au niveau des loci 7, 24D, et 26 (P<0,001) impliquant généralement l'échantillon d'Août 1994. En accord avec ces résultats, de très faibles taux de différenciation génétique ont été observés entre les différents isolats (Tableau 5: -0,0027<Fst<0,019 et -0,0102<Rst<0,0117). Seules huit estimations de Fst sur les 28 comparaisons effectuées se sont révélées statistiquement significatives (sur la base de 1000 permutations des génotypes entre populations), 5 d'entre elles restant significatives au seuil de 5% lorsque la procédure de Bonferroni est appliquée. Aucune hétérogénéité significative n'a été décelée par Rst. Une différenciation génétique très faible mais néanmoins significative a été révélée sur l'ensemble des données par Fst mais non par Rst (Fst=0,005, $P<10^{-4}$; Rst<0, P=0,76). Le court intervalle de temps considéré, en relation avec les faibles valeurs estimées de Rst non significatives, justifient l'hypothèse du rôle négligeable de la mutation dans cette étude (Slatkin, 1995). En accord avec le test d'homogénéité des fréquences alléliques, la population d'Août 1994 est impliquée dans tous les cas où une différenciation génétique significative a été révélée par Fst (Tableau 5).

Comme l'a souligné Waples (1989a et b, 1991), les échantillons collectés lorsque ce type d'approche est appliquée ne sont pas indépendants. En effet, chaque échantillon reflète en fait un certain nombre de générations de dérive génétique par rapport à ceux qui le précèdent. La dérive génétique peut à elle seule expliquer les résultats significatifs observés lorsque le test d'homogénéité des fréquences alléliques

Tableau 5: Evaluation de la divergence génétique entre différents échantillons de la population d'*An. arabiensis* de Barkedji prélevés au cours du temps, par Fst (sous la diagonale) et Rst (au-dessus de la diagonale).

	Août 1994	Décembre 1994	Août 1995	Décembre 1995	Août 1996	Décembre1996	Août 1997	Décembre 1997
Août 1994	******	-0,0056	-0,0093	0,0009	0,0004	0,0117	0,0012	-0,0043
Décembre1994	0,0048	******	-0,0003	-0,0006	-0,0031	0,0074	-0,0090	-0,0102
Août 1995	0,0052	0,0017	******	-0,0011	-0,0083	-0,0008	0,0008	-0,0070
Décembre 1995	**0,0170**	0,0094	0,0016	******	-0,0010	0,0083	-0,0032	0,0017
Août 1996	**0,0190**	0,0090	0,0006	-0,0028	******	-0,0017	-0,0017	-0,0032
Décembre 1996	**0,0149**	0,0047	-0,0013	-0,0025	-0,0027	******	0,0068	0,0031
Août 1997	**0,0144**	0,0032	0,0043	0,0046	0,0032	-0,0015	******	-0,0056
Décembre 1997	**0,0144**	0,0076	0,0033	0,0031	-0,0021	0,0022	0,0023	******

Global : Fst= 0,0052, $P<10^{-4}$; Rst= -0,0011, P=0,761.

Valeurs soulignées : P<0,05 ; Caractères **gras soulignés** : P<0,05 après application de la procédure de Bonferroni.

est effectué ou lors de l'estimation des indices de fixation car ces statistiques ne sont pas corrigées pour l'appariement entre échantillons; seul l'effet d'échantillonnage (effectif des échantillons) est pris en compte lorsque le test classique d'homogénéité est effectué (test du Chi-deux ou test exact de Fisher, voir Raymond et Rousset, 1995b). Le pourcentage de tests significatifs s'éloignera donc rapidement du seuil de 5% attendu sous le seul effet du hasard lorsque les échantillons sont indépendants. De plus, nous avons noté précédemment (Première Partie) que l'estimation de Fst était beaucoup plus sensible que Rst aux effets de la dérive génétique. Le fait que nous n'ayons observé que très peu de résultats significatifs en appliquant ces tests non corrigés à nos données, en dépit des biais engendrés par cette procédure, confirme donc la stabilité de la composition génétique de la population d'*An. arabiensis* de Barkedji au cours du temps.

L'ensemble de ces résultats plaide donc en faveur de la présence à Barkedji d'une population d'*An. arabiensis* panmictique, très stable dans le temps et soumise à l'action modérée de la dérive génétique ce qui contredit l'hypothèse de forts goulots d'étranglement démographiques récurrents.

3. 2. - Taille effective de la population

Les formules décrites précédemment (Paragraphe 2) ont été appliquées aux données concernant la population d'*An. arabiensis* de Barkedji. Les différentes estimations de la taille effective des populations sur chaque intervalle de temps ainsi que sur la durée totale du suivi (Août 1994 – Décembre 1997) sont reportées dans le Tableau 6.

Les valeurs numériques de Ne obtenues en utilisant Fc (Nei et Tajima, 1981) ou Fk (Pollak, 1983) pour estimer la variance standardisée des différences de fréquences

Tableau 6: Estimations de la taille effective de la population d'*An. arabiensis* de Barkedji pour l'intervalle de temps Août 1994 – Décembre 1997, en comptant 12 générations par an.

Intervalle de temps		Locus								Tous loci
		7	24D	147	141	26	803	45C	29C	
Août 94 – Déc 94	Ne (Fk)	65 (12 – 384)	144 (17 - ∞)	∞ (58 - ∞)	∞ (47 - ∞)	123 (25 - ∞)	∞ (29 -∞)	∞ (24 - ∞)	506 (1 - ∞)	209 (79 – 2317)
	Ne (Fc)	71 (12 – 469)	133 (16 - ∞)	∞ (59 - ∞)	∞ (59 - ∞)	125 (26 - ∞)	∞ (31 - ∞)	∞ (21 - ∞)	523 (1 - ∞)	228 (83 – 6230)
Déc 94 – Août 95	Ne (Fk)	816 (59 - ∞)	∞ (75 - ∞)	∞ (59 - ∞)	269 (32 - ∞)	513 (62 - ∞)	∞ (130 - ∞)	993 (41 - ∞)	∞ (1 - ∞)	1198 (249 - ∞)
	Ne (Fc)	676 (56 - ∞)	∞ (83 - ∞)	∞ (55 - ∞)	322 (35 - ∞)	556 (64 - ∞)	∞ (104 - ∞)	2114 (46 - ∞)	∞ (1 - ∞)	1571 (270 - ∞)
Août 95 – Déc 95	Ne (Fk)	1032 (36 - ∞)	∞ (55 - ∞)	191 (16 - ∞)	79 (12 - ∞)	182 (30 - ∞)	∞ (38 - ∞)	∞ (87 - ∞)	∞ (3 - ∞)	401 (107 - ∞)
	Ne (Fc)	2557 (40 - ∞)	∞ (56 - ∞)	209 (16 - ∞)	68 (11 – 1264)	214 (32 - ∞)	∞ (38 - ∞)	∞ (105 - ∞)	∞ (4 - ∞)	472 (114 - ∞)
Déc 95 – Août 96	Ne (Fk)	∞ (141 - ∞)	3453 (37 - ∞)	∞ (106 - ∞)	988 (59 - ∞)	1670 (95 - ∞)	∞ (116 - ∞)	441 (33 - ∞)	∞ (9 - ∞)	∞ (414 - ∞)
	Ne (Fc)	∞ (144 - ∞)	∞ (42 - ∞)	∞ (122 - ∞)	1657 (64 - ∞)	1677 (95 - ∞)	∞ (125 - ∞)	546 (36 - ∞)	∞ (9 - ∞)	∞ (497 - ∞)
Août 96 – Déc96	Ne (Fk)	197 (19 - ∞)	120 (13 - ∞)	18333 (20 - ∞)	∞ (59 - ∞)	∞ (75 - ∞)	∞ (35 - ∞)	172 (15 - ∞)	438 (1 - ∞)	634 (127 - ∞)
	Ne (Fc)	208 (19 - ∞)	166 (15 - ∞)	∞ (24 - ∞)	∞ (70 - ∞)	∞ (77 - ∞)	∞ (34 - ∞)	244 (17 - ∞)	453 (1 - ∞)	1148 (148 - ∞)
Déc 96 – Août 97	Ne (Fk)	508 (45 - ∞)	∞ (40 - ∞)	426 (23 - ∞)	∞ (89 - ∞)	682 (79 - ∞)	220 (23 - ∞)	609 (26 - ∞)	767 (1 - ∞)	693 (182 - ∞)
	Ne (Fc)	604 (49 - ∞)	∞ (87 - ∞)	683 (27 - ∞)	∞ (87 - ∞)	734 (80 - ∞)	282 (25 - ∞)	546 (25 - ∞)	793 (1 - ∞)	921 (203 - ∞)
Août 97 – Déc 97	Ne (Fk)	248 (28 - ∞)	490 (19 - ∞)	∞ (49 - ∞)	735 (21 - ∞)	269 (47 - ∞)	∞ (52 - ∞)	1545 (23 - ∞)	∞ (2 - ∞)	654 (145 - ∞)
	Ne (Fc)	173 (23 - ∞)	1075 (22 - ∞)	∞ (52 - ∞)	1970 (23 - ∞)	279 (47 - ∞)	∞ (57 - ∞)	2243 (23 - ∞)	∞ (2 - ∞)	760 (152 - ∞)
Août 94 – Déc 97	Ne (Fk)	456 (88 – 1828)	530 (62 – 6794)	∞ (601 - ∞)	1879 (159 - ∞)	1127 (259 – 209632)	930 (204 - ∞)	4758 (1457 - ∞)	∞ (198 - ∞)	1000 (479 – 2465)
	Ne (Fc)	443 (85 – 1735)	458 (56 – 3796)	∞ (416 - ∞)	2089 (166 - ∞)	1225 (272 - ∞)	1057 (120 - ∞)	3417 (195 - ∞)	∞ (198 - ∞)	1019 (486 – 2547)
Août 94 – Déc 97*	Ne (Fk)	263 (51 – 1051)	305 (36 – 3907)	∞ (1 - ∞)	1080 (91 - ∞)	648 (149 – 120539)	535 (117 - ∞)	2736 (843 - ∞)	∞ (114 - ∞)	575 (275 – 1417)
	Ne (Fc)	255 (49 – 997)	263 (32 – 2183)	∞ (239 - ∞)	1201 (95 - ∞)	704 (157 - ∞)	608 (69 - ∞)	1965 (112 - ∞)	∞ (114 - ∞)	586 (279 – 1464)

Ne (Fk): Estimation de Ne basée sur Fk (Pollak, 1983).
Ne (Fc): Estimation de Ne basée sur Fc (Nei et Tajima, 1981).
∞ : infini.
Les intervalles de confiance à 95% des valeurs estimées sont donnés entre parenthèses (calculés selon Waples, 1989a).
* Ne estimée en comptant seulement deux générations en saison sèche (intervalles Décembre – Août).

alléliques sont très similaires, conformément à ce qui est théoriquement attendu (Pollak, 1983; Waples, 1989a) et généralement observé empiriquement (Taylor *et al.*, 1993; Miller et Kapuscinski 1997; Lehmann *et al.*, 1998). Quelques différences de l'ordre d'un rapport de 2 à 3 ont tout de même été observées ponctuellement (par exemple au locus 141 sur l'intervalle Août-Décembre 1997) mais dans aucun des cas les estimations de Ne fournies par l'une ou l'autre des méthodes ne sont incompatibles (en dehors de l'intervalle de confiance à 95%). Aucune des deux méthodes ne produit des estimations de Ne systématiquement plus faibles ou plus fortes que l'autre.

De très nombreuses estimations de Ne basées sur un seul locus sont infinies, même lorsque l'intervalle de temps total est considéré (loci 29C et 147). Ces résultats ne sont pas surprenants au vu des faibles variations de fréquences alléliques observées, qui peuvent être totalement imputées à l'erreur d'échantillonnage due à la collecte des spécimens utilisés pour l'analyse génétique (Waples, 1989a). Dans la très grande majorité des cas, les intervalles de confiance à 95% ne sont pas bornés vers le haut. Lorsque l'information est combinée sur les huit loci (moyenne des F pondérée par le nombre d'allèles indépendants observés sur l'échantillon), les estimations de Ne pour chaque intervalle de temps varient sensiblement et tous les intervalles de confiance à 95% restent ouverts sur des valeurs infinies, à l'exception de l'intervalle Août-Décembre 1994. En conséquence du recouvrement important des intervalles de confiance, aucune comparaison n'a pu être effectuée entre les différentes estimations de Ne en saison sèche ou en saison des pluies. Les valeurs de Ne obtenues varient de 209 à l'infini, selon l'intervalle de temps considéré et la méthode d'estimation utilisée. Ces variations sont en accord avec les intervalles de confiance très larges obtenus. Aucune indication de taille effective réduite en saison sèche (intervalles Décembre-Août) n'est apparente à ce niveau de l'analyse. Il est intéressant de noter que la seule estimation de

Ne infinie qui subsiste à ce niveau est observée pour l'intervalle Décembre 1995 – Août 1996 qui encadre la saison sèche particulière lors de laquelle les gîtes larvaires sont restés en eau jusqu'en Février.

Lorsque l'intervalle de temps total est considéré, les estimations de la taille effective des populations obtenues en utilisant Fc ou Fk sont tout à fait comparables et les intervalles de confiance sont bien définis dans les deux cas et sensiblement identiques. Lorsqu'une valeur moyenne de 12 générations par an est considérée, les estimations de Ne obtenues par les deux méthodes sont Ne=1000∈ [479; 2465] et Ne=1019∈ [486; 2547] à partir de Fk et Fc respectivement. Ces valeurs qui reflètent la taille minimale de la population atteinte au cours de l'intervalle de temps considéré (4 saisons des pluies séparées par 3 saisons sèches) restent conséquentes.

Si les individus sont effectivement maintenus sur le site de Barkedji durant la saison sèche, l'absence de gîtes larvaires disponibles dans cette région sous-entend que les femelles persistent dans un état physiologique particulier qui ne permet pas l'oviposition (Omer et Cloudsley-Thompson, 1968, 1970). Nos estimations du nombre de générations doivent alors être modifiées en conséquence. Afin de prendre en compte ces données écologiques, nous avons recalculé les valeurs de Ne en ne comptant que deux générations en saison sèche (au lieu de 8 précédemment supposées en prenant une valeur constante sur l'année). Le nombre total de générations estimées pour l'intervalle Août 1994 – Décembre 1997 passe alors de 40 à 23 et les estimations de Ne obtenues sont réduites dans les mêmes proportions (le nombre de générations *t* étant seul au numérateur de l'équation permettant d'estimer Ne). Dans ces conditions, les valeurs obtenues restent élevées, Ne=575∈ [275; 1417] et Ne=586∈ [279; 1464] à partir de Fk et Fc respectivement, suggérant une population résiduelle relativement importante.

En conclusion, les résultats de l'analyse temporelle du polymorphisme des marqueurs microsatellites suggèrent que la taille effective de la population d'*An. arabiensis* de Barkedji ne descend pas en dessous d'un seuil de plusieurs centaines voir quelques milliers d'individus en saison sèche, malgré d'importantes fluctuations de la densité de moustiques actifs au cours de l'année.

3. 3. - Implications sur la biologie d'*An. arabiensis* en zone sahélienne

Plusieurs résultats importants découlent de l'analyse de la population d'*An. arabiensis* de Barkedji:

- la panmixie est respectée quelle que soit la date d'échantillonnage et aucune structure génétique interne n'a été mise en évidence,
- le taux de variabilité génétique est important (hétérozygotie moyenne et nombre d'allèles par locus élevés et comparables quelque soit la date testée) , suggérant une population de grande taille,
- la stabilité des fréquences alléliques au cours du temps se traduit par des estimations de taille effective élevées avec des intervalles de confiance à 95% le plus souvent non bornés vers le haut lorsque les périodes considérées englobent alternativement la saison des pluies (fortes densités) ou la saison sèche (très faibles densités),
- la taille effective de la population estimée sur l'intervalle de temps entier, englobant 4 saisons des pluies et 3 saisons sèches reste substantielle (plusieurs centaines à quelques milliers d'individus).

En conséquence, il apparaît que la population d'*An. arabiensis* de Barkedji ne soit pas soumise à des phénomènes récurrents de goulots d'étranglement démographiques d'une amplitude aussi importante que les données écologiques

pourraient le laisser penser. Nos résultats sont en parfait accord avec les estimations de Ne obtenues lors d'une étude précédente basée sur les variations temporelles des fréquences d'inversions chromosomiques chez *An. arabiensis* en Afrique de l'Ouest (Taylor *et al.*, 1993).

Pour réconcilier les paramètres «importante taille effective de la population» et «très faibles densités en saison sèche», nous devons faire appel aux notions de capacité d'estivation et de dispersion chez cette espèce.

3. 3. 1. - L'estivation

Les données concernant la capacité d'estivation chez les anophèles du complexe *An. gambiae* ont été décrites précédemment et mettent en jeu un mécanisme de diapause ou «dissociation gonotrophique», qui permet aux femelles de subsister lorsque les conditions climatiques deviennent très défavorables et qu'aucun gîte larvaire n'est disponible. Cet état physiologique particulier a été mis en évidence au Soudan par Omer et Cloudsley-Thompson (1968, 1970). Ce mécanisme a été également soupçonné par d'autres auteurs dans des régions aux caractéristiques climatiques similaires mais les observations sont plus nuancées (Holstein, 1954, au Burkina Faso; Coosemans *et al.*, 1989 au Burundi). Lors d'expériences de laboratoire reproduisant les conditions observées en saison sèche, Holstein (1954) a montré que des femelles pouvaient survivre plus de 150 jours dans des conditions arides en ne pondant que deux à trois fois. Privées de source de nourriture (sucre ou sang), certaines femelles sont également capables de survivre jusqu'à 70 jours, probablement grâce aux ressources trouvées dans leurs propres œufs. Toutes ces données sur la biologie des anophèles du complexe *An. gambiae* en conditions extrêmes suggèrent que ces moustiques sont capables de survivre

à la saison sèche sahélienne. Des investigations très poussées à Barkedji lors des mois d'Avril 1996 et Avril 1997 nous ont effectivement permis de collecter quelques spécimens d'*An. arabiensis* (12 et 13 respectivement, collectés sur 24 hommes-nuits, voir annexe 3) à la recherche d'un repas de sang alors qu'aucun moustique n'a été observé dans les chambres (après pulvérisation d'insecticide) ni dans les gîtes de repos naturels potentiels que nous ayons prospecté (terriers, termitières, cases inhabitées, silos, ...). Aucun spécimen mâle, ni *An. gambiae* s.s. n'a été trouvé. A cette période de l'année, les températures dépassent fréquemment les 40°C et l'hygrométrie descend sous la barre des 20% d'humidité relative. Aucun gîte larvaire n'est disponible à Barkedji ni dans la savane environnante, les rares collections d'eau accessibles étant les canaris utilisés par les populations locales pour l'eau de consommation qui sont quotidiennement remplis et utilisés (aucune larve de Culicidae n'y a été observée) et de rares puits très dispersés qui sont également soumis à une agitation constante. Cette observation écologique pourrait soutenir l'hypothèse du maintien des populations locales par le biais de femelles en estivation. En effet, sous cette hypothèse (en ne comptant que deux générations en saison sèche) nous avons estimé que la taille effective de la population d'*An. arabiensis* de Barkedji était de 580 individus (entre 275 et 1500 individus). Comme nous l'avons précédemment remarqué, cette estimation doit être proche du minimum atteint au cours de l'intervalle Août 1994 – Décembre 1997, vraisemblablement en saison sèche (à titre de comparaison, il nous est arrivé de collecter plus de 500 anophèles gorgés dans une seule chambre en saison des pluies à Barkedji). Etant donné la profusion de gîtes potentiels de repos que l'on peut trouver dans cette zone (maisons, troncs d'arbres creux, buses...), un tel nombre de moustiques en estivation en saison sèche n'est pas inconcevable et la population de Barkedji pourrait effectivement être maintenue par ce biais. Bien que nos prospections de ce

genre de gîtes n'aient pas permis la mise en évidence de spécimens en estivation, nous avons pu collecter un petit nombre de moustiques agressifs pour l'homme en pleine saison sèche ce qui constitue une preuve directe de la présence continue au cours de l'année d'*An. arabiensis* à Barkedji.

3. 3. 2. - La dispersion des individus

Les données écologiques et génétiques accumulées sur *An. arabiensis* plaident en faveur d'un taux de dispersion élevé chez cette espèce. Mis à part les zones de forêt équatoriale où seul *An. gambiae* est présent, *An. arabiensis* est ubiquitaire sur le continent Africain, de la bordure sud du Sahara jusqu'en Afrique du Sud, avec une répartition longitudinale ininterrompue (Gillies et de Meillon, 1968; Gillies et Coetzee, 1987, voir Figure 1). On le trouve également dans les zones d'altitude restrictives à l'établissement des populations d'*An. gambiae* en Afrique de l'Est ou à Madagascar (Mosha et Subra, 1982; Ralisoa Randrianasolo et Coluzzi, 1987; voir première partie). *An. arabiensis* est anthropophile préférentiel mais peut facilement se nourrir alternativement sur d'autres hôtes vertébrés ce qui le rend beaucoup moins dépendant de la présence humaine que son espèce jumelle. Il a en effet été trouvé dans des zones totalement sauvages où les populations humaines sont absentes ou très transitoires: c'est le cas du Parc National Kruger en Afrique du Sud (Braack *et al.*, 1994) ou de la vallée du Rift au Kenya (L. Kamau, communication personnelle). Ce complexe géographique aride très faiblement peuplé a pourtant été décrit comme une barrière aux flux de gènes chez *An. gambiae* (McLain *et al.*, 1989; Kamau *et al.*, 1998a; Lehmann *et al.*, en préparation). Les études indirectes basées sur la cytogénétique (Coluzzi *et al.*, 1979, 1985), le polymorphisme des marqueurs isoenzyme (Smits *et al.*, 1996) ou de l'ADN mitochondrial (Besansky *et al.*, 1997) n'ont pas permis de déceler de structure

génétique forte à l'intérieur ou entre populations d'*An. arabiensis*, contrairement à ce qui est observé chez *An. gambiae*, suggérant des flux de gènes importants qui relient les populations locales. Les expériences de marquage-recapture réalisées en saison des pluies s'accordent généralement pour estimer que le taux de dispersion des membres du complexe *An. gambiae* est fortement corrélé à la disponibilité locale en hôtes et en gîtes larvaires, à la géographie du terrain et à la densité de la population (Service, 1993; Costantini *et al.*, 1996; Touré *et al.*, 1998). Dans ces conditions, il est impossible d'extrapoler à l'ensemble de l'espèce les résultats obtenus dans certaines localités à des dates bien précises. Néanmoins, il apparaît clairement que la présence de nombreux gîtes larvaires et d'hôtes potentiels va favoriser un comportement sédentaire. A l'inverse, un environnement ouvert tel que celui rencontré dans les zones de savane sahélienne, associé à la raréfaction des gîtes larvaires et des troupeaux généralement déplacés vers des zones plus humides (voir annexe 3) pourrait favoriser une dispersion sur une échelle plus vaste. Dans ces conditions, la taille effective de la population reste large en saison sèche bien que les densités locales soient extrêmement faibles, les individus étant répartis sur une surface géographique importante. Dès lors, une forte structure génétique à l'échelle microgéographique, caractéristique des métapopulations est improbable. Nous avons testé cette hypothèse en comparant la population d'*An. arabiensis* de Barkedji à une population prélevée dans un autre village du Sénégal.

4. Structure génétique d'*An. arabiensis* au Sénégal

Afin d'estimer le degré de différenciation génétique des populations d'*An. arabiensis* sur une échelle géographique réduite, nous avons comparé la population de

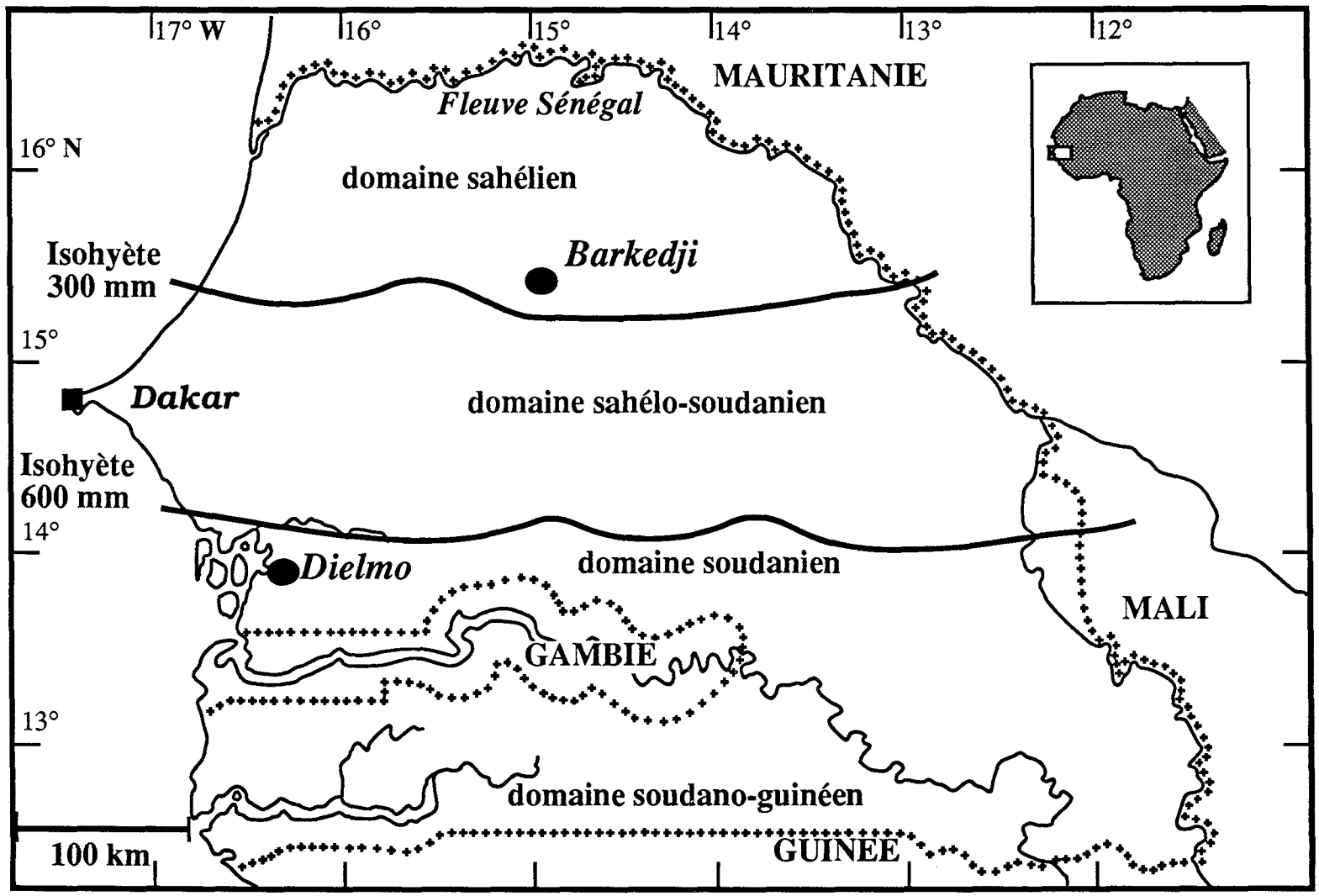


Figure 8: Localisation géographique du village de Dielmo, 250 km au Sud-Ouest de Barkedji.

Barkedji prélevée en Août 1996 à une population collectée à la même date dans un autre village du Sénégal, Dielmo, localisé à 250 km au sud-ouest de Barkedji, à la frontière Nord avec la Gambie (voir Figure 8). Dielmo est situé sur la rive d'une petite rivière pérenne, la Néma, qui fournit des gîtes larvaires tout au long de l'année aux trois espèces de vecteurs présents: *An. gambiae*, *An. arabiensis* et *An. funestus*. Par conséquent, bien que des fluctuations saisonnières de la densité des populations soient également observées à Dielmo, elles sont nettement plus limitées qu'à Barkedji. La biologie de la reproduction d'*An. arabiensis* fonctionne toute l'année selon la succession de cycles gonotrophiques classiques (Fontenille *et al.*, 1997).

Le polymorphisme génétique observé sur les huit loci polymorphes à Dielmo est comparable aux résultats obtenus pour la population de Barkedji (nombre moyen d'allèles par locus=7,25; hétérozygotie moyenne observée=0,515; hétérozygotie moyenne calculée=0,608; Tableau 7). Une forte déviation par rapport aux proportions d'Hardy-Weinberg a été observée aux loci 24D et 141, confirmant la présence possible d'allèles nuls au locus 24D. Aucun déséquilibre de liaison significatif entre paires de loci n'a été révélé à l'intérieur de la population.

Une différenciation génétique très faible mais néanmoins significative a été révélée entre ces deux populations par Fst mais pas par Rst (Fst=0,012, P=0,013; Rst=0,009, P=0,2; Tableau 8). Ce très faible taux de différenciation se traduit par des estimations de l'index de migration nettement supérieures à 1 (Nm=21,4 et Nm=28,1 dérivé des estimations de Fst et de Rst respectivement, voir Tableau 8) suggérant que ces deux populations soient reliées par un très fort taux de flux géniques.

Le profil observé soutient l'hypothèse de populations larges, réparties de façon continue dans l'espace, plus compatible avec un modèle démographique de voisinage (isolement par la distance) qu'avec un modèle en îles ou de stepping stone qui

Tableau 7: Variabilité génétique intra-population.

	Barkedji				Dielmo			
Locus	Nall	H_o	H_e	Fis	Nall	H_o	H_e	Fis
7	8	0,620	0,587	-0,057	8	0,653	0,692	0,057
24D	5	0,360	0,489	0,265	7	0,388	0,587	**0,342**
147	6	0,660	0,719	0,083	6	0,563	0,739	0,241
141	8	0,520	0,636	0,184	6	0,469	0,646	0,276
26	14	0,813	0,884	0,081	15	0,783	0,856	0,087
803	7	0,600	0,621	0,035	8	0,673	0,774	0,131
45C	6	0,460	0,419	-0,098	6	0,429	0,419	-0,022
29C	2	0,140	0,165	0,155	2	0,163	0,151	-0,079
Tous loci	7	0,522	0,565	0,078	7,25	0,515	0,608	**0,155**

Nall: Nombre d'allèles par locus; H_0: hétérozygotie observée; H_e: hétérozygotie calculée.
Valeurs soulignées P<0,05; caractères **gras soulignés** P<0,01.

Tableau 8: Différenciation génétique entre les populations d'*An. arabiensis* de Barkedji et de Dielmo.

Locus	Fst	P	Nm	Rst	P	Nm
7	0,005	0,47	51	-0,005	0,61	>>1
24D	0,029	0,14	8,4	-0,004	0,59	>>1
147	-0,007	0,93	>>1	-0,010	0,99	>>1
141	0,018	0,18	13,5	0,056	0,054	4,2
26	0,002	0,69	126	0,004	0,38	58
803	0,022	0,16	11,2	0,007	0,30	36
45C	0,009	0,31	26,1	0,013	0,27	19,5
29C	-0,010	0,87	>>1	-0,010	0,85	>>1
Tous loci	0,012	0,013	21,4	0,009	0,20	28,1

P: Signification statistique après permutation des génotypes entre populations.
Valeur soulignée P<0,05.

considèrent des populations locales isolées géographiquement. Le fait que nous n'ayons testé que deux populations ne nous permet cependant pas de pousser plus avant cette analyse. Une analyse plus fine de la structure des populations du complexe *An. gambiae* au Sénégal est actuellement en cours au laboratoire.

5. Conclusion

5. 1. - Fiabilité de l'estimation de Ne

Il nous paraît judicieux à ce niveau de revenir sur les hypothèses de base du modèle (recensées par Waples 1989a, 1991) afin de vérifier dans quelles mesures ces conditions peuvent être respectées dans notre étude et de souligner les biais éventuellement introduits. Ces différents points sont :

- L'échantillonnage aléatoire et homogène de toute la population

Cette condition semble remplie dans notre cas car aucune indication de déviation par rapport à la panmixie n'a été révélée dans la population de Barkedji, par le test de Hardy-Weinberg ou le test du déséquilibre de liaison entre paires de loci. Aucune structure génétique interne n'a été détectée quelque soit la date de collecte.

- Des générations discrètes

Si l'on considère le fort taux de croissance de la population en début de saison des pluies, cette condition a de grandes chances de ne pas être vérifiée. Néanmoins la méthode temporelle est relativement robuste par rapport à cette condition et les biais introduits sont négligeables si l'intervalle de temps considéré dépasse une génération (Waples, 1989a; Jorde et Ryman, 1996). De plus, si cette condition ne tient pas,

l'estimation de Ne obtenue sera biaisée vers le bas, ce qui confirme nos conclusions d'absence de minima extrêmes.

Un biais important peut par contre être lié à l'estimation incorrecte du nombre de générations *t* couvertes par l'intervalle de temps étudié. En effet, comme nous l'avons vu, *t* est seul au numérateur de la fraction permettant de déterminer Ne. Les variations sur l'estimation de Ne sont donc directement proportionnelles à la valeur de *t* qui ne peut être qu'estimé approximativement pour des populations en conditions naturelles. Le nombre de générations que nous avons estimé pour déterminer Ne (12 générations par an) est vraisemblablement la limite minimale de ce qui peut être observé dans cette région. En effet, dans des conditions optimales, en insectarium, un cycle complet de développement chez *An. arabiensis* peut être réalisé en 16 jours. Dans ce cas également, l'estimation de Ne présentée sera biaisée vers le bas, la taille effective réelle des population étant certainement supérieure.

- La neutralité des marqueurs

Comme nous l'avons dit en introduction, les marqueurs microsatellites sont généralement considérés comme neutres bien qu'ils soient soumis à différentes formes de contraintes dont l'influence reste encore mal connue. Néanmoins, l'existence d'une force de sélection constante n'aura que peu d'effet sur l'estimation de Ne résultant de l'analyse temporelle si t/Ne est faible (Pollak, 1983) car les variations de fréquences alléliques resteront imputables essentiellement à la dérive génétique et résulteront alors de fluctuations stochastiques autour de la valeur d'équilibre. De plus, l'influence de la sélection naturelle ne saurait toucher dans les mêmes proportions tous les loci étudiés.

- L'absence de migration

D'après nos résultats suggérant des flux de gènes importants entre Barkedji et Dielmo, cette condition est de toute évidence violée. L'estimation de Ne peut être

biaisée si des migrants provenant de sous-populations adjacentes où les fréquences alléliques sont différentes se sont établis au cours de l'intervalle de temps considéré car les variations décelées ne seront plus exclusivement dues à la dérive génétique. Dans notre cas, l'influence de la migration doit être limitée étant donné la grande similarité des fréquences alléliques observées entre les populations de Dielmo et de Barkedji, distantes de 250 km.

- L'absence de mutation

Il est raisonnable d'estimer que l'influence de la mutation est négligeable étant donné le faible intervalle de temps considéré (40 mois), même si l'on considère des loci avec un taux de mutation élevé tels que les microsatellites (Pollak, 1983; Waples, 1989a; Slatkin, 1995). De plus, les estimations de Rst obtenues lors des comparaisons des différents échantillons sont toujours négligeables et non significatives.

En conséquence, les estimations de la taille effective de la population de Barkedji que nous avons obtenues lors de cette étude peuvent être considérées comme une indication relativement fiable de la taille minimale atteinte en saison sèche. Il convient toutefois de rester prudent quant à l'estimation précise du nombre d'individus constituant la population résiduelle en saison sèche. Conformément aux résultats obtenus par Taylor *et al.* (1993) notre étude soutient que les populations d'*An. arabiensis* d'Afrique de l'Ouest soient maintenues lors de la saison sèche à des effectifs non négligeables impliquant plusieurs centaines à quelques milliers d'individus. La concordance entre ces estimations fournies par des marqueurs moléculaires différents est également un argument en faveur de la fiabilité des résultats obtenus.

5. 2. - Conclusion générale

L'analyse de la variation temporelle du polymorphisme des marqueurs microsatellites au sein de la population d'*An. arabiensis* de Barkedji suggère la présence permanente des vecteurs sur le site malgré de très fortes fluctuations saisonnières de la densité de population. Le dème, large et dense en saison des pluies reste conséquent en saison sèche, les individus étant vraisemblablement dispersés sur une surface géographique importante. Nos données génétiques et écologiques (indirectes et directes) semblent favoriser l'hypothèse d'estivation des femelles comme mécanisme permettant la survie de la population en saison sèche. Alternativement, étant donné le potentiel de dispersion des individus et la nature ouverte du terrain, les populations pourraient être maintenues par l'intermédiaire d'individus tirant profit de quelques rares points d'eau encore disponibles en saison sèche mais de tels gîtes n'ont pas été trouvés dans la zone étudiée. L'hypothèse d'extinction de la population en saison sèche et de recolonisation annuelle des gîtes temporaires n'est pas soutenue par nos données. Les populations séparées par plus de 200 km sont très peu différenciées génétiquement ce qui soutient l'hypothèse de dèmes larges aux frontières géographiques imprécises, interconnectés par des flux de gènes importants reflétant une forte capacité de dispersion des individus.

Ces considérations doivent être prises en compte pour parvenir à un contrôle optimal des populations d'*An. arabiensis*. En effet, il apparaît très improbable que des protocoles de lutte basés sur des traitements majoritairement intra-domiciliaires et centrés sur des zones géographiques limitées (à l'échelle du village ou d'une agglomération) permettent une action efficace à long terme sur les populations d'*An. arabiensis*. La présence toujours menaçante d'*An. arabiensis* à La Réunion et à Maurice en dépit d'une stratégie de lutte menée dans des conditions optimisées (voir première partie) témoigne de la nécessité de maintenir un suivi minutieux et régulier des sites

traités pour parvenir au contrôle de la transmission du paludisme par ce vecteur. De plus, la nature temporaire des gîtes larvaires exploités par cette espèce, la diversité et l'ubiquité des gîtes de repos (et d'estivation) extérieurs sont autant d'entraves à un traitement optimal par les moyens dont nous disposons actuellement. La capacité de ces vecteurs à survivre sans être étroitement associés aux populations humaines leur fournit un excellent moyen de contourner les tentatives d'éradication menées jusqu'à présent, en trouvant refuge dans des zones sauvages non traitées ou en adoptant un comportement alternatif limitant leur exposition aux insecticides. Nous pouvons citer pour exemple le «changement de comportement des vecteurs» noté à Madagascar suite aux campagnes de lutte des années 50 (Chauvet *et al.*, 1964; Chauvet et Rajaonarivelo, 1973; Ralisoa Randrianasolo et Coluzzi, 1987; Mouchet *et al.*, 1997) qui reflète plus certainement l'élimination d'*An. gambiae* et la sélection d'une population d'*An. arabiensis* au comportement plus exophile et zoophile.

Le taux de flux géniques important reliant des populations éloignées géographiquement est également un élément dont l'influence est primordiale en termes de gestion des campagnes de lutte. La dispersion des gènes de résistance aux insecticides est alors à craindre mais, parallèlement, le libre échange des gènes peut apparaître comme un avantage pour l'application de nouvelles stratégies de lutte basées sur l'utilisation d'individus génétiquement modifiés. Indépendamment d'autres facteurs limitant d'ordre pratiques (moléculaires) ou éthiques liés à la mise en œuvre de ces nouvelles technologies (Crampton *et al.*, 1992; Ashburner *et al.*, 1998), ces méthodes de contrôle indirectes reposant sur la diffusion de transgènes à l'intérieur des populations naturelles pourraient apparaître comme le moyen privilégié permettant un contrôle effectif des populations d'*An. arabiensis*.

TROISIEME PARTIE

Etude de la différenciation génétique entre les formes chromosomiques Mopti et Savane chez *Anopheles gambiae* s.s.

L'étude présentée dans cette troisième partie concerne exclusivement le taxon nominal *An. gambiae* s.s.. Comme nous l'avons vu précédemment, une structure génétique très hétérogène a été révélée au sein de cette espèce par l'analyse du polymorphisme des inversions paracentriques observées sur les chromosomes polytènes. Alors que chez *An. arabiensis*, ce système apparaît généralement en équilibre d'Hardy-Weinberg, une image beaucoup plus contrastée ressort des études menées sur *An. gambiae* (Coluzzi *et al.*, 1979; Coluzzi, 1984). Nous allons brièvement rappeler, dans un chapitre introductif les bases de cette hétérogénéité et présenter de nouveaux arguments ayant motivé notre investigation.

1. Introduction

1. 1. - Compléments sur les études cytogénétiques

En Afrique de l'Ouest, l'observation au sein de populations naturelles d'*An. gambiae* s.s. de forts déficits voire de l'absence complète de certains hétérozygotes impliquant des arrangements chromosomiques particuliers localisés notamment sur le chromosome 2 (bras R et L) ont conduit à la désignation de 5 formes, ou cytotypes différents correspondant à des unités de croisement *a priori* panmictiques (Touré *et al.*, 1983; Coluzzi *et al.*, 1985). La forme Forêt, adaptée aux zones de forêt tropicale ou de savane humide présente majoritairement l'arrangement standard $2R^+$-$2L^+$. La forte fréquence de l'inversion 2Rd dans les zones côtières est caractéristique de la forme Bissau. Dans les zones de savanes plus arides, trois formes peuvent être rencontrées: Savane, Mopti et Bamako. Ces trois formes présentent en forte fréquence l'arrangement 2La et se différencient par un polymorphisme important sur le bras 2R. Néanmoins, les caryotypes standards $2R^+$ et $2L^+$ sont également observés mais en très faible fréquence.

Tous ces arrangements sont donc polymorphes chez ces trois formes. La forme Savane, caractérisée par l'inversion 2Rb, présente une très vaste répartition géographique sur toute la ceinture de savanes sahéliennes et soudaniennes, au Sud du Sahara. C'est apparemment la seule forme chromosomique présente dans les savanes d'Afrique de l'Est. Tous les spécimens *An. gambiae* collectés à l'Est de la vallée du Rift appartiennent au cytotype Savane. Les formes Mopti et Bamako, caractérisées par les inversions 2Rbc/u et 2Rj respectivement, sont beaucoup plus localisées (Mali, Burkina Faso, Côte d'Ivoire et Guinée Conakry) et apparaissent souvent en sympatrie avec la forme Savane. La distribution spatiale et temporelle des différents arrangements polymorphes à l'intérieur de ces formes chromosomiques a été mise en relation avec certains paramètres environnementaux, notamment le degré d'aridité, et il est maintenant établi que ces inversions jouent effectivement un rôle non négligeable dans la capacité d'adaptation des individus à un environnement particulier (Coluzzi *et al.*, 1979, 1985; Bryan *et al.*, 1982; Touré *et al.*, 1983, 1998). A titre d'exemple, l'inversion 2La, fixée chez *An. arabiensis* et polymorphe chez *An. gambiae* semble associée à la capacité de coloniser un habitat relativement aride (Coluzzi *et al.*, 1985). Il en est de même pour l'inversion 2Rbc chez Mopti dont la fréquence au sein de populations du Mali varie énormément et très rapidement en fonction des conditions écologiques du milieu, jusqu'à fixation en saison sèche dans les zones les plus arides (Touré *et al.*, 1998).

Dans les zones de sympatrie, les formes Savane, Bamako et Mopti apparaissent reproductivement isolées. Toutefois, cet isolement génétique semble incomplet. En effet, seuls les hétérocaryotypes correspondant aux hybrides Mopti-Bamako n'ont jamais été observés alors que des hybrides naturels Mopti-Savane ont été reportés dans plusieurs études (Coluzzi *et al.*, 1985; Robert *et al.*, 1989; Lanzaro *et al.*, 1998). Pourtant, toutes ces formes sont parfaitement interfécondes au laboratoire et les

hybrides obtenus, parfaitement viables et fertiles, sont porteurs des hétérocaryotypes rarement ou jamais observés dans la nature ce qui suggère l'absence de barrière post-copulatoire (Di Deco *et al.*, 1980; Persiani *et al.*, 1984; Favia *et al.*, 1997). L'hypothèse d'une forte barrière pré-copulatoire, probablement d'ordre comportemental est alors la plus vraisemblable (Coluzzi *et al.*, 1985). Récemment, la mise au point d'un test diagnostique basé sur l'amplification spécifique par PCR d'un fragment d'ADN ribosomique localisé sur le chromosome X et de son hydrolyse par certains enzymes de restriction (PCR-RFLP) a montré que la plupart de ces hybrides observés sur le terrain peuvent être rattachés à l'une ou l'autre des formes parentales (Favia *et al.*, 1997). Les hybrides obtenus en insectarium présentent quant à eux les caractéristiques des deux parents. Cette nouvelle approche suggère que l'analyse cytogénétique surestime probablement la fréquence des croisements entre Mopti et Savane et soutient l'hypothèse de spéciation et d'isolement génétique entre les différentes formes chromosomiques, en conformité avec ce que l'on attend d'espèces différentes.

1. 2. - Nouvelles données et présentation de l'étude

Un argument supplémentaire confortant cette hypothèse de spéciation a très récemment été rapporté par Chandre *et al.* (1999a) lors d'une étude la prévalence de la mutation *kdr* (Knock Down Resistance) chez *An. gambiae* menée à l'échelle du continent Africain.

La mutation *kdr* confère aux insectes qui en sont porteurs une résistance croisée aux pyréthrinoides et au DDT, par modification de leur cible commune, le canal sodium voltage dépendant (Pauron *et al.*, 1989; Williamson *et al.*, 1996). Cette mutation a récemment été identifiée chez *An. gambiae* et un test PCR basé sur l'amplification spécifique des allèles sensible et résistant (muté) a été mis au point (Martinez-Torres *et*

al., 1998). Elle est autosomale mais sa localisation cytologique précise est encore inconnue.

Au cours de leurs investigations, Chandre *et al.* (1999a) ont remarqué un fort déficit, et dans certains cas l'absence complète d'hétérozygotes pour l'allèle *kdr*, en présence d'homozygotes sensibles et résistants dans certaines zones de savane de Côte d'Ivoire et du Burkina Faso où les formes Mopti et Savane sont sympatriques. L'identification des spécimens par le test PCR-RFLP (Favia *et al.*, 1997) a révélé que seuls les individus Savane étaient porteurs de l'allèle muté, l'absence d'hétérozygotes dans certaines localités résultant alors de la coexistence d'une population Savane homozygote pour l'allèle muté résistant et d'une population Mopti homozygote sensible. Sur un total de 18 localités prospectées, aucun individu Mopti n'a été trouvé porteur de la mutation *kdr* alors que les spécimens Savane peuvent être homozygotes résistants, hétérozygotes ou homozygotes sensibles, l'équilibre d'Hardy-Weinberg étant respecté au sein des populations Savane (Chandre *et al.*, 1999a). Cette distribution remarquable de la mutation *kdr* apparaît donc comme un argument très fort soutenant l'hypothèse de spéciation entre les formes Mopti et Savane. Parallèlement, aucun hybride Savane-Mopti n'a été identifié par PCR-RFLP parmi les spécimens collectés au cours de cette étude.

Les fréquences de la mutation *kdr* observées chez Savane dépassent souvent 90% (Chandre *et al.*, 1999a) suggérant une pression de sélection importante en faveur des individus résistants dans cette zone rizicole où les insecticides sont très utilisés, tant au niveau domestique qu'en agriculture. Il apparaît alors difficilement concevable dans ces conditions que Mopti n'ait pas acquis cette mutation si des flux de gènes ont effectivement lieu entre les deux formes chromosomiques. De plus, des croisements au laboratoire ont permis de vérifier que la mutation *kdr* peut être effectivement transmise de Mopti à Savane via des hybrides obtenus expérimentalement (Chandre *et al.*, 1999a).

Si l'explication de ce profil de distribution hétérogène de la mutation *kdr* réside effectivement dans l'existence réelle d'une forte tendance à la reproduction entre individus appartenant à la même forme chromosomique (assortative mating), l'influence de cette barrière au flux de gènes doit alors s'exercer de façon uniforme sur tout le génome. En conséquence, le polymorphisme neutre des microsatellites devrait indiquer une hétérogénéité dans la composition génétique de ces deux formes. Nous avons ainsi tenté de vérifier cette hypothèse en estimant le taux de différenciation génétique entre populations de Savane (résistantes) et de Mopti (sensibles) provenant de cette zone de sympatrie grâce aux 9 loci microsatellites que nous avons précédemment utilisés chez *An. arabiensis*.

Au total, 97 spécimens comprenant 60 Savane et 37 Mopti identifiés par le test PCR-RFLP, nous ont été fournis. Ces moustiques ont été collectés en Juillet 1998, sur sujets humains volontaires dans trois villages de Côte d'Ivoire: Kabolo, Piétonara et Ounantiékara. A Ounantiékara, seule la forme Savane a été trouvée. Mopti étant sensible aux insecticides, est extrêmement faiblement représenté dans les zones irriguées où l'utilisation des insecticides est intense (culture du riz et du coton). En conséquence, les populations sympatriques Mopti-Savane ont toutes été échantillonnées dans des zones où la pression d'insecticide est la plus faible, les zones rurales R0, c'est à dire sans culture de riz, selon la nomenclature de Chandre *et al.* (1999a).

2. Résultats

2. 1. - Distribution des fréquences alléliques

Contrairement à ce qui est observé chez *An. arabiensis*, le locus 49 s'est révélé très polymorphe au sein des deux formes chromosomiques Savane et Mopti (11 allèles

différents observés). Par contre, le locus 141, localisé sur le bras 2R, a révélé des profils de migration très ambigus qui ne nous ont pas permis de déterminer précisément le génotype des moustiques à ce locus. Zheng *et al.* (1996) l'avaient d'ailleurs classé comme locus monomorphe chez *An. gambiae* lors de l'identification originale des loci microsatellites chez cette espèce. Nous n'avons donc pas inclus ce locus dans l'analyse.

L'homogénéité de la répartition des fréquences alléliques entre les différents échantillons a été testée grâce au test exact de Raymond et Rousset (1995a) afin d'étudier l'influence relative du site de capture et de la forme chromosomique sur la répartition de la variabilité génétique. Les résultats de cette analyse sont présentés dans le Tableau 9. A l'évidence, les populations apparaissent plus homogènes au sein de la même forme chromosomique (Tableau 9, section «influence du site de capture») qu'au sein d'une même localité (Tableau 9, section «influence de la forme chromosomique»). En effet, alors que seules quelques différences significatives au seuil de 5% apparaissent lorsque les populations appartenant au même cytotype sont comparées entre lieux de capture, une hétérogénéité plus importante apparaît lorsque les différents cytotypes sont comparés aussi bien au niveau du village (Kabolo ou Piétonara car seuls des individus Savane ont été collectés à Ounantiékara) que sur la totalité de la zone, plusieurs différences dans la répartition des fréquences alléliques étant très significatives ($P<0,0034$). Il est intéressant de noter que ces différences semblent s'exprimer particulièrement sur le chromosome X, homogène entre Mopti et Savane au niveau des inversions mais siège du polymorphisme discriminant ces deux formes par PCR-RFLP (Favia *et al.*, 1997). Les loci 24D et 26, respectivement localisés à l'intérieur des inversions 2La et 2Rb montrent également une répartition hétérogène des fréquences alléliques entre formes chromosomiques, bien que cette hétérogénéité ne soit pas statistiquement significative lorsque la procédure de Bonferroni est appliquée. Dans tous les cas, les loci situés sur le chromosome 3 ne révèlent aucune hétérogénéité.

Tableau 9: Test d'homogénéité des fréquences allèliques entre populations.

Locus		Influence du site de capture				Influence de la forme chromosomique			
		Savane (3 localités)	Savane (2 localités)*	Mopti (2 localités)	Combiné (2 localités)*	Kabolo	Piétonara	Combiné (2 localités)*	Combiné (total)
	n[¥]	18/24/18	18/18	25/12	43/30	18/25	18/12	36/37	60/37
X									
	7	0,0738	0,0528	0,3419	0,6675	0,0530	**0,0034**	0,0164	**0,0018**
	49	0,3877	0,3535	1,0000	0,9017	**<0,0001**	0,0344	**<0,0001**	**<0,0001**
II									
	24D	0,0150	0,2095	0,6192	0,6547	0,0203	0,2885	0,0252	0,0314
	147	0,0644	0,0616	0,9083	0,2211	0,1177	0,8363	0,4222	0,1822
	26	0,0113	0,2378	0,0143	0,1509	**0,0002**	0,1108	**<0,0001**	0,0140
	803	0,9334	0,8260	0,1711	0,1089	0,6639	0,7375	0,2802	0,1351
III									
	45C	0,1550	0,6660	0,2441	0,3728	0,2503	0,3348	0,0855	0,0579
	29C	0,4808	0,4261	0,7875	0,8598	0,3437	0,7856	0,5814	0,3328
Tous loci[§]		**0,004**	0,1432	0,2687	0,5246	**<0,0001**	0,0239	**<0,0001**	**<0,0001**

Les données ont été analysées afin d'étudier l'effet du site de capture (différenciation microgéographique) ou des formes chromosomiques (différenciation interformes).

Les loci sont disposés selon le chromosome sur lequel ils sont localisés (linkage group).

* Test de l'homogénéité de la répartition des fréquences allèliques dans les deux localités (Kabolo et Pietonara) où les formes Savane et Mopti ont été prélevées simultanément (à Ounantiékara, seuls des individus Savane ont été collectés).

[¥] Rapport des effectifs des différents échantillons comparés.

[§] Test combiné de Fisher sur les données fournies par tous les loci (Raymond et Rousset, 1995a).

Valeurs soulignées : P<0,05.

Valeurs en **gras soulignées** : P<0,05 après application la procédure de Bonferroni.

2. 2. - Test de la panmixie

Afin de tester l'hypothèse de panmixie à l'intérieur des différents cytotypes ainsi que sur la population totale, une analyse hiérarchique de conformité par rapport aux proportions attendues à l'équilibre d'Hardy-Weinberg a été effectuée à l'échelle du village et sur la totalité de la zone échantillonnée. Les résultats sont présentés dans les Tableaux 10 A, B et C. Globalement, c'est lorsque les deux formes chromosomiques sont regroupées que l'on observe le plus grand nombre de déviations significatives, associées à de forts déficits en hétérozygotes (Tableau 10 C). Cependant, si ceci est particulièrement vrai dans le village de Kabolo, cette tendance est beaucoup plus limitée dans le village de Piétonara. De plus, les forts déficits en hétérozygotes observés sur la population totale aux loci 7 et 26, se retrouvent également lorsque les différentes formes chromosomiques sont considérées indépendamment, le locus 7 déviant significativement par rapport aux proportions attendues sous l'hypothèse de panmixie dans la population Savane (Tableau 10 A) et le locus 26 présentant le même profil au sein de la population Mopti (Tableau 10 B). D'autres déséquilibres mineurs sont observés ponctuellement aux loci 24D chez Mopti et 147 chez Savane, ainsi que sur la population totale.

Il apparaît alors que les déviations par rapport aux proportions attendues sous l'équilibre d'Hardy-Weinberg observées dans la population totale d'*An. gambiae* étudiée ici ne soient pas dues à un artefact expérimental résultant du regroupement des formes Mopti et Savane (effet Wahlund) puisque ces déviations sont effectivement observées lorsque ces sous-populations sont étudiées isolément, associées à des déficits en hétérozygotes d'amplitude comparable (valeur de Fis).

Néanmoins, certains déséquilibres semblent spécifiquement observés chez l'une ou l'autre des formes chromosomiques ce qui pourrait indiquer une hétérogénéité de la composition génétique des deux formes (influence des inversions ou fréquence des

Tableau 10 : Test exact de conformité aux proportions d'Hardy-Weinberg entre et au sein des différentes formes chromosomiques chez *An. gambiae*. A: au sein du cytotype Savane; B: au sein du cytotype Mopti; C: sans distinction du cytotype.

A.

		Cytotype Savane							
		Kabolo		Piétonara		Ounantiékara		Savane total	
Locus		P	Fis	P	Fis	P	Fis	P	Fis
	n	18		17		24		59	
X									
	7	0,0113	0,318	0,1367	0,248	**0,0048**	0,413	**<0,0001**	0,330
	49	0,5820	0,047	0,8354	-0,181	0,5494	-0,165	0,8491	-0,113
II									
	24D	0,1714	0,205	0,7229	0,173	0,0849	0,245	0,0635	0,220
	147	0,1403	0,141	**0,0023**	0,344	0,0982	0,223	0,0112	0,243
	26	0,3073	0,148	0,5629	0,098	0,2635	0,293	0,1895	0,207
	803	0,1299	0,074	0,3646	0,167	0,1817	0,075	0,0809	0,094
III									
	45C	0,0671	-0,074	0,1045	0,300	0,8159	-0,068	0,0938	0,044
	29C	0,1587	0,382	1,0000	0,029	0,0627	0,428	0,0471	0,268

B.

		Cytotype Mopti					
		Kabolo		Piétonara		total	
Locus		P	Fis	P	Fis	P	Fis
	n	25		12		37	
X							
	7	0,0334	0,284	0,3144	0,217	0,0514	0,265
	49	0,2426	0,062	0,6872	0,099	0,7367	0,061
II							
	24D	0,0230	0,156	0,7071	-0,031	0,0114	0,104
	147	0,4869	-0,052	0,0656	0,185	0,4018	0,014
	26	0,0190	0,327	0,0075	0,430	**<0,0001**	0,381
	803	0,2559	-0,120	0,1058	0,286	0,1671	0,034
III							
	45C	1,0000	-0,115	0,2702	-0,204	0,5688	-0,140
	29C	0,3770	0,217	1,0000	0,035	0,4558	0,150

C.

		An. gambiae s.s.					
		Kabolo		Piétonara		total	
Locus		P	Fis	P	Fis	P	Fis
	n	43		29		96	
X							
	7	**0,0004**	0,311	0,0111	0,257	**<0,0001**	0,319
	49	0,3937	0,133	0,7848	-0,033	0,4984	0,019
II							
	24D	**0,0034**	0,182	0,7709	0,124	0,0192	0,180
	147	0,6353	0,030	**0,0066**	0,271	0,0408	0,159
	26	**0,0040**	0,305	0,1070	0,230	**0,0078**	0,279
	803	0,1955	-0,038	0,0764	0,201	0,1305	0,075
III							
	45C	0,5457	-0,096	0,3617	0,106	0,1541	-0,018
	29C	0,1315	0,278	1,0000	0,016	0,0440	0,218

n = Effectif de l'échantillon.
P = Probabilité de conformité à l'équilibre d'Hardy-Weinberg calculée par le test exact de Raymond et Rousset (1995a). Valeurs soulignées : P<0,05. Valeurs en **gras soulignées** : P<0,05 après application de la procédure de Bonferroni.

allèles nuls par exemple). Cependant, les faibles effectifs dont nous disposions ainsi que l'absence de données cytogénétiques ne nous ont pas permis de préciser davantage cette éventualité.

Aucun déséquilibre de liaison significatif entre paires de loci n'a été mis en évidence tant sur la population totale qu'à l'intérieur des différents cytotypes (P>0,11) suggérant une répartition aléatoire des génotypes multi-loci entre et au sein des différentes formes chromosomiques.

En conséquence, à ce niveau de l'analyse, aucune indication ne semble soutenir de façon probante l'hypothèse de flux géniques restreints entre Mopti et Savane. Toutefois, la composition génétique de la population apparaît plus hétérogène lorsque aucune distinction entre cytotypes n'est effectuée qu'à l'intérieur des deux formes chromosomiques individualisées.

2. 3. - Différenciation génétique entre formes chromosomiques

Au vu des résultats présentés dans les paragraphes précédents, nous avons décidé de regrouper les individus selon leur cytotype et d'effectuer une analyse par les statistiques F et R afin d'estimer et de comparer le degré de différenciation génétique observé entre et au sein de chaque forme chromosomique. Les résultats de l'analyse hiérarchique de variance effectuée sur les fréquences alléliques (méthode F) ou sur la taille des allèles (méthode R) sont présentés dans le Tableau 11 A.

Ces résultats sont tout à fait comparables quelle que soit la méthode utilisée et indiquent que seule une très faible proportion de la variabilité totale observée est imputable à une différenciation Mopti/Savane (2,76 et 2,02% respectivement pour les méthodes F et R). Cette variation n'est pas significative (P>0,1). Toutefois, la variabilité intra-formes chromosomiques est sensiblement moindre,

Tableau 11: Analyse de la différenciation génétique chez *An. gambiae s.s.*

A. Analyse hiérarchique de la variance moléculaire (AMOVA) entre les cytotypes Mopti et Savane.

Méthode	Source	ddl	SSD	variance	% total	P
F						
	Entre cytotypes	1	10,421	0,0764	2,76	0,112
	Entre populations appartenant au même cytotype	3	10,044	0,0182	0,66	0,281
	Intra-populations	189	504,922	2,6715	96,58	<0,0001
	Totale	193	525,387	2,7661	-	
R						
	Entre cytotypes	1	74,36	0,487	2,02	0,101
	Entre populations appartenant au même cytotype	3	87,51	0,153	0,63	0,298
	Intra-populations	189	4440,24	23,493	97,35	0,056
	Totale	193	4602,11	24,133	-	

F: Analyse de variance effectuée sur les fréquences alléliques (Identité – Non Identité des allèles).
R: Analyse de variance effectuée sur la taille des allèles (nombre de répétitions du motif).
P: Signification statistique basée sur la permutation des génotypes entre populations aux différents niveaux hiérarchiques (Schneider *et al.*, 1997).

B. Indices de fixation.

Savane:	Fst = 0,00684	P = 0,2796	(3 sous-populations)
	Rst = 0,01589	P = 0,1408	
Mopti:	Fst = 0,00705	P = 0,289	(2 sous-populations)
	Rst = -0,0233	P = 0,935	
Savane x Mopti:	Fst = 0,0306	P<0,0001	(2 sous-populations)
	Rst = 0,023	P = 0,024	

expliquant environ 0,6% de la variabilité totale, l'essentiel de la variabilité étant généré au niveau intra-populationnel (variations entre individus au sein de la même population).

Les indices de fixation estimés au sein des cytotypes Savane d'une part et Mopti d'autre part, ainsi qu'entre les deux formes chromosomiques sont présentés dans le Tableau 11 B. Aucune différenciation génétique significative n'est révélée entre les différentes populations appartenant au même cytotype (Fst<0,007; P>0,27 et Rst<0,016; P>0,14). Par contre, une différenciation génétique, certes faible mais néanmoins statistiquement significative est observée entre Mopti et Savane (Fst=0,0306; $P<10^{-4}$ et Rst=0,023; P=0,024).

L'analyse de la contribution de chaque locus au profil observé est présentée dans le Tableau 12. Les loci portés par le chromosome X et, dans une moindre mesure par le chromosome 2 semblent déterminants pour la mise en évidence de différences entre les formes Mopti et Savane.

En accord avec le faible degré de différenciation génétique observé, les indices de migration (Nm) dérivés des estimations de Fst et de Rst sont nettement supérieurs à 1, suggérant des flux géniques importants entre Mopti et Savane. Les valeurs de Nm ainsi estimées sont considérablement plus élevées que celles attendues entre populations appartenant à des espèces différentes.

L'analyse de la variabilité génétique observée au niveau des huit loci microsatellites utilisés lors de cette étude n'a donc pas permis de mettre en évidence l'existence d'une forte structure génétique au sein des populations d'*An. gambiae* présentes dans notre zone d'étude, malgré les très fortes indications de structuration fournies par la cytogénétique (Coluzzi *et al.*, 1985), l'étude des profils de restriction de certaines régions d'ADN codant pour les ribosomes (Favia *et al.*, 1997) et la répartition

Tableau 12: Différenciation génétique entre les formes Savane et Mopti.

	Statistiques F			Statistiques R		
Locus	Fst	P	*Nm*	Rst	P	*Nm*
X						
49	**0,1306**	**<0,0001**	1,65	0,0331	0,139	7,3
7	0,0447	0,057	5,4	0,0022	0,400	113,5
II						
24D	0,0040	0,436	62,3	0,0108	0,317	23,0
147	0,0039	0,703	64,5	0,0397	0,129	6,0
26	0,0210	0,175	11,7	0,0494	0,089	4,8
803	0,0062	0,420	40	0,0429	0,107	5,6
III						
45C	0,0106	0,359	23,3	0,0007	0,420	372
29C	-0,0003	0,510	∞	-0,0003	0,493	∞
Chromosome X	**0,0902**	**<0,0001**	2,5	0,0188	0,122	13,1
Chromosome II	0,0092	0,070	26,9	0,0267	0,045	9,1
Chromosome III	0,0068	0,465	37,2	0,0006	0,428	427
Tous loci	**0,0306**	**<0,0001**	7,9	0,0230	0,024	10,6

Valeurs soulignées : P<0,05.
Valeurs en **gras soulignées** : P<0,0001.
Les loci sont disposés en fonction de leur localisation cytologique sur le génome. La moyenne des estimateurs du taux de différenciation génétique (Fst et Rst) est donnée pour chaque chromosome indépendamment ainsi que sur tous les loci utilisés.

hétérogène de la mutation *kdr* (Chandre *et al.*, 1999a). Néanmoins un degré d'isolement génétique plus important est révélé entre les cytotypes Savane et Mopti dans cette zone de sympatrie que ce qui a précédemment été observé entre deux populations d'*An. arabiensis* séparées par plus de 200 km au Sénégal (7 loci microsatellites sur 8 étant communs aux deux études). Ces résultats sont comparables à ceux obtenus entre les populations d'*An. gambiae* cytotype Savane du Sénégal et du Kenya (annexe 1).

Plusieurs interprétations peuvent être données à ces résultats, comme nous allons le voir maintenant.

3. Discussion

De nombreux arguments plaident en faveur de phénomènes de spéciation actuellement en cours ou très récents au sein de l'espèce *An. gambiae*. Cependant, les résultats que nous venons de présenter ne nous ont pas permis de soutenir cette hypothèse.

Des taux de différenciation comparables à ceux que nous avons observés ici ont été obtenus lors d'études utilisant les marqueurs isoenzymes pour étudier la différenciation génétique entre les mêmes populations Savane et Mopti (Fst = 0,042; *in* Chandre *et al.*, 1999a) ou entre les formes Mopti, Bamako et Savane collectées au Mali (Bullini et Coluzzi, 1978; Cianchi *et al.*, 1983, 1985; *in* Lanzarro *et al.*, 1998). Une étude basée sur des critères morphologiques (Di Deco *et al.*, 1983) a également révélé de très faibles différences entre les formes Mopti et Bamako qui, bien que statistiquement significatives, sont du même ordre de grandeur que les différences attendues entre sous-populations de la même espèce (Coluzzi *et al.*, 1985).

La seule étude de ce type utilisant les loci microsatellites a été publiée récemment par Lanzaro *et al.* (1998). Ces auteurs ont estimé le taux de différenciation génétique entre les formes Mopti et Bamako dans une zone de sympatrie au Mali grâce à 21 loci microsatellites répartis sur les trois chromosomes d'*An. gambiae*. Les taux de flux géniques estimés entre ces deux formes (Nm = 10,3 et 5,8 dérivés des estimations de Fst et de Rst respectivement) sont tout à fait comparables à ceux que nous avons obtenus entre Mopti et Savane (Tableau 12: Nm = 10,6 et 7,9 respectivement). L'utilisation d'un nombre plus important de loci a permis une analyse plus fine de la répartition de ces différences en fonction de la localisation cytologique des marqueurs. L'essentiel de la différenciation observée semble générée par les loci portés par le chromosome 2, siège des inversions ayant conduit à la désignation des différents cytotypes. Cette tendance a également été détectée au cours de notre étude bien que, dans notre cas, les loci situés sur le chromosome X se révèlent plus discriminants (Tableau 12).

La forte homogénéité des résultats obtenus par l'étude du polymorphisme de ces différents marqueurs qui peuvent être considérés comme très faiblement soumis à la sélection naturelle conforte la tendance mise en évidence par notre étude qui suggère des flux de gènes importants entre Mopti et Savane.

Les divergences dans la résolution de la structure génétique des populations d'*An. gambiae* observées en fonction des marqueurs utilisés (microsatellites, isoenzymes et critères morphologiques d'une part, inversions chromosomiques, mutation kdr et RFLP d'autre part) sont elles-mêmes très informatives et peuvent être reliées à des propriétés spécifiques des marqueurs utilisés et du modèle biologique étudié que nous allons maintenant aborder.

3. 1. - Le temps de divergence entre populations

La première hypothèse pouvant être avancée pour expliquer cette apparente contradiction est celle d'un manque de puissance des marqueurs «neutres», incluant les loci microsatellites. Or, nous avons vu au cours des chapitres précédents, que ces marqueurs très polymorphes se sont révélés particulièrement adaptés pour ce type d'investigations permettant la mise en évidence de différences importantes dans la composition génétique de populations appartenant à la même espèce (voir introduction et première partie). Mais, il a été démontré à la fois par des études expérimentales et par simulation (Slatkin *et al.*, 1995; Goldstein *et al.*, 1995b) que le potentiel résolutif des microsatellites était également très étroitement lié au temps de divergence des populations, en d'autres termes au temps écoulé depuis l'interruption (ou la forte diminution) des flux de gènes entre les différentes sous-populations considérées (voir première et deuxième parties). Il est en effet évident qu'un nombre important de générations doivent s'écouler afin que l'action combinée de la dérive génétique et des mutations génère des profils suffisamment différents pour qu'une hétérogénéité puisse être détectée, en dépit des forts taux de mutation caractéristiques des loci microsatellites.

Or, la forme chromosomique Mopti est considérée comme l'évolution la plus récente de l'espèce *An. gambiae* (mais voir paragraphe 3.3 ci-dessous) en regard de sa répartition géographique encore relativement limitée et surtout de son extraordinaire adaptation à l'environnement humain (Petrarca *et al.*, 1986, 1987; Robert *et al.*, 1989; Akogbeto *et al.*, 1992; Touré *et al.*, 1987, 1994). En effet, très anthropophile et endophile au stade adulte, ses larves se montrent particulièrement aptes à coloniser les gîtes pérennes ou sub-permanents résultant exclusivement d'actions anthropiques tels que les périmètres irrigués agricoles ou les rizières. Cette particularité lui permet de poursuivre son cycle de reproduction durant la saison sèche, alors qu'aucun gîte

plus disponible et représente ainsi un exemple d'adaptation extrêmement fine à un environnement aux caractéristiques très variables, doublée d'une aptitude très développée à tirer profit d'une étroite association avec l'homme (Touré *et al.*, 1994). Dès lors, si l'on accepte l'hypothèse de spéciation entre Mopti et Savane, force est de lui reconnaître une origine très récente.

De plus, indépendamment de la controverse introduite par Favia *et al.* (1997) au sujet du statut des spécimens porteurs des rares hétérocaryotypes potentiellement identifiés comme hybrides entre les formes Savane et Mopti par cytogénétique, le fait que ces deux formes soient effectivement totalement interfécondes en insectarium plaide également en faveur d'une spéciation très récente n'ayant pas encore entraîné d'incompatibilité génétique. Dans ces conditions, il est fort probable que le temps de divergence entre les formes Mopti et Savane ne soit pas suffisamment long pour permettre la détection de différences au niveau de la répartition du polymorphisme des microsatellites, et à plus forte raison des loci isoenzymes qui présentent des taux de mutation nettement moins élevés.

Dés lors, et si l'on s'en réfère à ce qui a été dit dans le chapitre introductif sur les contraintes possibles agissant sur la taille des allèles (l'homoplasie en particulier) qui limitent leur application à l'étude des phylogénies interspécifiques (Garza *et al.*, 1995), le potentiel résolutif des microsatellites pourrait n'être optimal que dans un intervalle relativement restreint de temps de divergence entre les populations. Dans le cas du système étudié ici, on se trouverait donc en dessous du temps de divergence minimal permettant la détection d'hétérogénéités significatives de la distribution des allèles entre populations. Les index de migration (Nm) supérieurs à 1 refléteraient alors le polymorphisme ancestral commun aux deux formes bien que les flux géniques contemporains soient fortement réduits, voire nuls comme le suggère la répartition hétérogène de la mutation *kdr*. Ceci implique que l'apparition de cette mutation

particulière chez *An. gambiae* soit encore plus récente que le phénomène de spéciation dans la mesure où elle semble spécifique à la forme Savane. Elle n'a encore jamais été décelée chez *An. arabiensis* ou *An. melas* (Chandre, 1998; Chandre *et al.*, 1999a).

Il est important de noter ici que ces considérations sur le manque de puissance des loci microsatellites ne sont valables que si l'on suppose que l'évènement de spéciation n'a pas été accompagné de phénomènes de réduction de taille des populations ce qui peut paraître incompatible avec le processus évolutif supposé être à l'origine des inversions (voir paragraphe 3.3 ci-dessous) mais en accord avec la taille effective des populations estimées chez *An. arabiensis* ou chez *An. gambiae* Savane (voir deuxième partie, Taylor *et al.*, 1993; Besansky *et al.*, 1997; Lehmann *et al.*, 1998). En d'autres termes, si le faible taux de différenciation entre ces populations s'explique par la rétention du polymorphisme ancestral, la spéciation entre les différentes formes chromosomiques a dû se faire graduellement au sein d'une population panmictique. La mutation *kdr* ne peut alors être apparue chez Savane qu'une fois la barrière de reproduction déjà bien établie entre les différents cytotypes, interdisant ainsi son transfert aux autres formes chromosomiques.

3. 2. - Différents régimes de sélection

Globalement, il apparaît que les marqueurs potentiellement les moins soumis à l'action directe de la sélection naturelle ne permettent pas de différencier les formes chromosomiques désignées par cytogénétique.

Comme nous en avons déjà fait mention plus haut, la corrélation établie au cours de nombreuses études entre les paramètres du milieu ou certains comportements et la fréquence de certains arrangements chromosomiques particuliers est un argument très fort contre la neutralité des inversions vis-à-vis de la sélection naturelle. En effet, ces réarrangements chromosomiques peuvent être vus comme de grosses sous-unités de

recombinaison permettant de préserver une certaine association entre plusieurs gènes co-adaptés à un biotope particulier ou favorisant un comportement spécifique, la recombinaison génétique étant très limitée à l'intérieur des inversions (Carson, 1982; Coluzzi, 1982).

Cette hypothèse est fortement soutenue par certains arguments d'ordre structural tels que la distribution non aléatoire des points de rupture sur les chromosomes (plusieurs inversions semblent partager les mêmes frontières) ou purement génétiques tels que la démonstration récemment effectuée d'une origine différente de la même inversion 2La fixée chez *An. arabiensis* et chez *An. merus* par Caccone *et al.* (1998). Certains conflits apparus dans les phylogénies retracées à partir des inversions et des séquences de différents loci ADN suggèrent l'introgression sélective de ces sortes de «super-gènes» d'une espèce à l'autre au sein du complexe (Besansky *et al.*, 1994; Mathiopoulos *et al.*, 1995; Caccone *et al.*, 1996, 1998; Garcia *et al.*, 1996; Della Torre *et al.*, 1997), appuyant le caractère adaptatif des inversions.

Ce rôle des inversions chromosomiques dans la capacité d'adaptation des individus a été particulièrement étudié chez la drosophile et leur importance dans les phénomènes de spéciation ainsi que leur transfert par introgression a été documenté, confortant ainsi les observations réalisées chez *An. gambiae* (Carson, 1969, 1982; Aquadro *et al.*, 1991; Krimbas et Powell, 1992; Rozas et Aguadé, 1994; Popadic et Anderson, 1994; Popadic *et al.*, 1995). Dans ces conditions, les inversions chromosomiques seraient alors de mauvais marqueurs pour l'étude des flux de gènes entre sous-populations, leur distribution étant majoritairement gouvernée par l'action de la sélection naturelle, indépendamment de toute structuration génétique des populations. Le profil de répartition des arrangements au sein des populations naturelles d'*An. gambiae* serait alors le résultat unique de la contre-sélection d'hétérocaryotypes présentant des combinaisons fortement délétères sans qu'il ne soit nécessaire d'avoir

recours à l'hypothèse de barrière à la reproduction entre porteurs des différents arrangements. Pour citer Avise (1994), ce profil «refléterait plutôt les conditions environnementales que le régime de flux géniques au sein de l'espèce». Une telle hypothèse a d'ailleurs été émise par Touré *et al.* (1994) au sujet de la fréquence des arrangements alternatifs chez Mopti dans le delta du fleuve Niger au Mali.

Cette explication est compatible avec le fait que la variabilité génétique semble moindre à l'intérieur des inversions (Mathiopoulos et Lanzaro, 1995) et qu'une plus forte différenciation soit observée lorsque les loci neutres sont situés à l'intérieur ou très proches de ces inversions (Lanzaro *et al.*, 1998; Kamau *et al.*, 1998b). Selon cette théorie, le profil de répartition de la mutation *kdr* s'expliquerait par le fait qu'elle se trouve en fort déséquilibre de liaison avec un des arrangements chromosomiques spécifiques à la forme Savane où elle semble être apparue récemment, très probablement à la suite de l'importante utilisation du DDT à la fin des années 60 en Afrique de l'Ouest et avoir été maintenue depuis l'arrêt de l'utilisation du DDT par la forte pression exercée par l'utilisation consécutive des pyréthrinoides auxquels elle confère une résistance croisée (Chandre *et al.*, 1999b et c). La détermination de la localisation cytologique de cette mutation chez *An. gambiae* est en cours d'investigation (P. Guillet, communication personnelle). Il serait en outre intéressant de comparer le génotype pour la mutation *kdr* des spécimens collectés et leur caryotype déterminé par cytogénétique afin de préciser cette éventualité, par l'étude des déséquilibres de liaisons entre les allèles *kdr* et les différents arrangements alternatifs.

Le profil de répartition de la mutation *kdr* peut également être le reflet de régimes de sélection différents appliqués sur les populations Mopti et Savane, dans le contexte récemment développé par Lenormand et Raymond (1998) de la «zone stable» pour l'étude de la diffusion des gènes de résistance au sein des populations. Ces auteurs ont démontré que la fréquence d'un allèle résistant dans une population soumise à un

traitement insecticide ainsi que l'influence de ce traitement sur la densité des populations locales sont étroitement corrélées à l'avantage sélectif procuré par le phénotype résistant, au coût de la résistance dans les zones non traitées et à la taille relative de la zone traitée par rapport aux zones non traitées adjacentes. En fonction des flux de gènes entre populations exposées et non exposées aux traitements, l'apparition d'un gène de résistance et sa montée en fréquence peut ainsi être retardée, voire évitée à l'intérieur d'une zone géographique donnée malgré l'utilisation localisée d'insecticides. Le faible degré d'isolement génétique entre Mopti et Savane que nous avons déterminé pourrait être suffisant pour influencer différemment l'évolution de la mutation *kdr* chez Mopti et chez Savane. Il est alors tout à fait possible qu'une zone stable soit établie pour Mopti dans cette région, évitant l'apparition de la mutation *kdr*, mais que la balance entre zones traitées et non traitées n'ait pas le même impact sur les populations de Savane, résultant en la quasi-fixation de l'allèle résistant.

Les zones de répartition géographique de Savane et Mopti sont en effet très différentes sur le continent Africain en général et dans cette région en particulier, bien qu'un certain nombre de zones de sympatrie existent (Coluzzi *et al.*, 1985; Touré *et al.*, 1998). Les études en cours tant sur la mutation *kdr* elle-même (coût et avantages associés à la présence de l'allèle résistant) que sur la génétique des populations et les échanges géniques entre et au sein des différentes formes chromosomiques pourront certainement permettre de préciser cette hypothèse.

3. 3. - Des populations en expansion

Selon Coluzzi (1982, 1985), les arrangements observés aujourd'hui auraient été générés à la suite de réorganisations stochastiques du génome dans des zones marginales (géographiques ou écologiques), les combinaisons particulières permettant

une adaptation des individus à leur nouvel environnement étant alors très fortement sélectionnées. Les porteurs de ces caryotypes avantageux pourraient alors, par migration, élargir la zone de répartition de l'espèce et/ou remplacer les populations ancestrales dans les zones où leur avantage adaptatif (fitness) est conséquent. Dans ces conditions, les formes évolutives récentes correspondant aux différents cytotypes individualisés chez *An. gambiae*, seraient alors en pleine expansion.

Cette hypothèse est en accord avec le fait que la majeure partie de la variabilité génétique observée chez *An. gambiae* résulte de variations intra-populations, seule une fraction minimale étant imputable aux variations inter-populations (voir Tableau 11A). Cette théorie d'expansion récente des populations à partir de foyers localisés a été particulièrement développée dans l'étude de l'histoire de l'Humanité pour expliquer une répartition analogue de la variabilité génétique totale, principalement due aux variations intra-populations (voir par exemple Cavalli-Sforza *et al.*, 1994; Goldstein *et al.*, 1995b; Barbujani *et al.*, 1997; Reich et Goldstein, 1998). Or, les estimations de Nm sont basées sur l'hypothèse de populations à l'équilibre (Wright, 1951; Slatkin, 1985a et b, 1995). Ces conditions ne sont pas vérifiées si les populations sont en expansion et les taux de flux géniques alors estimés sont biaisés.

Certains arguments issus de données écologiques sont en faveur de cette dynamique des populations comme en témoigne la colonisation des zones adjacentes au delta du fleuve Niger par Mopti, qui semble avoir pris l'avantage sur *An. arabiensis* et *An. gambiae* Savane via des phénomènes de compétition pour l'occupation quasi exclusive de cette niche écologique (Touré *et al.*, 1983; Coluzzi, 1984; Coluzzi *et al.*, 1985). En effet, la zone géographique de référence du cytotype Mopti est le delta intérieur du fleuve Niger, zone strictement naturelle qui présente des gîtes larvaires pérennes. On peut alors supposer que Mopti soit plus ancien que ce qui est généralement avancé mais qu'il ne soit sorti de son refuge écologique que récemment,

lorsque l'homme, par le biais de son agriculture irriguée, lui a fourni les gîtes de saison sèche auxquels il était adapté (et inféodé) dans son habitat original.

4. Conclusion

L'image globale qui ressort de l'analyse que nous venons d'effectuer concernant la structure génétique des populations d'*An. gambiae* est celle d'une organisation très complexe dont la description et la résolution varient en fonction des marqueurs considérés. Le génome de ce vecteur semble constitué d'une mosaïque hétérogène de fragments plus ou moins soumis à la sélection naturelle dont les mouvements entre sous-populations ne sont pas aléatoires.

L'analyse du polymorphisme des marqueurs microsatellites suggère que la fragmentation du taxon nominal *An. gambiae* s.s. en 5 sous-unités différentes reflète majoritairement l'action de la sélection naturelle sur certains fragments particuliers du génome mais que des échanges de gènes sont toujours possibles sur d'autres régions moins exposées. Néanmoins, l'hypothèse de spéciation consécutive à la création récente d'une barrière au flux de gènes dont le mécanisme serait essentiellement comportemental, lié à la reconnaissance mutuelle des partenaires avant la reproduction, ne peut être écartée. Seule la confrontation des informations fournies par des marqueurs indépendants les uns des autres, soumis à des mécanismes d'évolution différents et étudiés dans des contextes écologiques variables pourra permettre une description optimale de la structure génétique de ces populations.

Une attention particulière doit être portée sur le fait que les résultats expérimentaux obtenus en laboratoire et les données issues de collectes sur le terrain semblent contradictoires. Ces différences mettent en exergue le fait que les études de la génétique des populations naturelles représentent un élément fondamental à prendre en compte lors de l'élaboration de nouveaux moyens de lutte. En particulier, la stratégie actuellement préconisée de lutte par l'intermédiaire de moustiques transformés génétiquement ne pourra se baser exclusivement sur les résultats d'expérimentations réalisées en insectarium et la libre circulation d'un gène au laboratoire ne sera pas forcément vérifiée sur le terrain (cas de la mutation *kdr* par exemple). Il semble évident d'après ce que nous venons de discuter dans le paragraphe précédent que la cinétique de diffusion des gènes entre et au sein des populations naturelles fait intervenir plusieurs paramètres liés (i) à la structure proprement dite des populations révélée par les marqueurs neutres et (ii) au bénéfice ou au désavantage en termes de potentiel adaptatif que vont procurer certains arrangements ou mutations en fonction des conditions écologiques.

Chez *An. gambiae*, plusieurs arguments tendent à prouver que les flux géniques ne sont pas homogènes sur tout le génome: la localisation cytologique des gènes étudiés ou éventuellement transformés dans une optique d'altération de la compétence vectorielle sera donc d'une importance primordiale dans la capacité de diffusion entre et au sein des différentes formes chromosomiques et des espèces constitutives du complexe. Le chromosome 2 en particulier, porteur de la majorité des inversions polymorphes, pourrait jouer un rôle prédominant à ce niveau car il est très probable, au vu de l'influence de ces inversions sur la biologie et le comportement des vecteurs, qu'un certain nombre de loci d'un intérêt majeur s'y trouvent localisés. En effet, c'est le cas de loci impliqués dans la réponse immunitaire d'*An. gambiae* face à *Plasmodium cynomolgi* (Crews-Oyen *et al.*, 1993; Gorman *et al.*, 1997; Zheng *et al.*, 1997).

Cette hétérogénéité peut représenter un obstacle à la mise en pratique des méthodes de lutte basées sur le lâcher d'individus génétiquement modifiés. Parallèlement, l'absence de barrières au flux de gènes en laboratoire peut permettre de générer facilement des souches différentes d'insectes porteurs de modifications dont la capacité de reproduction (compatibilité avec les populations naturelles) sera certainement optimisée par rapport à des souches hybrides, pourtant très fertiles et compétitives au laboratoire. En d'autres termes, pour ce genre d'investigations, le recours à une stratégie généraliste risque fort de ne pas permettre d'obtenir de résultats probants sur le terrain. Les transposons ou symbiotes utilisés comme vecteurs de gènes modifiés apparaissent alors comme des outils à privilégier étant donné les indications de plus en plus nombreuses de flux de gènes effectivement autorisés sur certaines parties des chromosomes, même entre les deux vecteurs principaux *An. gambiae* et *An. arabiensis* (voir Besansky *et al.*, 1994, 1997; Della Torre *et al.*,1997).

CONCLUSIONS & PERSPECTIVES

Les études que nous venons de présenter sur la structure génétique des populations et les flux de gènes à l'intérieur du complexe *An. gambiae* ont permis (i) de confirmer le caractère hautement informatif des loci microsatellites pour l'investigation des relations entre sous-populations appartenant à la même espèce et leur utilisation possible et pertinente entre espèces jumelles, (ii) de mettre en évidence le rôle prépondérant de l'habitat et de l'histoire des populations sur l'établissement d'une structure génétique et (iii) de souligner l'extraordinaire complexité du système vectoriel étudié constitué de différentes sous-unités très spécialisées mais qui restent néanmoins interconnectées et ne peuvent être considérées comme totalement indépendantes les unes des autres. Au terme de ce travail, plusieurs conclusions et perspectives peuvent être dégagées concernant l'outil utilisé et le modèle étudié.

1. Les marqueurs microsatellites et leur utilisation pour l'étude de la génétique des populations d'anophèles

Au cours de notre étude, 9 loci microsatellites ont été utilisés pour étudier les variations du polymorphisme observé au sein de deux espèces jumelles de moustiques appartenant au complexe *An. gambiae*, système vectoriel responsable de la majeure partie de la transmission du paludisme en Afrique: *An. arabiensis* d'une part et le taxon nominal *An. gambiae* s.s. d'autre part.

Cette étude a révélé que les amorces permettant l'amplification par PCR des allèles chez *An. gambiae* pouvaient également être utilisées chez *An. arabiensis*, ce qui représente un avantage considérable permettant (i) une économie importante en temps et en moyens car la caractérisation des loci et la détermination des séquences des amorces permettant leur amplification est un travail long et fastidieux qui nécessite un investissement important, (ii) la standardisation des protocoles d'amplification et de

collecte des données nécessaires à l'analyse génétique impliquant (iii) une possibilité de comparer les résultats obtenus chez différentes espèces au niveau des mêmes loci.

Néanmoins, comme l'on peut s'y attendre suite aux phénomènes de spéciation, certains loci polymorphes chez l'une des espèces peuvent être fixés et non informatifs chez l'autre suite à des modifications moléculaires ou structurales au niveau de l'ADN. C'est ce que nous avons observé au niveau du locus 49, sur le chromosome X qui reste fixé chez toutes les populations d'*An. arabiensis* que nous ayons testées alors qu'il est très polymorphe chez *An. gambiae*. De même, il nous a été impossible de déterminer convenablement le génotype des spécimens *An. gambiae* au locus 141, localisé au niveau de l'inversion 2Rd alors que des profils de migration clairs et reproductibles ont été observés chez *An. arabiensis*. Parallèlement, aucun des 9 loci n'a pu être amplifié chez *An. funestus*, un autre vecteur important du paludisme souvent rencontré en sympatrie avec *An. gambiae* en Afrique, confirmant ainsi la spécificité des amorces déterminées et le statut taxonomique très proche des espèces *An. gambiae* et *An. arabiensis*.

L'avantage principal des marqueurs microsatellites réside dans leur fort taux de mutation leur conférant un polymorphisme important (voir annexe 1 et Lanzaro *et al.*, 1995 pour une comparaison avec les loci isoenzyme chez *An. gambiae*) et leur apparente neutralité. Ces particularités leur procurent un taux d'évolution rapide qui les rend peu susceptibles de garder les traces d'un polymorphisme ancestral commun et en fait de très bons indicateurs des taux de flux géniques contemporains entre sous-populations (Bossart et Pashley-Prowell, 1998). De plus, l'étude simultanée de plusieurs loci répartis sur tous les chromosomes permet d'obtenir une image générale du taux de divergence génétique entre sous-populations et limite les possibilités de méprise dues à des contraintes spécifiques agissant au niveau de loci particuliers (Barbujani *et al.*,

1997). Inversement, leur nombre et leur ubiquité sur le génome des Eucaryotes peut permettre une analyse très fine du comportement de différentes portions du génome et la mise en évidence de différences dans l'intensité des contraintes exercées sur certaines zones (par exemple, loci situés à l'intérieur où à l'extérieur des inversions chromosomiques chez *An. gambiae*). Néanmoins, nous avons vu que plusieurs types de contraintes peuvent affecter leur potentiel résolutif et certaines indications révélées par notre étude méritent d'être développées:

- à plusieurs reprises, de forts déficits en hétérozygotes ont été observés au niveau du locus 24D, localisé à l'intérieur de l'inversion 2La suggérant un comportement particulier de ce locus. Chez *An. arabiensis*, cette inversion est fixée à l'état homozygote. Par conséquent, la recombinaison est possible à ce niveau, les deux chromosomes étant parfaitement appariés lors de la méiose, et l'influence de cette inversion ne peut être retenue comme cause possible du déséquilibre observé. L'hypothèse d'un effet Wahlund ou de l'apparentement des individus (consanguinité) est également écartée car l'équilibre d'Hardy-Weinberg est généralement respecté aux autres loci. Les études du déséquilibre de liaison à l'intérieur des populations vont également dans ce sens, le locus 24D n'étant pas systématiquement impliqué dans les déséquilibres observés de façon ponctuelle. Dès lors, seules des contraintes spécifiques à ce locus peuvent être à l'origine de ce déficit récurrent en hétérozygotes.

Des données non présentées dans ce rapport concernant l'étude d'une fraction de la descendance de femelles *An. arabiensis* collectées sur le terrain et mises individuellement à pondre au laboratoire nous ont permis de révéler la présence effective d'allèles nuls à ce locus chez *An. arabiensis*. En effet, parmi la descendance de certaines femelles apparemment homozygotes à ce locus, certains individus sont apparus homozygotes pour un autre allèle que celui présent chez leur mère, allèle

paternel également présent chez d'autres spécimens issus de la même ponte. Plusieurs réplications de ces amplifications ont donné des résultats similaires et reproductibles. Ainsi, la présence d'allèles nuls au locus 24D a donc été confirmée chez *An. arabiensis*.

Leur influence sur le taux de différenciation génétique estimé par Fst ou Rst peut être importante lorsqu'ils sont présents à des fréquences variables au sein des différentes populations comparées. Néanmoins, ces différences font partie intégrante des divergences génétiques établies entre populations au cours de leur évolution et doivent être prises en compte dans l'estimation du taux de différenciation génétique. En conséquence, les informations fournies par ce locus ont été intégrées aux estimations moyennes données lors de cette étude. D'une manière générale, les valeurs de Fst et de Rst estimées à partir de ce locus ne se sont jamais révélées en désaccord avec les informations provenant des autres loci utilisés (voir Tableau 3 dans la première partie et Tableau 12 dans la troisième partie), laissant supposer une influence mineure de la présence de ces allèles nuls.

La présence d'allèles nuls à d'autres loci microsatellites a été révélée chez *An. gambiae* par plusieurs auteurs (voir annexe 1; Lanzaro *et al.*, 1995; Lehmann *et al.*, 1996; Kamau *et al.*, 1998a et b; Walton *et al.*, 1998; Wang *et al.*, 1999) suggérant que ce phénomène soit relativement répandu chez cette espèce. Les mutations ponctuelles généralement à l'origine de l'apparition de ces allèles nuls sont rares (10^{-9} mutation/gamète/génération, selon Lewin, 1994) et le fait qu'un nombre important de loci semble présenter de telles contraintes plaide en faveur d'une taille effective importante des populations constitutives du complexe *An. gambiae*. Ces considérations mettent en exergue la nécessité de tester d'une manière intensive les différents loci disponibles avant d'entamer une étude de génétique des populations à grande échelle avec ce type de marqueurs.

- malgré leurs avantages, les loci que nous avons utilisés au cours de cette étude ne se sont pas révélés plus efficaces que les isoenzymes pour détecter des différences entre les formes chromosomiques Mopti et Savane au sein du complexe *An. gambiae*. Ce résultat est surprenant si l'on considère les indications fournies par d'autres marqueurs (inversions chromosomiques et mutation *kdr* en particulier) qui suggèrent l'existence d'une forte barrière au flux de gènes entre ces formes chromosomiques. De même, les résultats présentés en annexe 1 révèlent qu'entre deux populations d'*An. gambiae* Savane séparées par plus de 6000 km, un taux de différenciation génétique très faible et tout à fait comparable a été estimé à partir des isoenzymes et des microsatellites, en dépit d'un polymorphisme beaucoup plus important observé chez ces derniers marqueurs.

Un manque de puissance des marqueurs microsatellites, impliquant des contraintes sur la taille des allèles ou, alternativement, un taux de mutation plus faible que ce qui est généralement avancé pourrait expliquer le faible taux de différenciation révélé au cours de ces deux études. Bien que les phénomènes d'homoplasie semblent plus courants au niveau inter-spécifique qu'à l'intérieur d'un complexe d'espèces jumelles (Estoup *et al.*, 1995a) et donc à plus forte raison entre les formes chromosomiques d'*An. gambiae*, Walton *et al.* (1998) font mention de variations au niveau des séquences flanquantes des répétitions pour le locus 147 entre les formes Mopti et Forêt, prélevées dans des zones géographiques différentes et éloignées. De telles variations ne peuvent être détectées suivant notre méthode d'investigation et pourraient en effet entraîner une surestimation du degré d'homologie entre les populations testées, si les allèles identiques par état (allèles de même taille dans notre cas) ne sont pas plus semblables entre eux que ne le sont des allèles de taille différente.

Néanmoins, le fait que nos résultats soient tout à fait comparables à ceux reportés par Lanzaro *et al.* (1998), utilisant d'autres loci, entre des populations Mopti et Bamako sympatriques, suggère que des flux géniques importants relient effectivement les différentes formes chromosomiques. Une étude plus poussée de la séquence des allèles amplifiés pourrait alors permettre de préciser dans quelle mesure ce phénomène d'homoplasie peut biaiser les estimations obtenues.

- il est fort probable que les profils détectés lors de notre étude sur *An. gambiae* traduisent effectivement des flux de gènes importants entre les formes chromosomiques Mopti et Savane car la preuve de la capacité des microsatellites à révéler des différences importantes au niveau intra-spécifique a été donnée au cours de la première partie, où des taux de différenciation génétique élevés et très significatifs ont été mis en évidence entre 4 populations d'*An. arabiensis*. De même, Kamau *et al.* (1998a) ont démontré leur potentiel résolutif chez *An. gambiae* par la mise en évidence de forts taux de différenciation entre populations du cytotype Savane séparées par la vallée du Rift au Kenya. Des valeurs également élevées de Fst et de Rst ont été estimées par Lanzaro *et al.* (1998) entre formes chromosomiques pour les loci situés à l'intérieur ou très proches des inversions polymorphes du chromosome 2. Toutes ces données confirment le caractère hautement informatif des loci microsatellites pour l'étude de la structure des populations d'anophèles et soulignent la complexité du système étudié.

Afin de pouvoir comparer les résultats obtenus par différents groupes de recherche et éviter la duplication des études, il serait souhaitable de standardiser les protocoles d'investigation entre laboratoires. Cette standardisation passe par la sélection consensuelle d'un certain nombre de loci microsatellites parmi ceux isolés et localisés

sur la carte génomique d'*An. gambiae* (Zheng *et al.*, 1996). Très récemment Wang, Kafatos et Zheng (1999) ont proposé l'utilisation de 24 marqueurs à motif dinucléotide sélectionnés sur la base de leur fort polymorphisme chez *An. gambiae* comme chez *An. arabiensis*, de leur localisation cytologique parfaitement déterminée et couvrant la totalité du génome, et de la reproductibilité des résultats préliminaires obtenus. La taille des allèles séquencés pour chacun de ces loci (80-196 pb) apparaît compatible avec une bonne résolution sur gel d'acrylamide à 10%, tels que ceux que nous avons utilisés lors de cette étude. Nous soutenons cette initiative qui devrait permettre une meilleure cohésion des résultats générés par les différents groupes impliqués dans ce champ de recherches.

D'un point de vue plus fondamental, nous avons vu que de nombreuses incertitudes demeurent concernant les taux et processus de mutation des séquences microsatellites. Parallèlement, les modèles d'analyse des données actuellement disponibles sont basés sur un certain nombre de conditions qui sont rarement respectées chez les populations naturelles. L'utilisation de plus en plus importante de ces loci chez différentes espèces et sous différentes conditions expérimentales et écologiques permettra des études plus fines des propriétés évolutives de ces marqueurs. De nouvelles statistiques, basées sur des simulations par ordinateur et leur confrontation aux données expérimentales collectées à partir de populations naturelles voient peu à peu le jour et soulignent la quantité importante d'informations générées par les variations de polymorphisme des microsatellites et la pertinence des indications qu'ils procurent (voir par exemple Goldstein et Pollock 1997; Reich et Goldstein, 1998; Waser et Strobeck, 1998; Goldstein *et al.*, 1999). Il ne fait aucun doute que ces marqueurs sont voués à un développement rapide, facilité par leur ubiquité chez les espèces animales et

végétales, l'optimisation et l'automatisation des moyens et des protocoles permettant d'accéder aux génotypes des individus, de traiter les données recueillies et de modéliser leur évolution.

2. Structure des populations et flux de gènes chez *An. gambiae*: implications sur les moyens de lutte antivectorielle

Le modèle biologique que nous avons étudié grâce au polymorphisme des marqueurs microsatellites apparaît particulièrement complexe et hétérogène.

La panmixie est respectée chez toutes les populations d'*An. arabiensis* que nous ayons testé confirmant ainsi l'homogénéité des populations présentes en un lieu donné à une époque donnée. Un très fort taux de différenciation génétique a été révélé entre populations séparées par une distance géographique considérable représentant les limites de la zone de répartition de l'espèce, et l'influence de l'Océan Indien comme barrière à la dispersion des individus a été démontrée. Les résultats obtenus sont en accord avec l'historique de l'implantation et de l'évolution indépendante des différents isolats géographiques bien que des flux de gènes extrêmement limités semblent se maintenir via des phénomènes occasionnels de migration assistée (cyclones ou transport associé aux activités humaines). Au Sénégal, les populations sont très peu différenciées et l'étude temporelle présentée en deuxième partie nous a permis de démontrer que la composition génétique reste stable malgré les très fortes variations saisonnières de la densité des populations en zone sahélienne. Nous avons ainsi vérifié que seuls des changements démographiques extrêmes pouvaient induire à court terme une variation conséquente de la diversité génétique au sein d'une population (Waples et Teel, 1990; Tessier et Bernatchez, 1999).

Il serait souhaitable de confronter les résultats obtenus par cette méthode indirecte à des données issues de mesures directes (écologiques) de la dispersion effectuées sur le terrain grâce à des expériences de marquage-recapture. Or, un obstacle majeur à la mise en pratique de ces méthodes réside en l'apparente disparition des populations vectorielles durant les mois les plus arides. Les conditions environnementales totalement différentes entre saison sèche et saison des pluies (humidité relative, hauteur de la végétation...), la disponibilité des gîtes d'oviposition et des hôtes potentiels, certains phénomènes liés à la densité de la population et peut-être également des paramètres liés à l'état physiologique des individus (dissociation gonotrophique, diapause) peuvent considérablement modifier le comportement migratoire des individus de sorte qu'il est très douteux d'extrapoler des résultats obtenus en saison des pluies à ce qui peut se passer en saison sèche (Service, 1993; Costantini *et al.*, 1996; Touré *et al.*, 1998). Il paraît alors difficile de valider nos résultats par des études directes de la dispersion dans les mêmes conditions. Néanmoins, nous avons pu démontrer qu'*An. arabiensis* était effectivement présent à Barkedji en pleine saison sèche, renforçant ainsi les conclusions auxquelles nous sommes parvenus par l'analyse génétique.

Dans leur ensemble, ces résultats suggèrent qu'*An. arabiensis*, espèce très adaptée aux zones de savane sèches et anthropophile préférentielle mais non obligatoire, possède des capacités d'estivation et de dispersion importantes qui en font un vecteur particulièrement difficile à contrôler par les méthodes de lutte actuellement disponibles. En effet, là où la transmission du paludisme a pu être enrayée suite à l'application des stratégies de contrôle essentiellement basées sur les pulvérisations intra-domiciliaires d'insecticide comme à Madagascar, La Réunion ou Maurice, *An. arabiensis*,

contrairement à *An. gambiae* ou *An. funestus*, s'est toujours maintenu par des phénomènes d'échappement essentiellement liés à sa capacité de survivre dans des conditions totalement sauvages (zoophilie, exophilie). Sa présence à La Réunion et Maurice en particulier représente toujours une menace pour le maintien de l'éradication du paludisme et contraint les autorités locales à une surveillance ininterrompue qui nécessite des moyens humains et financiers importants (DDASS-Réunion, 1997), une stratégie inapplicable sur le continent Africain.

De telles zones géographiques relativement confinées pourraient représenter des sites privilégiés pour étudier la mise en pratique de nouvelles méthodes de lutte antivectorielle basées sur l'utilisation de moustiques transformés génétiquement, permettant d'obtenir des renseignements précieux sur la diffusion de gènes introduits (ou modifiés) au sein de populations dans des conditions naturelles, données préliminaires indispensables et incontournables pour l'application de ces méthodes à l'échelle du continent Africain. Il pourrait ainsi être judicieux d'étudier par exemple la diffusion de gènes marqueurs portés par différents vecteurs moléculaires (transposons, symbiotes) au sein de ces populations afin de tester l'efficacité de ces différentes méthodes de transformation des populations naturelles (efficacité du mécanisme de «genetic drive», stabilité du système vecteur-insert, évolution de la fréquence du gène introduit) (Collins et James, 1996; Ashburner *et al.*, 1998; Curtis et Townson, 1998).

La dernière partie de notre travail, consacrée à l'étude de la différenciation génétique entre les formes chromosomiques Mopti et Savane du taxon *An. gambiae* s.s. dans une zone de sympatrie en Côte d'Ivoire nous a permis de mettre en évidence l'ambiguïté et la complexité de la structure génétique des populations au sein de cette espèce. L'influence des conditions environnementales (donc de la sélection) sur la

structuration génétique des populations a été mise en relief sur la base des résultats divergents révélés par les marqueurs neutres (isoenzymes et microsatellites) et les gènes ou fragments chromosomiques soumis à sélection (mutation *kdr*, inversions chromosomiques). Une telle structuration génétique au sein de populations sympatriques apparaît comme un obstacle à la libre circulation des gènes mais nous avons vérifié que des échanges géniques sont toujours possibles entre les différentes sous-unités identifiées chez *An. gambiae*. A l'évidence, ces échanges ne sont pas homogènes sur tout le génome (Favia *et al.*, 1997; Della Torre *et al.*, 1997; Lanzarro *et al.*, 1998) ce qui confirme que les forts index de migration détectés traduisent bien l'existence de flux géniques contemporains plutôt que la conservation d'un polymorphisme ancestral commun.

Cette description empirique de la structure génétique et des interactions entre les entités constitutives du complexe *An. gambiae* reflète l'hétérogénéité des profils épidémiologiques de la transmission du paludisme en Afrique et la diversité de la réponse des vecteurs aux moyens de lutte actuellement utilisés. Néanmoins, nous avons vu que la composition génétique des populations reste globalement très stable au cours du temps (deuxième partie; Taylor *et al.*, 1993; Lehmann *et al.*, 1998) tout en permettant le maintien d'un polymorphisme important étroitement lié aux conditions variables de l'environnement dans lequel ces vecteurs se développent (Coluzzi *et al.*, 1985; Touré *et al.*, 1994).

Le fait que les différentes populations restent potentiellement en contact les unes avec les autres soutient les espoirs basés sur la lutte par introduction de moustiques transgéniques. L'objectif ultime des méthodes de lutte par insectes génétiquement transformés est d'aboutir à la modification du patrimoine génétique des individus, non

seulement au niveau des populations locales mais également et inévitablement, à terme, au niveau de l'espèce entière. Ces résultats, pour être bénéfiques et justifier leur contribution à l'amélioration de la santé publique doivent de surcroît être obtenus dans un laps de temps raisonnable, très inférieur à l'échelle de temps associée aux processus de l'évolution. Cette stratégie peut ainsi paraître extrêmement ambitieuse. Cependant, l'exemple scientifiquement prouvé et reconnu de l'invasion récente et rapide (<50 ans) à l'échelle mondiale du génome de la drosophile par l'élément P (voir Clarck et Kidwell, 1997) prouve que de telles modifications sont possibles chez les Diptères. Les transposons pourraient ainsi se révéler d'excellents vecteurs pour la transformation rapide des populations d'insectes et leur potentiel d'utilisation est actuellement très étudié (Collins et James, 1996; Curtis et Townson, 1998; Ashburner *et al.*, 1998). Récemment, Jasinskiene *et al.* (1998) et Coates *et al.*, (1998) sont parvenus à obtenir des transformants stables chez *Aedes aegypti*, moustique vecteur de la fièvre jaune, grâce à l'utilisation des transposons *Hermes* et *Mariner*, donnant ainsi naissance aux premiers moustiques transgéniques.

D'autres études ont démontré, toujours chez la drosophile, que des symbiotes du genre *Wolbachia* pouvaient envahir très rapidement et uniformément des populations naturelles réparties sur une zone géographique très vaste (Turelli et Hoffman, 1991; Turelli *et al.*, 1992) et la possibilité d'utiliser de telles bactéries comme vecteurs de gènes pour altérer la capacité vectorielle a été démontrée chez la punaise *Rhodnius prolixus* (Durvasula *et al.*, 1997). Ces vecteurs moléculaires particuliers présentent un avantage évident pour permettre la diffusion des gènes au sein des populations dans la mesure où la cinétique de diffusion est principalement fonction des caractéristiques du vecteur utilisé plutôt que de l'avantage sélectif éventuellement procuré par le transféron lui-même, qui, comme nous en avons un exemple avec la mutation *kdr*, peut s'avérer un

très mauvais indicateur de la capacité de diffusion dans des conditions naturelles pourtant favorables (forte pression de sélection d'insecticides dans ce cas). De plus, contrairement aux gènes portés par les chromosomes, de très faibles taux de flux géniques peuvent permettre le passage de ces systèmes d'une sous-population à une autre (Sinkins *et al.*, 1997; Bourtis et O'Neil, 1998).

Les recherches sur ces différents mécanismes et leur application chez *An. gambiae* sont en cours et avancent à grands pas. Parallèlement, comme nous avons pu le mettre en évidence tout au long de ce mémoire, de nombreux groupes travaillent à l'identification de gènes cibles potentiels ou s'attachent à étudier la structure des populations naturelles de ces vecteurs. Ces méthodes de lutte faisant intervenir des organismes modifiés génétiquement ne pourront toutefois être appliquées qu'après des études très poussées de la stabilité des organismes relâchés ainsi que sous réserve de l'accord des comités d'éthiques compétents. Des études de terrain sont alors particulièrement indispensables afin de s'assurer que les modifications préconisées ne vont pas entraîner une augmentation de la capacité vectorielle de ces anophèles génétiquement modifiés pour d'autres agents pathogènes d'importance médicale ou vétérinaire. Les résultats des recherches entreprises dans cette optique permettront à plus court terme de mieux connaître et reconnaître ces vecteurs, leur histoire et leur biologie afin d'optimiser la mise en œuvre et la gestion des moyens de contrôle déjà disponibles.

Références Bibliographiques

Akogbeto M, Mukabayire O, Di Deco MA et Coluzzi M. 1992. Osservazioni sulla distribuzione ecogeografica e stagionale del complesso *Anopheles gambiae* nel nord del Borgou, Benin, Africa Occidentale. *Parassitologia* **34(Suppl)**: 20.

Aldous P. 1993. Malaria: focus on mosquito genes. *Sciences* **261**: 546-547.

Alonso PL, Smith T, Schellenberg JR, Masanja H, Mwankusye S, Urassa H, Bastos de Azevedo I, Chongela J, Kobero S, Menendez C *et al.*. 1994. Randomised trial of efficacy of SPf66 vaccine against *Plasmodium falciparum* malaria in children in southern Tanzania. *Lancet* **344**: 1175-1181.

Amos W, Sawcer SJ, Feakes RW et Rubinsztein DC. 1996. Microsatellites show mutational bias and heterozygote instability. *Nature Genetics* **13**: 390-391.

Aquadro CF, Weavers AL, Schaeffer SW et AndersonWW. 1991. Molecular evolution of inversions in *Drosophila pseudoobscura*: the amylase gene region. *Proc. Natl. Acad. Sci. USA* **88**: 305-309.

Ashburner M, Hoy MA et Peloquin JJ. 1998. Prospects for the genetic transformation of arthropods. *Insect. Mol. Biol.* **7(3)**: 201-213.

Ashley MV et Dow BD. 1994. The use of microsatellite analysis in population biology: backgrounds, methods and potential applications. In: Molecular Ecology and Evolution Approaches and Applications. Eds B. Schierwater, Streit B, Wagner GP & DeSalle R. Birkhauser Verlag, Basel, Switzerland: Pp 185-201.

Avise JC. 1994. Molecular Markers, Natural History and Evolution. Eds Chapmann & Hall.

Ayala FJ. (ed.). 1976. Molecular evolution. Sinauer Associates, Sunderland, Mass.

Ballou WR, Sherwood JA, Neva FA, Gordon DM, Wirtz RA, Wasserman GF, Diggs CL, Hoffman SL, Hollingdale MR, Hockmeyer WT, Schneider I, Young JF, Reeve P et Chulay JD. 1987. Safety and efficacy of a recombinant DNA *Plasmodium falciparum* sporozoite vaccine. *Lancet* **1**: 1277-1281.

Barat. 1869. Etude sur la fièvre épidémique qui a régné en 1868 à l'île de La Réunion. *Archives de Médecine Navale* **XII**: 422-440. Paris: Ballière & fils.

Barbujani G, Magagni A, Minch E et Cavalli-Sforza LL. 1997. An appointment of human DNA diversity. *Proc. Natl. Acad. Sci. USA* **94**: 4516-4519.

Beier JC, Perkins PV, Wirtz RA, Koros J, Diggs D, Gargan TPII et Koech DK. 1988. Bloodmeal identification by direct enzyme-linked immunosorbent assay (ELISA) tested on *Anopheles* (Diptera: Culicidae) in western Kenya. *J. Med. Entomol.* **33**: 613-618.

Beier JC, Copeland R, Oyaro C, Masinya A, Odago WO, Oduor S, Koech DK et Roberts CR. 1990. *Anopheles gambiae* complex egg-stage survival in dry soil from larval development sites in western Kenya. *J. Am. Mosq. Control Assoc.* **6(1)**: 105-109.

Bell CJ et Ecker JR. 1994. Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis. Genomics* **19**: 137-144.

Besansky NJ, Powell JR, Caccone A, Miles Hamm D, Scott JA et Collins FH. 1994. Molecular phylogeny of the *Anopheles gambiae* complex suggests genetic introgression between the principal malaria vectors. *Proc. Natl. Acad. Sci. USA* **91**: 6885-6888.

Besansky NJ, Lehmann T, Fahey GT, Fontenille D, Braack LEO, Hawley WA et Collins FH. 1997. Patterns of mitochondrial variation within and between African malaria vectors, *Anopheles gambiae* and *Anopheles arabiensis*, suggest extensive gene flow. *Genetics* **147**: 1817-1828.

Blanquer-Maumont A et Crouau-Roy B. 1995. Polymorphism, monomorphism and sequences in conserved microsatellites in primate species. *J. Mol. Evol.* **41**: 492-497.

Blazquez J. 1987. Airport malaria in Spain. *Med. Clin.* **88**: 41.

Bory de Saint-Vincent JB. 1804. Voyage dans les quatre principales îles des mers d'Afrique fait par ordre du Gouverneur pendant les années 9 et 10 de la République. Paris: F. Buisson Imprimeur, *tome I*, 228 pp.

Bossart JL et Pashley Prowell D. 1998. Genetic estimates of population structure and gene flow: limitations, lessons and new directions. *T.R.E.E.* **13(5)**: 202-206.

Bourtzis K et O'Neill S. 1998. Wolbachia infections and arthropod reproduction. *BioScience* **48(4)**: 287-293.

Bowcock AM, Ruiz Linares A, Tomfohrde J, Minch E, Kidd JR et Cavalli-Sforza LL. 1994. High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* **368**: 455-457.

Braack LEO, Coetzee M, Hunt RH, Biggs H, Cornel AJ et Gericke A. 1994. Bitting patterns and host seeking behaviour of *Anophleles arabiensis* (Diptera: Culicidae) in northeastern South Africa. *J. Med. Entomol.* **31**: 333-339.

Brahmachari SK, Meera G, Sarkar PS, Balagurumoorty P, Tripathi J, Raghavan S, Shaligram U et Pataskar S. 1995. Simple repetitive sequences in the genome: Structure and functional significance. *Electrophoresis* **16(9)**: 1705-1714.

Brinkmann B, Klintschar M, Neuhuber F, Hühne J et Rolf B. 1998. Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *Am. J. Hum. Genet.* **62**: 1408-1415.

Brookfield JFY. 1996. A simple new method for estimating null allele frequency from heterozygote deficiency. *Mol. Ecol.* **5**: 453-455.

Bryan JH, Di Deco MA, Petrarca V et Coluzzi M. 1982. Inversion polymorphism and incipient speciation in *Anopheles gambiae* s.str. in The Gambia, West Africa. *Genetica* **59**: 167-176.

Bufford MW et Wayne RK. 1993. Microsatellites and their application to population genetic studies. *Current Opinion in Genetics and Development* **3**: 939-943.

Bullini L et Coluzzi M. 1978. Applied and theoretical significance of electrophoretic studies in mosquitoes (Diptera: Culicidae). *Parassitologia* **20**: 7-21.

Bullini L. 1984. Enzyme variants in the identification of parasites and vectors: methodological aspects of the electrophoretic approach. *In*: New Approach to the Identification of Parasites and their Vectors. Eds Newton BN et Michal F. World Health Organization, Tropical Disease Research, Geneva. **Series 5**: 53-69.

Caballero A. 1994. Developments in the prediction of effective population size. *Heredity* **73**: 657-679.

Caccone A, Garcia BA et Powell JR. 1996. Evolution of the mitochondrial DNA control region in the *Anopheles gambiae* complex. *Insect. Mol. Biol.* **5**: 51-59.

Caccone A, Min G-S et Powell JR. 1998. Multiple origins of cytologically identical chromosome inversions in the *Anopheles gambiae* complex. *Genetics* **150**: 807-814.

Cairns MJ et Murray V. 1994. Rapid silver staining and recovery of PCR products separated on polyacrylamide gels. *Biotechniques* **17**: 915-919.

Callen DF, Thompson AD, Shen Y, Phillips HA, Richards RI, Mulley JC et Sutherland GR. 1993. Incidence and origin of "null" alleles in the (AC)n microsatellite markers. *Am. J. Hum. Genet.* **52**: 922-927.

Carlson DA et Service MW. 1980. Identification of mosquitoes of *Anopheles gambiae* species complex A and B by analysis of cuticular components. *Science* **207**: 1089-1091.

Carson HL. 1969. Parallel polymorphism in different species of Hawaiian Drosophila. *Am. Nat.* **113**: 323-330.

Carson HL. 1982. Speciation as a major reorganization of polygenic balances. In: Mechanisms of Speciation. Ed. Barigozzi C. Liss, New York: 411-433.

Cavalli-Sforza LL, Menozzi P et Piazza A. 1994. History an geography of human genes. Princeton University Press, Princeton.

Chakraborty R et Nei M. 1977. Bottleneck effects on average heterozygosity and genetic distance with the stepwise mutation model. *Evolution* **31**: 347-356.

Chakraborty R, De Andrade M, Daiger SP et Budowle B. 1992. Apparent heterozygote deficiencies observed in DNA typing data and their implications in forensic applications. *Annals of Human Genetics* **56**: 45-57.

Chakraborty R et Zhong Y. 1994. Statistical power of an exact test of Hardy-Weinberg proportions of genotypic data at a multiallelic locus. *Hum. Hered.* **44**: 1-9.

Chandre F. 1998. Résistance d'*Anopheles gambiae* Giles et de *Culex pipiens quinquefasciatus* Say aux insecticides en Afrique de l'Ouest et implications opérationnelles. Thèse de doctorat de l'Université de Paris XII – Val de Marne. France. 112 p.

Chandre F, Brengues C, Dossou Yovo J, Manguin S, Darriet F, Diabate A, Carnevale P et Guillet P. 1999a. Current distribution of pyrethroid resistance gene (*kdr*) in *Anopheles gambiae* complex from West Africa and further evidence for reproductive isolation of the Mopti form. « The malaria challenge after one hundred years of malariology », Malariology Centenary Conference *Parassitologia – Special issue*: sous presse.

Chandre F, Darriet F, Manguin S, Brengues C, Carnevale P et Guillet P. 1999b. Pyrethroid cross resistance spectrum among populations of *Anopheles gambiae* from Côte d'Ivoire. *J. Mosq. Control Assoc.* **15**: sous presse.

Chandre F, Darriet F, Manga L, Akogbeto M, Faye O, Mouchet J et Guillet P. 1999c. Status of pyrethroid resistance in *Anopheles gambiae* s.l. *Bull. Wld. Hlth. Org.* **17(2)**: sous presse.

Charlwood JD, Kihonda J, Sama S, Billingsley PF, Heiz B et Takken W. 1995. The rise and fall of *Anopheles arabiensis* (Diptera: Culicidae) in a Tanzanian village. *Bull. Ent. Res.* **85**: 37-44.

Chauvet G, Coz J, Gruchet H, Grjebine A et Lumaret R. 1964. Contribution à l'étude biologique des vecteurs du paludisme à Madagascar. Résultats de 5 années d'études (1958-1962). *Médecine Tropicale* **24(1)**: 26-43.

Chauvet G et Rajaonarivelo E. 1973. «Modification de comportement» d'une espèce et variation de la composition d'une population appartenant au complexe *An. gambiae* pendant et après les pulvérisations d'insecticides domiciliaires dans les environs de Tananarive (Madagascar). *Cah. ORSTOM, Sér. Ent. Méd. et Parasitol.* **11(3)**: 155-167.

Chung M, Ranum LPW, Duvick LA, Servadio A, Zoghbi HY et Orr HT. 1993. Evidence for a mechanism predisposing to intergenerational CAG repeat instability in spinocerebellar ataxia type I. *Nat. Genet.* **5**: 254-258.

Cianchi R, Villani F, Touré YT, Petrarca V et Bullini L. 1983. Electrophoresis study of different chromosomal forms within *Anopheles gambiae* s.s.. *Parassitologia* **25**: 239-241.

Cianchi R, Urbanelli S, Villani F, Sabatini A et Bullini L. 1985. Electrophoresis study in mosquitoes: Recent advances. Parassitologia **27**: 157-167.

Clark JB et Kidwell MG. 1997. A phylogenetic perspective on P transposable element evolution in Drosophila. *Proc. Natl. Acad. Sci. USA* **94**: 11428-11433.

Coates CJ, Jasinskiene N, Miyashiro L et James AA. 1998. *Mariner* transposition and transformation of the yellow fever mosquito, *Aedes aegypti*. *Proc. Natl. Acad. Sci. USA* **95**: 3748-3751.

Collins FH, Sakai RK, Vernick KD, Paskewitz SM, Seeley DC, Miller LH, Collins WE, Campbell CC et Gwadz RW. 1986. Genetic selection of a *Plasmodium*-refractory strain of the malaria vector *Anopheles gambiae*. *Science* **234**: 607-610.

Collins FH, Mendez MA, Rasmussen MO, Mehaffey PC, Besansky NJ et Finnerty V. 1987. A ribosomal RNA gene probe differentiates member species of the *Anopheles gambiae* complex. *Am. J. Trop. Med. Hyg.* **37(1)**: 37-41.

Collins FH. 1994. Prospects for malaria control through the genetic manipulation of its vectors. *Parasitology Today* **10(10)**: 370-371.

Collins FH, Besansky NJ et Paskewitz SM. 1994. Peut-on espérer interrompre la transmission du paludisme par des vecteurs génétiquement modifiés? *Annales de l'Institut Pasteur/actualités* **5(4)**: 324-329.

Collins FH et James AA. 1996. Genetic modification of mosquitoes. *Science & Medicine* (Nov-Dec): 52-61.

Collins FH, Zheng L, Paskewitz SM et Kafatos FC. 1997. Progress in the map-based cloning of the *Anopheles gambiae* genes responsible for the encapsulation of malarial parasites. *Ann. Trop. Med. Parasitol.* **91(5)**: 517-521.

Colson I et Goldstein DB. 1999. Evidence for complex mutations at microsatellite loci in Drosophila. *Genetics* **152**: 617-627.

Coluzzi M et Sabatini A. 1967. Cytogenetic observations on sp. A and B of the *Anopheles gambiae* complex. *Parassitologia* **9**: 73-88.

Coluzzi M et Sabatini A. 1968a. Cytogenetic observations on species C of the *Anopheles gambiae* complex. *Parassitologia* **10**: 155-165.

Coluzzi M et Sabatini A. 1968b. Cytogenetic observations on species C, *merus* and *melas* of the *Anopheles gambiae* complex. *Cah. ORSTOM Ent. Méd.* **6**: 211-213.

Coluzzi M et Sabatini A. 1969. Cytogenetic observations on the salt-water species *Anopheles merus* and *An. melas*, of the *gambiae* complex. *Parassitologia* **11**: 177-187.

Coluzzi M et Kitzmiller J. 1975. Anopheline mosquitoes. *In*: Handbook of genetics vol. III. Ed King RC. Plenum Press. Pp: 285-309.

Coluzzi M, Sabatini A, Petrarca V et Di Deco MA. 1979. Chromosomal differentiation and adaptation to human environments in the *Anopheles gambiae* complex. *Trans. R. Soc. Trop. Med. Hyg.* **73**: 483-497.

Coluzzi M. 1982. Spatial distribution of chromosomal inversions and speciation in Anopheline mosquitoes. In: Mechanisms of Speciation. Ed Barigozzi C. Liss, New York: 143-153.

Coluzzi M. 1984. Heterogeneities of the malaria vectorial system in tropical Africa and their significance in malaria epidemiology and control. *Bull. Wld. Hlth. Org.* **62(Suppl.)**: 107-113.

Coluzzi M, Petrarca V et Di Deco MA. 1985. Chromosomal inversion intergradation and incipient speciation in *Anopheles gambiae. Boll. Zool.* **52**: 45-63.

Coosemans M, Petrarca V, Barutwanayo M et Coluzzi M. 1989. Species of the *Anopheles gambiae* complex and chromosomal polymorphism in a rice-growing area of the Rusizi valley (Republic of Burundi). *Parassitologia* **31** : 113-122.

Coosemans M, Smits A et Roelants P. 1998. Intraspecific polymorphism of *Anopheles gambiae* in relation to environment, behavior, and malaria transmission in Southwestern Burkina Faso. *Am. J. Trop. Med. Hyg.* **58(1)**: 70-74.

Cornuet JM et Luikart G. 1996. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **144**: 2001-2014.

Costantini C, Li SG, Della-Torre A, Sagnon N'F, Coluzzi M et Taylor CE. 1996. Density, survival and dispersal of *Anopheles gambiae* complex mosquitoes in a west African sudan savanna village. *Med. Vet. Entomol.* **10**: 203-219.

Crampton JM, Morris AC, Lycett GJ, Warren A et Eggleston P. 1990. Transgenic mosquitoes: a future vector control strategy? *Parasitology Today* **6**: 31-36.

Crampton JM, Comley I, Eggleston P, Hill S, Hughes M, Knapp T, Lycett G, Urwin R et Warren A. 1992. Molecular biological approaches to the study of vectors in relation to malaria control. *Mem. Inst. Oswaldo Cruz, Rio de Janeiro* **87(Suppl. III)**: 43-49.

Crampton JM et Hill SM. 1993. DNA probes for identification of malaria vectors. *Parassitologia* **35**: 31-44.

Crampton JM. 1994. Approaches to vector control: new and trusted. 3. Prospects for genetic manipulation of insect vectors. *Trans. R. Soc. Trop. Med. Hyg.* **88(2)**: 141-143.

Crews-Oyen AE, Kumar V et Collins FH. 1993. Association of two esterase genes, a chromosomal inversion, and susceptibility to *Plasmodium cynomolgi* in the African malaria vector *Anopheles gambiae. Am. J. Trop. Med. Hyg.* **49(3)**: 341-347.

Curtis CF. 1994. The case for malaria control by genetic manipulation of its vectors. *Parasitology Today* **10(10)**: 371-374.

Curtis CF et Townson H. 1998. Malaria: existing methods of vector control and molecular entomology. *British Medical Bulletin* **54(2)**: 311-325.

D'Allessandro U, Leach A, Drakeley CJ, Bennett S, Olaleye BO, Fegan GW, Jawara M; Langerock P, George MO, Targett GAT *et al.*. 1995. Efficacy trial of malaria vaccine SPf66 in Gambian infants. *Lancet* **346**: 462-467.

D'Allessandro U et Coosemans M. 1997. Concerns on the long-term efficacy of an insecticide-treated bednet program on child mortality. *Parasitology Today* **13(3)**: 124-125.

Dallas JF. 1992. Estimation of microsatellite mutation rates in recombinant inbred strains of mouse. *Mamm. Genome* **3**: 452-456.

Danis M et Mouchet J. 1991. Paludisme. Paris, Universités francophones UREF. 240pp.

Daruty de Grandpre A et Emmerrez de Charmoy D d'. 1900. Les moustiques: anatomie et biologie. Contribution à l'étude des Culicidae et principalement des genres *Culex* et *Anopheles*, de leur rôle dans la propagation de la malaria et de la filariose et des moyens de s'en préserver. *Planters & Commercial Gazette*, Port-Louis, 59pp.

Davidson G. 1962. *Anopheles gambiae* complex. *Nature* **196**: 907.

Davidson G et Jakson CE. 1962. Incipient speciation in *Anopheles gambiae* Giles. *Bull. Wld. Hlth. Org.* **27**: 303-305.

Davidson G. 1964a. The five mating types in the *Anopheles gambiae* complex. *Riv. Malar.* **13**: 167-183.

Davidson G. 1964b. *Anopheles gambiae*, a complex of species. *Bull. Wld. Hlth. Org.* **31**: 625.

Davidson G et White GB. 1972. The crossing characteristics of a new, sixth species in the *Anopheles gambiae* complex. *Trans. R. Soc. Trop. Med. Hyg.* **66**: 531-532.

DDASS (Direction Départementale des Affaires Sanitaires et Sociales) de La Réunion – Service Santé et Environnement. 1997. Lutte contre la réintroduction du paludisme à La Réunion. Rapport officiel de Novembre 1997. DDASS-DDSS, 97408 Saint-Denis, La Réunion.

Declan B. 1997. Time to put malaria on the global agenda. *Nature* **386**: 535-541.

Deka R, Shriver MD, Yu LM, Aston CE, Chakraborty R et Ferrell RE. 1994. Conservation of human chromosome 13 polymorphic human (CA)n repeats in chimpanzees. *Genomics* **22**: 226-230.

Della Torre A, Favia G, Mariotti G, Coluzzi M et Mathiopoulos KD. 1996. Physical map of the malaria vector *Anopheles gambiae. Genetics* **143**: 1307-1311.

Della Torre A, Merzagora L, Powell JR et Coluzzi M. 1997. Selective introgression of paracentric inversions between two sibling species of the *Anopheles gambiae* complex. *Genetics* **146**: 239-244.

Detinova TS. 1962. Age grouping methods in Diptera of medical importance, with special reference to some vectors of malaria. World Health Organization, Geneva.

Di Deco MA, Petrarca V, Villani F et Coluzzi M. 1980. Polimorfismo cromosomico da inversioni paracentriche ed eccesso degli eterocariotipi in ceppi di *Anopheles* allevati in laboratorio. *Parassitologia* **22**: 304-306.

Di Deco MA, Sabatinelli G, Camiz S et Touré YT. 1983. Studio biometrico di due nuove entità del complesso *Anopheles gambiae*. Parassitologia **25**: 260-267.

Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M et Freimer NB. 1994. Mutational processes of simple-sequences repeat loci in human populations. *Proc. Natl. Acad. Sci. USA* **91**: 3166-3170.

Diagne N, Fontenille D, Konate L, Faye O, Traore-Lamizana M, Molez JF et Trape JF. 1994. Les Anophèles du Sénégal: liste commentée et illustrée. *Bull. Soc. Path. Exo.* **87**: 267-277.

Dib C, Fauré S, Fizaures C, Samson D, Drouot N, Vigual A, Millasseau P, Marc S, Hazan J, Sebaun E, Lathrop M, Gyapay G, Morissette J et Weissenbach J. 1996. A comprehensive genetic map of the human genome based on 5264 microsatellites. *Nature* **380**: 152-154.

Digoutte JP. 1993. Rapport annuel du centre collaborateur OMS de référence et de recherche pour les arbovirus. Institut Pasteur, Dakar, Sénégal.

Dimopoulos G, Zheng L, Kumar V, Della Torre A, Kafatos FC et Louis C. 1996. Integrated genetic map of *Anopheles gambiae*: use of RAPD polymorphisms for genetic, cytogenetic and STS landmarks. *Genetics* **143**: 953-960.

Dover G.1995. Slippery DNA runs on and on and on.... *Nature Genetics* **10**: 254-256.

Doyle JJ, Morgane M, Tingey SV et Powell W. 1998. Size homoplasy in chloroplast microsatellites of wild perennial relatives of Soybean (*Glycine* Subgenus *Glycine*). *Mol. Biol. Evol.* **15(2)**: 215-218.

Durvasula RV, Gumbs A et Panackal A, Kruglov O, Aksoy S, Merrifield RB, Richards FF et Beard CB. 1997. Prevention of insect-borne disease: an approach using transgenic symbiotic bacteria. *Proc. Natl. Acad. Sci. USA* **94**: 3274-3278.

Edwards A, Hammond HA, Jin L, Caskey CT et Chakraborty R. 1992. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics* **12**: 241-253.

Engers HD et Godal T. 1998. Malaria vaccine development: current status. *Parasitology Today* **14(2)**: 56-64.

Epperson BK. 1993. Recent advances in correlation studies of spatial patterns of genetic variation. *In*: Evolutionary Biology. Ed Hecht MK. New-York, Plenum Press.

Epplen C, Melmer G, Siedlaczck I, Scwaiger FW, Mäueler W et Epplen JT. 1993. On the essence of "meaningless" simple repetitive DNA in eukaryote genomes. *In*: DNA

Fingerprint: State of the Science. Eds Pena SDJ, Chakraborty R, Epplen JT and Jeffreys AJ. Brikhauser Verlag, Basel. Pp 29-45.

Estoup A, Garnery L, Solignac M et Cornuet JM. 1995a. Microsatellite variation in honey bee (*Apis mellifera* L.) populations: hierarchical genetic structure and test of the infinite allele and stepwise mutation models. *Genetics* **140**: 679-695.

Estoup A, Tailliez C, Cornuet JM et Solignac M. 1995b. Size homoplasy and mutational process of interrupted microsatellites in two bee species, *Apis mellifera* and *Bombus terrestris* (Apidae). *Mol. Biol. Evol.* **12(6)**: 1074-1084.

Ewens WJ et Gillespie JH. 1974. Some simulation results for the neutral allele model with interpretations. *Theor. Popul. Biol.* **6**: 35-57.

Excoffier L, Smouse P et Quattro J. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**: 479-491.

Failloux AB, Raymond M, Ung A, Chevillon C et Pasteur N. 1997. Genetic differentiation associated with commercial traffic in the Polynesian mosquito, *Aedes polynesiensis* Marks 1951. *Biological Journal of the Linnean Society* **60**: 107-118.

Favia G, Dimopoulos G et Louis C. 1994. Analysis of the *Anopheles gambiae* genome using RAPD markers. *Insect. Biol. Mol.* **3**: 149-157.

Favia G, Della Torre A, Bagayoko M, Lanfrancotti A, Sagnon N'F, Touré YT et Coluzzi M. 1997. Molecular identification of sympatric chromosomal forms of *Anopheles gambiae* and further evidence of their isolation. *Insect Molecular Biology* **6(4)**: 377-383.

Faye O, Fontenille D, Herve JP, Diak PA, Diallo S et Mouchet J. 1993. Le paludisme en zone sahélienne du Sénégal. 1- Données entomologiques sur la transmission. *Ann. Soc. Belge Med. Trop.* **73**: 21-30.

Feldmann AM et Ponnudurai T. 1989. Selection of *Anopheles stephensi* for refractoriness and susceptibility to *Plasmodium falciparum. Med. Vet. Entomol.* **3**: 41-52.

Ferreira M. 1996. Malaria vaccine: the missing qualitative data. *Immunology and Cell Biology* **74(4)**: 301-305.

Fisher RA et Ford EB. 1947. The spread of a gene in natural conditions in a colony of the moth *Panaxia dominula. Heredity* **1**: 143-174.

Foltz DW. 1986. Null alleles as a possible cause of heterozygote deficiencies in the oyster *Crassostera virginica* and other bivalves. *Evolution* **40**: 869-870.

Fontenille D, Lepers JP, Campbell GH, Coluzzi M, Rakotoarivony I et Coulanges P. 1990. Malaria transmission and vector biology in Manarintsoa, High Plateaux of Madagascar. *Am. J. Trop. Med. Hyg.* **43(2)**: 107-115.

Fontenille D. 1992. Hétérogénéité de la transmission des paludismes à Madagascar. *Mem. Soc. r. Belge Ent.* **35**: 129-132.

Fontenille D, Faye O, Konate L, Sy N and Collins FH. 1993. Comparaison des techniques PCR et cytogénétique pour la détermination des membres du complexe *Anopheles gambiae* au Sénégal. *Ann. Parasitol. Hum. Comp.* **68(5/6)**: 239-240.

Fontenille D, Lochouarn L, Diagne N, Sokhna C, Lemasson JJ, Diatta M, Konate L, Faye F, Rogier C et Trape JF. 1997. High annual and seasonal variations in malaria transmission by anophelines and vector species composition in Dielmo, a holoendemic area in Senegal. *Am. J. Trop. Med. Hyg.* **56(3)**: 247-253.

FritzSimmons NN, Moritz C et Moore SS. 1995. Conservation and dynamics of microsatellite loci over 300 million years of marine turtle evolution. *Mol. Biol. Evol.* **12**: 432-440.

Frizzi G et Holstein M. 1956. Etude cytogénétique d'*Anopheles gambiae*. *Bull. Wld. Hlth. Org.* **15**: 425-435.

Garcia BA, Caccone A, Mathiopoulos KD et Powell JR. 1996. Inversion monophyly in African Anopheline malaria vectors. *Genetics* **143**: 1313-1320.

Garza JC, Slatkin M et Freimer NB. 1995. Microsatellite allele frequencies in Humans and Chimpanzees, with implications for constaints on allele size. *Mol. Biol. Evol.* **12(4)**: 594-603.

Garza JC et Freimer NB. 1996. Homoplasy for size at microsatellite loci in Humans and Chimpanzees. *Gen. Res.* **6**: 211-217.

Gillies MT et De Meillon B. 1968. The Anophelinae of Africa south of the Sahara. *Pub. South Afr. Inst. Med. Res.* **54**: 343 pp.

Gillies MT et Coetzee M. 1987. A supplement to the Anophelinae of Africa south of the Sahara. *Pub. South Afr. Inst. Med. Res.* **55**: 143 pp.

Goldstein DB, Linares AR, Feldman MW and Cavalli-Sforza LL. 1995a. An evaluation of genetic distances for use with microsatellite loci. *Genetics* **139**: 463-471.

Goldstein DB, Linares AR, Cavalli-Sforza LL et Feldman MW. 1995b. Genetic absolute dating based on microsatellites and the origin of modern humans. *Proc. Natl. Acad. Sci. USA* **92**: 6723-6727.

Goldstein DB et Pollock DD. 1997. Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. *J. Hered.* **88**: 335-342.

Goldstein DB, Roemer GW, Smith DB, Reich DE, Bergman A et Wayne R. 1999. The use of microsatellite variation to infer population structure and demographic history in a natural model system. *Genetics* **151**: 797-801.

Gorman MJ, Cornel AJ, Collins FH et Paskewitz SM. 1996. A shared mechanism for melanotic encapsulation of CM-Spehadex beads and a malaria parasite, *Plasmodium cynomolgi* B, in the mosquito, *Anopheles gambiae*. *Exp. Parasitol.* **84**: 380-386.

Gorman MJ et Paskewitz SM. 1997. A genetic study of melanization response to Sephadex beads in *Plasmodium*-refractory and -susceptible strains of *Anopheles gambiae*. *Am. J. Trop. Med. Hyg.* **56(4)**: 446-451.

Gorman MJ, Severson DW, Cornel AJ, Collins FH et Paskewitz SM. 1997. Mapping a quantitative trait locus involved in melanotic encapsulation of foreign bodies in the malaria vector, *Anopheles gambiae*. *Genetics* **146**: 965-971.

Goudet J, Raymond M, De Meeüs T et Rousset F. 1996. Testing differentiation in diploid populations. *Genetics* **144**: 1933-1940.

Green CA. 1972. Cytological maps for the practical identification of females of the three freshwater species of the *Anopheles gambiae* complex. *Ann. Trop. Med. Parasitol.* **66**: 143-147.

Green CA. 1980. Lack of recombination between X-chromosomes in interspecific hybrids of the *Anopheles gambiae* group of species. *Heredity* **44**: 423-424.

Grimaldi MC et Crouau-Roy B. 1997. Microsatellite allelic homoplasy due to variable flanking sequences. *J. Mol. Evol.* **44**: 336-340.

Guo SW et Thompson EA. 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* **48**: 361-372.

Gwadz R et Collins FH. 1996. Anopheline mosquitoes and the agents they transmit. *In*: The biology of disease vectors. Eds Beaty BJ and Marquardt WC. University Press of Colorado, USA. Pp: 73-84.

Gysin J et Doury J-C. 1997. Recherche et Paludisme: de l'échec aux espoirs. *Med. Trop.* **57**: 243-244.

Haldane JBS. 1954. An exact test for randomness of mating. *J. Genet.* **52**: 631-635.

Hamada H, Petrino MG et Takunaga T. 1982. A novel repeated element with Z-DNA-forming potential is widely found in evolutionary diverse eukaryotic genomes. *Proc. Natl. Acad. Sci. USA* **79**: 6465-6469.

Hamilton RJ et Service MW. 1983. Value of cuticular and internal hydrocarbons for the identification of *Anopheles gambiae* Giles, *Anopheles arabiensis* Patton and *Anopheles melas* Theobald. *Ann. Trop. Med. Parasitol.* **77**: 203-210.

Hamon J, Adam JP et Grjebine A. 1956. Observations sur la répartition et le comportement des anophèles de l'Afrique équatoriale française, du Cameroun et de l'Afrique occidentale. *Bull. Wld Hlth Org.* **15**: 549-591.

Hardy GH. 1908. Mendelian proportions in a mixed population. *Science* **28**: 49-50.

Hartl DL et Clark AG. 1989. Principles of population genetics, 2nd Edition. Eds Sinauer Associates, Sunderland, Mass.

Hedrick PW. 1980. Hitchhiking: a comparison of linkage and partial selfing. *Genetics* **94**: 791-808.

Herrington DA, Clyde DF, Losonsky G, Cortesia M, Murphy JR, Davis J, Baqar S, Felix AM, Heimer EP, Gillessen D, Nardin E, Nussenzweig RS, Nussenzweig V, Hollingdale MR et Levine MM. 1987. Safety and immunogenicity in man of a synthetic peptide malaria vaccine against *Plasmodium falciparum* sporozoites. *Nature* **328**: 257-259.

Hervy JP, Le Goff G, Geoffroy B, Hervé JP, Manga L et Brunhes J. 1998. Les anophèles de la région afrotropicale. CD-ROM, ORSTOM-éditions, Paris.

Holland MM. 1993. Short tandem repeat loci: application to forensic and human remains identification. *In*: DNA Fingerprint: State of the Science. Eds Pena SDJ, Chakraborty R, Epplen JT and Jeffreys AJ. Brikhauser Verlag, Basel. Pp 267-274.

Holm S. 1979. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* **6**: 65-70.

Holstein MH. 1954. Biology of *Anopheles gambiae*, research in French West Africa. WHO, Geneva. 173 pp.

Holvoet G, Michielsen P et Vandepitte P. 1983. Autochtonous *falciparum* malaria in Belgium. *Ann. Soc. Belge Med. Trop.* **63**: 111-117.

Hunt RH, Coetzee M et Fettene M. 1998. The *Anopheles gambiae* complex: a new species from Ethiopia. *Trans. R. Soc. Trop. Med. Hyg.* **92**: 231-235.

Irwin DM, Kocher TD et Wilson AC. 1991. Evolution of cytochrome b gene of mammals. *J. Mol. Evol.* **32** : 128-144.

Jarne P et Lagoda JL. 1996. Microsatellites, from molecules to populations and back. *Trends Ecol. Evol.* **11(10)**: 424-429.

Jasinskiene N, Coates CJ, Benedict MQ, Cornel AJ, Salazar-Rafferty C, James AA et Collins FH. 1998. Stable transformation of the yellow fever mosquito, *Aedes aegypti*, with the *Hermes* element from the housefly. *Proc. Natl. Acad. Sci. USA* **95**: 3743-3747.

Jin L et Chakraborty R. 1995. Population structure, stepwise mutations, heterozygote deficiency and their implications in DNA forensics. *Heredity* **74**: 274-285.

Jorde PE et Ryman N. 1996. Demographic genetics of brown trout (*Salmo trutta*) and estimation of effective population size from temporal change of allele frequencies. *Genetics* **143**: 1369-1381.

Julvez J, Mouchet J et Ragavoodoo C. 1990. Epidemiologie historique du paludisme dans l'archipel des Mascareignes (Océean Indien). *Ann. Soc. belge Med. Trop.* **70**: 249-261.

Julvez J et Mouchet J. 1994a. Le peuplement culicidien des îles du sud-ouest de l'Océan Indien: l'action de l'homme dans l'importation des espèces d'intérêt médical. *Ann. Soc. Entomol. France* **30(4)**: 391-401.

Julvez J et Mouchet J. 1994b. Epidémiologie historique de la filariose lymphatique dans les îles du sud-ouest de l'Océan Indien. *Bull. Soc. Path. Ex.* **87(3)**: 194-201.

Julvez J, Ragavoodoo C, Gopaul AR et Mouchet J. 1998. Maladies humaines transmises par les culicidés dans les îles du sud-ouest de l'Océan Indien. *Bull. Soc. Path. Ex.* **91(1)**: 99-103.

Kamau L, Lehmann T, Hawley WA, Orago ASS et Collins FH. 1998a. Microgeographic genetic differentiation of *Anopheles gambiae* mosquitoes from Asembo Bay, western Kenya: a comparison with Kilifi in Coastal Kenya. *Am. J. Trop. Med. Hyg.* **58(1)**: 64-69.

Kamau L, Hawley WA, Lehmann T, Orago ASS, Cornel A, Ke Z et Collins FH. 1998b. Use of short tandem repeats for the analysis of genetic variability in sympatric populations of *Anopheles gambiae* and *Anopheles arabiensis*. *Heredity* **80**: 675-682.

Kaslow DC. 1997. Transmission-blocking vaccines - uses and current status of development. *International Journal of Parasitology* **27(2)**: 183-189.

Kidwell MG et Ribeiro JMC. 1992.Can transposable elements be used to drive disease refractoriness genes into vector populations? *Parasitology today* **8**: 325-329.

Kimmel M, Chakraborty R, Stivers DN et Deka R. 1996. Dynamics of repeat polymorphism under a forward-backward mutation model: within- and between-population variability at microsatellite loci. *Genetics* **143**: 549-555.

Kimura M. 1953. «Stepping stone» model of populations. *Ann. Rep. Natl. Inst. Genet. Japan* **3**: 62-63.

Kimura M et Crow JF. 1964. The number of alleles that can be maintained in a finite population. *Genetics* **49**: 725-738.

Kimura M et Ohta T. 1978. Stepwise mutation model and distribution of allelic frequencies in a finite population. *Proc. Natl. Acad. Sci. USA* **75(6)**: 2868-2872.

Koorey DJ, Bishop GA et McCaughan GW. 1993. Allele non-amplification: a source of confusion in linkage studies employing microsatellite polymorphisms. *Human molecular Genetics* **2**: 289-291.

Krimbas CB et Powell JR (Eds). 1992. Drosophila Inversion Polymorphism. CRC Press, Boca Raton FL, USA.

Krimbas KB et Tsakas S. 1971. The genetics of *Dacus oleae*. V. Changes of esterase polymorphism in a natural population following insecticide control: selection or drift? *Evolution* **25**: 454-460.

Kruglyak S, Durett RT, Schug MD et Aquadro CF. 1998. Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. *Proc. Natl. Acad. Sci. USA* **95(18)**: 10774-10778.

Lanzaro GC, Zheng L, Touré YT, Traoré SF, Kafatos FC et Vernick KD. 1995. Microsatellite DNA and isozyme variability in a west African population of *Anopheles gambiae*. *Insect Mol. Biol.* **4(2)**: 105-112.

Lanzaro GC, Touré YT, Carnahan J, Zheng L, Dolo G, Traoré S, Petrarca V, Vernick KD et Taylor CE. 1998. Complexities in the genetic structure of *Anopheles gambiae* populations in west Africa as revealed by microsatellite DNA analysis. *Proc. Natl. Acad. Sci. USA* **95**: 14260-14265.

Latter BDH. 1973. The island model of population differentiation: a general solution. *Genetics* **73**: 147-157.

Laventure S, Mouchet J, Blanchy S, Marrama L et Rajaonarivelo E. 1996. Le riz, source de vie et de mort sur les Plateaux de Madagascar. *Cahiers Santé* **6**: 79-86.

Lehmann T, Hawley WA et Collins FH. 1996. An evaluation of evolutionary constraints on microsatellite loci using null alleles. *Genetics* **144**: 1155-1163.

Lehmann T, Besansky NJ, Hawley WA, Fahey TG, Kamau L et Collins FH. 1997. Microgeographic structure of *Anopheles gambiae* in western Kenya based on mtDNA and microsatellite. *Mol. Ecol.* **6**: 243-253.

Lehmann T, Hawley WA, Grebert H et Collins FH. 1998. The effective population size of *Anopheles gambiae* in Kenya: Implications for population structure. *Mol. Biol. Evol.* **15(3)**: 264-276.

Lengeler C, Smith TA et Armstrong Schellenberg JR. 1997. Focus on the effect of bednets on malaria morbidity and mortality. *Parasitology Today* **13(3)**: 123-124.

Lengeler C. 1998. Insecticide bednets and curtains for malaria control (Cochrane Review). In: The Cochrane Library, Issue 3. Oxford: Update software.

Lenormand T et Raymond M. 1998. Resistance management: the stable zone strtegy. *Proc. R. Soc. Lond. B* **265**: 1983-1988.

Levins R. 1969. Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bulletin of the Entomological Society of America* **15**: 237-240.

Levins R. 1970. Extinction. *In*: Some mathematical problems in biology. Ed. Gesternhaber M. Providence R.I.. American Mathematical Society: Pp 77-107.

Levinson G et Gutman GA. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* **4**: 203-221.

Lewin B. 1994. Genes IV. Ed. John Willey & Sons, New York.

Luikart G et Cornuet JM. 1997. Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. *Conservation Biology* **12(1)**: 228-237.

Luikart G. 1997. Usefulness of molecular markers for detecting population bottlenecks and monitoring genetic change. Ph.D thesis. University of Montana, Missoula, USA.

Mahon RJ, Green CA et Hunt RH. 1976. Diagnostic allozymes for routine identification of adults of the *Anopheles gambiae* complex (Diptera: Culicidae). *Bull. Ent. Res.* **66**: 25-31.

Marchand RP et Mzanva AEP. 1985. A field test of biochemical key to identify members of the *Anopheles gambiae* group of species in north-east Tanzania. *Am. J. Trop. Med. Hyg.* **88**: 205-210.

Martinez-Torres D, Chandre F, Williamson MS, Darriet F, Bergé JB, Devonshire AL, Guillet P, Pasteur N et Pauron D. 1998. Molecular characterization of pyrethroid knockdown resistance (*kdr*) in the major malaria vector *Anopheles gambiae s.s.*. *Insect Mol. Biol.* **7(2)**: 179-184.

Maruyama T et Fuerst PA. 1985a. Population bottlenecks and nonequilibrium models in population genetics. II. Number of alleles in a small population that was formed by a recent bottleneck. *Genetics* **111**: 675-689.

Maruyama T et Fuerst PA. 1985b. Population bottlenecks and nonequilibrium models in population genetics. III. Genic homozygosity in populations which experience periodic bottlenecks. *Genetics* **111**: 691-703.

Mathiopoulos KD et Lanzaro GC. 1995. Distribution of genetic diversity in relation to chromosomal inversions in the malaria mosquito *Anopheles gambiae. J. Mol. Evol.* **40**: 578-584.

Mathiopoulos KD, Powell JR et McCutchan TF. 1995. An anchored restriction-mapping approach applied to the genetic analysis of the *Anopheles gambiae* malaria vector complex. *Mol. Biol. Evol.* **12**: 103-112.

McLain DK, Collins FH, Brandling-Bennett AD et Were JBO. 1989. Microgeographic variation in rDNA intergenic spacers of *Anopheles gambiae* in western Kenya. *Heredity* **62**: 257-264.

Mekuria Y, Petrarca V et Tesfamariam T. 1982. Cytogenetic studies on the malaria vector mosquito *Anopheles arabiensis* in the Awash valley, Ethiopia. *Parassitologia* **24**: 238-243.

Michalakis Y et Excoffier L. 1996. A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. *Genetics* **142**: 1061-1064.

Miles SJ. 1978. Enzyme variations in the *Anopheles gambiae* Giles group of species (Diptera, Culicidae). *Bull. Entomol. Res.* **68**: 85-96.

Miles SJ. 1979. A biochemical key to adult members of the *Anopheles gambiae* group of species (Diptera: Culicidae). *J. Med. Entomol.* **15**: 297-299.

Miller LH, Good MF et Kaslow DC. 1997. The need for assays predictive of protection in development of malaria bloodstage vaccines. *Parasitology Today* **13(2)**: 46-47.

Miller LM et Kapuscinski AR. 1997. Historical analysis of genetic variation reveals low effective population size in a Northern Pike (*Esox lucius*) population. *Genetics* **147**: 1249-1258.

Mnzava AEP, Rwegoshora RT, Wilkes TJ, Tanner M et Curtis CF. 1995. *Anopheles arabiensis* and *An. gambiae* chromosomal inversion polymorphism, feeding and resting behaviour in relation to insecticide house-spraying in Tanzania. *Med. Vet. Entomol.* **9**: 316-324.

Mons B. 1997. The problem of multiplicity in malaria vaccine development. *Ann. Trop. Med. Parasitol.* **91(1)**: 55-58.

Mosha FW et Subra R. 1982. Ecological studies on *An. gambiae* complex sibling species in Kenya. I. Preliminary observations on their geographical distribution and chromosomal polymorphic inversions. WHO/VBC/82.867. Geneva: Switzerland.

Mouchet J, Carnevale P, Coosemans M, Fontenille D, Ravaonjanahary C, Richard A et Robert V. 1993. Typologie du paludisme en Afrique. Cahiers Santé ORSTOM **3(4)**: 220-238.

Mouchet J, Laventure S, Blanchy S, Fioramonti R, Rakotonjanabelo A, Rabarison P, Sircoulon J et Roux J. 1997. La reconquête des hautes terres de Madagascar par le paludisme. *Bull. Soc. Path. Ex.* **90(3)**: 162-168.

Muirhead-Thomson RC. 1951. Studies on salt-water and fresh-water *Anopheles gambiae* on the East African coast. *Bull. Entomol. Res.* **41**: 487-502.

Mukiama T. 1987. Genetic variation in wild *Anopheles arabiensis* Patton of Mwea irrigation scheme, Kenya. *Insect Sci. Applic.* **8(2)**: 245-249.

Mulla MS, 1994. Mosquito control then, now, and in the future. *J. Am. Mosq. Control Assoc.* **10(4)**: 574-584.

Nàjera JA et Hempel J. 1996. The burden of malaria. Unpublished report CTD/MAL/96.10. World Health Organization. Geneva: Switzerland.

Nauta MJ et Weissing FJ. 1996. Constraints on allele size at microsatellite loci: implications for genetic differentiation. *Genetics* **143**: 1021-1032.

Neel JV. 1973. «Private» genetic variants and the frequency of mutation among South American Indians. *Proc. Natl. Acad. Sci. USA* **70**: 3311-3315.

Nei M, Maruyama T et Chakraborty R. 1975. The bottleneck effect and genetic variability in populations. *Evolution* **29(1)**: 1-10.

Nei M. 1978. Estimation of the average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**: 583-590.

Nei M et Tajima F. 1981. Genetic drift and estimation of effective population size. *Genetics* **98**: 625-640.

Nei M et Chesser RK. 1983. Estimation of fixation indices and gene diversities. *Ann. Hum. Genet.* **47**: 253-259.

Nei M. 1987. Molecular evolutionary genetics. New-York, Columbia University Press, USA.

Nordheim A et Rich A. 1983. The sequence (dC-dA)n.(dG-dT)n forms left-handed Z-DNA in negatively supercoiled plasmids. *Proc. Natl. Acad. Sci. USA* **80**: 1821-1825.

Nosten F, Luxemburger C, Kyle DE, Ballou WR, Wittes J, Wah E, Chongsuphajaisiddhi T, Gordon DM, White NJ, Sadoff JC *et al.*. 1996. Randomised double-blind placebo-controlled trial of SPf66 malaria vaccine in children in northwestern Thailand. *Lancet* **348**: 701-707.

Omer SM et Cloudsley-Thompson JL. 1968. Dry season biology of *Anopheles gambiae* Giles in the Sudan. *Nature* **217**: 879-880.

Omer SM et Cloudsley-Thompson JL. 1970. Survival of females *Anopheles gambiae* Giles through a 9-month dry season in Sudan. *Bull. Wld. Hlth. Org.* **42**: 319-330.

OMS. 1996. Insecticide impregnated materials in the African region. Report of a meeting in Brazzaville, 18-20 March 1996. Brazzaville: World Health Organization.

OMS. 1998. Evaluation de la santé. *In*: Rapport sur la Santé dans le Monde. La vie au 21e siècle, une perspective pour tous. World Health Organization (ed), Genève, Suisse. Pp: 43-65.

Otha T et Kimura M. 1973. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genetical Research* **22**: 201-204.

Paetkau D et Strobeck C. 1995. The molecular basis and evolutionary history of a microsatellite null allele in bears. *Mol. Ecol.* **4**: 519-520.

Pardue ML, Lowenhaupt K, Rich A et Nordheim A. 1987. (dC-dA)n.(dG-dT)n sequences have evolutionary conserved chromosomal locations in dropsophila with implications for roles in chromosome structure and function. *EMBO J.* **6**: 1781-1789.

Patarroyo ME, Amador R, Clavijo P, Moreno A, Guzman F, Romero P, Tascon R, Franco A, Murillo LA, Ponto G et Turillo G. 1988. A synthetic vaccine protects humans against challenge with asexual blood stages of *Plasmodium falciparum* malaria. *Nature* **332**: 158-161.

Paterson HE. 1964. Direct evidence for the specific distinctness of Forms A, B and C of the *Anopheles gambiae* complex. *Riv. Malar.* **43**: 191-196.

Pauron D, Barhanin J, Amichot M, Pralavorio M, Bergé JB et Lazdunski M. 1989. Pyrethroid receptor in the insect Na^+ channel: alteration of its properties in pyrethroid-resistant fies. *Biochemistry* **28**: 1673-1677.

Pemberton JM, Slate J, Bancroft DR et Barrett JA. 1995. Nonamplifying alleles at microsatellite loci: a caution for parentage and population studies. *Mol. Ecol.* **4**: 249-252.

Pépin L, Amiques Y, Lepingle A, Berthier J, Bensaid A et Vaiman D. 1995. Sequence conservation of microsatellites between *Bos taurus* (cattle), *Capra hircus* (goat) and related species. Examples of use in parentage testing and phylogeny analysis. *Heredity* **74**: 53-61.

Persiani A, Di Deco MA et Petrangeli G. 1984. Osservazioni di laboratorio su polimofismi da inversioni originati da incroci tra popolazioni diverse di *Anopheles gambiae* s.s.. *Atti XIII Congr. Soc. It. Parassitologia*. Roma. 3-5 dicembre 1984.

Petrarca V, Carrara GC, Di Deco MA et Petrangeli G. 1983. Il complesso *Anopheles gambiae* in Guinea Bissau. *Parassitologia* **25**: 29-39.

Petrarca V, Petrangeli G, Rossi P et Sabatinelli G. 1986. Etude chromosomique d'*Anopheles gambiae* et d'*Anopheles arabiensis* à Ouagadougou (Burkina Faso) et dans quelques villages voisins. *Parassitologia* **28**: 41-61.

Petrarca V, Vercruysse J et Coluzzi M. 1987. Observations on the *Anopheles gambiae* complex in the Senegal River Basin, West Africa. *Med. Vet. Entomol.* **1**: 303-312.

Petrarca V et Beier JC. 1992. Intraspecific chromosomal polymorphisms in the *Anopheles gambiae* complex as a factor affecting malaria transmission in the Kisumu area of Kenya. *Am. J. Trop. Med. Hyg.* **46**: 229-237.

Pollak E. 1983. A new method for estimating the effective population size from allele frequency changes. *Genetics* **104**: 531-548.

Popadic A et Anderson WW. 1994. The history of a genetic system. *Proc. Natl. Acad. Sci. USA* **91**: 6819-6823.

Popadic A, Popadic D et Anderson WW. 1995. Interchromosomal transfer of genetic information between gene arrangements on the third chromosome of *Drosophila pseudoobscura*. *Mol. Biol. Evol.* **12**: 938-943.

Primmer CR, Moller AP et Ellegren H. 1996a. A wide-range survey of cross-species microsatellite amplification in birds. *Mol. Ecol.* **5**: 365-378.

Primmer CR, Ellegren H, Saino N et Moller AP. 1996b. Directional evolution in germline microsatellite mutations. *Nature Genetics* **13**: 391-393.

Queller DC, Strassmann JE et Hughes CR. 1993. Microsatellites and kinship. *Trends Ecol. Evol.* **8**: 285-288.

Ralisoa Randrianasolo BO et Coluzzi M. 1987. Genetical investigations on zoophilic and exophilic *Anopheles arabiensis* from Antananarivo (Madagascar). *Parassitologia* **29**: 93-97.

Rannala B et Hartigan JA. 1996. Estimating gene flow in island populations. *Genetical Research Camb.* **67**: 147-158.

Raymond M et Rousset F. 1995a. GENEPOP Version 1.2. A population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**: 248-249.

Raymond M et Rousset F. 1995b. An exact test for population differentiation. *Evolution* **49**: 1280-1283.

Raymond M. 1996. Les marqueurs génétiques en biologie des populations: vers une écologie moléculaire. *Médecine/Sciences* **12(n° spécial)**: 153-157.

Reich DE et Goldstein DB. 1998. Genetic evidence for a paleolithic human population expansion in Africa. *Proc. Natl. Acad. Sci. USA* **95**: 8119-8123.

Rice WR. 1989. Analysing tables of statistical tests. *Evolution* **43(1)**: 223-225.

Richardson BJ, Baverstock PR et Adams M. 1986. Allozyme electrophoresis. A handbook for animal systematics and population studies. Academic, Sydney.

Robert V et Carnevale P. 1984. Les vecteurs du paludisme en Afrique sub-saharienne. *Etudes Médicales* **2**: 79-90.

Robert V, Petrarca V, Carnevale P, Ovazza L et Coluzzi M. 1989. Analyse cytogénétique du complexe *Anopheles gambiae* dans la région de Bobo-Dioulasso (Burkina Faso). *Ann. Parasit. Hum. Comp.* **64**: 290-311.

Robert V et Trape JF. 1997. Lutter contre le paludisme en réduisant sa transmission: présentation de la controverse. *MS-Medecine Sciences* **13(5)**: 678-682.

Rodhain F et Perez C. 1985. Précis d'entomologie médicale et vétérinaire. Paris.

Rosci MA, Paglia MG, de Felici A, Antonucci G, Armignacco O, Cardini A, Osci L et Savioli L. 1987. A case of *falciparum* malaria acquired in Italy. *Trop. Geogr. Med.* **39**: 77-79.

Rousset F et Raymond M. 1995. Testing heterozygote excess and deficiency. *Genetics* **140**: 1413-1419.

Rozas J et Aguadé M. 1994. Gene conversion is involved in the transfer of genetic information between naturally occuring inversions of *Drosophila. Proc. Natl. Acad. Sci. USA* **91**: 11517-11521.

Rubinsztein DC, Amos W, Leggo J, Goodburn S, Jain S, Li SH, Margolis RL, Ross CA et Ferguson-Smith MA. 1995. Microsatellite evolution - evidence for directionality and variation in rate between species. *Nature Genetics* **10**: 337-343

SAS Institute Inc. 1990. SAS language: references. Version 6, 1st edn. Cary, NC.

Schlötterer C, Amos B et Tautz D. 1991. Conservation of polymorphic simple sequence loci in cetacean species. *Nature* **354**: 63-65.

Schlötterer C, Ritter R, Harr B et Brem G. 1998. High mutation rate of a long microsatellite allele in *Drosophila melanogaster* provides evidence for allele-specific mutation rates. *Mol. Biol. Evol.* **15(10)**: 1269-1274.

Schneider S, Kueffer JM, Roessli D et Excoffier L. 1997. Arlequin ver. 1.1: A software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Scott JA, Brogdon WG et Collins FH. 1993. Identification of single specimens of *Anopheles gambiae* complex by the polymerase chain reaction. *Am. J. Trop. Med. Hyg.* **49**: 520-529.

Service MW. 1993. Mosquito Ecology: Field Sampling Methods, 2nd edition. Elsevier Applied Science Publishers. London.

Service MW. 1997. Mosquito (Diptera: Culicidae) dispersal - the long and short of it. *J. Med. Entomol.* **34(6)**: 579-588.

Shahabuddin M, Lemos FJA, Kaslow DC et Jacobs-Lorena M. 1996. Antibody-mediated inhibition of *Aedes aegypti* midgut trypsins blocks sporogonic development of *Plasmodium gallinaceum. Infection and Immunity* **64(3)**: 739-743.

Shahabuddin M, Cociancich S et Zieler H. 1999. The search for novel malaria transmission-blocking targets in the mosquito midgut. *Parasitology Today* **14(12)**: 493-497.

Shaw CR et Prasad R. 1970. Strach gel electrophoresis of enzymes: a compilation of recipes. *Biochem. Gen.* **4**: 297-320.

Shpilberg O, Shaked Y, Sjroh CM, Ehrenfeld M et Samra Y. 1988. Suspected airport malaria in Israel. *Harefuah* **115**: 117-119.

Shriver MD, Jin L, Chakraborty R et Boerwinkle E. 1993. VNTR allele frequency distribution under the Stepwise Mutation Model: a computer simulation approach. *Genetics* **134**: 983-993.

Sinkins SP, Curtis CF et O'Neill S. 1997. The potential application of inherited symbiont systems to pest control. *In:* Influential Passengers: Inherited Microorganisms and Arthropod Reproduction. Eds O'Neill S, Werren JH and Hoffmann AA. Oxford: Oxford University Press. Pp 155-175.

Slatkin M. 1985a. Gene flow in natural populations. *Annual Review of Ecology and Systematics* **16**: 393-430.

Slatkin M. 1985b. Rare alleles as indicators of gene flow. *Evolution* **39(1)**: 53-65.

Slatkin M. 1987. Gene flow and the geographical structure of natural populations. *Science* **236**: 787-792.

Slatkin M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**: 457-462.

Smits A, Roelants P, Van Bortel W et Coosemans M. 1996. Enzyme polymorphism in the *Anopheles gambiae* (Diptera: Culicidae) complex related to feeding and resting behavior in the Imbo valley, Burundi. *J. Med. Entomol.* **33(4)**: 545-553.

Snow RW, Omumbo JA, Lowe B, Molineux CS, Obiero JO, Palmer A, Weber MW, Pinder M, Nahlen B, Obonyo C, Newbold C, Gupta S et Marsh K. 1997. Relation between severe malaria morbidity in children and level of *Plasmodium falciparum* transmission in Africa. *Lancet* **349**: 1650-1654.

Soper FL and Wilson DB. 1943. *Anopheles gambiae* in Brazil 1930-1940. Rockefeller Foundation, New York.

Stallings RL. 1995. Conservation and evolution of $(CT)_n$-$(GA)_n$ microsatellite sequences at orthologous positions in diverse mammalian genomes. *Genomics* **25**: 107-113.

Sutherlands GR et Richards RI. 1995. Simple tandem DNA repeats and human genetic disease. *Proc. Natl. Acad. Sci. USA* **92**: 3636-3641.

Swofford DL et Selander RB. 1989. BIOSYS-1. A computer program for the analysis of allelic variation in population genetics and biochemical systematics. Release 1.7. University of Illinois, Urbana, IL.

Tajima F et Nei M. 1984. Note on genetic drift and estimation of effective population size. *Genetics* **106**: 569-574.

Tautz D et Rentz M. 1984. Simple sequences are ubiquitous components of eukaryotic genomes. *Nucleic Acids Research* **12(10)**: 4127-4138.

Tautz D et Schlötterer C. 1994. Simple sequences. *Current Opinion in Genetics and Development* **4**: 832-837.

Taylor CE, Touré YT, Coluzzi M et Petrarca V. 1993. Effective population size and persistence of *Anopheles arabiensis* during the dry season in West Africa. *Med. Vet. Entomol.* **7**: 351-357.

Tessier N et Bernatchez L. 1999. Stability of population structure and genetic diversity across generations assessed by microsatellites among sympatric populations of landlocked Atlantic salmon (*Salmo salar* L.). *Molecular Ecology* **8**: 169-179.

Thathy V, Severson DW et Christensen BM. 1994. Reinterpretation of the genetics of susceptibility of *Aedes aegypti* to *Plasmodium gallinaceum*. *J. Parasitol.* **80**: 705-712.

Touré YT, Petrarca V et Coluzzi M. 1983. Nuove entità del complesso *Anopheles gambiae* in Mali. *Parassitologia* **25**: 367-370.

Touré YT, Petrarca V et Coluzzi M. 1987. Patterns of geographical and seasonal distributions of the «Mopti» karyotypes of *Anopheles gambiae* in Mali. Proceedings of the 3rd International Conference on Malaria and Babesiosis, Annecy, France: 172.

Touré YT, Petrarca V, Traoré SF, Coulibaly A, Maïga HM, Sow M, Di Deco MA et Coluzzi M. 1994. Ecological genetic studies in the chromosomal form Mopti of *Anopheles gambiae* s.str. in Mali, West Africa. *Genetica* **94**: 213-223.

Touré YT, Dolo G, Petrarca V, Traore SF, Bouare M, Dao A, Carnahan J et Taylor CE. 1998. Mark-release-recapture experiments with *Anopheles gambiae s.l.* in Banambani village, Mali, to determine population size and structure. *Med. Vet. Entomol.* **12**: 74-83.

Trape JF et Rogier C. 1996. Combatting malaria morbidity and mortality by reducing malaria transmission. *Parasitology Today* **12(6)**: 236-240.

Turelli M et Hoffmann AA. 1991. Rapid spread of an inherited incompatibility factor in California *Drosophila*. *Nature* **353**: 440-442.

Turelli M, Hoffman AA et McKechnie SW. 1992. Dynamics of cytoplasmic incompatibility and mtDNA variation in natural *Drosophila simulans* populations. *Genetics* **132**: 713-723.

Valdes AM, Slatkin M et Freimer NB. 1993. Allele frequencies and microsatellite loci: the stepwise mutation model revisited. *Genetics* **133**: 737-749.

Valero MV, Amador LR, Galindo C, Figueroa J, Bello MS, Murillo LA, Mora AL, Patarroyo G, Rocha CL, Rojas M *et al.*. 1993. Vaccination with SPf66, a chemically synthesised vaccine, against *Plasmodium falciparum* malaria in Colombia. *Lancet* **341**: 705-710.

Valero MV, Amador LR, Aponte JJ, Narvaez A, Galindo C, Silva Y, Rosas J, Guzman F et Patarroyo ME. 1996. Evaluation of SPf66 malaria vaccine during a 22-month follow-up field trial in the Pacific coast of Colombia. *Vaccine* **14(15)**: 1466-1470.

Van Treuren R. 1998. Estimating null allele frequencies at a microsatellite locus in the oystercatcher (*Haematopus ostralegus*). *Mol. Ecol.* **7**: 1413-1417.

Vernick KD et Collins FH. 1989. Association of a *Plasmodium*-refractory phenotype with an esterase locus in *Anopheles gambiae*. *Am. J. Trop. Med. Hyg.* **40(6)**: 593-597.

Vernick KD, Collins FH et Gwadz RW. 1989. A general system of resistance to malaria infection in *Anopheles gambiae* controlled by two main genetic loci. *Am. J. Trop. Med. Hyg.* **40(6)**: 585-592.

Viard F. 1996. Autofécondation et migration en population subdivisées: apport des marqueurs microsatellites chez le gastéropode *Bulinus truncatus*. Thèse. Université de Montpellier II – Sciences et Techniques du Languedoc. France. 87p.

Walton C, Thelwell NJ, Priestman A et Butlin RK. 1998. The use of microsatellites to study gene flow in natural populations of *Anopheles* malaria vectors in Africa: potential and pitfalls. *J. Mosq. Control Assoc.* **14(3)**: 266-272.

Wang R, Kafatos FC et Zheng L. 1999. Microsatellite markers and genotyping procedures for *Anopheles gambiae*. *Parasitology Today* **15(1)**: 33-37.

Waples RS. 1989a. A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics* **121**: 379-391.

Waples RS. 1989b. Temporal variation in allelic frequencies: testing the right hypothesis. *Evolution* **43(6)**: 1236-1251.

Waples RS et Teel DJ. 1990. Conservation genetics of Pacific salmon. I. Temporal changes in allele frequency. *Conservation Biology* **4**: 144-156.

Waples RS. 1991. Genetic methods for estimating the effective population size of cetacean populations. In Hoelzel AR, ed. Genetic ecology of whales and dolphins. International Whaling Commission (Special Issue n°13), Cambridge, UK. Pp 279-300.

Waser PM et Strobeck C. 1998. Genetic signatures of interpopulation dispersal. *Trends Ecol. Evol.* **13(2)**: 43-44.

Weber JL et Wong C. 1993. Mutation at human short tandem repeats. *Human Molecular Genetics* **2(8)**: 1123-1128.

Weinberg W. 1908. Über den Nachweis der Vererbung beim Menschen. Jaresh. *Verein f. vaterl. Naturk. In Wruttemberg* **64**: 368-382.

Weir BS et Cockerham CC. 1984. estimating F-statistics for the analysis of population structure. *Evolution* **38**: 1358-1370.

White GB. 1970. Chromosomal evidence for natural interspecific hybridization by mosquitoes of the *Anopheles gambiae* complex. *Nature* **231**: 184-185.

Whitlock MC. 1992. Temporal fluctuations in demographic parameters and the genetic variance among populations. *Evolution* **46**: 608-615.

Wierdl M, Dominska M et Petes TD. 1997. Microsatellite instability in yeast: dependence on the length of the microsatellite. *Genetics* **146**: 769-779.

Williamson MS, Martinez-Torres D, Hick CA et Devonshire AL. 1996. Identification of mutations in the housefly para-type sodium channel gene associated with knockdown resistance (*kdr*) to pyrethroid insecticides. *Mol. Gen. Genet.* **252**: 51-60.

Wright S. 1931. Evolution in Medelian populations. *Genetics* **16**: 97-159.

Wright S. 1938. Size of populations and breeding structure in relation to evolution. *Sciences* **87**: 430-431.

Wright S. 1943. Isolation by distance. *Genetics* **28**: 114-138.

Wright S. 1951. The genetical structure of populations. *Ann. Eugen.* **15**: 323-354.

Wright S. 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution* **19**: 395-420.

Yan G, Christensen BM et Severson DW. 1997. Comparisons of genetic variability and genome structure among mosquito strains selected for refractoriness to a malaria parasite. *Journal of Heredity* **88**: 187-194.

Zahar AR. 1985. Vector bionomics in the epidemiology and control of malaria. Part I. The WHO African region and the southern WHO eastern mediterranean region. Section III: Vector bionomics, malaria epidemiology and control by geographical areas. (D) East Africa. (E) Eastern outer islands. (F) Southwestern Arabia. Unpublished report WHO/CTD/VBC/85.3-MAP/85.3. World Health Organization, Geneva, Switzerland.

Zheng L, Saunders RDC, Fortini D, della Torre A, Coluzzi M, Glover DM et Kafatos FC. 1991. Low-resolution genome map of the malaria mosquito *Anopheles gambiae. Proc. Natl. Acad. Sci. USA* **88**: 11187-11197.

Zheng L, Collins FH, Kumar V et Kafatos FC. 1993. A detailed genetic map for the X chromosome of the malaria vector, *Anopheles gambiae. Science* **261**: 605-608.

Zheng L, Benedict MQ, Cornel AJ, Collins FH et Kafatos FC. 1996. An integrated genetic map of the african human malaria vector mosquito, *Anopheles gambiae. Genetics* **143**: 941-952.

Zheng L, Cornel AJ, Wang R, Erfle H, Voss H, Ansorge W, Kafatos FC et Collins FH. 1997. Quantitative trait loci for refractoriness of *Anopheles gambiae* to *Plasmodium cynomolgi* B. *Science* **276**: 425-428.

Zouros E. 1979. Mutation rates, populations sizes and amounts of electrophoretic variation of enzyme loci in natural populations. *Genetics* **92**: 623-646.

Annexe 1

Lehmann T, Hawley WA, Kamau L, Fontenille D, Simard F et Collins FH. 1996. Genetic differentiation of *Anopheles gambiae* populations from East and West Africa: a comparison of microsatellite and allozyme loci. *Heredity* 77: 192-208.

Heredity **77** (1996) 192–208 *Received 8 November 1995*

Genetic differentiation of *Anopheles gambiae* populations from East and West Africa: comparison of microsatellite and allozyme loci

TOVI LEHMANN*†, WILLIAM A. HAWLEY†‡, LUNA KAMAU‡, DIDIER FONTENILLE§, FREDERIC SIMARD§ & FRANK H. COLLINS†¶

†Division of Parasitic Diseases, Centers for Disease Control and Prevention, MS F22, 4770 Buford Highway, Chamblee, GA 30341, U.S.A., ‡Kenya Medical Research Institute, Clinical Research Centre, Nairobi, Kenya, §Laboratoire ORSTOM de Zoologie Medicale, Institut Pasteur, BP 220, Dakar, Senegal and ¶Department of Biology, Emory University, Atlanta, GA 30322, U.S.A.

Genetic variation of *Anopheles gambiae* was analysed to assess interpopulation divergence over a 6000 km distance using short tandem repeat (microsatellite) loci and allozyme loci. Differentiation of populations from Kenya and Senegal measured by allele length variation at five microsatellite loci was compared with estimates calculated from published data on six allozyme loci (Miles, 1978). The average Wright's F_{ST} of microsatellite loci (0.016) was lower than that of allozymes (0.036). Slatkin's R_{ST} values for microsatellite loci were generally higher than their F_{ST} values, but the average R_{ST} value was virtually identical (0.036) to the average allozyme F_{ST}. These low estimates of differentiation correspond to an effective migration index (*Nm*) larger than 3, suggesting that gene flow across the continent is only weakly restricted. Polymorphism of microsatellite loci was significantly higher than that of allozymes, probably because the former experience considerably higher mutation rates. That microsatellite loci did not measure greater interpopulation divergence than allozyme loci suggested constraints on microsatellite evolution. Alternatively, extensive mosquito dispersal, aided by human transportation during the last century, better explains the low differentiation and the similarity of estimates derived from both types of genetic markers.

Keywords: allozymes, *Anopheles gambiae,* gene flow, microsatellites, population genetic structure, population genetics.

Introduction

In sub-Saharan Africa *Anopheles gambiae* is the principal vector of human malaria, a disease which continues to inflict immense misery despite substantial efforts to bring it under control. An understanding of the genetic structure of *A. gambiae* populations is critical in evaluating the possibility of genetic manipulation of this species to block malaria transmission (e.g. Collins & Besansky, 1994; Crampton *et al.,* 1994). Moreover, such understanding could also aid control based on currently available technology, such as in the management of insecticide resistance.

*Correspondence.

Microsatellite loci have been described as powerful markers for measuring intraspecies differentiation because of their high polymorphism, codominance, abundance throughout the genome, and relative ease of scoring (Bowcock *et al.*, 1994; Estoup *et al.*, 1995). A microsatellite survey in *A. gambiae* (Lanzaro *et al.*, 1995) demonstrated the above features and concluded that microsatellite loci are superior to allozymes for studies of population structure. Currently several groups are using microsatellite loci to assess gene flow and related phenomena in the *A. gambiae* complex of species. The forces that shape allele composition at such loci are poorly understood, however (Edwards *et al.*, 1992; Di Rienzo *et al.*, 1994; FitzSimmons *et al.*, 1995; Garza *et al.*, 1995), and some evidence suggests that these forces include biased mutation rates (Garza *et*

©1996 The Genetical Society of Great Britain.

al., 1995) and/or selection acting on allele size (Epplen *et al.*, 1993). If such forces strongly influence allele composition at these loci, estimates of differentiation between populations and rates of gene flow will be misleading, because they are derived on the basis of genetic drift, migration, and random mutation as the main forces in operation.

Using five polymorphic microsatellite loci, we measured differentiation between populations of *A. gambiae* from Kenya and Senegal that represent two geographical extremes (6000 km) in the range of this species. Assuming that gene flow is restricted by distance, genetic differentiation between these populations would be expected to be near the maximum possible for the species. Additionally, we compared differentiation based on microsatellite loci, using both Wright's F_{ST} and Slatkin's R_{ST}, to differentiation based on the allozyme data (Miles, 1978) using Wright's F_{ST} derived from populations in Kenya and The Gambia. The rationale for this comparison was to evaluate the possibility that constraints on microsatellite loci such as biased mutation rates or selection on allele size could yield lower estimates of population differentiation than those based on allozyme data.

Materials and methods

Study sites

The Asembo Bay area in western Kenya is located on the northern shores of Lake Victoria. It is a relatively flat, densely populated landscape, traversed by semipermanent streams. During the major rainy season (April to July), many mosquito breeding sites are available. Mosquito populations are much reduced during the dry season, when larval breeding sites are scarce. The village of Barkedji in north Senegal is located in the Sahelian region, and *A. gambiae* can be found only during the June to December rainy season. In both sites, mosquitoes were collected from houses less than 2 km apart to minimize the possibility of sampling members of different demes.

Mosquito collection

In Kenya, mosquitoes were aspirated at dawn from bed nets hung the previous evening over the beds of sleeping volunteers. The nets were hung in a manner to leave a space for the mosquitoes to enter. Thus, samples consisted of blood-fed and blood-seeking females. In Senegal, mosquitoes were aspirated after landing on human volunteers or collected from the floor after spraying the interior of houses with pyrethrum insecticide early in the morning. Collections were carried out in Kenya between 28 June and 6 July 1994 and in Senegal on 5–6 October 1994. Only the savanna cytotype of *A. gambiae* (Coluzzi *et al.*, 1985; Fontenille, unpublished data) and *A. arabiensis* of the *A. gambiae* complex, were present in both study sites. Only *A. gambiae* were included in the analysis, however, after species identification (Scott *et al.*, 1993).

DNA extraction and genotype scoring

DNA from individual specimens (or parts of a specimen) was extracted as described by Collins *et al.* (1987) and resuspended in water or TE buffer (Sambrook *et al.*, 1989). Loci *33C1*, *29C1*, *1D1* and *2A1* were identified in cloned *A. gambiae* genes (Table 1). Locus *33C1* is from the *dopa decarboxylase* (*Ddc*) gene (P. Romans, unpublished data), *29C1* is from the *xanthine dehydrogenase* gene (F. Collins, unpublished data), *1D1* is from the *actin1D* gene (Salazar *et al.*, 1993), and *2A1* is from the *white* gene (Besansky *et al.*, 1995). Locus *AG2H46* was isolated from an *A. gambiae* chromosome division-specific library (Zheng *et al.*, 1991) by probing with a labelled GT-repeat oligonucleotide. Microsatellite alleles were PCR amplified and viewed by autoradiography using incorporation of alpha-^{33}P dATP (Amersham) into the PCR product or using a primer end-labelled with gamma-^{33}P dATP (Dupont). Both techniques provided identical results. Standard PCR in 20 μL reaction volume was run in a Perkin-Elmer 9600 thermal cycler. For incorporation of radiolabelled dATP, a mixture containing 0.2 mM each of dGTP, dCTP, dTTP; 0.05 mM of dATP (Perkin-Elmer) and 0.4 μL of alpha-^{33}P dATP at 1000–3000 Ci/mmol; 5 ng/μL of each primer (approx. 15 pmol); 1 × reaction buffer (Boehringer Mannheim); and 0.035 units of Taq polymerase (Boehringer Mannheim) was used. For PCR reactions with one end-labelled primer, all dNTP concentrations were 0.2 mM, 2 pmol radiolabelled primer with 8 pmol of the same primer, which was not radiolabelled, 10 pmol of the complementary primer and the other components unchanged. An equivalent of 1/100 or less of genomic DNA extracted from a whole mosquito was used. PCR conditions were: denaturation at 94°C for 5 min, followed by 30 cycles consisting of 94°C for 25 s, 55°C for 28 s, and 72°C for 30 s. The last elongation step was at 72°C for 5 min. The PCR product was mixed (3:2) with formamide stop solution (Amersham), denatured at 94°C for 5 min before loading

© The Genetical Society of Great Britain, *Heredity*, 77, 192–208.

Table 1 Microsatellite loci in *Anopheles gambiae*: cytological location, repeat sequence and primer sequences

Locus	Cytol. Location*	Repeat	Primers†
AG2H46	IIR:7A	GT	CGC CCA TAG ACA ACG AAA GG
			TGT ACA GCT GCA GAA CGA GC
			CAA AGA AAG CGC CCA TAG AC
			CGC TGT GTT TTC GTC TTG TA
33C1	IIIR:33C	AGC	TTG CGC AAC AAA AGC CCA CG
			ATG AAA CAC CAC GCT CTC GG
29C1	IIIR:29C	TGA	ATG TTC CAG AGA CGA CCC AT
			TGT TGC CGG TTT GTT GCT GA
1D1	X:1D	CCA	TAA TGG TCC CAA ATC GTT GC
			GTT ATC CAC TGC GCA TCA TG
2A1	X:2A	10/6 bases	GAA TTC GTT TAG AGT CTT TC
			GTA TAC AGG CCT TTG TTT CC

*The cytological position of each locus was determined by polytene chromosome *in situ* hybridization (Kumar & Collins, 1993).
†For locus *AG2H46* the upper pair refers to the original primers and the lower pair to alternative primers (see Materials and methods).

3 μL of the mixture onto a 6 per cent acrylamide, 7 M urea sequencing gel (Life Technologies) in parallel to a 2-lane-ladder standard, which was loaded every 10–20 lanes. The standard, constructed by sequencing an AT-rich region of *A. gambiae* mtDNA with the dideoxy terminators ddA and ddT (Beard *et al.*, 1993), allowed exact determination of allele size. Autoradiographs of gels, developed after overnight exposure, were visually inspected and allele size was determined. Alleles were distinguished from occasional artefacts by intensity and size. These criteria were tested by scoring the alleles at loci *1D1* and *AG2H46* in laboratory-reared progeny of parent mosquitoes with known genotypes. All scored genotypes were in complete concordance with the genotypes expected from the test crosses.

At locus *AG2H46*, no PCR product was visible in repeated reactions for approx. 5 per cent of the specimens that were scored successfully for other loci. Using alternative primers, which flank the sites of the original primers (Table 1), PCR products were obtained. Subsequent sequencing revealed a mutation in one of the original primer annealing regions. For this locus 2/3 of the homozygotes scored with the original primers were found to be heterozygotes when PCR amplification was carried out with the alternative primers. No 'hidden' heterozygotes were found in the other microsatellite loci using alternative primers.

Allozyme data

To identify allozyme alleles that could be used to identify the different cryptic species in the *A. gambiae* complex, Miles (1978) examined variation at 18 loci from *A. gambiae* complex populations from different parts of Africa. Allele frequency data based on sample sizes larger than six mosquitoes were summarized only for the loci *alpha-naphthyl acetate esterases* (*EST-1, EST-2, EST-3*), *octanol dehydrogenase* (*ODH*) and *phosphoglucomutases* (*PGM-1, PGM-2*). These data for the populations from Kenya (Chulaimbo) and Gambia (Mandinari, having the largest sample size) were used. Additional analyses including samples from East, West, and Central Africa were carried out to evaluate the consistency of these results (see Discussion). We have assumed that the population structure of *A. gambiae* across the continent has not changed significantly between the time of Miles's study and the present.

Data analysis

Goodness of fit tests of genotype distributions with Hardy–Weinberg expectations in each population were performed for each microsatellite locus, after pooling rare alleles to achieve expected values per cell higher than two. Because genotype data were not available for the published allozyme data, analy-

© The Genetical Society of Great Britain, *Heredity*, 77, 192–208.

sis was carried out on allele frequencies, assuming random mating in each population. F-statistics were calculated based on Wright (1978) for microsatellite and allozyme data using BIOSYS (Swofford & Selander, 1989). This method adjusts for sampling variation and does not require genotype frequencies. R_{ST} (Slatkin, 1995), a statistic related to F_{ST} developed specifically for microsatellite loci, accounts for sampling variation and especially for the different mutation process thought to occur in microsatellite loci (high mutation rate and partial dependence of the mutant allele size on the original allele size). Slatkin's method relies on the assumptions of no constraints on allele size and that the mutation process is similar across all allele sizes. Calculation of the repeat number for each allele was based on a known sequence, where the length of the regions flanking the repeated motif was subtracted from the total allele length, and the result was divided by the length of the repeat unit. The sizes of a few alleles at locus *33C1* differed by an amount that was not equal to the size of a repeat (Fig. 1) and R_{ST} calculation proceeded based on the following assumptions, namely that these alleles were created by an insertion/deletion of one nucleotide that occurred outside the repeat region, allowing rounding of the repeat number to the nearest integer. These noncanonical alleles had low frequency, comprising together eight per cent in both populations, thus even if these assumptions were wrong, the effect would be small. Locus *2A1* had a complex series of allele sizes (Fig. 1) resulting from several repeat motifs of different sizes. Calculation of R_{ST} for locus *2A1* was based on the assumption of a closer relationship between alleles of similar size, and thus rounding was performed based on the smallest motif size (6 bases). However, as the effect of this procedure was not known, average R_{ST} values for the set of microsatellite loci were calculated with and without this locus (Table 4).

Significance of the F_{ST} was evaluated based on a chi-square test of the contingency table of allele frequencies by populations, after pooling rare alleles such that expected cell counts would be higher than 2, and no more than 20 per cent of the cell counts would be lower than 5. Significance of the R_{ST} (and R_{IS}) was evaluated based on an F-test in a nested ANOVA on the repeat number in a model including the individual and the population as factors (Slatkin, 1995). The average R_{ST} was calculated from the averages of the within-population and total variance components across loci. Estimates of Nm were derived from F_{ST} for two populations according to Slatkin (1995): $Nm = 1/4\ (1/F_{ST} - 1)$. Nm was derived from R_{ST} by substituting R_{ST} for F_{ST}. Calculations not available in BIOSYS were carried out by programs written in the SAS language (SAS Institute, 1990).

Results

Microsatellite loci were highly polymorphic in both populations, with an average of 7.8 alleles (ranges of 2–13) per locus per population. The average unbiased heterozygosity was 0.63 (range of

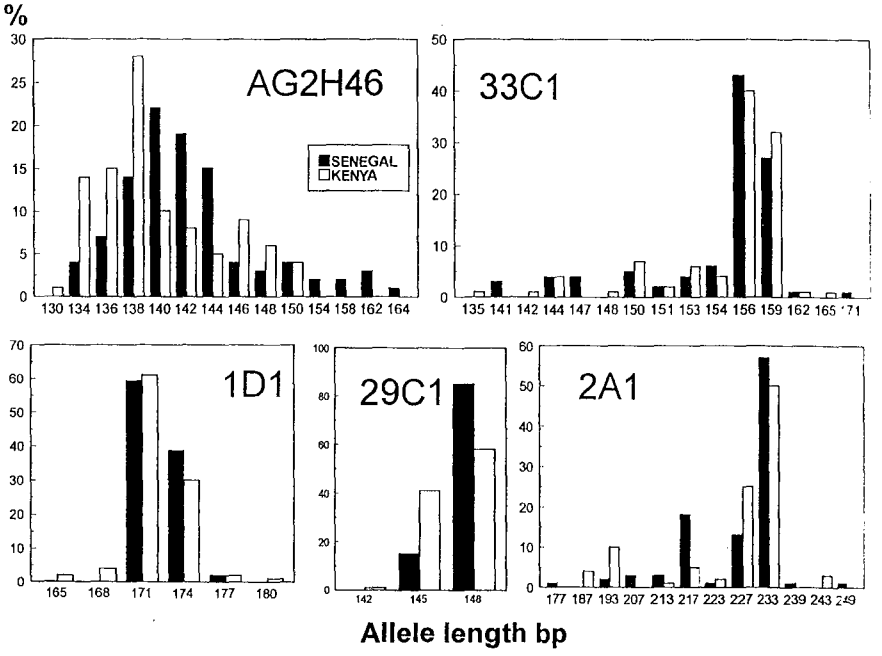


Fig. 1 Allele composition at microsatellite loci in East and West African *Anopheles gambiae*.

© The Genetical Society of Great Britain, *Heredity*, 77, 192–208.

0.26–0.87, Table 2 and Fig. 1). Allozyme loci were moderately polymorphic with an average of 3.2 alleles (between 1 and 6) per locus, and the average unbiased heterozygosity was 0.38 (range of 0.0–0.76) (Table 2 and Fig. 2). Although the average unbiased heterozygosities of both microsatellite and allozyme loci were slightly higher in East Africa, the differences were not significant (signed rank paired test by locus = 15, d.f. = 10, $P > 0.21$ and Table 3), suggesting no large difference in effective population sizes

Table 2 Polymorphism of microsatellite and allozyme loci in populations of *Anopheles gambiae* from Kenya and Senegambia

	Kenya*					Senegambia*				
Locus	*N*	No. of alleles	Common allele %	H_E†	H_O‡	*N*	No. of alleles	Common allele %	H_E†	H_O‡
Microsatellites										
AG2H46	50	10	28	0.856	0.760	50	13	22	0.870	0.860
33C1	50	12	40	0.732	0.580	50	11	43	0.737	0.720
29C1	50	3	58	0.500	0.480	50	2	85	0.258	0.260
1D1	50	6	61	0.541	0.500	49	3	59	0.504	0.449
2A1	50	8	50	0.679	0.640	50	10	57	0.629	0.640
Average	50	7.8	47.4	0.662	0.592	49.8	7.8	53.2	0.600	0.586
Allozymes										
EST-1	24	3	75	0.415	—	22	3	48	0.591	—
EST-2	24	3	50	0.550	—	22	4	68	0.481	—
EST-3	24	6	42	0.755	—	22	6	52	0.671	—
ODH	22	1	100	0.000	—	22	2	98	0.045	—
PGM1	22	3	75	0.409	—	22	3	77	0.385	—
PGM2	22	2	89	0.206	—	22	2	96	0.089	—
Average	23	3.0	71.1	0.389	—	22	3.3	87.8	0.377	—

*For microsatellites, the Kenyan population was obtained from Asembo Bay and the Senegalian population from Barkedji; for allozymes (Miles, 1978), the Kenyan population was obtained from Chulaimbo and the Gambian population from Mandinari.
†Unbiased heterozygosity (Nei, 1978).
‡Observed heterozygosity based on direct count of heterozygotes.

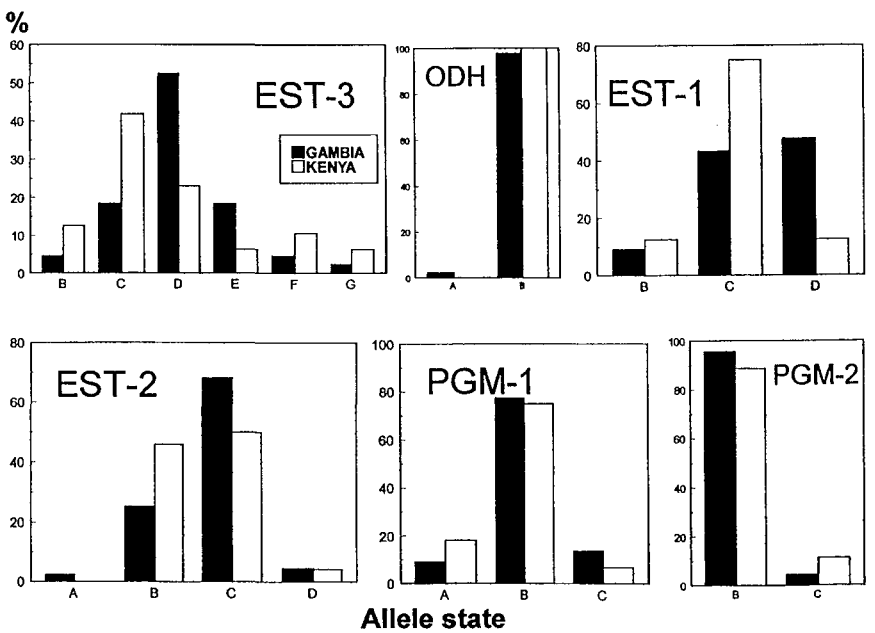


Fig. 2 Allele composition at allozyme loci in East and West African *Anopheles gambiae*.

© The Genetical Society of Great Britain, *Heredity*, 77, 192–208.

between these populations from East and West Africa. The heterozygosity of microsatellite loci, however, was significantly higher than that of allozymes (Table 3). This difference was expected, based on known higher mutation rates reported for microsatellite loci (e.g. Dallas, 1992; Weber & Wong, 1993).

Deviations from Hardy–Weinberg expectations were not significant ($P>0.05$) for all microsatellite loci in both populations (data not shown). Similarly, the within-individual variance component of the number of repeats tested by nested ANOVA (Slatkin, 1995) was not significant for all loci (data not shown). No evidence was obtained for nonrandom mating in the populations.

The allele compositions of both microsatellite and allozyme loci were very similar between West and East Africa (Figs 1 and 2). Every allele with a frequency higher than five per cent in a given population was also found in the other population, and the most common allele was the same in both populations for four of the five microsatellite loci (Fig. 1) and four of the six allozyme loci (Fig. 2). Differentiations of the populations, as measured by F_{ST} and R_{ST} were accordingly low (0.0–0.09, Table 4) and all *Nm* values were larger than 1. The R_{ST} values of microsatellite loci were higher than their corresponding F_{ST} (except for locus *2A1*, see Materials and methods), as expected of loci with high mutation rates, and with high likelihood that the same allele could be produced by independent mutations (Slatkin, 1995). Nevertheless, the R_{ST} (and F_{ST}) values of *AG2H46* and *29C1* were significant, whereas only the F_{ST} but not the R_{ST} of locus *2A1* was significant (Table 4). Likewise, the F_{ST} values of two of the six allozyme loci were significant.

Table 3 The differences in locus unbiased heterozygosity (H_E) between genetic markers and locales (as indicators of N_e): ANOVA results

Source	d.f.	Mean square	*P*
Model*	2	0.170	0.060
Error	19	0.052	—
Genetic marker	1	0.334	0.020
Population	1	0.007	0.724

Model $R^2 = 26\%$.
See Table 2 for group means.
*Interaction term was removed after it was found to be not significant when included ($P>0.8$).

Discussion

A remarkable similarity in allele profile of *A. gambiae* populations 6000 km apart was evident at microsatellite and allozyme loci. The significance of the divergence indices in two (or three if F_{ST} is used) of the five microsatellite loci and in two of the six allozyme loci implies that a degree of separation between the gene pools does exist. However, low estimates of interpopulation differentiation were measured by Wright's F_{ST} and Slatkin's R_{ST}, corresponding to high estimates of the average migration index ($Nm>3$) across this enormous distance. Genetic differentiation at 'neutral' loci because of genetic drift is expected if $Nm<1$ but not if $Nm>1$ (Slatkin, 1987). The consistency of the differentiation indices across loci, i.e. all 11 estimates fall in a narrow range of 0–0.087, implies that heterogeneity among loci in each marker group was not large. Accordingly, considering the three loci with the

Table 4 Differences between *Anopheles gambiae* populations from Kenya and Senegambia based on microsatellite and allozyme loci

Microsatellites			Allozymes	
Locus (N_k,N_s)†	Wright's F_{ST} (*Nm*)	Slatkin's R_{ST} (*Nm*)	Locus (N_g,N_k)	Wright's F_{ST} (*Nm*)
AG2H46 (50,50)	0.013** (9.5)	0.0624*** (1.9)	*EST-1* (22,24)	0.087*** (1.3)
33C1 (50,50)	0.0 NS (≫1)	0.0000 NS (≫1)	*EST-2* (22,24)	0.018 NS (6.8)
29C1 (50,50)	0.077*** (1.5)	0.08620*** (1.3)	*EST-3* (22,24)	0.037** (3.3)
1D1 (50,49)	0.0 NS (≫1)	0.0097 NS (12.7)	*ODH* (22,22)	0.0 NS (≫1)
2A1 (50,50)	0.01** (12.4)	0.0029 NS (42.5)	*PGM-1* (22,22)	0.0 NS (≫1)
			PGM-2 (22,22)	0.0 NS (≫1)
Average 5 loci	0.016 (7.7)	0.0358 (3.4)	Average	0.036 (3.3)
4 loci		0.0469 (2.5)		

†Sample size (mosquitoes): N_k, Kenya; N_s, Senegal; and N_g, Gambia.
NS $P>0.05$, **$P<0.01$, ***$P<0.001$.

© The Genetical Society of Great Britain, *Heredity*, 77, 192–208.

highest differentiation estimates from each group increased the average R_{ST} of microsatellite loci to 0.061 (corresponding to an $Nm = 1.9$) and the average divergence of allozymes to only 0.047 (corresponding to an $Nm = 2.5$). Thus, even these estimates imply extensive gene flow across the continent. Furthermore, regional analyses of Miles's (1978) six-loci data including either 10 or six localities divided into West (Gambia), East (Kenya and Tanzania), and Central Africa (Nigeria and Cameroon) with samples larger than 12 or 20 mosquitoes, respectively, provided slightly lower estimates of between-region F_{ST} (0.015 and 0.021, data not shown).

For comparison, microsatellite loci in the honey bee, *Apis mellifera*, measured far greater divergence between populations within lineages, with an average F_{ST} value of 0.34 (Estoup *et al.*, 1995), whereas allozyme variation hardly exists. Pairwise average F_{ST} values for populations within the least differentiated honey bee subspecies, separated by less than 2500 and 2000 km (two from France and one from Sweden), were 0.083 and 0.042, respectively, vs. 0.016 in the populations of the present study (Table 4). The mosquito *Aedes aegypti* showed a complex structure with substantial differentiation on a worldwide scale (Powell *et al.*, 1980), although extensive gene flow ($Nm > 9$) was observed across approx. 150 km in Puerto Rico (Apostol *et al.*, 1995). Even greater geographical differentiation was observed in *Aedes albopictus* in Malaysia and Borneo (Black *et al.*, 1988).

Because mark-release-recapture experiments have suggested that active dispersal in *A. gambiae* is restricted to a few kilometers (Gillies & De Meillon, 1968), it is unlikely that active dispersal accounts for the low differentiation. An alternative explanation is that *A. gambiae*, a highly domestic mosquito, has expanded to its current range in recent evolutionary time in response to the major expansion in human population density that followed the development of agriculture in West Africa about 2000–5000 years ago (Coluzzi *et al.*, 1985). It is conceivable that with population expansion occurring within the past few thousand years and with large effective population size, the divergence of populations across the African continent could remain low today.

The similarity of estimates of divergence measured by marker systems so different in their time-scale resolution is surprising. Microsatellite loci evolve considerably faster than allozyme loci, and have mutation rates of 10^{-3}–10^{-4} per locus per gamete per generation (Dallas, 1992; Weber & Wong, 1993) vs. 10^{-6}–10^{-9} for allozyme loci (Ayala, 1976). Thus they are expected to detect differentiation between populations that occurred over shorter periods. The lower polymorphism of the allozymes as compared with the microsatellite loci (Tables 2 and 3) is consistent with the mutation rates of the microsatellite loci being higher than those of the allozyme loci. In spite of mutation rates at least one scale order higher than that of allozymes, the R_{ST} estimate for the microsatellite loci was not higher than the F_{ST} for the allozymes. The failure of microsatellite loci to detect greater divergence than allozyme loci suggests the existence of constraints on their evolution, such as biased mutation rates (Garza *et al.*, 1995) and/or selection for certain allele sizes (Epplen *et al.*, 1993).

Alternatively, dispersal via both active and passive transportation of *A. gambiae* during the past century might be much higher than generally appreciated. Surprisingly high numbers of *A. gambiae* have been captured in trains and boats in Africa (Holstein, 1954). The rapid geographical spread of this mosquito after its introduction into Brazil in 1930 also indicates considerably higher rates of dispersal than commonly appreciated. The spread of *A. gambiae* in Brazil from its point of introduction in 1930 to the extreme extent of its range in 1938–40 was at an average minimal rate of 50 km y^{-1}, based on Soper & Wilson (1943). This is probably an underestimate because (1) it is based on the detection of new, fully established populations rather than small numbers of migrants, (2) the estimate assumes a constant rate of movement over the entire 10-year period, and (3) populations could successfully spread only in the north-westerly direction. If migration at such a rate is also occurring in Africa, it may be sufficient to explain the low levels of genetic divergence between populations in East and West Africa. Furthermore, such high rates of migration would also explain our finding of similar estimates of divergence obtained by the two different marker systems, without the need to invoke constraints on their evolution.

The low levels of genetic differentiation between populations from East and West Africa would seem to be incompatible with the very high levels of chromosomal-inversion-based genetic population structure recorded in *A. gambiae* populations, particularly from West Africa (Coluzzi *et al.*, 1985; Lanzaro *et al.*, 1995). Several distinct chromosomal forms and a series of climato-geographically stable clines in the frequencies of certain chromosomal inversions are characteristic of many populations in West Africa, whereas they have not been found in East Africa despite extensive karyotype data from

© The Genetical Society of Great Britain, *Heredity*, 77, 192–208.

Kenya and Tanzania. These reports of highly structured West African populations might suggest restricted gene flow in sympatric populations across dimensions other than geographical distance — the only dimension explored in this study. It is important to note that the Savanna chromosomal form is the only form known to occur in Kenya (Coluzzi *et al.*, 1985) and Barkedji (Fontenille, unpublished data). Additionally, the strong selection operating on many of these chromosomal inversions that is evident by regular seasonal or spatial-climatological change in karyotype frequencies may render them unsuitable for gene flow evaluations.

At this point there is insufficient information to explain the low differentiation of *A. gambiae* populations across Africa by ruling out either the hypothesis of recent expansion, which is coupled with constraints on microsatellite evolution, or the hypothesis of migration augmented by transportation. Information based on additional genetic loci, different genetic markers, and populations on micro- and macrogeographical scales is necessary before final conclusions can be reached.

Acknowledgements

We are grateful to William C. Black IV, Nora J. Besansky, and Michel Tibayrenc for valuable suggestions throughout this study and useful comments on earlier versions of this manuscript. We thank Brian Holloway and staff of the NCID Biotechnology Core Facility for synthesizing the oligonucleotide primers and Diane M. Hamm for valuable advice on the laboratory work. T.L. was supported by an American Society for Microbiology Fellowship. Additional financial support was provided by the John D. and Catherine T. MacArthur Foundation and by the UNDP/WORLD BANK/WHO Special Programme for Research and Training in Tropical Diseases.

References

APOSTOL, B. L., BLACK, W. C., IV., REITER, P. AND MILLER, B. R. 1995. Population genetics with RAPD-PCR markers: the breeding structure of *Aedes aegypti* in Puerto Rico. *J. Med. Entomol.*, (in press).

AYALA, F. J. (ed.) 1976. *Molecular Evolution.* Sinauer Associates, Sunderland, MA.

BEARD, C. B., HAMM, D. M. AND COLLINS, F. H. 1993. The mitochondrial genome of the mosquito *Anopheles gambiae*: DNA sequence, genome organization, and comparisons with mitochondrial sequences of other insects. *Insect Mol. Biol.*, **2**, 103–124.

BESANSKY, N. J., BEDELL, J. A., BENEDICT, M. Q., MUKABAYIRE, O., HILFIKER, D. AND COLLINS, F. H. 1995. Cloning and characterization of the *white* gene from *Anopheles gambiae*. *Insect Mol. Biol.*, **4**, 217–231.

BLACK, W. C., IV, HAWLEY, W. A., RAI, K. S. AND CRAIG, G. B., Jr. 1988. Breeding structure of a colonizing species: *Aedes albopictus* (Skuse) in peninsular Malaysia and Borneo. *Heredity*, **61**, 439–446.

BOWCOCK, A. M., RUIZ-LINARES, A., TOMFOHRDE, J., MINCH, E., KIDD, J. R. AND CAVALLI-SFORZA, L. L. 1994. High resolution of human evolutionary trees with polymorphic microsatellites. *Nature*, **368**, 455–457.

COLLINS, F. H. AND BESANSKY, N. J. 1994. Vector biology and control of malaria in Africa. *Science*, **264**, 1874–1875.

COLLINS, F. H., MENDEZ, M. A., RASMUSSEN, M. O., MEHAFFEY, P. C., BEASANSKY, N. J. AND FINNERTY, V. 1987. A ribosomal RNA gene probe differentiates member species of *Anopheles gambiae* complex. *Am. J. Trop. Med. Hyg.*, **37**, 37–41.

COLUZZI, M., PETRARCA, V. AND DI DECO, M. 1985. Chromosomal inversion intergradation and incipient speciation in *Anopheles gambiae*. *Boll. Zool.*, **52**, 45–63.

CRAMPTON, J. M., WARREN, A., LYCETT, G. J., HUGHES, M. A., COMLEY, I. P. AND EGGLESTON, P. 1994. Genetic manipulation of insect vectors as a strategy for the control of vector-borne disease. *Ann. Trop. Med. Parasitol.*, **88**, 3–12.

DALLAS, J. F. 1992. Estimation of microsatellite mutation rates in recombinant inbred strains of mouse. *Mamm. Genome*, **3**, 452–456.

DI RIENZO, A., PETERSON, A. C., GARZA, J. C., VALDES, A. M., SLATKIN M. AND FREIMER, N. B. 1994. Mutational processes of simple-sequence repeat loci in human populations. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 3166–3170.

EDWARDS, A., HAMMOND, H. A., JIN, L., CASKEY, C. T. AND CHAKRABORTY, R. 1992. Genetic variation of five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics*, **12**, 241–253.

EPPLEN, C., MELMER, G., SIEDLACZCK, I., SCHWAIGER, F.-W., MÄUELER, W. AND EPPLEN, J. T. 1993. On the essence of "meaningless" simple repetitive DNA in eukaryote genomes. In: Pena, S. D. J., Chakraborty, R., Epplen, J. T., and Jeffreys, A. J. (eds) *DNA Fingerprinting: State of the Science,* pp. 29–45. Birkhauser Verlag, Basel.

ESTOUP, A., GARNERY, L. SOLIGNAC, M. AND CORNUET, J.-M. 1995. Microsatellite variation in honey bee (*Apis mellifera* L.) populations: hierarchical genetic structure and test of the infinite allele and stepwise mutation models. *Genetics*, **140**, 679–695.

FITZSIMMONS, N. N., MORITZ, C. AND MOORE, S. S. 1995. Conservation and dynamics of microsatellite loci over 300 million years of marine turtle evolution. *Mol. Biol. Evol.*, **12**, 432–440.

GARZA, J. C., SLATKIN, M. AND FREIMER, N. B. 1995. Microsatellite allele frequencies in humans and chimpanzees, with implications for constraints on allele size. *Mol. Biol. Evol.*, **12**, 594–603.

GILLIES, M. T. AND DE MEILLON, B. 1968. *The Anophelinae*

© The Genetical Society of Great Britain, *Heredity*, **77**, 192–208.

of Africa South of the Sahara, 2nd edn. Publications of the South African Institute for Medical Research no. 54. Johannesburg, South Africa.

HOLSTEIN, M. H. 1954. *Biology of* Anopheles gambiae. *Research in French West Africa*. WHO, Geneva.

KUMAR, V. K. AND COLLINS, F. H. 1993. A technique for nucleic acid *in situ* hybridization to polytene chromosomes of mosquitoes in the *Anopheles gambiae* complex. *Insect Mol. Biol.*, **3**, 41–47.

LANZARO, G. C., TOURE, Y. T., ZHENG, L., KAFATOS, F. C. AND VERNICK, K. D. 1995. Microsatellite DNA and isozyme variability in a West African population of *Anopheles gambiae*. *Insect Mol. Biol.*, **4**, 105–112.

MILES, S. J. 1978. Enzyme variation in the *Anopheles gambiae* Giles group of species (Diptera: Culicidae). *Bull. ent. Res.*, **68**, 85–96.

NEI, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**, 583–590.

POWELL, J. R., TABACHNICK, W. J. AND ARNOLD, J. 1980. Genetics and the origin of a vector population: *Aedes aegypti*, a case study. *Science*, **208**, 1385–1387.

SALAZAR, C. E., HAMM, D. M., WESSON, D. M., KUMAR, V. K., BEARD, C. B. AND COLLINS, F. H. 1993. A cytoskeletal actin gene in the malaria vector *Anopheles gambiae*. *Insect Mol. Biol.*, **3**, 1–14.

SAMBROOK, J., FRITSCH, E. F. AND MANIATIS, T. 1989. *Molecular Cloning, a Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, NY.

SAS INSTITUTE INC. 1990. *SAS language: references.* Version 6, 1st edn. Cary, NC.

SCOTT, J. A., BROGDON, W. G. AND COLLINS, F. H. 1993. Identification of single specimens of the *Anopheles gambiae* complex by the polymerase chain reaction. *Am. J. Trop. Med. Hyg.*, **49**, 520–529.

SLATKIN, M. 1987. Gene flow and the geographical structure of natural populations. *Science*, **236**, 787–792.

SLATKIN, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 457–462.

SOPER, F. L. AND WILSON, D. B. 1943. Anopheles gambiae *in Brazil 1930–1940*. Rockefeller Foundation, New York.

SWOFFORD, D. L. AND SELANDER, R. B. 1989. BIOSYS-1. *A Computer Program for the Analysis of Allelic Variation in Population Genetics and Biochemical Systematics.* Release 1.7. University of Illinois, Urbana, IL.

WEBER, J. L. AND WONG, C. 1993. Mutation of human short tandem repeats. *Hum. Mol. Genet.*, **2**, 1123–1128.

WRIGHT, S. 1978. *Evolution and the Genetics of Populations.* vol. 4, *Variability Within and Among Natural Populations.* University of Chicago Press, Chicago.

ZHENG, L., SOUNDERS, R. D. C., FORTINI, D., TORRE, A. D., COLUZZI, M., GLOBER, D. M. AND KAFATOS, F. C. 1991. Low-resolution genome map of the malaria mosquito *Anopheles gambiae*. *Proc. Natl. Acad. Sci. U.S.A.*, **88**, 11187–11191.

© The Genetical Society of Great Britain. *Heredity*, **77**, 192–208.

Annexe 2

Simard F, Fontenille D, Lehmann T, Girod R, Brutus L, Gopaul R, Dournon C et Collins FH. 1999. High amounts of genetic differentiation between populations of the malaria vector *Anopheles arabiensis* from West Africa and Eastern Outer Islands. *Am. J. Trop. Med. Hyg.* **60(6)**: 1000-1009.

High amounts of genetic differentiation between populations of the malaria vector *Anopheles arabiensis* from West Africa and Eastern Outer Islands

Frederic Simard[1], Didier Fontenille[1], Tovi Lehmann[2], Romain Girod[3], Laurent Brutus[4], Raheem Gopaul[5], Christian Dournon[6] and Frank H. Collins[7]

[1] Laboratoire ORSTOM de Zoologie Médicale, Institut Pasteur, BP 220, Dakar, Senegal.

[2] Division of Parasitic Diseases, Centers for Diseases Control and Prevention, Mailstop F22, 4770 Buford Highway, Chamblee, GA, 30341, USA.

[3] DDASS (Direction Départementale des Affaires Sanitaires et Sociales), Service de lutte Antipaludique, 60, rue du Général de Gaulle, 97400 SAINT-DENIS de La Réunion, France.

[4] Programme "Recherche Appliquées à Madagascar sur la Santé et l'Environnement", ORSTOM, BP 434, Antananarivo 101 Madagascar.

[5] Medical Entomology Division, Candos, Mauritius.

[6] EA 2401 Génétique et Interactions Cellulaires en Reproduction, Laboratoire de Biologie Experimentale-Immunologie, Université Henri Poincaré, Domaine Scientifique Victor Grignard, BP 239, 54506 Vandoeuvre-les-Nancy Cedex, France.

[7] Department of Biological Sciences, Galvin Life Sciences, University of Notre Dame, Notre Dame, IN 46556, USA.

Corresponding author:

Frederic SIMARD, Laboratoire ORSTOM de Zoologie Médicale, Institut Pasteur, BP 220, Dakar, Senegal.

Tel: 221-823-48-74, Fax: 221-839-92-10, E-mail: frederic.simard@orstom.sn

Keywords: *Anopheles arabiensis*, islands, malaria, microsatellites, population genetic structure, F-statistics, R-statistics.

Running head: Genetic differentiation in *Anopheles arabiensis*.

ABSTRACT

Polymorphism at nine microsatellite loci was examined to assess the level of genetic differentiation between four *Anopheles arabiensis* populations from Senegal, high plateaux of Madagascar, Reunion and Mauritius islands. Eight out of nine loci showed great polymorphism (2 to 16 alleles/locus) and significant genetic differentiation was revealed between all four populations by F- and R-statistics, with Fst estimates ranging from 0.080 to 0.215 and equivalent Rst values between 0.022 and 0.300. These high amounts of genetic differentiation are discussed in relation with geographical distance including large bodies of water, and history of mosquito settlement and insecticide use on the islands. The results suggest that historical events of drift rather than mutation are probably the forces generating genetic divergence between these populations, homogenization of the gene pool by migration being drastically restricted across the ocean.

INTRODUCTION

Members of the *Anopheles gambiae* complex are the principal human malaria vectors occurring in sub-Saharan Africa and surrounding islands. Among the seven sibling species belonging to the complex, *An. gambiae* Giles and *An. arabiensis* Patton are the most efficient malaria vectors and widely coexist throughout the continent. The latter species, however, is more adapted to dryer environments [1] and remains active significantly longer than *An. gambiae* after the end of the sahelian rainy season.[2] While both species are highly anthropophilic, *An. arabiensis* shows a greater tendency to feed and rest outdoors, particularly in East Africa, and is accordingly less susceptible than *An. gambiae* to the effect of house-spraying,[3] the most commonly used method to control adult mosquito populations.

Alternative strategies for efficient malaria control are desperately needed and current efforts tend to focus on a long-term goal of using transgenic mosquitoes with reduced malaria competence.[4, 5] Provided reliable techniques for arthropods transformation and efficient, stable and safe trasgenes are available,[6] a comprehensive knowledge of the target mosquito population's genetic structure appears absolutely necessary to properly evaluate the outcome of field releases of transgenic vectors. The area of insecticide management will also benefit from the study of gene flow which allows prediction of the spread of resistance genes. Conflicting results have been obtained for *An. gambiae* populations in West and East Africa using different genetic markers. [7-12] These include chromosomal inversions, allozymes, Random Amplified Polymorphic DNA (RAPD), mitochondrial DNA sequences and microsatellite loci analysis. Some of these markers were found to correlate with environmental parameters. For example, spatial and temporal variability of chromosomal inversion frequencies suggest important constraints acting on these markers that could lead to erroneous observations in term of gene flow. [7]

Microsatellites are relatively short tracts of tandemly repeated DNA sequences of 2 to 6 bp. [13] These loci have been described in recent studies as being powerful markers for population genetics because of their abundance throughout Eukaryote genomes, high polymorphism, codominance and relative ease of scoring. [14, 15] While they are considered selectively neutral, evidence for constraints acting on some of these loci include biased mutation rate [16] or selection on allele size and distribution. [17-19] However, the forces that shape allele composition at microsatellite loci remain poorly understood and the intensity of constraints they experience may differ from one locus to another. Nevertheless, because of high mutation rates estimated to range between 10^{-2} to 10^{-5} mutation per generation [20, 21] due mostly to replication slippage (e.g. Levinson and Gutman [22] for a review), those fast evolving loci revealed appropriate for fine scale population genetics studies and should provide most accurate descriptions of the actual levels of gene flow accessible by indirect studies (see Bossart *et al.* [23] and references therein).

While *An. gambiae s.s.* has been extensively studied, very little work has been conducted on *An. arabiensis* population genetics despite its importance as a malaria vector. Furthermore, this species should be a more suitable candidate than its sibling *An. gambiae* to test for efficiency of transgenic mosquitoes control strategies applications in the fields to reduce malaria burden. Indeed, malaria control by antivectorial means should prove more efficient in low or unstable malaria areas where transmission could be reduced in a sufficient amount to allow subsequent epidemiological impact. [24, 25] Such areas, located on the edges of Sub-Saharan Africa (dry savannas on the south border of the Sahara desert, northern South-Africa or highlands) are generally colonized by *An. arabiensis* only, environmental conditions being highly restrictive to establishment of *An. gambiae* populations. In the present study, genetic variation at nine microsatellite loci was investigated in four geographically well-isolated *An. arabiensis* populations from Senegal, Madagascar, Reunion and Mauritius. The purpose of this work was to assess the following points. (1) Are microsatellite loci isolated from *An. gambiae* DNA a suitable tool to study the population genetics of its sibling species *An. arabiensis*? (2) What is the level of genetic variation in *An. arabiensis* natural populations? (3) Is this variability a global trait over all samples or does it depend on specific characteristics of the population under study? (4) What could be the factors affecting the genetic variability distribution among populations and what is their influence in terms of genetic structuring and gene flow? Results were analysed in relation with geographical distance including barriers to migration (ocean) and historical events of colonization and extinction on the islands (initial founder effects and bottlenecks due to insecticide use).

MATERIALS AND METHODS

Study sites

Senegal:

The village of Barkedji (15°17'N, 14°53'W) is located in the Sahelian region. The rainy season is short, extending from July to October, with little annual variation in the amount of rainfall (300 to 400 mm). An extensive study of malaria vectors ecology and population dynamics in this location showed that the vectors are *An. arabiensis* and *An. gambiae*. [2] Because no mosquito breeding sites are available during the dry season, malaria transmission is highly seasonal.

Madagascar:

Fenoarivo (19°47'S, 46°34'E) is a village on the high plateaux of Madagascar, the largest island in the Indian Ocean (587000 km^2) located 350 km away from the east coast of the african continent. At an elevation of 1235m, Fenoarivo is surrounded by flooded rice fields. The hot rainy season lasts from November to April with an annual rainfall of around 1,400 mm. A malaria eradication program based on DDT spraying and chemoprophylaxis was held from 1949 until elimination of the disease in the early 1960s, some local irregular DDT spraying being kept until 1975 (for a review, see Mouchet and others [26]). However, malaria was still occurring in certain particular areas in the island and, due to the lack of control measures, the situation worsened until the occurrence of an epidemic

malaria in 1986-1988. Yearly DDT spraying was then reintroduced from 1993 to the present. In Madagascar, malaria transmission is carried out by the *An. gambiae* complex and by *An. funestus*, [27] but *An. arabiensis* and *An. funestus* are the only two malaria vectors present on the high plateaux. [28]

Mascarene Islands:

Reunion (2500 km^2) and Mauritius (2000 km^2), two islands 240 km apart, belong to the Mascarene archipelago. They are located in the Indian Ocean at 20 to 25°S latitude and 55°E longitude, respectively 800 and 1000 km away from the east coast of Madagascar. Mosquitoes were collected in a single location on the northwestern part of Reunion called « Etang de Saint Paul » while mosquitoes from Mauritius were collected from all over the island (17 locations) and pooled. The annual amount of rainfall ranges from 5000 mm in upwind locations to 500 mm in downwind ones. Cyclones often occur during the December to April rainy season. Following strong malaria vector control measures and intensive quinine prophylaxis campaigns in the early fifties, malaria was declared eradicated from Mauritius in 1973 and from Reunion in 1979 by WHO. [29] Presently, these islands are in an « anophelism without malaria » state with *An. arabiensis* as the only potential malaria vector. [30, 31] In order to prevent reemergence of the disease, malaria situation rests on a careful epidemiological and entomological surveillance. *Anopheles merus* Dönitz, another member of the *An. gambiae* complex is also present in Mauritius, where salt-water breeding sites are available.

Mosquito collection

Female specimens of the *An. gambiae s.l.* complex from the four study sites were collected either by landing on human volunteers or after indoor pyrethrum spraying early in the morning. Mosquito collections were carried out in August of 1994 in Senegal, in February of 1995 in Reunion, in November-December of 1996 in Madagascar and in March-April of 1997 in Mauritius.

All the mosquitoes were stored individually in 1.5 ml tubes with desiccant and kept at -20°C until further analyzed.

DNA extraction and Microsatellite PCR processing

DNA was extracted from a leg or wing of each individual mosquito as described by Collins and others [32] and resuspended in 100 µl sterile distilled water. Only *An. arabiensis* were included in the analysis after species identification using the PCR technique described by Scott and others. [33] A set of about 50 specimens per location was investigated, excepted for the Mauritius population where only 44 specimens were available.

Microsatellites loci were selected from published *An. gambiae* sequence data [34, 35] and nine of them were chosen for further analysis based on high polymorphism exhibited in *An. gambiae*, allele sizes ranging from 100 to 200 bp (to optimize detection using 10% acrylamide gels), and successful preliminary amplification of *An. arabiensis* DNA (Table 1).

Standard PCR was run in a Perkin-Elmer 9600 thermal cycler in 25 µl reaction volume containing 2.5 µl 10x PCR buffer (Pharmacia Biotech) with $MgCl_2$ final concentration of 1.5

mM, 0.2 mM each dNTP, 100 ng each primer (approx. 15 pmol), 1 Taq polymerase unit (Pharmacia Biotech) and 5 µl DNA extraction. PCR conditions were: denaturation at 94°C for 2 min, followed by 30 cycles consisting of 94°C for 25 s, standard annealing temperature for all primers at 55°C for 28 s, and 72°C for 30 s. The last elongation step was at 72°C for 10 min.

PCR products were then mixed with 6x loading buffer containing glycerol, bromophenol and xylene cyanol blue. A 10 µl aliquot of this mixture was loaded onto a 10 per cent non denaturing polyacrylamide gel. The allelic bands were visualized after rapid silver staining. [36] A few samples that provided the whole spectrum of alleles present at a particular locus were amplified and mixed together to form an allelic ladder for this locus. All samples were scored against such allelic ladders and sequencing of one of those alleles allowed exact determination of the molecular length of each allele at each locus.

Data analysis

Genetic variability parameters (namely allelic frequencies, observed and expected heterozygosity under Hardy-Weinberg's equilibrium and mean number of alleles per locus) were assessed for each locus in each population using BIOSYS-1 software. [37]

Genotype frequencies were tested against Hardy-Weinberg expectations for each locus in the pooled population and for each location. Exact goodness-of-fit tests were performed using GENEPOP 3.1. [38] Possible non-random allelic associations between pairs of polymorphic loci were assessed within each population using Fisher's exact test on contingency tables. The same procedure was then performed across all populations (Fisher's method for a global test). [38] The significance level for each test was adjusted to take into account the other tests using the sequential Bonferroni procedure. [39, 40] Differentiation between populations was examined by F-statistics (Wright [41], calculated according to Weir and Cockerham [42]) and their microsatellite's equivalent, R-statistics. [43] The latter was assessed using a program written in the SAS language [44] by T. Lehmann. [12] Significance of Fis and Fst was assessed using the exact probability test and the exact test of a contingency table testing homogeneity of genotypic frequencies among populations, respectively (available in GENEPOP [45]). Significance of Ris and Rst was evaluated by an F-test in a nested ANOVA on the allele length in a model including the individual and the population as factors. [43] The average Ris across loci and/or populations was estimated by doing the simple means of individual Ris estimates, allowed by sufficient sample sizes (between 44 and 52 individuals per population) and relative homogeneity of the results. Average Rst was calculated following Slatkin [43] from the averages of the within-population and total variance components across loci.

The number of effective migrants (Nm) was estimated using three different ways, either by Slatkin private alleles method [46] (namely, alleles found in only one population as referred to in Neel [47]), and from F- and R-statistics of each locus according to Wright's formula $Nm = (1/Fst - 1)/4$. [48] This formula assumes independence between loci and an island model of migration. Nm was derived from Rst by substituting Rst for Fst and modifying Wright's formula according to Slatkin [43] to incorporate

sampling considerations. Nm values derived from X-linked loci were adjusted to their lower effective population size (Ne: see Wright, Hartl and Clark). [41, 49]

RESULTS

Genetic variability and Hardy-Weinberg equilibrium

Microsatellite loci were highly polymorphic with a number of distinct alleles per locus ranging from 2 for locus 29C up to 16 for locus 26 (Table 1, mean number of alleles per locus on pooled populations = 7.6). Locus 49 was found to be fixed in most *An. arabiensis* populations with a single allele of 87 bp. In the Senegal population, two other alleles were found only once (frequency < 0.01 for each one). This locus was therefore considered monomorphic and was not taken into account in further analyses. Loci 7 and 29C were also found to be monomorphic in the Reunion population. Figure 1 shows the allele composition and distribution of frequencies for the 8 polymorphic microsatellite loci in the four *An. arabiensis* populations tested.

The observed number of alleles per locus (Figure 1 and Table 2) was significantly different among the four populations (F-test on a One-Way ANOVA: F=5.706, P=0.0035) but no significant difference was observed when average unbiased heterozygosities under Hardy-Weinberg equilibrium were compared (F=2.264, P=0.103). Pairwise comparisons of the observed number of alleles per locus showed hierarchical clustering of the four populations studied in the order: Senegal-Madagascar-Mauritius-Reunion. The most common allele was the same for all 4 populations for 5 of 8 loci. Locus 26 showed an interesting pattern of allele distribution with very restricted spectrums for Mauritius and Reunion in comparison to Senegal and Madagascar; the former populations possessing only the longer alleles. Furthermore, the most common allele shared by the Senegalese and Madagascar populations was different than that of the Mascarene island populations (Figure 1).

Hardy-Weinberg predictions were significantly rejected (P<0.05) for each locus when considering the pooled samples as belonging to a single population with significant deficit of heterozygotes (i.e. positive values of Fis, that is equal to Fit here) in all cases (Table 3), indicating a strong Wahlund effect (pooling of separate gene pools). At the population level, slight deviations from Hardy-Weinberg expectations occurred with unpooled data. They remained significant only in two cases after the sequential Bonferroni procedure was applied: for locus 24D in the Senegal population and for locus 147 in the Madagascar population. The occurrence of null alleles due to mutation within the DNA sequence complementary to the oligoprimers, preventing their binding and inhibiting DNA amplification by the Taq polymerase, could explain such a pattern of excess homozygosity. [50] It is an unlikely explanation in this particular case because no homozygote for this « null allele » has been recorded. Nevertheless, L. Kamau and T. Lehmann (personal communication) also observed comparable values of heterozygote deficiency at locus 147 when studying Kenyan *Anopheles gambiae* populations. These two single instances suggesting heterozygote

deficiency were not strengthened by the rest of the dataset and do not suggest obvious violation of Hardy-Weinberg equilibrium. Moreover, all loci conformed when differences were assessed with a chi-square test on pooled alleles (data not shown). No subdivision of the gene pool was then evidenced at the population level.

Linkage disequilibrium

Exact tests for linkage disequilibrium within each population resulted in 5 significant values out of 112 comparisons (5.6 significant values are expected at the 5% level) with P-values between 0.0046 and 0.0388. No pair of loci appeared in linkage disequilibrium in more than one population, suggesting the absence of physical linkage between loci. On the other hand, when the Fisher tests were done on the pooled populations, all but four loci pairs (out of 28 possible combinations) showed highly significant P-values, ranging from $P<10^{-6}$ to 0.045.

Alleles of different loci are in linkage disequilibrium if they occur together in the same individual with a probability higher or lower than would be expected by chance alone. It is an indicator of non-random mating or of selection for certain genotypes, thus being reliable to separation of the gene pool in a single population. [49, 51] The Wahlund effect detected in the above section when the different populations are fused has generated the pattern observed here. These results corroborate our findings of homogeneity of the gene pool within each of the four populations that can be considered as distinct gene pools.

Genetic differentiation

Pairwise Fst and Rst and global values over all loci and over all populations were assessed using GENEPOP version 3.1 and SAS programs, respectively. Estimations are shown in Table 4, together with their corresponding Nm values.

High amounts of genetic differentiation were generally observed in this study. The mean Fst estimate based on the whole dataset (8 loci and 4 populations) was 0.169 and was very significant ($P<10^{-4}$ by Fisher's combined probability test). The corresponding Rst value was slightly higher with an estimate of 0.213.

In order to compare each population to the others, pairwise estimates of F- and R-statistics were assessed. Mean pairwise Fst over all loci between different locations ranged from 0.080 (between Madagascar and Mauritius) to 0.215 (between Reunion and Mauritius) and were very significant (probability of homogeneity < 0.005). Mean pairwise Rst estimates were accordingly lower than their Fst equivalent in all cases, excepted between Senegal and Reunion where it reached 0.302 (corresponding Fst estimate = 0.206). The associated Nm values ranged from 0.91 to 2.88 using Fst method, from 0.29 to 5.61 using Rst values and from 0.50 to 1.35 using Slatkin's private alleles method. As a result, all four populations were found to be significantly differentiated from one another with subsequent gene flow (Nm>1) arising only between Senegal and Madagascar and between Madagascar and Mauritius.

Each locus was then studied independently in order to investigate their contribution to the

global Fst and Rst values observed. Mean locus-specific Fst estimates over all populations were all very significant (probability of homogeneity < 0.005) and varied from 0.038 for locus 141 up to 0.246 for locus 45C. Rst estimates were in good agreement with their Fst equivalents although most were lower. Only five out of 48 locus-specific pairwise Fst estimates were found to be non-significant (p>0.05). Three of them occurred between Reunion and Mauritius, populations that share very similar allele compositions and poor polymorphism at those particular loci (namely locus 141, 803 and 29C). Rst estimates were also non-significant in these particular cases and in 10 more pairwise comparisons.

DISCUSSION

In this study, four *Anopheles arabiensis* populations from different locations were investigated for variation in polymorphism and allele distribution at 9 microsatellite DNA loci. PCR primers described from *An. gambiae* proved very useful to amplify *An. arabiensis* DNA and showed high polymorphism of the loci in this sibling species. The observed number of alleles per locus ranged from 1 for locus 49 up to 16 for locus 26. Subdivision of the gene pool was revealed by both a highly significant deviation from Hardy-Weinberg equilibrium with heterozygote deficits and subsequent linkage disequilibrium between pairs of loci when all samples were considered as belonging to one single population in panmixia. At the population level, no outstanding deviation from Hardy-Weinberg equilibrium was observed. High amounts of genetic divergence were measured by Wright's Fst and Slatkin's Rst, leading to low estimates of the average migration index (Nm<1 in most of the cases) suggesting the occurrence of a strong genetic structuring of the gene pool.

Genetic differentiation between populations

High levels of genetic variability as observed in the present study are very favorable to detecting separation of the gene pool. Indeed, significant differentiation was detected between all four populations we investigated both with F- and R-statistics. This indicates that *An. arabiensis* populations from those four locations have diverged from one another and that the genetic markers we used are prone to detect geographically differentiated units of the same species.

Failloux and others [52] found comparable estimates of Fst when studying *Aedes aegypti* populations from islands of French Polynesia using allozyme markers. In contrast, much lower values were obtained in a study comparing *An. gambiae* populations from Senegal and Kenya using both allozyme and microsatellite markers. [12] The authors found a mean Fst of 0.016 and equivalent Rst of 0.036, suggesting a high level of gene flow in this species across the African continent. Although these populations were 6000 km apart, they were both carriers of the savanna cytotype and a continuous savanna-type habitat bridges them. Such a difference in amounts of genetic differentiation described in this study and ours underlines the expected major influence of the sea acting as a barrier to gene flow. Our high estimates of interpopulation differentiation correspond to low estimates of the migration index (Nm) between pairs of

populations, indicating restricted migration of individuals. The three methods used to estimate Nm gave comparable results and significant correlation was found between Nm estimates from Fst and from Rst (r^2=0.811, P=0.014).

The highest Fst and Rst values were observed between Reunion and each of any other location while Mauritius and Madagascar showed the lowest degree of differentiation. This suggests no correlation between the level of genetic differentiation and geographical distance. Moreover, distance alone, even across large bodies of water, does not appear to be a significant hindrance to mosquito migrations. Similar pattern of absence of significant correlation between genetic differentiation and geographic distance was observed in a recent study of island *Aedes polynesiensis* populations. [53] Interestingly, the authors found significant correlation between gene flow (Nm estimates) and commercial traffic by planes and/or boats between islands: modern intensification of maritime and air travel could then thwart in a non negligible manner the effect of geographic barriers to insect dispersion.

In light of the history of malaria and human population migrations in the Mascarene islands and Madagascar, vector settlement in these particular sites are thought to be due to active transportation by steamship lines, first from the African continent to Madagascar and then from Madagascar to Mauritius and Reunion. [54-56] Also, transport of vectors by wind during a cyclone from Mauritius to Reunion cannot be excluded. Indeed, Mauritius, Reunion and the highlands of Madagascar were originally described as free of malaria vectors. Their introduction was allowed after environmental reshaping due to human settlement (deforestation, development of agriculture) suggesting a strong founder effect that occurred very recently (Mauritius and Reunion were settled by humans at the beginning of the XVIth century). Later, intensive insecticide use on both islands during the seventies considerably diminished malaria vectors density, thus constraining local populations to experience a severe bottleneck. Then, after eradication of malaria from Mauritius and Reunion, great efforts were maintained on autochtonous mosquito population control. Today, particular attention is still paid to avoid unintentional importation of foreign mosquitoes by focusing on areas around harbors and airports and by conducting entomological control in airplanes and boats. These strong control efforts are likely to be one of the major components responsible for such an isolation of the Reunion population and for the overall high amount of genetic differentiation observed in this study by constantly maintaining a reduced population size of the malaria vectors on the islands and by limiting the possibilities of migration and the introduction of foreign alleles. This observation is supported by the high level of genetic differentiation observed between Reunion and Mauritius populations which, despite a comparable entomological history, revealed significantly different patterns of allelic variation at the microsatellite loci studied here.

The heterogeneity of allelic variation patterns observed between Mauritius and Reunion populations has also a more fundamental implication as indicating no evidence for overall constraints acting on allele size or for selection

for one (or few) particular genotype(s) despite fairly high insecticide selection pressure. Nevertheless, three out of eight microsatellite loci showed no significant differentiation between these populations (locus 141, 803 and 29C, see Table 4) and could therefore reflect the occurrence of such constraints acting on at least some of those loci.

The above section has shown great influence of both genetic drift and restricted migration opportunities on a large scale genetic structuring of *An. arabiensis* populations. We will now consider their effect at a local population level.

Genetic variability within locale An. arabiensis *populations*

One of the 9 loci we investigated, locus 49, located on the X-chromosome was found to be fixed for the same allele in all four populations and was therefore not informative in *An. arabiensis* populations, despite high polymorphism observed in *An. gambiae* test samples. This locus should therefore have experienced different molecular evolution since speciation occurred between *An. arabiensis* and *An. gambiae*. Locus 7, located less than 1 cM away from locus 49 [35] was found to be monomorphic in the Reunion population but showed great polymorphism in the three other samples. If we assume 1:1 sex ratio in natural diploid species, the X-chromosome is accordingly 25% less represented than the autosomes. As such, loci of this particular chromosome are subject to lower effective population sizes than their autosomal homologues, and therefore are more susceptible to random genetic drift. Such a high level of genetic drift occurring on those particular loci together with high insecticide selection pressure could have led to the loss of variability observed on locus 7 in the Reunion population, as effective population size seems to be reduced on this island, according to the overall low genetic variability observed (Table 2). Indeed, as shown in Figure 1, out of 62 distinct alleles across eight loci, no « private » allele (e.g. diagnostic allele) was found for the Reunion population, 1 for Mauritius (namely allele 102 at locus 7), 3 for Madagascar and 11 for Senegal, the remaining 47 alleles being shared by two or more populations. This pattern of variability would be expected if alleles are lost by enhanced genetic drift in the former populations and is consistent with lower Ne of the Mascarene populations.

Analysis of genotype frequencies and distribution across loci suggest random mating within each of the four populations investigated as no obvious indication of inbreeding nor of genotypic selection was shown by both Hardy-Weinberg and linkage disequilibrium exact probability tests.

Furthermore, even when genetic variability was low and putative selection pressure by intensive use of insecticides was high (e.g. in Reunion and Mauritius populations), no evidence for selection of certain genotypic combinations was observed. Up to now, no resistance to the different insecticides used in those locations (neither to Temephos in the larvae nor to DDT or other organophosphorous compounds in the adults) has been reported. The outcome of such an intensive effort, rather than selecting one particular insecticide-resistant strain of malaria vector, probably contributes to genetic isolation (e.g. limiting gene flow with other continental or

island populations) and to restriction in density (and therefore effective population size) of the locale population of vectors.

It should be noted here that a possible bias towards the results we obtained may have been generated by the way our sampling was realized. Indeed, mosquitoes were collected at times from August 1994 to April 1997 but also at different geographic scales (e.g. in one location in Reunion and from all over the island in Mauritius). Collection in one location at one time point may severely limit the genetic diversity observed, thus increasing the amount of genetic differentiation detected. However, homogeneity of the gene pool was verified within each of the four samples and comparable studies on western Kenyan *An. gambiae* populations failed to detect significant genetic differentiation at the microgeographic level (up to50 km in diameter). [57, 58] It is possible also that allelic and gametic frequencies changes have occurred during the 1994 to 1997 time interval. Studies are currently conducted to assess for temporal variations within *An. arabiensis* natural populations.

Conclusion

Our survey showed that island populations of the malaria vector *Anopheles arabiensis* have significantly diverged from each other essentially by founder effect and enhanced genetic drift due to permanent limitation of their effective population size. Homogenization of the gene pool by exchange of migrants between the different demes is strongly restricted by the presence of a geographic barrier limiting passive dispersion, and intensive control of the different man-made active means of transportation. Further studies are needed to confirm the high amounts of genetic differentiation observed and to investigate the temporal behavior of the genetic structure within and among populations. Comparative studies and estimation of the effective population sizes of these four populations have to be investigated in order to check for the accuracy of the assumptions used to estimate F- and R-statistics: populations have to be at mutation-drift equilibrium and care should be taken not to overlook significant differences in Ne between them that would lead to an overestimation of the level of genetic differentiation. Further studies on linkage disequilibrium and temporal variation in allelic frequencies and heterozygosities should prove very helpful to investigate putative differences in effective population sizes. [59] Furthermore, such studies should help to quantify and appreciate in a greater extent the amplitude of the founder effect that allowed colonization of the Mascarene islands by *An. arabiensis* populations and to assess more precisely the effect of intensive control measures on locale vector populations.

Acknowledgments

We are grateful to Sadjo SOW and all the inhabitants of Barkedji (Senegal) for their help in collecting mosquitoes. We thank Modou DIAGNE and Amadou SALL for sequence data.

Financial support

This work was supported by the Institut Français de Recherche Scientifique pour le Développement en Coopération (ORSTOM), as part of grant n° CNRS 96/C/08. Frederic Simard gratefully acknowledge financial support from Mr François LACOSTE.

LITERATURE CITED:

1. Gillies MT and Coetzee M, 1987. A supplement to the anophelinae of Africa south of the Sahara. *Publ S Afr Inst Med Res n° 55*, 143pp.
2. Lemasson JJ, Fontenille D, Lochouarn L, Dia I, Simard F, Ba K, Diop A, Diatta M and Molez JF, 1997. Comparison of behaviour and vector efficiency *of Anopheles gambiae* and *An. arabiensis* (Diptera: Culicidae) in Barkedji, a sahelian area of Senegal. *J Med Entomol 34(4):* 396-403.
3. Service MW, Joshi GP and Pradhan GD, 1978. A survey of *Anopheles gambiae* (sp. A) *and An. arabiensis* (sp. B) of the *An. gambiae* Giles complex in the Kisumu area of Kenya following insecticidal spraying with OMS-43 (fenitrothion). *Ann Trop Med Parasitol 72:* 377-386.
4. Collins FH and Besansky NJ, 1994. Vector biology and control of malaria in Africa. *Science. 264:* 1874-1875.
5. Crampton JM, Warren A, Lycett GJ, Hughes MA, Comley IP and Eggleston P, 1994. Genetic manipulation of insect vectors as a strategy for the control of vector-borne disease. *Ann Trop Med Parasitol 88:* 3-12.
6. Ashburner M, Hoy MA and Peloquin JJ, 1998. Prospects for the genetic transformation of arthropods. *Insect Molecular Biology 7(3):* 210-213.
7. Coluzzi M, Sabatini A, Petrarca V and Di Deco MA, 1979. Chromosomal differentiation and adaptation to human environment in *Anopheles gambiae* complex. *Trans R. Soc Trop Med Hyg 73:* 483-497.
8. Coluzzi M, Petrarca V and Di Deco MA, 1985. Chromosomal inversion intergradation and incipient speciation in *Anopheles gambiae. Bolletino di Zoologia 52:* 45-63.
9. Cianchi R, Villani F, Toure YT, Petrarca V and Bullini L, 1983. Electrophoretic study of different chromosomal forms within *Anopheles gambiae s.s. Parassitologia 25:* 239-241.
10. Toure YT, Petrarca V, Traore SF, Coulibaly A, Maiga HM, Sankare O, Sow M, Di Deco MA and Coluzi M, 1994. Ecological genetic studies in the chromosomal form Mopti of *Anopheles gambiae* s. str. in Mali, West Africa. *Genetica 94:* 213-223.
11. Lanzaro GC, Toure YT, Zheng L, Kafatos FC and Vernick KD, 1995. Microsatellite DNA and isozyme variability in a West African population of *Anopheles gambiae. Insect Molecular Biology 4:* 105-112.
12. Lehmann T, Hawley WA, Kamau L, Fontenille D, Simard F and Collins FH, 1996. Genetic differentiation of *Anopheles gambiae* populations from East and West Africa: comparison of microsatellite and allozyme loci. *Heredity 77:* 192-208.
13. Queller DC, Strassmann JE, and Hughes CR, 1993. Microsatellites and kindship. *Trends Ecol Evol 8:* 285-288.
14. Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR and Cavalli-Sforza LL, 1994. High resolution of human evolutionary trees with polymorphic microsatellites. *Nature 368:* 455-457.
15. Estoup A, Garnery L, Solignac M and Cornuet JM, 1995. Microsatellite variation in honey bee (*Apis mellifera* L.) populations: hierarchical genetic structure and test of the infinite allele and stepwise mutation models. *Genetics 140:* 679-695.
16. Garza JC, Slatkin M and Freimer NB, 1995. Microsatellite allele frequencies in humans and chimpanzees, with implications for constraints on allele size. *Mol Biol Evol 12:* 432-440.
17. Epplen C, Melmer G, Siedlaczck I, Schwaiger FW, Mäueler W and Epplen JT, 1993. On the essence of « meaningless » simple repetitive DNA in eukaryote genomes. Pena SDJ, Chakraborty R, Epplen JT and Jeffreys AJ, eds. *DNA Fingerprints: State of the Science.* Birkhauser Verlag, Basel, 29-45.
18. Estoup A, Tailliez C, Cornuet JM and Solignac M, 1995. Size homoplasy and mutational processes of interrupted microsatellites in two bee species: *Apis mellifera* and *Bombus terrestris* (Apidae). *Mol Biol Evol 12:* 1074-1084.
19. Lehmann T, Hawley WA and Collins FH, 1996. An evaluation of constraints on microsatellite loci using null alleles. *Genetics 114:* 1155-1163.
20. Dallas JF, 1992. Estimation of microsatellite mutation rate in recombinant inbred strains of mouse. *Mamm Genome 3:* 452-456.
21. Weber JL and Wong C, 1993. Mutation of human short tandem repeats. *Hum Mol Genet 2:* 1123-1128.
22. Levinson G and Gutman .A, 1987. Slipped-strand mispairing: A major mechanism for DNA sequence evolution. *Mol Biol Evol 4(3):* 203-221.
23. Bossart JL and Pashley Prowell D, 1998. Genetic estimates of population structure and gene flow: limitations, lessons and new directions. *TREE 13(5):* 202-206.

24. Trape JF and Rogier C, 1996. Combating malaria morbidity and mortality by reducing transmission. *Parasitology Today 12(6):* 236-240.
25. Trape JF, 1997. Which strategy for malaria control in Africa? *Parasitology Today 13(3):* 125-126.
26. Mouchet J, Laventure S, Blanchy S, Fioramonti R, Rakotonjanabelo A, Rabarison P, Sircoulon J and Roux J, 1997. La reconquête des hautes terres de Madagascar par le paludisme. *Bull Soc Path Ex 90(3):* 162-168.
27. Fontenille D, 1992. Hétérogénéité de la transmission des paludismes à Madagascar. *Mém Soc r belge Ent 35:* 129-132.
28. Fontenille D, Lepers JP, Campbell GH, Coluzzi M, Rakotoarivony I and Coulanges P, 1990. Malaria transmission and vector biology in Manarintsoa, High Plateaux of Madagascar. *Am J Trop Med Hyg 43(2):* 107-115.
29. Zahar AR, 1985.Vector bionomics in the epidemiology and control of malaria. Part I. The WHO African region and the southern WHO Eastern Mediterranean region. Section III: Vector bionomics, malaria epidemiology and control by geographical areas. (D) East Africa. (E) Eastern outer islands. (F) Southwestern Arabia. VBC/85.3 - MAP:85.3.
30. Girod R, Salvan M and Denys JC, 1995. Lutte contre la réintroduction du paludisme à la Réunion. *Cahiers Santé AUPELF-UREF 5:* 397-400.
31. Gopaul R, 1995. Surveillance entomologique à Maurice. *Cahiers Santé AUPELF-UREF 5:* 401-405.
32. Collins FH, Mendez MA, Rasmussen MO, Mehaffey PC, Besansky NJ and Finnerty V, 1987. A ribosomal RNA gene probe differentiates member species of *Anopheles gambiae* complex. *Am J Trop Med Hyg 37:* 37-41.
33. Scott JA, Brogdon WG and Collins FH, 1993. Identification of single specimens *of Anopheles gambiae* complex by the polymerase chain reaction. *Am J Trop Med Hyg 49:* 520-529.
34. Zheng L, Collins FH, Kumar V and Kafatos FC, 1993. A detailed genetic map for the X chromosome of the malaria vector *Anopheles gambiae. Science 261:* 605-608.
35. Zheng L, Benedict MQ, Cornel AJ, Collins FH and Kafatos FC, 1996. An integrated genetic map of the african human malaria vector mosquito, *Anopheles gambiae. Genetics 143:* 941-952.
36. Cairns MJ and Murray V, 1994. Rapid silver staining and recovery of PCR products separated on polyacrylamide gels. *Biotechniques 17(5):* 915-919.
37. Swofford DL and Selander RB, 1989. BIOSYS-1. A computer program for the analysis of allelic variation in population genetics and biochemical systematics. Release 1.7. University of Illinois, Urbana, IL.
38. Raymond M and Rousset F, 1995. GENEPOP Version 1.2. A population genetics software for exact tests and ecumenicism. *Journal of Heredity 86:* 248-249.
39. Holm S, 1979. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics 6:* 65-70.
40. Rice WR, 1989. Analysing tables of statistical tests. *Evolution 43(1):* 223-225.
41. Wright S, 1978. Evolution and genetics of populations, Volume 4. Variability within and among natural populations. University of Chicago Press, Chicago, IL.
42. Weir BS and Cockerham CC, 1984. Estimating F-statistics for the analysis of population structure. *Evolution 38:* 1358-1370.
43. Slatkin M, 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics 139:* 457-462.
44. SAS Institute Inc. 1990. SAS language: references. Version 6, 1st edn. Cary, NC.
45. Raymond M and Rousset F, 1995. An exact test for population differentiation. *Evolution 49:* 1280-1283.
46. Slatkin M, 1985. Rare alleles as indicator of gene flow. *Evolution 39:* 53-65.
47. Neel JV, 1973. « Private » genetic variants and the frequency of mutation among South American Indians. *Proc Nat Acad Sci USA. 70:* 3311-3315.
48. Wright S, 1969. Evolution and genetics of populations, Volume 2. The theory of gene frequencies. University of Chicago Press, Chicago, IL.
49. Hartl D and Clark AG, 1989. Principles of population genetics. 2nd edn. Sinauer Associates, Sunderland, Mass.
50. Callen DF, Thompson AD, ShenY, Phillips H, Richards RI, Mulley JC and Sutherland GR, 1993. Incidence and origin of 'null' alleles in the (AC)n microsatellite markers. *American Journal of Human Genetics 52:* 922-927.
51. Nei M and Li WH, 1973. Linkage disequilibrium in subdivided populations. *Genetics 75:* 213-219.

52. Failloux AB, Darius H and Pasteur N, 1995. Genetic differentiation of *Aedes aegypti*, the vector of dengue virus in French Polynesia. *Journal of the American Mosquito Control Association 11(4):* 457-462.
53. Failloux AB, Raymond M, Ung A, Chevillon C and Pasteur N, 1997. Genetic differentiation associated with commercial traffic in the Polynesian mosquito, *Aedes polynesiensis* Marks 1951. *Biological Journal of the Linnean Society 60:* 107-118.
54. Julvez J, Mouchet J and Ragavoodoo C, 1990. Epidémiologie historique du paludisme dans l'archipel des Mascareignes (Océan Indien). *Ann Soc belge Med Trop 70:* 249-261.
55. Julvez J et Mouchet J, 1994. Le peuplement culicidien des îles du sud-ouest de l'Océan Indien. L'action de l'homme dans l'importation des espèces d'intérêt médical. *Ann Soc Entomol Fr 30(4):* 391-401.
56. Julvez J, 1995. Historique du paludisme insulaire dans l'Océan Indien (partie sud-ouest). Une approche éco-épidémiologique. *Cahiers Santé AUPELF-UREF 5:* 353-357.
57. Lehmann T, Besansky NJ, Hawley WA, Fahey TG, Kamau L and Collins FH, 1997. Microgeographic structure of *Anopheles gambiae* in western Kenya based on mtDNA and microsatellite loci. *Molecular Ecology 6:* 243-253.
58. Kamau L, Lehmann T, Hawley WA, Orago ASS and Collins FH, 1998. Microgeographic genetic differentiation of *Anopheles gambiae* mosquitoes from Asembo Bay, western Kenya: A comparison with Kilifi in coastal Kenya. *Am. J. Trop. Med. Hyg. 58(1):* 64-69.
59. Lehmann T, Hawley WA, Grebert H and Collins FH, 1998. The effective population size of *Anopheles gambiae* in Kenya: implications on population structure. *Mol Biol Evol 15(3):* 264-276.
60. Besansky NJ, Hamm DM and Collins FH, 1994. An *Anopheles gambiae* cDNA predicts a protein similar to a yeast *Sui1* translation factor. *Gene 141:* 299-300.
61. Nei M, 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics 89:* 583-590.

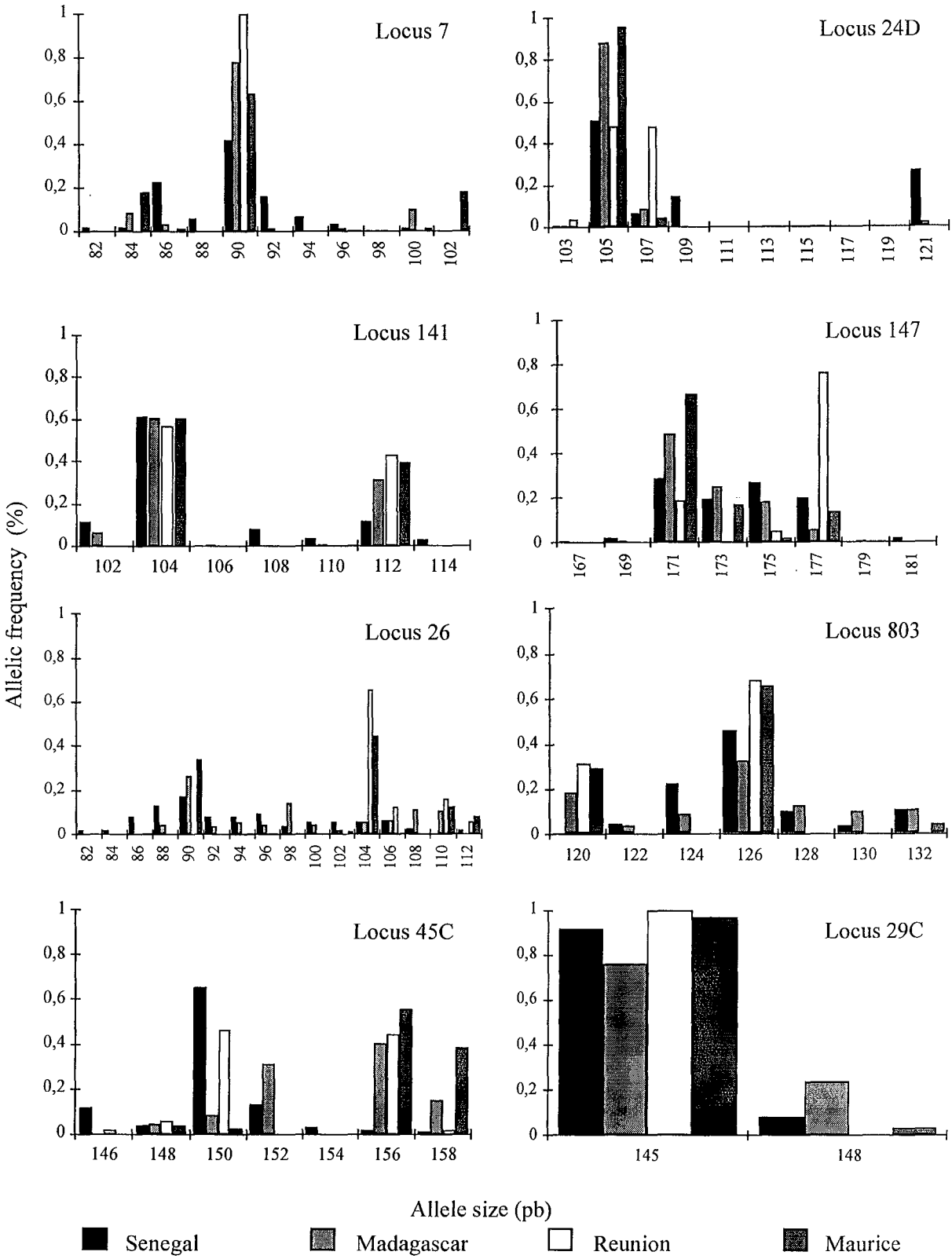


Figure 1 : Observed allelic frequencies in the four *An. arabiensis* populations for the eight loci studied. Alleles are denoted by their size in base pairs.

Table 1: *Anopheles arabiensis* microsatellite loci: cytological location, repeat motif and primer sequences.

Locus	Cytological Location	Repeat	Primers	Number of alleles in this study	Allele size range (bp)
24D [1,3]	2 La : 24D (marker 11/*Sui1*)	CT	5'GGCGAGCAGTTCATTCAAGT 3' 5'CGTCTGGAAGTTTCGTTGAG 3'	3	103 - 121
141 [1]	2 Rd	GT	5'CGGAGCAAATCTGAACCGTG 3' 5'CCTTGGCCACAACAACATCG 3'	7	102 - 114
26 [1]	2 Rb : 12	GT	5' GGTTCCTGTTACTTCCTGCC 3' 5'CCGGCAACACAAACAATCGG 3'	16	82 - 112
7 [2]	X : 1C	GT	5'CACGATGGTTTTCGGTGTGG 3' 5'ATTTGAGCTCTCCCGGGTG 3'	11	82 - 102
49 [2]	X : 1D	GT	5'CAGCGCCTCCATATAGAACG 3' 5'GATCATTCAGCTGAACCTGC 3'	1	87
803 [1]	2 R : 7B	TG	5'CTCGATAAATCCCGTCGGTG 3' 5'GTCGGTTTGAGGTTGTAAAGC 3'	7	120 - 132
45C	3 L : 45C (ATPase unit2-like gene)	GA	5'AAAAGTGGTGACCGAGTGAC 3' 5'ATCTTCAACACTTCAGCACG 3'	7	146 - 158
29C	3 R : 29C	TGA	5'ATGTTCCAGAGACGACCCAT 3' 5'TGTTGCCGGTTTGTTGCTGA 3'	2	145 - 148
147 [1]	2 R : 19	GT	5'CTGCTGTTGCTGCCAAAATG 3' 5'AGCTTCACGGAAAGCAAAGG 3'	8	167 - 181

[1] data from Zheng and others.[35] [2] data from Zheng and others.[34] [3] data from Besansky and others.[60]

Table 2: Genetic variability at 8 microsatellite loci for each population.

Population	Mean sample size per locus	Mean no. of alleles per locus	Percentage of loci polymorphic *	Mean heterozygosity	
				Direct-count	Hardy-Weinberg expected **
Senegal	53.8 (0.2)	7.3 (1.3)	100.0	0.558 (0.068)	0.636 (0.080)
Madagascar	52.9 (0.1)	5.9 (0.1)	100.0	0.549 (0.084)	0.571 (0.083)
Reunion	51.9 (0.1)	2.6 (0.5)	75.0	0.383 (0.087)	0.373 (0.084)
Mauritius	43.8 (0.2)	3.4 (0.5)	100.0	0.406 (0.074)	0.425 (0.079)

* A locus is considered polymorphic if more than one allele was detected,

** Unbiased estimate, [61]

( ) Standard errors in parentheses.

Table 3: Hardy-Weinberg Expectations

Locus		Pooled Population [a]	Senegal	Madagascar	Reunion	Mauritius	All Populations [b]
	P HW	**< 10^{-4}**	**0.0056**	1	0.0377	1	**< 0.005**
24D	Fis	**+ 0.320**	**+ 0.259**	- 0.09	+ 0.149	- 0.036	**+ 0.156**
	Ris	**+ 0.275**	+ 0.079	- 0.031	+ 0.296	- 0.036	+ 0.077
	P HW	**0.0156**	0.299	0.3301	0.1652	0.3524	0.69
141	Fis	**+ 0.004**	+ 0.128	+ 0.055	- 0.205	- 0.156	- 0.026
	Ris	- 0.002	+ 0.216	+ 0.053	- 0.204	- 0.156	- 0.023
	P HW	**< 10^{-4}**	0.261	0.0507	0.6893	0.3542	0.085
26	Fis	**+ 0.147**	+ 0.090	- 0.013	- 0.085	+ 0.123	+ 0.033
	Ris	**+ 0.291**	+ 0.043	+ 0.008	- 0.089	+ 0.027	- 0.003
	P HW	**< 10^{-4}**	0.0225	0.3614	-	0.5854	0.015
7	Fis	**+ 0.236**	+ 0.195	- 0.021	-	+ 0.072	+ 0.1074
	Ris	+ 0.128	+ 0.161	+ 0.032	-	+ 0.140	+ 0.111
	P HW	**< 10^{-4}**	0.6247	0.0434	0.2111	0.4202	0.44
803	Fis	**+ 0.072**	+ 0.076	+ 0.094	- 0.189	- 0.134	- 0.001
	Ris	+ 0.129	- 0.018	+ 0.245	- 0.189	+ 0.032	+ 0.017
	P HW	**< 10^{-4}**	0.9034	0.0665	0.0764	0.1531	0.14
45C	Fis	**+ 0.235**	+ 0.004	- 0.059	+ 0.093	+ 0.210	+ 0.047
	Ris	**+ 0.497**	- 0.111	+ 0.097	+ 0.237	+ 0.187	+ 0.102
	P HW	**0.0024**	1	0.1288	-	1	0.05
29C	Fis	**+ 0.199**	- 0.082	+ 0.224	-	- 0.024	+ 0.117
	Ris	**+ 0.199**	- 0.082	+ 0.224	-	- 0.024	+ 0.039
	P HW	**< 10^{-4}**	0.2764	**0.0025**	1	0.5561	0.025
147	Fis	**+ 0.263**	+ 0.138	**+ 0.107**	+ 0.025	+ 0.106	+ 0.1043
	Ris	**+ 0.346**	+ 0.069	**+ 0.368**	+ 0.041	+ 0.049	+ 0.132
All	P HW [c]	**< 10^{-4}**	**0.0107**	**0.0015**	0.0848	0.7513	**< 10^{-4}**
loci	Fis	**+ 0.214**	**+ 0.122**	**+ 0.039**	- 0.028	+ 0.044	**+ 0.055**
	Ris	+ 0.233	+ 0.045	+ 0.124	+ 0.015	+ 0.027	+ 0.054

P HW: Exact goodness-of-fit test probability (see text).
Underlined characters: P<0.05,
Bold characters indicate values that remain significant after the sequential Bonferroni procedure was applied to take into account the other tests.
[a] Fis=Fit and Ris=Rit.
[b] P refers to one-sided test probability of obtaining by chance a value as large or larger than the observed one (permutation of alleles within samples).
[c] P HW refers to Fisher's combined probality test across loci.
« All » sections: Ris is the simple mean of individual Ris (see text).

Table 4: Genetic differentiation between populations

Locus		Senegal x Madagascar	Senegal x Reunion	Senegal x Mauritius	Madagascar x Reunion	Madagascar x Mauritius	Reunion x Mauritius	All Populations
24D	Fst (Nm)	0.202 ** (0.99)	0.176 ** (1.17)	0.262 ** (0.70)	0.297 ** (0.59)	*0.012* ^{NS} (21.12)	0.379 ** (0.41)	0.243 ** (0.78)
	Rst (Nm)	0.162 ** (0.65)	0.144 ** (0.74)	0.184 ** (0.55)	*0.001* ^{NS} (12.76)	*0.006* ^{NS} (20.01)	0.162 ** (0.65)	0.218 ** (0.90)
141	Fst (Nm)	0.031 ** (7.79)	0.094 ** (2.41)	0.074 ** (3.11)	0.011 * (22.67)	0.001 * (227.03)	*0.0000* ^{NS} (inf.)	0.038 ** (6.23)
	Rst (Nm)	*0.010* ^{NS} (12.37)	0.059 ** (1.99)	0.042 ** (2.87)	*0.012* ^{NS} (9.95)	*0.003* ^{NS} (35.60)	*0.0000* ^{NS} (inf.)	0.031 ** (7.90)
26	Fst (Nm)	0.022 ** (11.06)	0.241 ** (0.79)	0.128 ** (1.70)	0.249 ** (0.75)	0.110 ** (2.02)	0.122 ** (1.80)	0.150 ** (1.42)
	Rst (Nm)	0.060 ** (1.96)	0.517 ** (0.12)	0.137 ** (0.79)	0.295 ** (0.30)	0.016 * (7.59)	0.157 ** (0.67)	0.275 ** (0.66)
7	Fst (Nm)	0.148 ** (1.91)	0.357 ** (0.60)	0.111 ** (2.68)	0.142 ** (2.02)	0.058 ** (5.40)	0.280 ** (0.86)	0.183 ** (1.49)
	Rst (Nm)	*0.010* ^{NS} (16.50)	*0.005* ^{NS} (33.20)	0.025 * (6.50)	*0.001* ^{NS} (166)	*0.0007* ^{NS} (238)	*0.016* ^{NS} (10.25)	0.016 * (15.20)
803	Fst (Nm)	0.036 ** (6.75)	0.155 ** (1.37)	0.128 ** (1.71)	0.126 ** (1.74)	0.101 ** (2.23)	*0.0000* ^{NS} (inf.)	0.095 ** (2.37)
	Rst (Nm)	*0.0001* ^{NS} (inf.)	0.141 ** (0.76)	0.086 ** (1.33)	0.064 ** (1.83)	0.032 * (3.72)	*0.0000* ^{NS} (inf.)	0.076 ** (3.00)
45C	Fst (Nm)	0.300 ** (0.58)	0.172 ** (1.20)	0.441 ** (0.32)	0.158 ** (1.33)	0.115 ** (1.92)	0.223 ** (0.87)	0.246 ** (1.83)
	Rst (Nm)	0.406 ** (0.18)	0.181 ** (0.57)	0.696 ** (0.05)	0.055 ** (2.15)	0.150 ** (0.71)	0.302 ** (0.29)	0.421 ** (0.34)
29C	Fst (Nm)	0.074 * (3.14)	0.074 * (3.13)	*0.011* ^{NS} (21.62)	0.225 ** (0.86)	0.141 * (1.52)	*0.027* ^{NS} (8.97)	0.120 ** (1.83)
	Rst (Nm)	0.039 ** (3.08)	0.038 ** (3.16)	*0.005* ^{NS} (24.90)	0.128 ** (0.85)	0.077 ** (1.50)	*0.014* ^{NS} (9.07)	0.094 ** (2.42)
147	Fst (Nm)	0.039 * (6.12)	0.255 ** (0.73)	0.131 ** (1.65)	0.386 ** (0.40)	0.046 * (5.22)	0.424 ** (0.34)	0.222 ** (0.88)
	Rst (Nm)	0.053 ** (2.23)	0.136 ** (0.79)	0.09 ** (1.26)	0.340 ** (0.24)	*0.004* ^{NS} (28.30)	0.382 ** (0.40)	0.258 ** (0.72)
All loci	Fst° (Nm)	0.109 ** (2.04)	0.206 ** (0.96)	0.188 ** (1.08)	0.214 ** (0.92)	0.080 ** (2.87)	0.215 ** (0.91)	0.169 ** (1.23)
	Rst° (Nm)	0.095 (1.19)	0.302 (0.29)	0.169 (0.61)	0.174 (0.59)	0.022 (5.61)	0.132 (0.82)	0.213 (0.69)
	Nm (PA)	1.35	0.56	0.69	0.59	0.87	0.50	1.33

Significance of Fst and Rst are indicated as follow: *NS* non significant at the 5% level; * P< 0.05; ** P< 0.005
Fst°: Probability of homogeneity over all loci was estimated by Fisher's combined probability test.
Rst°: Estimated as described in Slatkin (1995).[43]
Nm (PA): Nm estimates using Slatkin's Private alleles method (1985).[46]

AMERICAN JOURNAL OF TROPICAL MEDICINE AND HYGIENE
3088 Briarcliff Road, Suite A-11 TEL: (404) 636-3621
Atlanta, GA 30329 USA FAX: (404) 633-5737

E-mail (Copy Editor) tgryczan@ajtmh.org

January 21, 1999

MS#: 98-180

RRH: GENETIC DIFFERENTIATION IN *AN. ARABIENSIS*

Dr. Frederic Simard
Laboratoire ORSTOM de Zoologie Medicale
Institut Pasteur
BP 220
Dakar, SENEGAL

Dear Dr. Simard:

It is a pleasure to advise you that the above-listed manuscript has been accepted for publication in the journal. The enclosure with this letter is a copyedited version prepared from your revised manuscript; a similar version will be used by Allen Press to produce the final copy.

Please proofread the copy carefully, marking thereon any changes necessary. To the extent possible, these will be incorporated into the final document. Also, please respond to all queries marked in red. The marked copy should be returned as quickly as possible (preferably within five days of receipt). **Failure to return the marked copy will result in a delay in the editorial process.** A late response and/or significant modifications may also delay publication. Please do not return a new version of the manuscript; make all changes on the enclosed copy (additional pages can be added if needed). Later, you will receive actual page proofs from the printer. Changes to these page proofs, other than correcting typographical errors, can be costly and will be the responsibility of the author.

We look forward to working with you towards final publication at the earliest possible date.

Sincerely yours,

James W. Kazura, MD
Editor-in-Chief

Annexe 3

Lemasson JJ, Fontenille D, Lochouarn L, Dia I, Simard F, Ba K, Diop A, Diatta M et Molez J-F. 1997. Comparison of behavior and vector efficiency of *Anopheles gambiae* and *An. arabiensis* (Diptera: Culicidae) in Barkedji, a sahelian area of Senegal. *J. Med. Entomol.* **34(4)**: 396-403.

Comparison of Behavior and Vector Efficiency of *Anopheles gambiae* and *An. arabiensis* (Diptera: Culicidae) in Barkedji, a Sahelian Area of Senegal

JEAN-JACQUES LEMASSON,[1] DIDIER FONTENILLE,[1,2] LAURENCE LOCHOUARN,[1] IBRAHIMA DIA,[1] FREDERIC SIMARD,[1] KHALILOU BA,[1] ABDOULAYE DIOP,[3] MATHURIN DIATTA,[1] AND JEAN-FRANCOIS MOLEZ[1]

J. Med. Entomol. 34(4): 396–403 (1997)

ABSTRACT The ecology, population dynamics, and malaria vector efficiency of *Anopheles gambiae* and *An. arabiensis* were studied for 2 yr in a Sahelian village of Senegal. Anophelines were captured at human bait and resting indoors by pyrethrum spray. Mosquitoes belonging to the *An. gambiae* complex were identified by polymerase chain reaction. Of 26,973 females, *An. arabiensis* represented 79% of the mosquitoes captured and remained in the study area longer than *An. gambiae* after the rains terminated. There were no differences in nocturnal biting cycles or endophagous rates between *An. gambiae* and *An. arabiensis*. Based on an enzyme-linked immunosorbent assay test of bloodmeals, the anthropophilic rate of these 2 vectors were both ≈60%, when comparisons were made during the same period. Overall, 18% of the resting females had patent mixed bloodmeals, mainly human–bovine. The parity rates of *An. gambiae* and *An. arabiensis* varied temporally. Despite similar behavior, the *Plasmodium falciparum* circumsporozoite protein (CSP) rates were different between *An. gambiae* (4.1%) and *An. arabiensis* (1.3%). *P. malariae* and *P. ovale* only represented 4% of the total *Plasmodium* identified in mosquitoes. Transmission was seasonal, occurring mainly during 4 mo. The CSP entomological inoculation rates were 128 bites per human per year for the 1st yr and 100 for the 2nd yr. Because of the combination of a high human biting rate and a low CSP rate, *An. arabiensis* accounted for 63% of transmission. Possible origin of differences in CSP rate between *An. gambiae* and *An. arabiensis* is discussed in relation to the parity rate, blood feeding frequency, and the hypothesis of genetic factors.

KEY WORDS *Anopheles gambiae*, *Anopheles arabiensis*, malaria transmission, bloodmeal sources, population dynamics, Senegal

THE MAIN VECTORS responsible for malaria transmission in Africa are *Anopheles gambiae* Giles and *An. arabiensis* Patton, 2 of 6 species comprising the *An. gambiae* complex, as well as *An. funestus* Giles. In numerous regions of West Africa, *An. gambiae* and *An. arabiensis* are sympathric (Robert et al. 1989, Lindsay et al. 1993, Toure et al. 1996).

With a long-term goal of using transgenic mosquitoes with reduced malaria vector competences for control, data are being collected on intra- and interspecific gene flow within the *An. gambiae* complex (Lanzaro et al. 1995, Lehmann et al. 1996). The Sahelian area, with its short rainy season, could be a useful area for studying the population genetics of malaria vectors, because vector populations here probably are more isolated from each other both spacially and temporally than in regions where mosquitoes are present throughout the year.

Few studies have been conducted in the Sahelian region to compare population dynamics, biology, behavior and vector efficiency among *Anopheles* species. Some surveys evaluated malaria transmission (Hamon et al. 1965, Omer and Cloudsley-Thompson 1970, Haridi 1972, Vercruysse 1985, Gazin et al. 1988, Faye et al. 1993, Taylor et al. 1993). Following the definition of the WHO, vector efficiency is defined as the "ability of a mosquito species, in comparison to another species in a similar climatic environment, to transmit malaria in nature. A rough estimate of relative efficiency may be made by comparison of sporozoite rates taken in comparable conditions." (WHO 1963). This is, of course, different from vectorial capacity which does not take into account the sporozoite rate (Garret-Jones and Shidrawi 1969).

Generally, *An. arabiensis* is considered to have an anthropophilic rate, life expectancy, and vectorial capacity lower than *An. gambiae* (Coluzzi 1984, Gillies and Coetzee 1987). Because of the methods of capture, the methods of identification of species

[1]Institut Francais de Recherche Scientifique pour le Développement en coopération (ORSTOM), Institut Pasteur, BP 220, Dakar, Sénégal.

[2]To whom reprint requests should be addressed.

[3]Département de Biologie Animale, Universite Cheikh Anta Diop, Dakar, Sénégal.

0022-2585/97/0396–0403$02.00/0 © 1997 Entomological Society of America

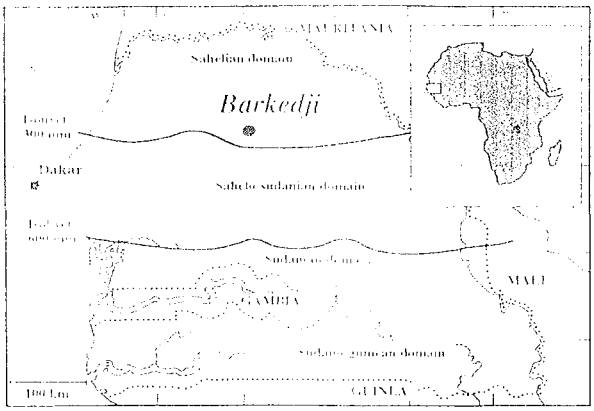


Fig. 1. Location of study village in Senegal.

of the *An. gambiae* complex, variations in vector density, and the availability of vertebrate hosts over time, comparison between *An. gambiae* and *An. arabiensis* may be distorted, as it was shown in Kenya (Petrarca et al. 1991) and in Diolmo, another region of Senegal (unpublished data).

The aim of our longitudinal study was to assess the differences in biology, behavior, and circumsporozoite protein (CSP) rate of *An. gambiae* and *An. arabiensis* in relation to malaria transmission, and to estimate the intensity and periodicity of transmission, and respective roles of each vector over time. Availability of a polymerase chain reaction (PCR) technique, which allowed identification of all the specimens of the *An. gambiae* complex, allowed us to reach these objectives.

Materials and Methods

Study Area. The study was carried out in the village of Barkedji (15° 17′ N, 14° 53′ W) (Fig. 1). This village of 700 inhabitants is situated in the Ferlo, in the Sahelian region of Senegal. It is representative of villages in the area. The dominant ethnic group is Peuhl, mainly cattle and sheep farmers, and some Wolof farmers. The rainy season is short, extending from July to October. Rainfall varies annually: 215 mm in 1991, 347 mm in 1992, 335 mm in 1993, 301 mm in 1994, and 404 mm in 1995, with extraordinary downpours of 26 mm in December 1995. The village of Barkedji is situated in the fossil valley river bed of the Ferlo and is surrounded by clay hollows which collect water as soon as the rains start. During the last 2 yr, these temporary ponds dried by November 1994 and by January 1996. The average minimum and maximum monthly temperature, recorded by the national meteorological service, were constantly high, ranging from 16.4–32.3°C in January 1995 to 22.6–41.2°C in May 1995.

Mosquito Collections. Adult mosquitoes were captured biweekly from June to December 1994 and from July 1995 to March 1996. A survey was conducted in June 1995 at the end of the dry season. The following 2 collection techniques were used: (1) Hourly human bait collections were made on adult volunteers from 1900 to 0700 hours at the same sites for 2 consecutive nights. One indoor collector and outdoor collector were positioned at each site. In total, 8 human-nights were done every 2 wk. The human biting rate was expressed as the number of mosquito bites per person per night during each 2-wk sample. (2) During 1995 pyrethrum spray collections were made in the early morning inside 16 bedrooms, which were different from houses used for human bait collections. Houses were divided into 4 groups of 4 compounds with 2 bedrooms per compound, situated in different parts of the village. Only 6 of these were used during the following year.

Field Processing of Anophelines. Anophelines were identified using the morphological characteristics and identification key of Gillies and De Meillon (1968). Ovaries from a portion of female anophelines captured on human bait were dissected to determine parity (Detinova 1962). All the mosquitoes from the *An. gambiae* complex, dissected or not, were stored individually in a numbered tube with dessicant for laboratory processing in Dakar.

Laboratory Processing of Anophelines. Bloodmeal sources of a sample of females captured biweekly by pyrethrum spray were identified by an enzyme-linked immunosorbent assay (ELISA) (Beier et al. 1988). The technique identified human, bovine, ovine or caprine (sheep and goat), equine (horse and donkey), or chicken hosts.

The head and thorax of female anophelines were tested for circumsporozoite protein of *Plasmodium falciparum*, *P. malariae*, and *P. ovale* by the ELISA as described by Burkot et al. (1984) and modified by Wirtz et al. (1987). The circumsporozoite protein rate, and the 95% CI were calculated. *P. vivax* is not present in this region of Africa. The entomological inoculation rate was calculated by multiplying the human biting rate by the circumsporozoite protein rate for each 2-wk period.

Females belonging to the *An. gambiae* complex were identified to species using the PCR technique described by Scott et al. (1993). A leg or a wing was placed directly into the reaction mixture containing the species-specific primers, dNTPs, buffer and polymerase. The length of the amplified sequences was 315 nucleotides for *An. arabiensis*, 390 for *An. gambiae*, and 464 for *An. melas*. This technique has been validated in West Africa (Fontenille et al. 1993). When >20 anophelines of the *An. gambiae* complex were captured during each 2-wk period by each of the capture methods, a minimum of 20 specimens were randomly identified by PCR. The probable number of specimens per species for each method of capture then was extrapolated. All mosquitoes found positive by circumsporozoite protein ELISA were identified by PCR.

Table 1. Number and percentage of malaria vectors caught from July 1994 to March 1996 by different methods in Barkedji

No. or %	Feeding on humans outdoors	Feeding on humans indoors	Resting in bedrooms	Total
No. *An. gambiae* s.l. captured	3,309	4,451	19,213	26,973
No. *An. gambiae* s.l. tested by PCR	504	601	1,728	2,833
Estimated No. *An. gambiae*	616	796	4,285	5,697
Estimated No. *An. arabiensis*	2,693	3,655	14,928	21,276
% *An. gambiae*	18.6	17.9	22.3	21.1
% *An. arabiensis*	81.4	82.1	77.7	78.9

Results

Capture of Mosquitoes. From July 1994 to March 1996, 26,973 *Anopheles* were collected by spray collections in bedrooms, and 264 man-nights on human volunteers. No *An. funestus* were caught. Of the *An. gambiae* complex females captured, 2,833 were identified by PCR. The number and percentage of females of *An arabiensis* and *An. gambiae* captured by each method are shown in Table 1. Non malaria-vector species captured included *An. pharoensis* Theobald, *An. rufipes* (Gough), *An. ziemanni* Grünberg, *An. coustani* Laveran, and *An. squamosus* Theobald.

Considerable differences were observed each year. During 1994, *An. gambiae* represented 27.3% of total *An. gambiae* complex females captured; but during 1995, only 12.3%, using the same techniques. The number of resting mosquitoes varied greatly among rooms. During 1994, 140 females of the *An. gambiae* complex were captured by 12 pyrethrum spray collections in the least productive room (maximum 54 females in October), whereas the most productive room produced 1,901 anophelines (maximum 555 females in October). The percentage of *An. gambiae* was significantly higher among females collected resting than biting indoors or outdoors ($\chi^2 = 6.7$, df = 1, $P < 0.01$). No anophelines were found during the survey conducted in June 1995 at the end of the dry season.

Seasonality and Biting Cycles. The human biting rate for each species varied temporally depending on the rainy season (Fig. 2). In 1994, the biting rate was maximum in October for *An. gambiae*, with a peak of 27 bites per human per night during the 1st wk of October. For *An. arabiensis* the maximum rate was in November with a peak of 107 bites per human per night during the 2nd wk of November. In 1995, the human biting rate for *An. gambiae* was always low, with a maximun average of 7.5 bites per human per night during the last week of October. For *An. arabiensis* the maximum rate was in October with a peak of 104 during the last week of the month. *An. gambiae* disappeared from collections earlier in the year than *An. arabiensis*.

Night-biting cycles were similar for *An. gambiae* and *An. arabiensis* (Fig. 3), and did not vary during the year (data not shown). A biting peak was observed consistently between 0300 and 0500 hours for both species.

Host-Seeking Behavior. Overall, 56.4% of *An. gambiae* and 57.6% of *An. arabiensis* were cap-

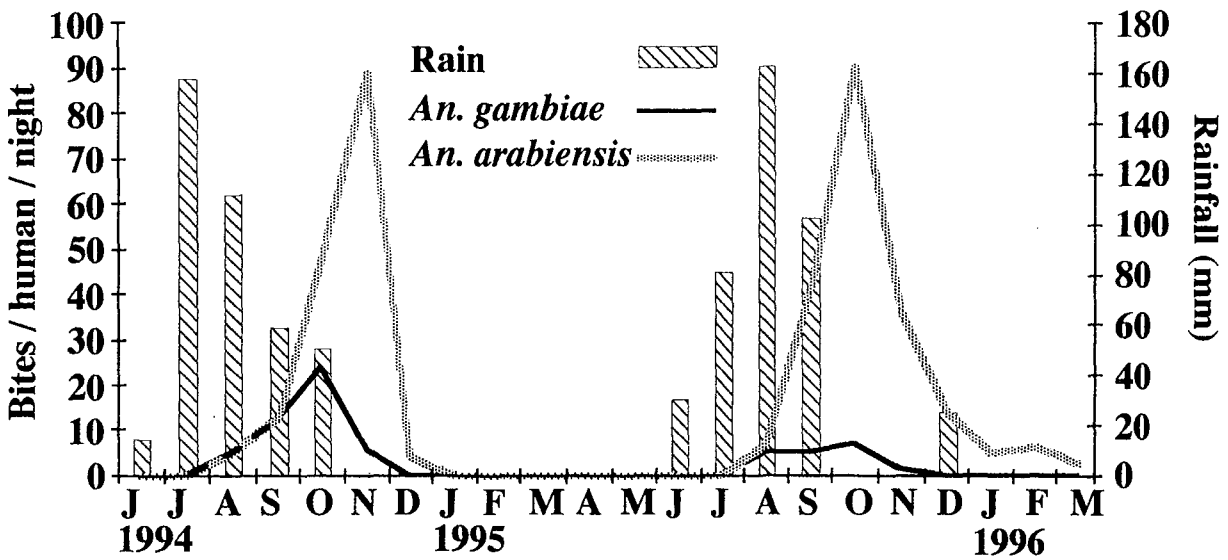


Fig. 2. Monthly total rainfall and human biting rate for *An. gambiae* and *An. arabiensis* in Barkedji from June 1994 to March 1996.

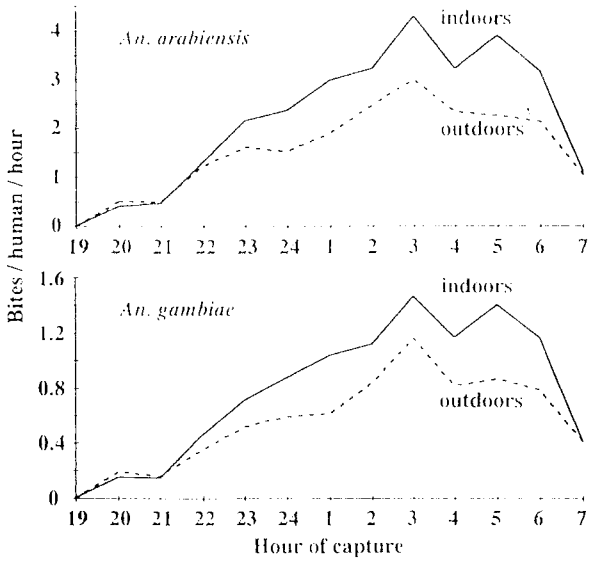


Fig. 3. Average number of bites per human per hour indoors and outdoors for *An. gambiae* and *An. arabiensis* during the rainy season in Barkedji.

tured on humans indoors. The endophagous rate was similar for both species (χ^2 = 0.68, df = 1, P = 0.41).

In total, 2,029 bloodmeals from resting females of the *An. gambiae* complex were tested by ELISA, including 264 from PCR identified *An. gambiae* and 859 *An. arabiensis* (Table 2). The anthropophilic rate of *An. gambiae* and *An. arabiensis* were similar only during the rainy season (61.9% versus 58.5%, χ^2 = 0.71, df = 1, P = 0.40) (Fig. 4). Over 2 yr, including rainy and dry seasons, the anthropophilic rate was 62.1% for *An. gambiae* and 68.0% for *An. arabiensis* (χ^2 = 3.12, df = 1, P = 0.08). In addition 35.2% of *An. gambiae* and 28.1% of *An. arabiensis* fed on cattle. Overall, 8% of anophelines had fed on ovine and 11.7% on equine. Only 2 females (0.1%) had taken chicken bloodmeals. Overall, 17.7% of bloodmeals were taken from 2 different host species. The most frequent combinations were human–cow bloodmeals (6.4% for *An. gambiae* and 7.5% for *An. arabiensis*), then human–equine bloodmeals (1.1% for *An. gambiae* and 4.4% for *An. arabiensis*). Four mosquitoes fed on 3 different species (0.2%). The percentage of mixed bloodmeals was the same for 1994 and 1995 (18.1% versus 17.4%, χ^2 = 0.18, df = 1, P = 0.67). Over the 2-yr study, 17.0% of *An. gambiae* tested had taken patent mixed bloodmeals compared with 18.6% for *An. arabiensis* (χ^2 = 0.28, df = 1, P = 0.60).

Parity Rates of Vectors. In 1994 the parity rate was 79.4% (n = 107) for *An. gambiae* (CI = 71.4–87.4), and 60.7% (n = 211) for *An. arabiensis* ($CI_{95\%}$ = 54.1–67.3). This difference was highly significant (χ^2 = 11.3, df = 1, P < 0.001). In 1995 the parity rate was 40.4% (n = 57) for *An. gambiae* ($CI_{95\%}$ = 27.7–53.1), and 80.2% (n = 323) for *An. arabiensis* ($CI_{95\%}$ = 75.9–84.5). This difference was highly significant (χ^2 = 40.2, df = 1, P < 0.001).

Circumsporozoite Protein Rates. The circumsporozoite protein rate was calculated monthly for each species. Overall, 96% of identified *Plasmodium* were *P. falciparum* (Table 3). In 1994 4.5% ($CI_{95\%}$ = 3.5–5.9) of *An. gambiae* and 1.3% (0.9–1.8) of *An. arabiensis* tested positive for *P. falciparum*. This difference was highly significant (χ^2 = 39, df = 1, P < 0.0001). In 1995, 2.8% (CI = 1.3–5.1) of *An. gambiae* and 1.3% (0.9–1.8) of *An. arabiensis* tested positive for *P. falciparum*. This difference also was significant (Fisher exact test, P = 0.047). If the circumsporozoite protein rates were calculated only with mosquitoes captured from July to October, when *An. gambiae* and *An. arabiensis* both were present, the circumsporozoite protein rates were 2.9 and 1.8% for *An. gambiae* and *An. arabiensis*, respectively, and were not different (χ^2 = 2.1, df = 1, P = 0.15).

Adjusting for years, this difference was highly significant (4.1% for *An. gambiae* versus 1.5% for *An. arabiensis*, Mantel–Haenszel chi-square for stratified analysis = 40.7, df = 1, P < 0.0001). For each species the circumsporozoite protein rates over the 2 yr were not significantly different (4.5% versus 2.8% for *An. gambiae*, χ^2 = 2.0, df = 1, P = 0.16, and 1.3% versus 1.3% for *An. arabiensis*, χ^2 = 2.10^{-5}, df = 1, P > 0.99). Little *P. malariae*

Table 2. Percentage of indoor resting mosquitoes fed on each vertebrate host during the rainy season from July to October when *An. gambiae* and *An. arabiensis* were present simultaneously, as well as over the whole year (in brackets) (Barkedji July 1994 to March 1996)

Mosquito species	No. of mosquitoes	Mosquitoes fed on each vertebrate host, %					% mixed bloodmeal
		Human	Bovine	Ovine	Chicken	Equine	
An. gambiae	257	61.9	35.4	11.3	0.0	8.6	17.1
	(264)	(62.1)	(35.2)	(11.0)	(0.0)	(8.7)	(17.0)
An. arabiensis	376	58.5	41.5	7.2	0.0	11.4	18.6
	(859)	(68.0)	(28.1)	(7.9)	(0.1)	(14.6)	(18.6)
Not determined *An. gambiae* s.l.	750	66.1	35.2	6.5	0.0	9.7	17.6
	(906)	(69.1)	(30.9)	(7.2)	(0.1)	(9.8)	(17.1)
Total	1,383	63.3	36.9	7.6	0.0	10.0	17.8
	(2,029)	(67.7)	(30.3)	(8.0)	(0.1)	(11.7)	(17.7)

Mixed bloodmeal: 2 different vertebrate species detected in the bloodmeal of the same mosquito.

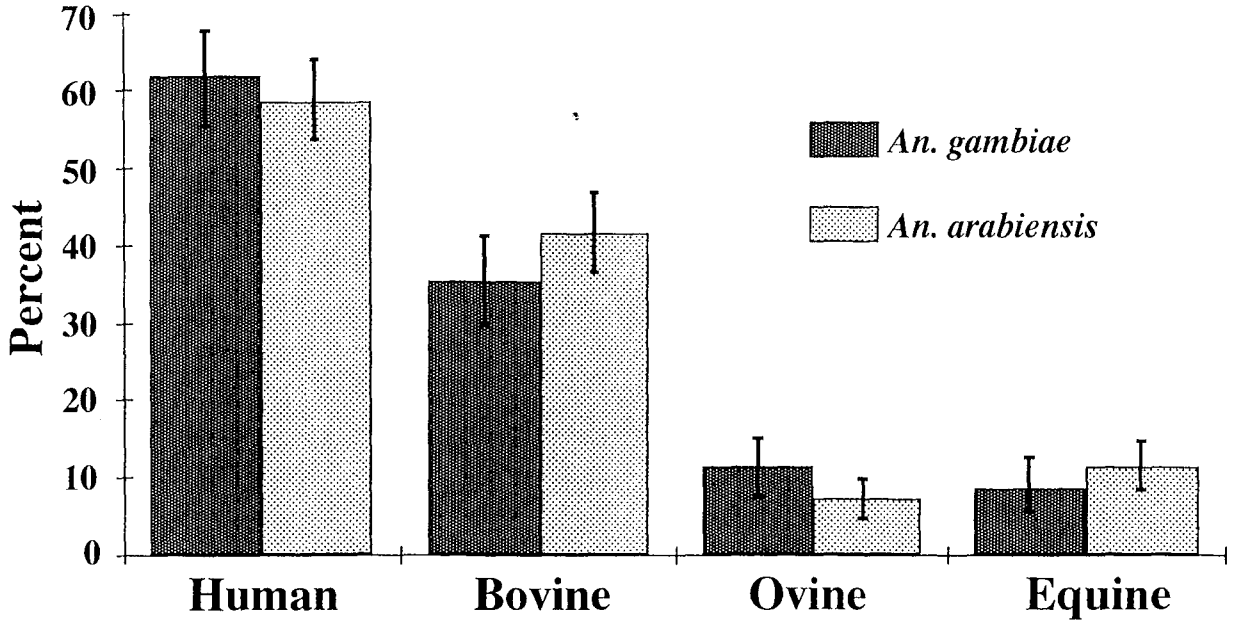


Fig. 4. Percentage of boodmeals taken on different sources for *An. gambiae* and *An. arabiensis* captured by pyrethrum spray in bedrooms during the rainy seasons (July–October) in Barkedji.

and *P. ovale* was detected. A mixed infection of both *P. falciparum* and *P. malariae* was found in 1 *An. gambiae* captured on human bait in August 1995.

Entomological Inoculation Rates. The annual entomological inoculation rate was 128 during 1994 and 100 during 1995 (Table 4). Transmission took place mainly from August to December (Fig. 5). In 1994 *An. gambiae* was responsible for 53% of *P. falciparum* transmission, but only for 15% in 1995. However, *An. gambiae* was the major vector during the 1st mo of transmission during both years, and was responsible for 70% of transmission in September 1994 and 82% in August 1995. In 1994, transmission reached its peak in October with 40 infecting bites per human during the first 15 d of the month. In 1995 this peak was reached at the end of September with 37 infecting bites per human during the last 15 d.

Discussion

Although *An. funestus* is a mosquito that may be found in certain regions with a Sahelian climate, none were captured in Barkedji. High peripheral vegetation, which is favorable for oviposition by this species, does not grow in the temporary pools around Barkedji.

During 1995–1996, anophelines were captured until March 1996, whereas in 1994–1995 none were found after January 1995 (Fig. 2). The explanation for this could be the extended duration of temporary pools after the rains in December 1995. However, only *An. arabiensis* was found after December. The fact that no anophelines were captured during the dry season raised the question of how these mosquitoes survive from one rainy season to another. Did some females remain in hibernation or did recolonization take place at the beginning of each rainy season? Omer and Cloudsley-Thompson (1970) demonstrated that a few autochtonous females remained during the dry season. Taylor et al. (1993), studying gene frequencies of *An. arabiensis* in various villages in Mali and Burkina Faso, suggested that populations were maintained continuously, but with seasonal variation.

During 1994 the number of resting anophelines varied greatly among the 16 bedrooms studied. However, the maximum and minimum number of

Table 3. Circumsporozoite protein rates with confidence interval (in brackets) determined by ELISA from head-thoraces of mosquitoes captured on human bait (Barkedji July 1994 to March 1996)

Mosquito species	1st yr				2nd yr			
	No. tested	P.f.	P.m.	P.o.	No tested	P.f.	P.m.	P.o.
An. gambiae	1,065	4.5 (3.5–5.9)	0.09 (0.5–0)	0	327	2.8 (1.3–5.1)	0.3 (1.7–0)	0.3 (1.7–0)
An. arabiensis	2,998	1.3 (0.9–1.8)	0.03 (0.18–0)	0	3,302	1.3 (0.9–1.8)	0.03 (0.17–0)	0

P.f., Infection rate for *P. falciparum*; P.m., Infection rate for *P. malariae*; P.o., Infection rate for *P. ovale*.

Table 4. Annual entomological inoculation rates estimated by ELISA, for the 3 *Plasmodium* species, for each vector species (Barkedji July 1994 to March 1996)

Species	1st yr			2nd yr		
	P.f.	P.m.	P.o.	P.f.	P.m.	P.o.
An. gambiae	86.7	1.25	0	14.6	1.4	1.1
An. arabiensis	38.7	1.90	0	82.1	2.0	0
Total	125.4	3.15	0	96.8	3.4	1.1

P.f., *P. falciparum*; P.m., *P. malariae*; P.o., *P. ovale*. One *An. gambiae* specimen captured in August 1995 was positive for *P. falciparum* and *P. malariae*.

specimens always were captured in the same bedrooms. We were unable to explain this disparity by factors such as distance from breeding sites, number of persons per bedroom, and number of animals in the surroundings. Significant differences also were observed within the same compound, where all other parameters were equal. The type of housing construction did not affect the results, contrary to what was observed elsewhere (Gamage-Mendis et al. 1991, Lindsay et al. 1995), because the 16 houses selected were identical.

The ratio of *An. gambiae* to *An. arabiensis* was higher among resting than among host-seeking collections. *An. gambiae* therefore seemed to have a higher degree of endophily than *An. arabiensis*. However, the rate of endophagy was not statistically different, and the night biting cycles were similar for *An. gambiae* and *An. arabiensis*.

The study of vector blood feeding behavior should be approached cautiously. Comparisons between *An. gambiae* and *An. arabiensis* should only be made during the rainy season when the 2 species were present simultaneously. From November onward, *An. arabiensis* became significantly more abundant than *An. gambiae*, and at the same time there was a change in host accessibility and in the proportion of bloodmeals on humans and animals. During the rainy season, all animals spend the night around the houses, except some of the horses and donkeys, which remained outside the village in the farming camps (i.e., the most accessible hosts were humans and cattle). However, at the end of the rainy season cattle also were moved to surrounding pasture, and the farmers returned to the village with their horses and donkeys. This explained the increase in human and equine bloodmeals, and the decrease in cattle bloodmeals for *An. arabiensis* after the rainy season. Comparison between human and bovine blood indices calculated during the same time did not show any significant differences between *An. gambiae* and *An. arabiensis*, contrary to what was observed in Nigeria (White and Rosen 1973) and Kenya (Petrarca et al. 1991).

The frequency of mixed bloodmeals in this region was at least 17.8%, but probably much higher, because it was not possible to detect the number of bloodmeals taken on 2 different hosts of the same vertebrate species using ELISA. Feeding patterns also were opportunistic, depending on host availability. Likewise, the rate of patent mixed bloodmeals, mainly human–cattle, was the same for *An. gambiae* and *An. arabiensis*.

To our knowledge, a high percentage of mixed bloodmeals rarely has been observed for malaria vectors (Burkot et al. 1988). Beier et al. (1988) showed that it was possible to identify host blood

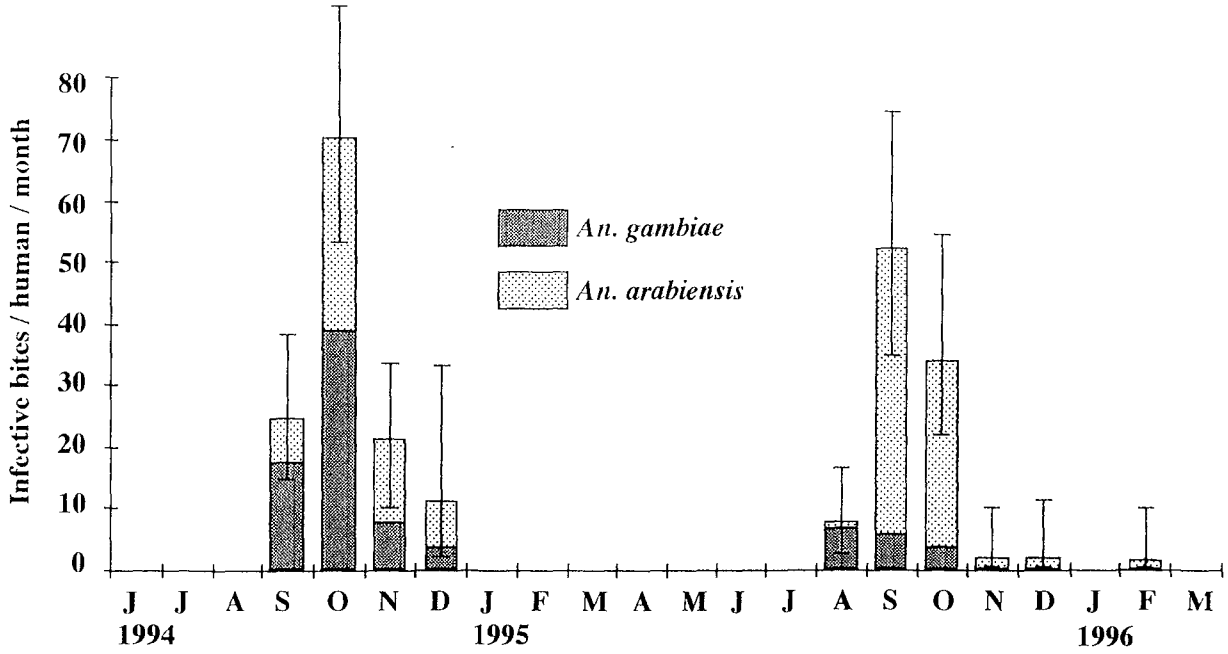


Fig. 5. Monthly entomological inoculation rate for *An. gambiae* and *An. arabiensis* in Barkedji from June 1994 to March 1996.

by ELISA up to 30 h after blood feeding. In Barkedji it was not possible to know if mixed bloodmeals were caused by the interrupted bloodmeals during the same night or multiple feeding on several nights during a single gonotrophic cycle. It was shown experimentally that *An. gambiae* females may take >1 bloodmeals within 6–24 h (Briegel and Hörler 1993). Such multiple feedings also were observed in the field (Beier 1996).

Despite a lower circumsporozoite protein rate, *An. arabiensis* was responsible for 63% of malaria transmission, because it was more abundant than *An. gambiae* and remained active within the village for a long time. To explain the higher circumsporozoite protein rate of *An. gambiae*, the following 3 hypothesis are proposed: (1) *An. gambiae* has a higher parity rate than *An. arabiensis* as observed during 1994–1995. However, data collected during 1995–1996 showed that the parity rate of *An. gambiae* was significantly lower than *An. arabiensis*, but there was no difference in the circumsporozoite protein rate, (2) The feeding frequency differed between the 2 species. If *An. gambiae* fed more frequently than *An. arabiensis*, its circumsporozoite protein rate might be elevated because of more host contacts, (3) Genetic susceptibility to *Plasmodium* differed. Differences may exist between genetically characterized populations within the *An. gambiae* complex, as reported in Kenya by Petrarca and Beier (1992). Differences in genetic susceptibility to *Plasmodium* may account for the higher circumsporozoite protein rate in *An. gambiae* than *An. arabiensis*.

Acknowledgments

We thank Mamoudou Diallo, Henri Manga, and Sadjo Sow for their technical assistance, Frank Collins (Centers for Disease Control and Prevention) for providing the *Plasmodium* monoclonal antibodies and for advice on PCR, Pauline Roussilhon for help in the English translation, Andre Spiegel, Christophe Rogier, and 2 anonymous reviewers for very helpful suggestions, and the villagers in Barkedji for their cooperation throughout the survey. This work was supported by the Institut Français de Recherche Scientifique pour le Développement en Coopération (ORSTOM).

References Cited

Beier, J. C. 1996. Frequent blood-feeding and restrictive sugar-feeding behavior enhance the malaria vector potential of *Anopheles gambiae* s.l. and *An. funestus* (Diptera: Culicidae) in western Kenya. J. Med. Entomol. 33: 613–618.

Beier, J. C., P. V. Perkins, R. A. Wirtz, J. Koros, D. Diggs, T.P.I.I. Gargan, and D. K. Koech. 1988. Bloodmeal identification by direct enzyme-linked immunosorbent assay (ELISA) tested on *Anopheles* (Diptera: Culicidae) in Kenya. J. Med. Entomol. 25: 9–16.

Briegel, H., and E. Hörler. 1993. Multiple blood meals as a reproductive strategy in *Anopheles* (Diptera: Culicidae). J. Med. Entomol. 30: 975–985.

Burkot, T. R., J. L. Williams, and I. Schneider. 1984. Identification of *Plasmodium falciparum* infected mosquitoes by a double antibody enzyme-linked immunosorbent assay. Am. J. Trop. Med. Hyg. 33: 783–788.

Burkot, T. R., P. M. Graves, R. Paru, and M. Lagog. 1988. Mixed blood feeding by the malaria vectors in the *Anopheles punctulatus* complex (Diptera: Culicidae). J. Med. Entomol. 25: 205–213.

Coluzzi, M. 1984. Heterogeneities of the malaria vectorial system in tropical Africa and their significance in malaria epidemiology and control. Bull. W.H.O. 62: 107–113 (suppl.).

Detinova, T. S. 1962. Age grouping methods in Diptera of medical importance, with special reference to some vectors of malaria. World Health Organization, Geneva.

Faye, O., D. Fontenille, J. P. Herve, P. A. Diak, S. Diallo, and J. Mouchet. 1993. Le paludisme en zone sahélienne du Sénégal. 1-Données entomologiques sur la transmission. Ann. Soc. Belge Med. Trop. 73: 21–30.

Fontenille, D., O. Faye, L. Konate, N. Sy, and F. H. Collins. 1993. Comparaison des techniques PCR et cytogénétique pour la détermination des membres du complexe *Anopheles gambiae* au Sénégal. Ann. Parasitol. Hum. Comp. 68: 239–240.

Gamage-Mendis, A. C., R. Carter, C. Mendis, A.P.K. De Zoysa, P.R.J. Herath, and K. N. Mendis. 1991. Clustering of malaria infections within an endemic population: risk of malaria associated with the type of housing construction. Am. J. Trop. Med. Hyg. 45: 77–85.

Garret-Jones, C., and G. R. Shidrawi. 1969. Malaria vectorial capacity of a population of *Anopheles gambiae*. Bull. W.H.O. 40: 531–545.

Gazin, P., V. Robert, M. Cot, J. Simon, J. M. Halna, F. Darriet, D. Legrand, P. Carnevale, and P. Ambroise-Thomas. 1988. Le paludisme dans l'Oudalan, région sahélienne du Burkina Faso. Ann. Soc. Belge Méd. Trop. 68: 255–264.

Gillies, M. T., and B. De Meillon. 1968. The Anophelinae of Africa south of the Sahara, 2nd ed. South African Institute of Medical Research, Johannesburg.

Gillies, M. T., and M. Coetzee. 1987. A supplement to the Anophelinae of Africa south of the Sahara. South African Institute of Medical Research, Johannesburg.

Hamon, J., J. Coz, S. Sales, and C. S. Ouedraogo. 1965. Etudes entomologiques sur la transmission du paludisme humain dans une zone de steppe boisée, la région de Dori (République de Haute-Volta). Bulletin de l'Institut Fondamental d'Afrique Noire Ser. A. 27: 1115–1150.

Haridi, A. M. 1972. Partial exophily of *Anopheles gambiae* species B in the Khashm Elgirba area in eastern Sudan. Bull. W.H.O. 46: 39–46.

Lanzaro, G. C., L. Zheng, Y. T. Toure, S. F. Traore, F. C. Kafatos, and K. D. Vernick. 1995. Microsatellite DNA and isozyme variability in a West African population of *Anopheles gambiae*. Insect Mol. Biol. 4: 105–112.

Lehmann, T., W. A. Hawley, L. Kamau, D. Fontenille, F. Simard, and F. H. Collins. 1996. Genetic differentiation of *Anopheles gambiae* populations from East and West Africa: comparison of microsatellite and allozyme loci. Heredity 77: 192–200.

Lindsay, S. W., P. L. Alonso, J.R.M. Armstrong Schellenberg, J. Hemingway, P. J. Thomas, F. C. Shenton, and B. M. Greenwood. 1993. A malaria control trial using insecticide-treated bed nets and targeted chemoprophylaxis in a rural area of The Gambia, West Africa. 3. Entomological characteristics of the study area. Trans. R. Soc. Trop. Med. Hyg. 87: 19–23 (suppl. 2).

Lindsay, S. W., J.R.M. Armstrong Schellenberg, H. A. Zeiler, R. J. Daly F. M. Salum, and H. A. Wilkins. 1995. Exposure of Gambian children to *Anopheles gambiae* malaria vectors in an irrigated rice production area. Med. Vet. Entomol. 9: 50–58.

Omer, S. M., and J. L. Cloudsley-Thompson. 1970. Survival of female *Anopheles gambiae* Giles through a 9-month dry season in Sudan. Bull. W.H.O. 42: 319–330.

Petrarca, V., and J. C. Beier. 1992. Intraspecific chromosomal polymorphism in the *Anopheles gambiae* complex as a factor affecting malaria transmission in the Kisumu area of Kenya. Am. J. Trop. Med. Hyg. 46 : 229–237.

Petrarca, V., J. C. Beier, F. Onyango, J. Koros, C. Asiago, D. K. Koech, and C. R. Roberts. 1991. Species composition of the *Anopheles gambiae* complex (Diptera: Culicidae) at two sites in Western Kenya. J. Med. Entomol. 28: 307–313.

Robert, V., V. Petrarca, P. Carnevale, L. Ovazza, and M. Coluzzi. 1989. Analyse cytogénétique du complexe *Anopheles gambiae* dans la région de Bobo-Dioulasso (Burkina Faso). Ann. Parasitol. Hum. Comp. 4: 290–311.

Scott, J. A., W. G. Brogdon, and F. H. Collins. 1993. Identification of single specimens of the *Anopheles gambiae* complex by polymerase chain reaction. Am. J. Trop. Med. Hyg. 49: 520–529.

Taylor, K. A., J. K. Koros, J. Nduati, R. S. Copeland, F. H. Collins, and A. D. Brandling-Bennett. 1990. *Plasmodium falciparum* infections rates in *Anopheles gambiae*, *An. arabiensis*, and *An. funestus* in Western Kenya. Am. J. Trop. Med. Hyg. 43 : 124–129.

Taylor, E. C., Y. T. Toure, M. Coluzzi, and V. Petrarca. 1993. Effective population size and persistence of *Anopheles arabiensis* during the dry season in West Africa. Med. Vet. Entomol. 7: 351–357.

Toure, Y. T., S. F. Traore, O. Sankare, M. Y. Sow, A. Coulibaly, F. Esposito, and V. Petrarca. 1996. Perennial transmission of malaria by the *Anopheles gambiae* complex in a North Sudan Savanna area of Mali. Med. Vet. Entomol. 10: 197–199.

Vercruysse, J. 1985. Etude entomologique sur la transmission du paludisme humain dans le bassin du fleuve Sénégal (Sénégal). Ann. Soc. Belge Med. Trop. 65: 171–179 (suppl. 2).

White, G. B., and P. Rosen. 1973. Comparative studies on sibling species of the *Anopheles gambiae* Giles complex (Dipt., Culicidae). II. Ecology of species A and B in savanna around Kadura, Nigeria, during transition from wet to dry season. Bull. ent. Res. 62: 613–625.

WHO. 1963. Terminology of Malaria and of Malaria Eradication. Report of a Drafting Committee. Geneva.

Wirtz, R. A., F. Zavala, Y. Charoenvit, G. H. Campbell, T. R. Burkot, I. Schneider, K. M. Esser, R. L. Beaudoin, and R. G. Andre. 1987. Comparative testing of monoclonal antibodies against *Plasmodium falciparum* sporozoites for ELISA development. Bull. W.H.O. 65: 39–45.

Received for publication 10 June 1996; accepted 3 January 1997

Annexe 4

Simard F, Lehmann T, Lemasson JJ, Diatta M et Fontenille D. Persistence of *Anopheles arabiensis* during severe dry-season conditions in West Africa: an indirect approach using microsatellite loci. En préparation.

Persistence of *Anopheles arabiensis* during the severe dry-season conditions in Senegal: an indirect approach using microsatellite loci.

Frédéric Simard[1], Tovi Lehmann[2], Jean-Jacques Lemasson[1], Mathurin Diatta[1] and Didier Fontenille[3].

1- Laboratoire de Zoologie Médicale, French Institut de Recherche pour le Développement (IRD, former ORSTOM), Institut Pasteur, BP 220, Dakar, Senegal.

2- Division of Parasitic Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia, U.S.A.

3- O.C.E.A.C., BP 288, Yaoundé, Cameroun.

Corresponding author:

Frédéric Simard
Laboratoire de Zoologie Médicale
IRD – Institut Pasteur
BP 220
Dakar
Sénégal
E-mail: frederic.simard@ird.sn

Keywords: *Anopheles arabiensis*, malaria, dry season, estivation, dispersal, microsatellites, effective population size, Senegal.

Running Head: Population genetics of *An. arabiensis* in Senegal.

ABSTRACT

Allelic composition at nine microsatellite loci was investigated in two *An. arabiensis* populations from Senegal, 250 km apart. Eight loci were polymorphic and revealed very similar allele distribution and frequencies in both populations, leading to low estimates of genetic differentiation indices (Fst=0.0115, P=0.013 and Rst=0.0088, P>0.2). Consequently, high amounts of gene flow were determined between these populations (Nm>>1).

Allelic frequencies were monitored in a long dry season location, the village of Barkedji, from August 1994 to December 1997 on a biannual sampling basis (mosquitoes were collected once at the beginning and once at the end of the rainy season each year). Genetic variability remained stable during the time interval under study, despite extreme seasonal fluctuations in density and observation of very low numbers of mosquitoes during the dry season. Accordingly, estimations of the effective population size (Ne) derived from changes in allele frequencies were high, ranging from 209 to infinity and no obvious difference between estimates encompassing the dry or the rainy seasons was detected. When the whole time interval is considered (including 4 rainy seasons and 3 dry seasons), Ne was estimated to range between 275 and 2,465 (95% confidence intervals limits) suggesting a continuous presence of *An. arabiensis* throughout the year in this sahelian location.

These inferences have several implications on *An. arabiensis* bio-ecology and provide information on the genetic structure at a moderate geographical scale within this species that are relevant for actual and future control strategies.

INTRODUCTION

Malaria transmission in sub-Saharan Africa is mostly due to mosquito vectors belonging to the *Anopheles gambiae* complex of species. Among the seven species actually described (Gillies & Coetzee, 1987; Hunt *et al.*, 1998), *An. gambiae* and *An. arabiensis* are the most efficient malaria vectors. In the dry savannas of Africa, those vector populations typically display strong seasonal fluctuations in abundance: while being present in large numbers as soon as the rains start, populations drop to very low levels when temporary breeding sites dry up (Taylor *et al.*, 1993; Charlwood *et al.*, 1995; Lemasson *et al.*, 1997). In some areas, no specimen can be found during the dry season. Nevertheless, soon after the onset of rains, when free water becomes available, those sites are recolonized with remarkable rapidity. It is still unclear how this phenomenon occurs. It has been thought that these populations experience severe bottlenecks or even become locally extinct during the dry months and that, each year, temporary breeding sites are colonized by new immigrating females from neighboring areas, where permanent breeding is allowed. In this case, vector populations do behave as a metapopulation, where geographically and genetically differentiated sub-units arise by migration of founder individuals, grow for a while and eventually become extinct (Levins, 1968; Hanski & Gilpin, 1991). Alternatively, another hypothesis is that large populations do survive locally. As low tolerance of eggs to desiccation has been documented for anophelines (Beier *et al.*, 1990), these populations may be maintained throughout the dry season either by extensive mobility of adults that would take opportunistic advantage of sparsely dispersed breeding sites (e.g. around wells, streams or gardens), or by estivating females in an adaptative state such as gonotrophic dissociation (Omer and Cloudsley-Thompson, 1968, 1970). A better knowledge of the dry season biology and population dynamics of these mosquitoes would be of considerable interest for conventional malaria vector control programs management, but also for the study of parasite panmixia via their vector (Gupta & Day, 1994). These studies are also relevant for guiding the development of new vector control strategies using genetically engineered mosquitoes with altered vector competence (e.g. Collins & Besansky, 1994; Crampton *et al.*, 1994).

In this study, we focused our attention on *An. arabiensis*, a species that typically breeds in such areas of dry sahelian savannas. Two different approaches are commonly used to investigate populations structure and dynamics: a direct (ecological) method that mostly refers to Mark-Release-Recapture experiments (MRR, see for example Costantini *et al.*, 1996; Touré *et al.*, 1998), and an indirect (genetic) method that consider spatial or temporal fluctuations in gene frequencies (Taylor *et al.*, 1993; Lehmann *et al.*, 1998). Both techniques have their own advantages but, indirect methods are more suitable for studying the dry season biology of mosquitoes because of very low population densities at that time, practically hindering application of MRR experiments (Touré *et al.*, 1998).

The indirect approach relies on the simple logic that large stochastic variations in allelic frequencies, mostly due to the action of genetic drift, are expected in small populations, while only minor changes would occur in large ones (Waples, 1991; Taylor *et al.*, 1993). It appears clearly from above that one of the most important parameter to estimate is the effective population size (Ne), a factor that will determine the rate of various evolutionary processes. The effective population size Ne is defined as the size of an ideal population that experiences genetic changes at the same rate as the natural population under consideration (Wright, 1931, 1938; Avise, 1994). Under the appropriate assumptions (listed in Waples 1989a, 1991), most processes caused by genetic drift are inversely proportional to Ne, particularly the amount of changes in gene frequency. Thus, conversely, measuring the extent of these changes and assuming they are due to the action of drift alone (i.e. neglecting the effect of selection, mutation, migration and peculiar reproductive system), would theoretically provide a means to estimate Ne. This indirect temporal approach has been widely used and refined since the first attempt conducted by Krimbas and Tsakas (1971). The present work aims to apply this strategy to improve knowledge about the dry season biology of *An. arabiensis* in the sahelian region of Senegal.

We used microsatellite DNA markers to estimate the level genetic variability within *An. arabiensis* populations, its spatial distribution and to follow its evolution in time. According to the literature, those markers are probably the most appropriate tool currently available to

estimate actual levels of gene flow and describe the genetic structure of natural populations, including members of the *An. gambiae* complex (Bossart & Pashley Prowell, 1998; Lanzaro *et al.*, 1995, 1998; Goldstein *et al.*, 1999; Wang *et al.*, 1999). This particularly refers to intrinsic properties of those markers such as great polymorphism (due to high mutation rates), codominance of alleles and selective neutrality.

Genetic variation at nine microsatellite loci was analyzed in field collected *An. arabiensis* from Senegal. First, the spatial distribution of the genetic diversity was assessed by comparing mosquito populations sampled in two locations, 250 km apart. Then, the temporal method was applied to one of these locations, subject to yearly episodes of drought and flood, that was sampled during four consecutive years. This work was undertaken in order to provide insights to the long debated question of how do mosquito survive the dry season. More precisely, the following questions were assessed: (1) Do *An. arabiensis* form structured (subdivided) populations in Senegal? (2) How different are *An. arabiensis* populations from Senegal and what is the level of gene flow between them? (3) Does the indirect approach provide genetic clues for dry season bottlenecks or extinction of local populations as seems to be observed from ecological studies? and (4) what is the effective population size of *An. arabiensis* in such a fluctuating environment? Conclusions are drawn in terms of populations bio-ecology and vector control management.

MATERIALS AND METHODS

Study sites and sampling :

The village of Barkedji (15°17'N, 14°53'W), 700 inhabitants, is situated in the fossil valley river bed of the Ferlo, in the Sahelian region of Senegal (Figure 1). The rainy season is short, extending from July to October. Rainfall varies from one year to another ranging from 200 to 400 mm per year. Barkedji is surrounded by clay hollows which collect water as soon as the rains start and dry up around January. These temporary pounds are the only anopheline breeding sites encountered in the area. *An. gambiae* and *An. arabiensis* are very common in the wet season. In the dry season, numbers of mosquitoes drop to very low levels and become impossible to find. They reappear after the onset of rains, when conditions again become favorable. At this time, population growth is very

high and vector density rapidly increases. Population dynamics and ecology of these vectors in this location have been extensively described elsewhere (Lemasson *et al.*, 1997).

Female specimens of the *Anopheles gambiae* complex were collected after indoor pyrethrum spraying early in the morning. The same bedrooms were sampled in Barkedji from August 1994 to December 1997. Each year, collections were conducted twice: in August, when temporary breeding sites get flooded and in December, just before they dry out.

Using the same collection technique, an additional location was also sampled in August 1996, for within-Senegal geographical comparison purposes. Mosquitoes were collected in the village of Dielmo (13°45'N, 16°25'W) located in the Sahelo-Soudanian region of Senegal called Saloum, 250 km south-west to Barkedji (Figure1). Detailed entomological study and precise description of this site have been published recently (Fontenille *et al.*, 1997). Malaria transmission is due to *An. arabiensis*, *An. gambiae* and *An. funestus*. Unlike Barkedji, Dielmo is located on the marshy bank of a small permanent stream which permits the persistence of anophelines larval development sites and malaria transmission all year round.

All mosquitoes were stored individually in numbered tubes with desiccant for laboratory processing in Dakar.

Microsatellite Genotype scoring:

DNA was extracted from a leg or a wing of each individual mosquito using the technique described by Collins and other (1987). Only *An. arabiensis* specimens were included in the analysis, after species identification was carried out by the diagnostic PCR described by Scott *et al.* (1993).

Nine microsatellite loci (24D, 147, 141, 26 and 803 on chromosome II; 45C and 29C on chromosome III; and 49 and 7 on chromosome X) were analyzed as previously described (Simard *et al.*, 1999), except that PCR reactions were performed in 12.5 µl instead of 25 µl. PCR products were loaded on 10% non denaturing polyacrylamide gels and the allelic bands were visualized after rapid silver staining. Multilocus genotypes of 50 specimens per sampling date and location were scored.

Data analysis:

For each population and sampling date, we calculated the allelic

frequencies, the mean number of alleles per locus, the observed and expected heterozygosity under Hardy-Weinberg's equilibrium using BIOSYS-1 software (Swofford & Selander, 1989). Tests for deviations from Hardy-Weinberg expectations and for linkage disequilibrium between loci were computed using exact tests available in GENEPOP 3.1 (Raymond & Rousset, 1995a). The sequential Bonferroni procedure was applied to evaluate significance when multiple tests were performed (Holm, 1979). Exact tests were also used to perform pairwise comparisons of allelic frequencies distribution at individual locus between two consecutive sampling dates in the Barkedji population. To quantify the association between allele frequencies from one date to the other, we computed Pearson's correlation coefficients between frequencies of each allele observed in the sample at those dates bounding the time interval under study.

Genetic differentiation among the Barkedji population at each sampling date and between geographic isolates sampled in August 1996 was examined by F statistics (Wright, 1978), calculated according to Weir and Cockerham (1984), and their microsatellite's equivalent R statistics (Slatkin, 1995). Estimates of Fst and Rst were computed using the ARLEQUIN 1.1 software package (available at http://anthropologie.unige.ch/arlequin/), developed by Schneider and collaborators (1997). Both Fst and Rst were tested for statistical significance by permuting individual genotypes among populations. Estimates of the gene exchange index, Nm, were derived from Fst and Rst according to equations 15b and 15a in Slatkin (1995), respectively. Nm values derived from loci on the X-chromosome, were adjusted for their lower effective population size (assuming three-fourth of that of the autosomes).

Estimation of the effective population size:

We estimated Ne of the *An. arabiensis* population from Barkedji based on temporal variations in allelic frequencies. Waples (1989a), according to Nei and Tajima (1981), recommended the use of two different ways of calculation, depending on the sampling scheme: samples for genetic analyses may have been taken before reproduction and not replaced (sampling plan II) or after reproduction (sampling plan I) and therefore contribute (or not) to the genetic composition of the next generation. Because specimens were sampled from a putative very large

population, and not replaced, we followed the methods drawn for sampling prior to reproduction, sampling plan II. Thus, estimates of Ne were obtained using equation 11 in Waples (1989a):

$$Ne = \frac{t}{2\left(F - \frac{1}{2S_0} - \frac{1}{2S_t}\right)}$$

where S_0 and S_t represent sample sizes (number of individuals) at generation 0 and t, respectively; and F estimates the standardized variance of allele frequency change. Several methods of computing F have been proposed (reviewed in Waples, 1989a). These methods generally lead to very similar results (Waples, 1989a; Taylor *et al.*, 1993; Miller & Kapuscinski, 1997; Lehmann *et al.*, 1998). We also verify this accordance, therefore, only results based on F_k (Pollak, 1983) will be presented. Pollak's estimator for the change in allele frequency from one time period to another at one locus is:

$$F_k = \frac{1}{K-1}\sum_{i=1}^{K}\frac{(x_i - y_i)^2}{(x_i + y_i)/2}$$

where K is the number of alleles, and x_i and y_i represent the frequency of allele *i* at generation 0 and t, respectively. Waples (1989a) advocated that extreme allele frequencies could introduce a bias in the estimation of F_k leading to an overestimation of Ne. Therefore, alleles with frequencies lower than 2% at both time points of the considered interval were pooled into one class (Lehmann *et al.*, 1998). For estimation of F_k over all loci, we computed weighted means of single-locus values as:

$$F_{kall} = \sum (K_j - 1)\, Fk_j \,/\, \sum (K_j - 1)$$

where *j* stands for the different loci. This weighted mean was then used to estimate Ne based on all loci information combination. Calculation of the 95% confidence intervals (CI) followed equation 16 of Waples (1989a).

Like in previous studies (Taylor *et al.*, 1993; Lehmann *et al.*, 1998), we conservatively assumed that twelve (discrete) generations occurred yearly for *An. arabiensis* in Barkedji. Estimates of Ne are also given for the total time period (between August 1994 and December 1997) assuming 4 generations between August and December (rainy season) and only 2 generations during the dry season (between December and August), based on the hypothesis that populations are effectively maintained by estivating females.

RESULTS:

Microsatellite loci revealed highly polymorphic in all the samples tested, except locus 49 that behaves in a

non-canonical manner in *An. arabiensis* and was therefore discard from further analysis. Within-populations genetic diversity parameters are summarized in Table 1, for each locus and overall. The number of distinct alleles per locus varied greatly from one locus to the other (from 3 at locus 29C up to 16 at locus 26) but the averaged values over all polymorphic loci were comparable in all the samples tested (Table 1). Allele frequencies were estimated from genotypes scored at each locus for each individual (available upon request to corresponding author). Their distribution was very similar in both locations studied, as well as at the different time points within the Barkedji population. This suggests spatial and temporal homogeneity of the distribution of the genetic diversity within *An. arabiensis* in Senegal.

Spatial comparison between two populations collected in August 1996:

Most of the loci conformed to Hardy-Weinberg proportions in both populations collected in August 1996. Nevertheless, significant deviations were observed after the sequential Bonferroni procedure was applied to take into account the other tests at locus 45C in the Barkedji population and at loci 24D and 141 in the Dielmo population (Table1). Deviation at locus 45C was associated with heterozygote excess (negative Fis value) while a significant excess of homozygotes was detected at loci 24D and 141 in the Dielmo population. No significant association between loci was detected by linkage disequilibrium analyses, neither in Barkedji nor in Dielmo, suggesting random association of gametes within each population and physical independence of the loci (data not shown). When the exact test was performed on the pooled data, two significant associations were evidenced out of 28 tests (for loci 24D and 141, and 45C and 147, both probabilities being $< 10^{-4}$). This result is consistent with the deviations from Hardy-Weinberg proportions observed at those loci. Therefore, no obvious indication of subdivision of the gene pool between Barkedji and Dielmo populations was evidenced.

Accordingly, single locus Fst and Rst values were very low and non significant leading to very high Nm estimates (Table 2). All single locus differentiation indices were homogenous across the eight loci ($Fst \leq 0.022$ and $Rst \leq 0.056$) indicating that the pattern observed reflects genome-wide characteristics. When averaged over all loci, Fst estimate was

0.0115 and was statistically significant (P=0.013) and Rst value was 0.0088 but was not significant (P>0.2). Derived Nm values were 21.4 and 28.1 (from Fst and Rst values, respectively) suggesting high amounts of gene flow between *An. arabiensis* populations in Senegal.

Temporal evolution in Barkedji:

No obvious deviation from Hardy-Weinberg proportions was detected, indicating no evidence for subdivision of the gene pool in Barkedji. Nevertheless, some casual deviations were observed once at loci 26, 45C and 147, and in three out of eight time points at locus 24D (Table 1). While single deviations (i.e. those that arise at one particular locus at one time point) could be due to the effect of chance alone, the pattern observed for locus 24D is more likely the result of the presence of null alleles, as significant excess homozygosity appears in three out of eight samples tested. Moreover, these unique deviations were not strengthen by other loci and should reflect some local constraints inherent to the microsatellite locus itself, rather than genetic sub-structuring of the population. The presence of null alleles has also been suspected for locus 147 in former studies (Kamau *et al.*, 1998; Walton *et al.*, 1998; Simard *et al.*, 1999).

The temporal evolution of the average observed number of alleles per locus (Nall), average observed heterozygosity (Hobs) and average expected heterozygosity (Hexp) over all loci is shown in Figure 2 for the Barkedji population. All these parameters that describe the amount of genetic variability within one particular population (or sample) appeared relatively stable during the time period under study (see Figure 2). Paired-samples Wilcoxon signed-ranks tests (between two consecutive time points for each locus) were not significant for each of the variable tested (df = 7, P>0.07 for both the observed and expected heterozygosities and P>0.27 for the number of alleles). Nevertheless, when testing differences in heterozygosity between samples from August 1994 and December 1997, the results were marginally significant (df = 7, P = 0.0499 in both cases). Thus, a slight decrease in heterozygosity was evidenced in the Barkedji population between August 1994 (64%) and December 1997 (57%). This decline suggests that one or more episode(s) of low effective population size occurred in Barkedji during the time period under study.

Similarly, allele frequencies distribution appeared very stable between temporal samples, as shown in Figure 3. No significant difference was observed over the whole dataset for loci 147, 141, 803, 45C and 29C (Fisher's exact test: P>0.16). Nevertheless, loci 24D, 26 and 7 showed significant differences in allele frequencies (P<0.001) between samples, particularly when the collection of August 1994 was included in pairwise comparisons. The frequencies of individuals alleles for all loci were highly correlated between temporal samples, with pairwise Pearson's correlation coefficients ranging from 0.962 to 0.987 (61<n<64) between two consecutive dates of sampling and 0.955 (n=62) between August 1994 and December 1997.

Accordingly, such stability of allelic frequencies resulted in very low levels of genetic differentiation estimated when comparing each time point to the others, with pairwise Fst estimates ranging between –0.0027 and 0.0190 and equivalent Rst between –0.0102 and 0.0117. Eight out of 28 pairwise Fst estimates but none of the Rst estimates were significant, based on the permutation test (Table 3). Waples (1989b) pointed out that classical statistics such as the exact test of homogeneity of allele (or genotype) frequencies or Fst and Rst that have been designed to investigate genetic differentiation between local populations should reveal biased when temporally spaced samples of the same population are compared, because genetic drift alone is expected to produce changes in allele distribution and frequencies. Therefore, assumptions of independence between samples and the island model of migration are not respected and these tests are expected to reveal statistically significant differences at a rate much higher than the 5% assumed. This trend was indeed observed in our data as 8 out of 28 pairwise Fst values were significant while none of the Rst values, even between the more extreme time points. This pattern suggests that mutation does not play a important role in the evolution of the genetic variability within the *An. arabiensis* population from Barkedji at the time scale under study, and that the differences observed are essentially due to a genetic drift component.

Estimation of the effective populations size in Barkedji:

Estimates of Ne based on Fk calculations are shown in Table 4 for each interval as well as for the total time period, together with their 95%

confidence intervals. Due to the very similar distributions of allelic frequencies among time points, numerous estimates suggest infinite Ne and all but one single locus estimates are unbounded upwards. Combining the information across the eight loci yielded similar results, the 95% CI being definable in only one case out of seven. Nevertheless, when estimated across the entire time interval (i.e. between August 1994 and December 1997), Ne estimates proved to be much more consistent and bounded 95% CI was obtained when all loci were considered.

Ne estimates were also computed assuming a minimum of two discrete generations between December and August each year, in an attempt to take into account eventual estivation of the females that would survive throughout the dry season without laying any egg batch. The resulting Ne value we obtained was very similar to our previous estimate assuming 12 generations a year (575 and 1,000, respectively), and the 95%CIs substantially overlapped (Table 4). The ratio of Ne estimates obtained this way (1,000/575=1.74) is identical with the ratio of the numbers of generations assumed to be encompassed by the time interval studied (40/23=1.74), as t (the estimated time length, expressed in number of generations elapsed between both dates bounding the interval) appears alone in the numerator of the equation used to estimate Ne. The lowest and highest bounds of the 95%CI of Ne obtained for the total time length were 275 and 2,465 respectively. Consecutively, Ne was estimated to range from several hundreds to a few thousand individuals during this four years time period in Barkedji and no evidence for strong and significant population's dry season bottlenecks was revealed.

DISCUSSION

In this study, we investigated spatial and temporal variations in the allelic composition at eight microsatellite loci of *An. arabiensis* populations in Senegal. Despite a relatively high geographical distance between two collection sites (250 km) and very strong seasonal variations in density observed in a long dry season habitat in the sahelian area of this part of Africa, allele frequencies and distributions revealed very similar at this geographical scale. Longitudinal sampling in the village of Barkedji allowed for determination of the average effective population size that was shown to range in the hundreds up to a few thousands individuals. Such

high values, derived from the quantification of allelic frequencies changes between time points, do not support the hypotheses of recurrent extinction of this vectorial population during the dry season each year. Indeed, allelic frequencies remained very stable in the time interval under study suggesting continuous presence of *An. arabiensis* throughout the year in this location. These results obtained by an indirect approach aiming to resolve the populations genetic structure of *An. arabiensis* in Senegal have several implications on the biology and ecology of this species in this area that will now be assessed.

At the geographical level first, two *An. arabiensis* populations from Senegal revealed very low levels of genetic differentiation suggesting high rates of gene flow over distances as much as 250 km. Indeed, no significant hindrance to dispersal seems to occur in such an open environment where local populations could therefore exchange migrants at a high rate. This part of central to northern Senegal is totally devoid of geographical disruption that could represent any barrier to individual dispersal as no mountains nor even hills are encountered along the distance between both sampling locations. Vegetation mostly consists of sparsely dispersed shrubby trees and baobabs that do not cluster to form dense forests. Thus, mosquito dispersal is likely to be important, or at least unrestricted by environmental barriers. Mark-Release-Recapture (MRR) experiments have suggested that active dispersal in this mosquito complex of species is limited to a few kilometers (Gillies, 1961; Costantini *et al.*, 1996; Touré *et al.*, 1998) but other evidences prompt that these values could be underestimated and that passive migration, augmented by transportation due to human activities and/or wind could play a significant role in anophelines spread over large geographical distances (Service, 1993, 1997; Lehmann *et al.*, 1996a; Simard *et al.*, 1999). In addition, Costantini *et al.* (1996) reported dispersal capacity estimates as high as 350-650 m per day per adult mosquito, a value that provides support for an extensive mobility of individuals. It is worth remembering that in the sahelian area of Africa where dry seasons are long and yearly drought are severe, MRR experiments are only applicable during the rainy season, when population densities are high. It has been noted that one important parameter driving members of the *An. gambiae* complex dispersal seems to be the availability of breeding sites and, to a

lower extent, of hosts (Thomson *et al.*, 1995; Costantini *et al.*, 1996; Touré *et al.*, 1998). Therefore, as breeding sites dry up during the dry season, it is likely that some switch in mosquitoes behavior towards higher propensity for dispersion would occur. As underlined by Touré *et al.* (1998), direct MRR results should not be extrapolated to other areas and/or time periods. Such experiments are actually under way in the Dielmo area.

Alternatively, the similarity of allele distribution evidenced between the Barkedji and Dielmo *An. arabiensis* populations could reflect retention of ancestral polymorphism and range expansion of the species in the recent past (Coluzzi *et al.*, 1985; Lehmann *et al.*, 1996a, 1997; Besansky *et al.*, 1997). However, at the geographical scale explored here, the hypothesis of high contemporary amounts of gene flow seems a more believable explanation as the resolving power of the eight microsatellite markers we used has been demonstrated in a former study comparing four widely isolated *An. arabiensis* populations from continental Africa and offshore islands (Simard *et al.*, 1999). Then, *An. arabiensis* seems to form a rather homogenous gene pool in this area, local populations being interconnected by extensive levels of gene flow.

Temporal stability of the genetic diversity was revealed in Barkedji despite extreme seasonal demographic changes and observation of very low numbers of mosquitoes during the dry season (Lemasson *et al.*, 1997). Indeed, relatively high average effective population size estimates have been obtained for the *An. arabiensis* population of Barkedji during our four years longitudinal sampling. As recently underlined by Tessier and Bernatchez (1999), this result does not necessarily imply that variations in abundance had no impact on the genetic integrity of the population but, instead, strengthens the thinking that temporal changes in intrapopulation genetic composition are not easily depicted unless effective population size drops to a very low level or changes are monitored over a large number of generations (Waples 1989a, 1991; Waples & Teel, 1990; Richards & Leberg, 1996; Miller & Kapuscinski, 1997). No evidence for lower effective population size in dry season versus rainy season was obtained as a consequence of overlapping 95% confidence intervals of these estimates. This leads to the inference that *An. arabiensis* is probably maintained continuously in the Barkedji area, a

conclusion already reached by Taylor *et al.* (1993) in other parts of Western Africa (i.e. Mali, Nigeria and Burkina Faso) and by Lehmann *et al.* (1998) for *An. gambiae* (savanna cytotype) in Kenya.

We should address here the possible bias introduced in the way we estimated Ne, in order to give insight to the accuracy of these estimates. According to Waples (1989a, 1991), the temporal method assumes selective neutrality of the genetic markers used, random sampling of a homogenous gene pool and negligible mutation and migration. Microsatellite markers are generally considered as selectively neutral, being devoid of known biological function except for rare cases where they were shown to correlate with genetic diseases in humans (Brahmachari *et al.*, 1995), they are usually found in non-coding regions of the genome. Nevertheless, some constraints on allele size, existence of null alleles or biased mutation rates have been evidenced for these markers (e.g. Bowcock *et al.*, 1994; Garza *et al.*, 1995; Lehmann *et al.*, 1996b) and could have introduced a bias in our Ne estimates to an unknown extent. It is noteworthy, however, that all these constraints are locus-specific and would not affect all the loci in the same way. A better understanding of the mutation processes involved in microsatellite evolution would allow for a more accurate assessment of the selective neutrality of these markers. No evidence for subdivision of the gene pool was evidenced in Barkedji, both by Hardy-Weinberg proportions test and by linkage disequilibrium studies suggesting that the assumption of homogeneity of the gene pool is met in *An. arabiensis* from this location. It also seems reasonable to assume that mutations do not have any significant effect on the genetic variability of the population at the time scale investigated (40 generations), a trend confirmed by the lack of significant Rst estimates between temporal samples. The assumption of negligible migration probably does not hold, as suggested by the high migration index revealed between Barkedji and Dielmo. Nevertheless, the effect of immigration may be limited because allele frequencies and distributions appeared very similar at all the loci we studied between these two sampling sites. However, this statement needs further consideration and a more detailed study of the pattern of gene flow at the microgeographic level in *An. arabiensis*, including more populations should be devised. Studies addressing

this issue are actually in progress in the Dielmo area. We have seen previously that the number of estimated discrete generations spanned by the time period under study is a parameter of paramount importance for the estimation of Ne. We gave two estimates of the average population size of *An. arabiensis* in Barkedji considering either 40 or 23 discrete generations, depending on biological assumptions (see results section). These values are likely to represent the lowest estimates of the number of generations elapsed during this time period as it has been shown that one generation of *An. arabiensis* could be obtained in no more than 16 days under laboratory-controlled ideal conditions. Thus, our estimates of Ne are conservative with regard to this parameter, as increasing the number of generations would lead to higher estimates of Ne, therefore corroborating our inferences of large population size. Generations are also assumed to be discrete, a condition that is probably not met in mosquito natural populations, but the temporal method is robust to overlapping generations, provided that the time period under study encompasses more than one generation (Waples, 1989a; Jorde & Ryman, 1995). Nevertheless, this could have introduced a bias in our estimations of Ne for "seasonal" intervals, but estimates based on the total time period should be less affected by deviations from this assumption.

However, an important question is just how large a geographical area should be associated with this estimate of Ne? As advocated in the first part of this discussion, high amounts of gene flow occur between populations as far as 250 km apart. Therefore, the geographical area associated with the deme (i.e. the basic reproductive unit within which mating is random and panmixia is respected) is obviously large. Thus, large populations scattered over a wide geographical area will result in low density of adults within a locale during the dry season. This phenomenon could reconcile the findings of very low dry-season densities observed by direct methods (i.e. entomological sampling) and high effective population size estimated by indirect methods (i.e. genetic analysis). In such a "diffused deme" (Lehmann *et al.*, 1998), large populations could be maintained locally, either through female mosquitoes surviving the dry season in a physiologically altered state such as gonotrophic dissociation (Omer and Cloudsley-Thompson, 1968, 1970) or by permanent breeding in a few dispersed sites and extensive mobility

of adults. However, our field investigations did not allow for observation of any specimen despite extensive search in potential estivation sites such as those described by Omer and Cloudsley-Thompson in Sudan (i.e. rodent burrows, trees holes, human dwellings or dry wells) nor for identification of permanent breeding sites in this area. In contrast, we obtained direct evidence of the effective presence of *An. arabiensis* in Barkedji during the dry season by collecting couples of females seeking for bloodmeals during fields investigations conducted in March 1996 and March 1997, when no breeding sites were available in the area (F. Simard and D. Fontenille, unpublished data). Unfortunately, the very low samples sizes obtained precluded genetic analysis. Were these individuals opportunistic immigrants from adjacent locations where permanent breeding is allowed or estivating females in an altered state is still an open question and further ecological studies are required. Whatever the major mechanism of dry season surviving, it appears clearly that *An. arabiensis* populations are not clustered to geographically restricted areas. The high level of genetic diversity observed in Barkedji and its temporal stability together with low levels of genetic differentiation between remote locations evidenced in this area strengthens the view that long-distance dispersal is common in *An. arabiensis* (Costantini *et al.*, 1996), a pattern that could have considerable impact on the genetic composition and level of geographical spread of genes within this species (Slatkin, 1987).

This ability to occupy space in a continuous manner together with other behavioral components such as an easy switch to alternative feeding on non-human hosts (Gillies and De Meillon, 1968; Chauvet and Rajaonarivelo, 1973; Gillies and Coetzee, 1987) make *An. arabiensis* a difficult target for actually available control methods essentially based on local insecticide treatments. With this respect, insecticide campaigns management requires much caution to avoid selection and rapid spread of insecticide resistance genes within *An. arabiensis* natural populations. If confirmed by other studies in different ecological conditions, this unrestricted pattern of gene flow on large geographical scales would appear as a favorable element for implementation of new means of struggle such as the release of transgenic mosquitoes with altered vector competence (Collins and Besansky, 1994; Crampton *et al.*, 1994).

Acknowledgements:

We are grateful to Sadjo SOW and all the inhabitants of Barkedji and Dielmo for their help in collecting mosquitoes. We thank Habibatou Diallo-Agne for technical assistance in the laboratory.

Financial support:

This work was supported by the French Institut de Recherche pour le Développement (IRD), as part of grant n° CNRS 96/C/08. Frederic Simard gratefully acknowledge financial support from Mr François LACOSTE.

References:

Avise JC, 1994. Molecular markers, natural history and evolution. ed Chapman and Hall, London.

Beier JC, Copeland R, Oyaro C, Masinya A, Odago WO, Oduor S, Koech DK and Roberts CR, 1990. *Anopheles gambiae* complex egg-stage survival in dry soil from larval development sites in western Kenya. *J. Am. Mosq. Control Assoc.* **6(1)**: 105-109.

Besansky NJ, Lehmann T, Fahey GT, Fontenille D, Braack LEO, Hawley WA and Collins FH, 1997. Patterns of mitochondrial variation within and between African malaria vectors, *Anopheles gambiae* and *An. arabiensis*, suggest extensive gene flow. *Genetics* **147**: 1817-1828.

Bossart JL and Pashley Prowell D, 1998. Genetic estimates of population structure and gene flow: limitations, lessons and new directions. *Trends Ecol. Evol.* **13(5)**: 202-206.

Bowcock AM, Ruiz Linares A, Tomfohrde J, Minch E, Kidd JR and Cavalli-Sforza LL, 1994. High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* **368**: 455-457.

Brahmachari SK, Meera G, Sarkar PS, Balagurumoorty P, Tripathi J, Raghavan S, Shaligram U and Pataskar S, 1995. Simple repetitive sequences in the genome: structure and functional significance. *Electrophoresis* **16(9)**: 1705-1714.

Charlwood JD, Kihonda J, Sama S, Billingsley PF, Heiz B and Takken W, 1995. The rise and fall of *Anopheles arabiensis* (Diptera: Culicidae) in a Tanzanian village. *Bull. Ent. Res.* **85**: 37-44.

Chauvet G and Rajaonarivelo E, 1973. "Modification de comportement" d'une espèce et variation de la composition d'une population appartenant au complexe *Anopheles gambiae* pendant et après les pulvérisations d'insecticides domiciliaires dans les environs de Tananarive (Madagascar). *Cah. O.R.S.T.O.M., Sér. Ent. Méd. et Parasitol.* **11(3)**: 155-167.

Collins FH and Besansky NJ, 1994. Vector biology and control of malaria in Africa. *Science* **264**: 1874-1875.

Collins FH, Mendez MA, Rasmussen MO, Mehaffey PC, Besansky NJ and Finnerty V, 1987. A ribosomal RNA gene probe differentiates member species of *Anopheles gambiae* complex. *Am. J. Trop. Med. Hyg.* **37**: 37-41.

Coluzzi M, Petrarca V and Di Deco MA, 1985. Chromosomal inversion intergradation and incipient speciation in *Anopheles gambiae. Boll. Zool.* **52**: 45-63.

Costantini C, Li SG, Della Torre A, Sagnon N'F, Coluzzi M and Taylor CE, 1996. Density, survival and dispersal of *Anopheles gambiae* complex mosquitoes in a West African Sudan savanna village. *Med. Vet. Entomol.* **10**: 203-219.

Crampton JM, Warren A, Lycett GJ, Hughes MA, Comley IP and Eggleston P, 1994. Genetic manipulation of insect vectors as a strategy for the control of vector-borne disease. *Ann. Trop. Med. Parasitol.* **88**: 3-12.

Fontenille D, Lochouarn L, Diagne N, Sokhna C, Lemasson JJ, Diatta M, Konate L, Faye F, Rogier C and Trape JF, 1997. High annual and seasonal variations in malaria transmission by anophelines and vector species composition in Dielmo, a holoendemic area in Senegal. *Am. J. Trop. Med. Hyg.* **56(3)**: 247-253.

Garza JC, Slatkin M and Freimer NB, 1995. Microsatellite allele frequencies in Humans and Chimpanzees, with implications for constraints on allele size. *Mol. Biol. Evol.* **12(4)**: 594-603.

Gillies MT and De Meillon B, 1968. The Anophelinae of Africa south of the Sahara. *Pub. South Afr. Inst. Med. Res.* **54**: 343 pp.

Gillies MT and Coetzee M, 1987. A supplement to the Anophelinae of Africa south of the Sahara. 2 nd Edition. *Pub. South Afr. Inst. Med. Res.* **55**: 143 pp.

Goldstein DB, Roemer GW, Smith DB, Reich DE, Bergman A and Wayne R, 1999. The use of microsatellite variation to infer population structure and demographic history in a natural model system. *Genetics* **151**: 797-801.

Gupta S and Day KP, 1994. A strain theory of malaria transmission. *Parasitology Today* **10** : 476-481.

Hanski I and Gilpin M, 1991. Metapopulation dynamics: brief history and conceptual domain. *Biological Journal of the Linnean Society* **42** : 3-16.

Holm S, 1979. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* **6**: 65-70.

Hunt RH, Coetzee M and Fettene M, 1998. The *Anopheles gambiae* complex: a new species from Ethiopia. *Trans. R. Soc. Trop. Med. Hyg.* **92**: 231-235.

Jorde PE and Ryman N, 1995. Demographic genetics of brown trout (*Salmo trutta*) and estimation of effective population size from temporal change in allele frequencies. *Genetics* **143**: 1369-1381.

Kamau L, Lehmann T, Hawley WA, Orago ASS and Collins FH, 1998. Microgeographic genetic differentiation of *Anopheles gambiae* mosquitoes from Asembo Bay, western Kenya : a comparison with Kilifi in coastal Kenya. *Am. J. Trop. Med. Hyg.* **58(1)** : 64-69.

Kimmel M, Chakraborty R, Stivers DN and Deka R, 1996. Dynamics of repeat polymorphism under a forward-backward mutation model: within- and between-population variability at microsatellite loci. *Genetics* **143**: 549-555.

Krimbas CB and Tsakas S, 1971. The genetics of *Dacus oleae* V. Changes of esterase polymorphism in a natural population following insecticide control: selection or drift? *Evolution* **25**: 454-460.

Lanzaro GC, Toure YT, Carnahan J, Zheng L, Dolo G, Traore S, Petrarca V, Vernick KD and Taylor CE, 1998. Complexities in the genetic structure of *Anopheles gambiae* populations in west Africa as revealed by microsatellite DNA analysis. *Proc. Natl. Acad. Sci. USA* **95**: 14260-14265.

Lanzaro GC, Zheng L, Toure YT, Traore SF, Kafatos FC and Vernick KD, 1995. Microsatellite DNA and isozyme variability in a West African population of *Anopheles gambiae*. *Insect Mol. Biol.* **4**: 105-112.

Lehmann T, Hawley WA, Kamau L, Fontenille D, Simard F and Collins FH, 1996a. Genetic differentiation of *Anopheles gambiae* populations from East and West Africa: a comparison of microsatellite and allozyme loci. *Heredity* **77**: 192-208.

Lehmann T, Hawley WA and Collins FH, 1996b. An evaluation of evolutionary constraints on microsatellite loci using null alleles. *Genetics* **144**: 1155-1163.

Lehmann T, Besansky NJ, Hawley WA, Fahey GT, Kamau L and Collins FH, 1997. Microgeographic structure of *Anopheles gambiae* in western Kenya based on mtDNA and microsatellite. *Mol. Ecol.* **6**: 243-253.

Lehmann T, Hawley WA, Grebert H and Collins FH, 1998. The effective population size of *Anopheles gambiae* in Kenya: implications for population structure. *Mol. Biol. Evol.* **15(3)**: 264-276.

Lemasson JJ, Fontenille D, Lochouarn L, Dia I, Simard F, Ba K, Diop A, Diatta M and Molez JF, 1997. Comparison of behaviour and vector efficiency of *Anopheles gambiae* and *An. arabiensis* (Diptera: Culicidae) in Barkedji, a sahelian area of Senegal. *J. Med. Entomol.* **34(4)**: 396-403.

Levins R, 1968. Evolution in changing environments. Princeton, Princeton University Press.

Miller LM and Kapuscinski AR, 1997. Historical analysis of genetic variation reveals low effective population size in a Northern Pike (*Esox lucius*) population. *Genetics* **147**: 1249-1258.

Nei M, 1978. Estimation of the average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**: 583-590.

Nei M and Tajima F, 1981. Genetic drift and estimation of effective population size. *Genetics* **98**: 625-640.

Omer SM and Cloudsley-Thompson JL, 1968. Dry season biology of *Anopheles gambiae* Giles in the Sudan. *Nature* **217**: 879-880.

Omer SM and Cloudsley-Thompson JL, 1970. Survival of female *Anopheles gambiae* Giles through a 9-month dry season in Sudan. *Bull. Wld. Hlth. Org.* **42**: 319-330.

Pollak E, 1983. A new method for estimating effective population size from allele frequency changes. *Genetics* **104**: 531-548.

Raymond M and Rousset F, 1995a. GENEPOP Version 1.2 A population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**: 248-249.

Raymond M and Rousset F, 1995b. An exact test for population differentiation. *Evolution* **49**: 1280-1283.

Richards C and Leberg PL, 1996. Temporal changes in allele frequencies and a population's history of severe bottlenecks. *Conservation Biology* **10**: 832-839.

Schneider S, Kueffer JM, Roessli D and Excoffier L, 1997. ARLEQUIN, a software package for population genetics. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, Switzerland.

Scott JA, Brogdon WG and Collins FH, 1993. Identification of single specimens of *Anopheles gambiae* complex by the polymerase chain reaction. *Am. J. Trop. Med. Hyg.* **49**: 520-529.

Service MW, 1997. Mosquito (Diptera: Culicidae) dispersal – the long and short of it. *J. Med. Entomol.* **34(6)**: 579-588.

Service MW, 1993. Mosquito Ecology: Field Sampling Methods, 2nd edition. Elsevier Applied Science Publishers. London.

Simard F, Fontenille D, Lehmann T, Girod R, Brutus L, Gopaul R, Dournon C and Collins FH, 1999. High amounts of genetic differentiation between populations of the malaria vector *Anopheles arabiensis* from West Africa and Eastern Outer Islands. *Am. J. Trop. Med. Hyg.* **60(1)**: 1000-1009.

Slatkin M, 1987. Gene flow and the geographical structure of natural populations. *Science* **236**: 787-792.

Slatkin M, 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**: 457-462.

Swofford DL and Selander RB, 1989. BIOSYS-1. A computer program for the analysis of allelic variation in population genetics and biochemical systematics. Release 1.7. University of Illinois, Urbana, IL.

Taylor CE, Toure YT, Coluzzi M and Petrarca V, 1993. Effective population size and persistence of *Anopheles arabiensis* during the dry season in West Africa. *Med. Vet. Entomol.* **7**: 351-357.

Tessier N and Bernatchez L, 1999. Stability of population structure and genetic diversity across generations assessed by microsatellites among sympatric populations of landlocked Atlantic salmon (*Salmo salar* L). *Molecular Ecology* **8**: 169-179.

Thomson MC, Connor SJ, Quinones ML, Jawara M, Todd J and Greenwood BM, 1995. Movement of Anopheles gambiae s.l. malaria vectors between village in the Gambia. *Med. Vet. Entomol.* **9**: 413-419.

Touré YT, Dolo G, Petrarca V, Traore SF, Bouare M, Dao A, Carnahan J and Taylor CE, 1998. Mark-release-recapture experiments with *Anopheles gambiae* s.l. in Banambani village, Mali, to determine population size and structure. *Med. Vet. Entomol.* **12**: 74-83.

Walton C, Thelwell NJ, Priestman A and Butlin RK, 1998. The use of microsatellites to study gene flow in natural populations of Anopheles malaria vectors in Africa: potential and pitfalls. *J. Am. Mosq. Control Assoc.* **14(3)**: 266-272.

Waples RS, 1989a. A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics* **121**: 379-391.

Waples RS, 1989b. Temporal variation in allelic frequencies: testing the right hypothesis. *Evolution* **43(6)**: 1236-1251.

Waples RS and Teel DJ, 1990. Conservation biology of Pacific salmon. I. Temporal changes in allele frequency. *Conservation Biology* **4**: 144-156.

Waples RS, 1991. Genetic methods for estimating the effective population size of cetacean populations. PP 279-300 in Hoelzel AR, ed. Genetic ecology of whales and dolphins. International Whaling Commission (Special Issue n°13), Cambridge, UK.

Weir BS and Cockerham CC, 1984. Estimating F-statistics for the analysis of population structure. *Evolution* **38**: 1358-1370.

Wright S, 1978. Evolution and genetics of populations, Volume 4. Variability within and among natural populations. University of Chicago Press, Chicago, IL.

Wright S, 1931. Evolution in Mendelian populations. *Genetics* **16**: 97-159.

Wright S, 1938. Size of populations and breeding structure in relation to evolution. *Science* **87**: 430-431.

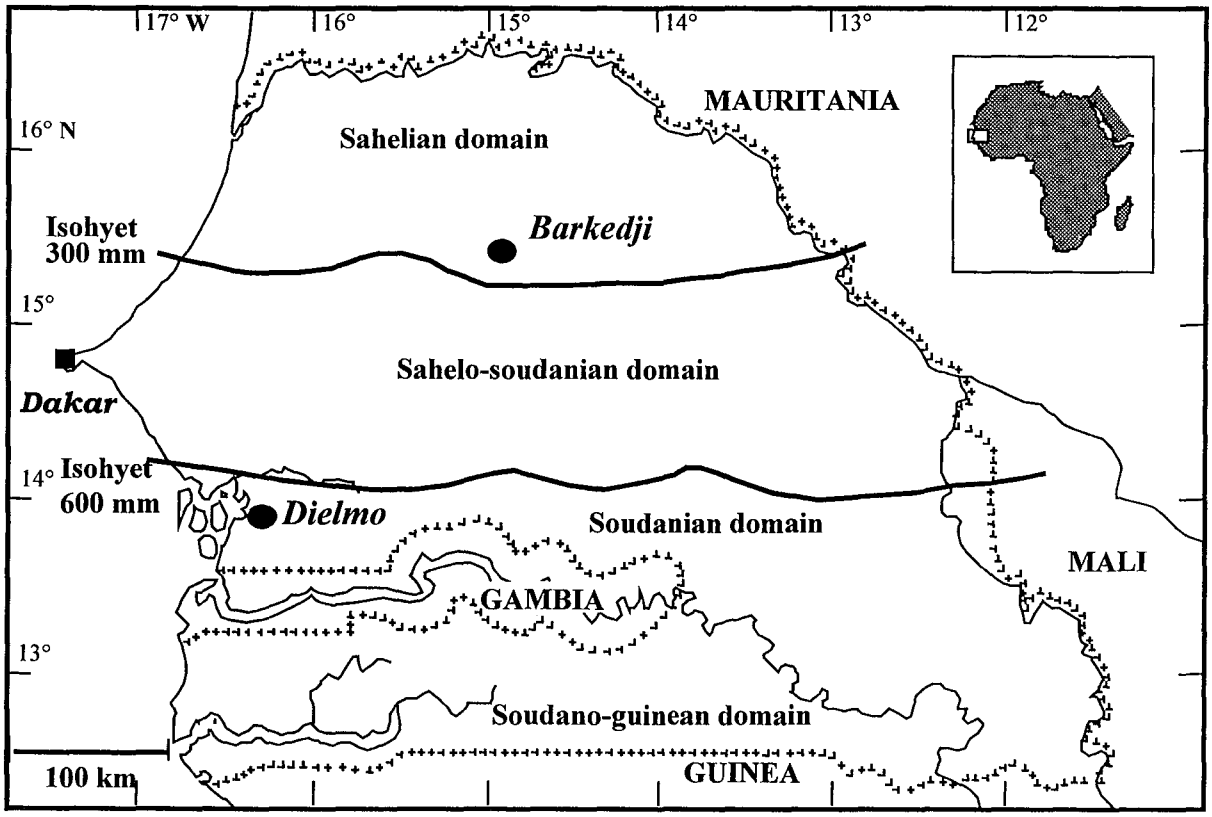

17° W
16°
15°
14°
13°
12°
16° N
15°
14°
13°
MAURITANIA
Sahelian domain
Isohyet
300 mm
Barkedji
Sahelo-soudanian domain
Dakar
Isohyet
600 mm
Dielmo
Soudanian domain
MALI
GAMBIA
Soudano-guinean domain
GUINEA
100 km

Figure 1

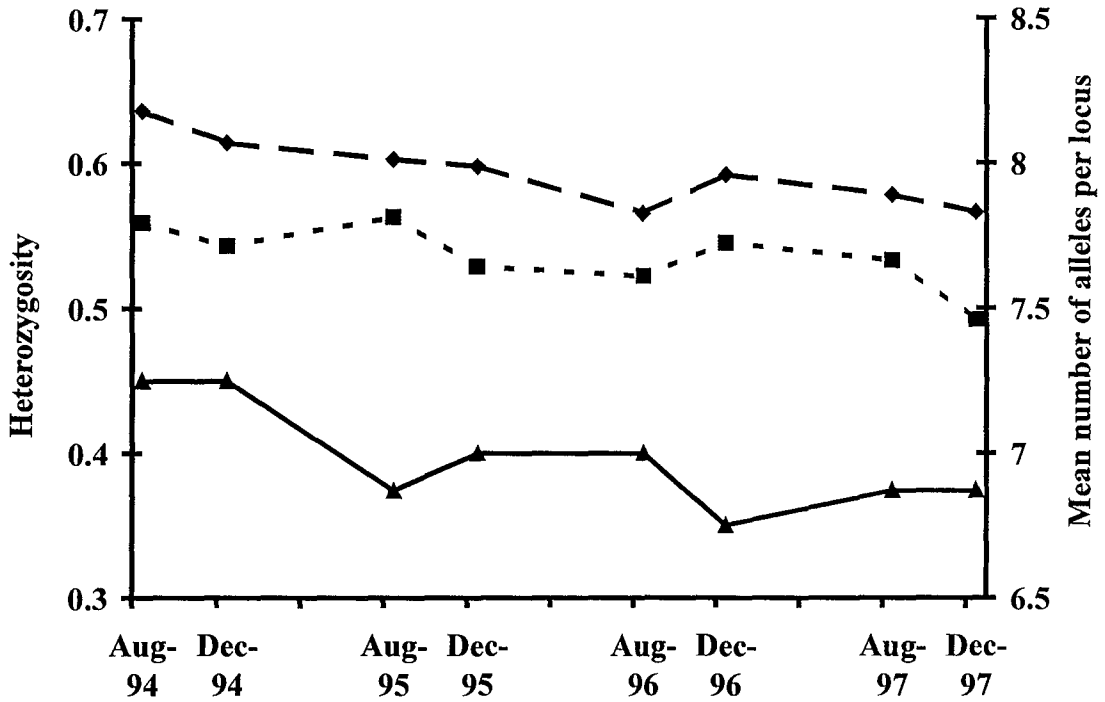

0.7
0.6
0.5
0.4
0.3
Heterozygosity
8.5
8
7.5
7
6.5
Mean number of alleles per locus
Aug-94
Dec-94
Aug-95
Dec-95
Aug-96
Dec-96
Aug-97
Dec-97

Figure 2

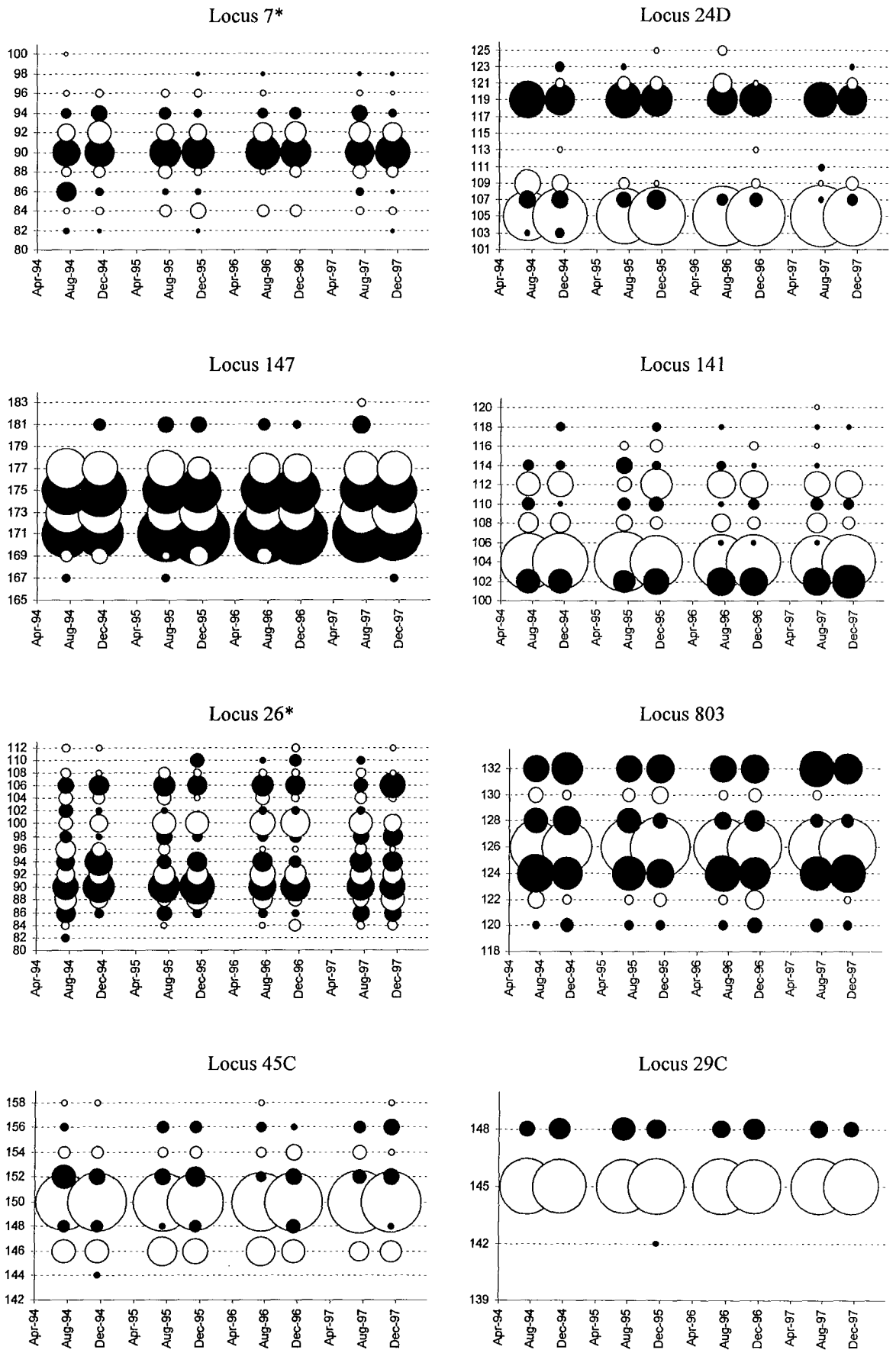

Locus 7*
Locus 24D
Locus 147
Locus 141
Locus 26*
Locus 803
Locus 45C
Locus 29C
Apr-94
Aug-94
Dec-94
Apr-95
Aug-95
Dec-95
Apr-96
Aug-96
Dec-96
Apr-97
Aug-97
Dec-97

Figure 3

Figure Legends :

Figure 1: Geographic location of the study sites in Senegal, the villages of Barkedji and Dielmo.

Figure 2: Temporal evolution of the mean observed heterozygosity (- ■ -), mean expected heterozygosity under Hardy-Weinberg equilibrium (—♦—) and mean number of alleles per locus (—▲—) averaged over all 8 polymorphic loci, in the *An. arabiensis* population from Barkedji.

Figure 3: Schematic representation of relative allelic frequencies observed at 8 microsatellite loci in the *An. arabiensis* population from Barkedji, at 8 consecutive sampling dates from August 1994 to December 1997 (X axis). "April" months stand as dry season indicators. Each chart represents data from one locus, named in the upper part. Alleles are denoted by their total size in base pairs on the Y axis. The surface area of each circle is directly proportional to the frequency of the allele in the sample at the sampling date considered. *: surface area of circles has been halved when too many alleles were present at one locus to optimize clarity.

Table 1: Genetic variability within *An. arabiensis* populations from Barkedji (eight time points) and Dielmo (one time point).

Population			Locus								Mean
			7	24D	147	141	26	803	45C	29C	over loci
		Nall	9	5	7	6	15	7	7	2	7.25
	Aug 94	H obs	0.604	0.481	0.667	0.519	0.833	0.660	0.537	0.167	0.558
		H exp	0.749	0.648	0.773	0.594	0.915	0.714	0.539	0.154	0.636
		Fis	+0.195	**+0.259**	+0.138	+0.128	+0.090	+0.076	+0.004	-0.082	0.122
		Nall	8	8	6	7	13	7	7	2	7.25
	Dec 94	H obs	0.600	0.380	0.680	0.460	0.826	0.680	0.460	0.260	0.543
		H exp	0.722	0.577	0.768	0.583	0.861	0.736	0.435	0.228	0.614
		Fis	+0.170	**+0.343**	+0.115	+0.213	+0.041	+0.077	-0.057	-0.140	0.117
		Nall	7	6	7	7	13	7	6	2	6.87
	Aug 95	H obs	0.680	0.560	0.520	0.420	0.841	0.660	0.520	0.300	0.563
		H exp	0.685	0.561	0.757	0.535	0.876	0.674	0.477	0.258	0.603
B		Fis	+0.007	+0.001	**+0.315**	+0.216	+0.041	+0.021	-0.090	-0.167	0.068
A		Nall	9	6	6	8	11	7	6	3	7
	Dec 95	H obs	0.640	0.460	0.580	0.540	0.694	0.600	0.520	0.200	0.529
R		H exp	0.654	0.531	0.712	0.687	0.846	0.625	0.511	0.216	0.598
		Fis	+0.021	+0.134	+0.186	+0.216	+0.181	+0.041	-0.018	+0.073	0.115
K											
		Nall	8	5	6	8	14	7	6	2	7
E	Aug 96	H obs	0.620	0.360	0.660	0.520	0.813	0.600	0.460	0.140	0.522
		H exp	0.587	0.489	0.719	0.636	0.884	0.621	0.419	0.165	0.565
D		Fis	-0.057	+0.265	+0.083	+0.184	+0.081	+0.035	**-0.098**	+0.155	0.078
J		Nall	5	6	5	8	15	7	6	2	6.75
	Dec 96	H obs	0.640	0.360	0.680	0.500	0.800	0.640	0.460	0.280	0.545
I		H exp	0.668	0.464	0.705	0.619	0.874	0.709	0.451	0.243	0.592
		Fis	+0.042	+0.226	+0.036	+0.193	+0.085	+0.098	-0.021	-0.153	0.079
		Nall	8	5	5	10	14	6	5	2	6.87
	Aug 97	H obs	0.776	0.400	0.540	0.551	0.837	0.640	0.340	0.180	0.533
		H exp	0.738	0.413	0.724	0.689	0.899	0.660	0.337	0.165	0.578
		Fis	-0.051	+0.033	+0.256	+0.202	+0.070	+0.031	-0.010	-0.089	0.079
		Nall	9	6	6	6	13	6	7	2	6.87
	Dec 97	H obs	0.500	0.300	0.700	0.520	0.735	0.600	0.420	0.160	0.492
		H exp	0.573	0.485	0.754	0.630	0.890	0.624	0.423	0.149	0.566
		Fis	+0.128	**+0.384**	+0.072	+0.176	**+0.176**	+0.039	+0.008	-0.077	0.132
Mean		Nall	7.87	5.87	6	7.5	13.5	6.75	6.25	2.12	7
over		H obs	0.645	0.423	0.636	0.490	0.803	0.624	0.438	0.210	0.533
sampling		H exp	0.669	0.545	0.731	0.631	0.887	0.661	0.426	0.196	0.593
dates		Fis	**+0.061**	**+0.211**	**+0.155**	**+0.191**	**+0.097**	**+0.054**	-0.035	-0.069	**+0.099**
		Nall	8	7	6	6	15	8	6	2	7.25
DIELMO		H obs	0.653	0.388	0.563	0.469	0.783	0.673	0.429	0.163	0.515
		H exp	0.692	0.587	0.739	0.646	0.856	0.774	0.419	0.151	0.608
		Fis	+0.057	**+0.342**	+0.241	**+0.276**	+0.087	+0.131	-0.022	-0.079	**+0.155**

Nall: number of alleles per locus,
H obs: observed heterozygosity,
H exp: expected heterozygosity (Nei, 1978),
Fis is calculated according to Weir & Cockerham, 1984.
Bolded values: P<0.05.
Bolded underlined values: P<0.05 after application of the sequential Bonferroni procedure (Holm, 1979).

Table 2: Genetic differentiation between *An. arabiensis* populations from Barkedji and Dielmo.

	Fst			Rst		
Locus	Fst	P	Nm	Rst	P	Nm
7 [a]	0.0049	0.467	68	-0.0049	0.608	∞
24D	0.0288	0.145	8.4	-0.0044	0.586	∞
147	-0.0067	0.926	∞	-0.0103	0.998	∞
141	0.0182	0.182	13.5	0.0561	0.054	4.2
26	0.0020	0.695	126.6	0.0043	0.379	58
803	0.0218	0.161	11.2	0.0069	0.304	35.8
45C	0.0095	0.305	26.1	0.0127	0.269	19.5
29C	-0.0098	0.869	∞	-0.0098	0.849	∞
All loci	0.0115	0.013	21.4	0.0088	0.202	28.1

P refers to the statistical significance of Fst and Rst estimates, based on 1000 permutations procedure (Schneider *et al.*, 1996).

[a] Nm values for locus 7 (on the X chromosomes) are adjusted assuming that Ne for the X chromosome is three-fourth that for the autososmes.

Table 3: Pairwise Fst (below diagonal) and Rst (above diagonal) among temporal samples of the An. arabiensis population from Barkedji.

	August 1994	December 1994	August 1995	December 1995	August 1996	December1996	August 1997	December 1997
August 1994	******	-0.0056	-0.0093	0.0009	0.0004	0.0117	0.0012	-0.0043
December1994	0.0048	******	-0.0003	-0.0006	-0.0031	0.0074	-0.0090	-0.0102
August 1995	0.0052	0.0017	******	-0.0011	-0.0083	-0.0008	0.0008	-0.0070
December 1995	0.0170*	0.0094*	0.0016	******	-0.0010	0.0083	-0.0032	0.0017
August 1996	0.0190*	0.0090*	0.0006	-0.0028	******	-0.0017	-0.0017	-0.0032
December 1996	0.0149*	0.0047	-0.0013	-0.0025	-0.0027	******	0.0068	0.0031
August 1997	0.0144*	0.0032	0.0043	0.0046	0.0032	-0.0015	******	-0.0056
December 1997	0.0144*	0.0076*	0.0033	0.0031	-0.0021	0.0022	0.0023	******

* : $P<0.05$ Global Fst= 0.0052 ; $P<10^{-4}$; Global Rst= -0.0011 ; P=0.761.

Table 4: Effective population size (Ne) estimates based on Fk computations (Pollak, 1983). Twelve generations a year was assumed.

Time Interval	Locus								All loci
	7	24D	147	141	26	803	45C	29C	
Aug 94 – Dec 94	65 (12 – 384)	144 (17 - ∞)	∞ (58 - ∞)	∞ (47 - ∞)	123 (25 - ∞)	∞ (29 -∞)	∞ (24 - ∞)	506 (1 - ∞)	209 (79 – 2,317)
Dec 94 – Aug 95	816 (59 - ∞)	∞ (75 - ∞)	∞ (59 - ∞)	269 (32 - ∞)	513 (62 - ∞)	∞ (130 - ∞)	993 (41 - ∞)	∞ (1 - ∞)	1,198 (249 - ∞)
Aug 95 – Dec 95	1,032 (36 - ∞)	∞ (55 - ∞)	191 (16 - ∞)	79 (12 - ∞)	182 (30 - ∞)	∞ (38 - ∞)	∞ (87 - ∞)	∞ (3 - ∞)	401 (107 - ∞)
Dec 95 – Aug 96	∞ (141 - ∞)	3,453 (37 - ∞)	∞ (106 - ∞)	988 (59 - ∞)	1,670 (95 - ∞)	∞ (116 - ∞)	441 (33 - ∞)	∞ (9 - ∞)	∞ (414 - ∞)
Aug 96 – Dec96	197 (19 - ∞)	120 (13 - ∞)	18,333 (20 - ∞)	∞ (59 - ∞)	∞ (75 - ∞)	∞ (35 - ∞)	172 (15 - ∞)	438 (1 - ∞)	634 (127 - ∞)
Dec 96 – Aug 97	508 (45 - ∞)	∞ (40 - ∞)	426 (23 - ∞)	∞ (89 - ∞)	682 (79 - ∞)	220 (23 - ∞)	609 (26 - ∞)	767 (1 - ∞)	693 (182 - ∞)
Aug 97 – Dec 97	248 (28 - ∞)	490 (19 - ∞)	∞ (49 - ∞)	735 (21 - ∞)	269 (47 - ∞)	∞ (52 - ∞)	1,545 (23 - ∞)	∞ (2 - ∞)	654 (145 - ∞)
Aug 94 – Dec 97	456 (88 – 1,828)	530 (62 – 6,794)	∞ (601 - ∞)	1,879 (159 - ∞)	1,127 (259 – 209,632)	930 (204 - ∞)	4,758 (1,457 - ∞)	∞ (198 - ∞)	1,000 (479 – 2,465)
Aug 94 – Dec 97†	263 (51 – 1,051)	305 (36 – 3,907)	∞ (1 - ∞)	1,080 (91 - ∞)	648 (149 – 120,539)	535 (117 - ∞)	2,736 (843 - ∞)	∞ (114 - ∞)	575 (275 – 1,417)

95% confidence intervals are shown in parenthesis.
∞ : infinity.
† Ne has been estimated assuming two generations during the December to August dry seasons time intervals.

SOMMAIRE (résumé)

INTRODUCTION **11**
1. La génétique des populations 14
2. Le complexe *Anopheles gambiae* 36
3. Structure génétique et spéciation au sein du complexe *An. gambiae* 49

METHODOLOGIE GENERALE **56**
1. Collecte et identification des moustiques sur le terrain 57
2. Détermination de l'espèce et du génotype 62
3. Analyse des données 67

PREMIERE PARTIE
Polymorphisme des marqueurs microsatellites chez *Anopheles arabiensis*: étude de la répartition spatiale de la variabilité génétique **70**
1. Description des sites d'étude 73
2. Analyse génétique 78
3. Structure génétique des populations d'*An. arabiensis*: le rôle des facteurs écologiques et historiques 86
4. Retour sur les conditions d'application des modèles 93
5. Conclusion 97

DEUXIEME PARTIE
Suivi temporel de la variabilité génétique dans une population d'*Anopheles arabiensis* en zone sahélienne **99**
1. Dynamique des populations d'anophèles à Barkedji 101
2. L'approche temporelle : estimation de la taille effective des populations grâce aux variations de fréquences alléliques 108
3. Application de l'approche temporelle à la population d'*An. arabiensis* de Barkedji 113
4. Structure génétique au Sénégal 128
5. Conclusion 132

TROISIEME PARTIE
Etude de la différenciation génétique entre les formes Mopti et Savane chez *Anopheles gambiae* s.s. **137**
1. Introduction 138
2. Résultats 142
3. Discussion 151
4. Conclusion 160

CONCLUSIONS & PERSPECTIVES **163**
1. Les marqueurs microsatellites et leur utilisation pour l'étude de la génétique des populations d'anophèles 164
2. Structure des populations et flux de gènes au sein du complexe *An. gambiae*: implications sur les moyens de lutte antivectorielle 171

REFERENCES BIBLIOGRAPHIQUES **177**
ANNEXES **198**

UNIVERSITE NANCY 1 BIBLIOTHEQUE SCIENCES

Résumé : Variabilité Génétique et Flux de Gènes chez les Moustiques *Anopheles arabiensis* et *Anopheles gambiae*, Vecteurs du Paludisme en Afrique. Apport des Loci Microsatellites.

Bien connaître les systèmes vectoriels responsables de la transmission de pathologies humaines est d'une importance primordiale pour mieux comprendre l'épidémiologie de la maladie et en optimiser les moyens de contrôle. Dans ce but, nous avons étudié la génétique des populations d'*Anopheles arabiensis* et dans une moindre mesure d'*Anopheles gambiae*, deux espèces jumelles de moustiques responsables de la plus grande part de la transmission du paludisme en Afrique. L'outil moléculaire préconisé a été l'utilisation de marqueurs microsatellites récemment isolés chez *An. gambiae*. Neuf loci polymorphes ont été sélectionnés.
L'analyse de la répartition de la variabilité génétique entre et au sein des populations naturelles d'*An. arabiensis* a été conduite (i) dans l'espace et (ii) dans le temps. Une forte différenciation génétique a été révélée entre populations prélevées au Sénégal, à Madagascar et sur deux îles de l'archipel des Mascareignes, La Réunion et l'île Maurice, mettant en exergue le rôle majeur de facteurs historiques souvent anthropiques (colonisation, campagnes de lutte antivectorielle) et des composantes géographiques (barrières à la dispersion) sur la structuration des populations. Au Sénégal, les populations apparaissent peu différenciées géographiquement. De plus, la composition allélique est très stable dans le temps, en dépit des fortes fluctuations saisonnières en densité observées dans les zones sahéliennes. Chez *An. gambiae*, une très faible différenciation génétique a été révélée entre les cytotypes Mopti et Savane collectés dans une zone de sympatrie en Côte d'Ivoire, alors que l'isolement reproductif semble important.
Le caractère informatif des loci microsatellites et la pertinence des informations qu'ils procurent pour l'étude de la structure génétique des populations au sein de ce système vectoriel sont discutés. Une attention particulière est portée sur les implications de ces résultats dans le domaine de la lutte antivectorielle.

Mots clés: Génétique, population, *Anopheles arabiensis*, *Anopheles gambiae*, paludisme, microsatellites, flux de gènes.

Abstract: Genetic Variability and Gene Flow within the African Malaria Vector Mosquitoes *Anopheles arabiensis* and *Anopheles gambiae*. Use of Microsatellite Loci.

Accurate knowledge of vectorial systems involved in human diseases transmission is of paramount importance for a better understanding of diseases epidemiology and improvement of control means. With respect to this rationale, we studied population genetics of *Anopheles arabiensis* and, to a lesser extent of *An. gambiae*, two mosquito sibling species responsible for most malaria transmission throughout continental Africa. Nine polymorphic microsatellite loci, recently developed in *An. gambiae*, were selected for this purpose.
Analysis of allele distributions among and within *An. arabiensis* natural populations has been conducted (i) at the spatial (geographical) level and (ii) at the temporal level. High amounts of genetic differentiation were revealed between populations collected in Senegal, Madagascar and two islands of the Mascarene archipelago, La Reunion and Mauritius. This result underlined the major contribution of both historical demographic events, most of which are a consequence to human activities (colonization, entomological control), and geographical components acting as barriers to dispersal, in populations genetic structuring. Within Senegal, a very low level of genetic differentiation was observed. In addition, allelic composition appears very stable in a long dry season location sampled during four consecutive years, despite strong seasonal fluctuations in density. Within *An. gambiae*, very low levels of genetic differentiation were observed between sympatric Mopti and Savanna chromosomal forms in Ivory Coast, while reproductive isolation seems likely.
Microsatellites' usefulness and throughput for the study of the genetic structure of this vectorial system are discussed. Particular attention is paid on the outputs of such studies in terms of anti-vectorial control strategies.

Keywords: Genetic, population, *Anopheles arabiensis*, *Anopheles gambiae*, malaria, microsatellites, gene flow.