

Sommaire

I.	INTRODUCTION.....	6
1.	CONTEXTE DE L'ETUDE	6
a.	<i>Biologie du Criquet pèlerin, une espèce ravageuse</i>	6
b.	<i>Le polyphénisme de phase : un trait conditionnant la nuisibilité de l'espèce</i>	7
c.	<i>Deux sous-espèces décrites comme ayant des traits de phase différents</i>	8
2.	APPORT DE LA GENETIQUE DES POPULATIONS ET DEMARCHE SUIVIE LORS DE CE STAGE	8
a.	<i>Une approche comparative des mécanismes de la grégarisation</i>	8
b.	<i>Etude de l'histoire évolutive et des processus neutres</i>	9
c.	<i>Questions posées dans le cadre du stage et démarche</i>	10
II.	BIBLIOGRAPHIE ET HYPOTHESES.....	11
1.	HYPOTHESES DEMOGRAPHIQUES POUR NOTRE ETUDE	11
a.	<i>Histoire paléoclimatique africaine.....</i>	11
b.	<i>Etude de la structuration locale du Criquet pèlerin</i>	11
3.	SPECIFICITES DES MARQUEURS MICROSATELLITES UTILISES	12
a.	<i>Caractéristiques des marqueurs microsatellites des Orthoptères.....</i>	12
b.	<i>Spécificités des marqueurs microsatellites dérivés de séquences transcrites.....</i>	13
III.	MATERIELS ET METHODES.....	14
1.	PRESENTATION DES MARQUEURS UTILISES ET GENOTYPAGE	14
a.	<i>Populations étudiées</i>	14
b.	<i>Génotypage des allèles.....</i>	14
c.	<i>Détermination du taux de mutation des marqueurs génomiques.....</i>	15
2.	DESCRIPTION DE L'« ETAT » GENETIQUE DE L'ESPECE : SIGNATURES DE DESEQUILIBRE, DIVERSITE GENETIQUE ET STRUCTURE GENETIQUE.....	16
a.	<i>Conformité aux équilibres d'Hardy-Weinberg et de liaison.....</i>	16
b.	<i>Test de marqueurs sous sélection.....</i>	17
c.	<i>Diversité au sein des populations et différenciation entre populations.....</i>	17
d.	<i>Analyse d'excès d'hétérozygotie, signature de goulot d'étranglement.....</i>	17
3.	TEST D'ISOLEMENT DEMOGRAPHIQUE DES DEUX SOUS-ESPECES.....	18
4.	ESTIMATION DE L'HISTOIRE EVOLUTIVE NEUTRE DES SOUS-ESPECES : TEMPS DE DIVERGENCE ET TAILLES EFFICACES.....	18
a.	<i>Présentation de l'analyse utilisée</i>	18
b.	<i>Le modèle mutationnel : information « réaliste » des paramètres de nuisance.....</i>	19
c.	<i>Le modèle populationnel : un modèle de divergence des deux sous-espèces.....</i>	20
d.	<i>Analyse de sensibilité du modèle à la variation de taille efficace des populations</i>	20
IV.	RESULTATS.....	22
1.	DESCRIPTION DE LA VARIATION GENETIQUE DES POPULATIONS ET CONSTRUCTION DU JEU DE DONNEES POUR L'ETUDE DE LA DIVERGENCE.....	22
a.	<i>Conformité aux équilibres d'Hardy-Weinberg et de liaison.....</i>	22
b.	<i>Test de marqueurs sous sélection.....</i>	22
c.	<i>Diversité génétique des populations.....</i>	23
d.	<i>Structuration génétique entre populations</i>	23
2.	ESTIMATION DE L'HISTOIRE EVOLUTIVE NEUTRE DES SOUS-ESPECES : TEMPS DE DIVERGENCE ET TAILLES EFFICACES	23
a.	<i>Comparaison des modèles démographiques.....</i>	23
b.	<i>Estimation des valeurs de paramètres pour le modèle démographique retenu.....</i>	24

c.	<i>Information des paramètres mutationnels : confirmation expérimentale des taux de mutation des marqueurs microsatellites utilisés.....</i>	25
V.	DISCUSSION	26
1.	CONCLUSIONS SUR LA STRUCTURATION GENETIQUE	26
a.	<i>Absence de structuration au sein des sous-espèces</i>	26
b.	<i>Une différenciation entre sous-espèces probablement sous-estimée.....</i>	26
2.	CONCLUSIONS SUR LE MODELE ET LES PARAMETRES DEMOGRAPHIQUES	26
3.	DEUX SOURCES DE BIAIS POTENTIELLES DANS NOS RESULTATS	27
a.	<i>La non prise en compte de la migration</i>	27
b.	<i>Des résultats sensibles au modèle mutationnel</i>	27
4.	UNE ESTIMATION COHERENTE DU TAUX DE MUTATION DE NOS MARQUEURS	28
5.	UNE MEILLEURE CONNAISSANCE DES PROCESSUS NEUTRES POUR CALIBRER LES ETUDES SUR LES TRAITS LIES A LA PHASE	28
6.	PERSPECTIVES METHODOLOGIQUES POUR MIEUX INFORMER L'HISTOIRE EVOLUTIVE DES DEUX SOUS-ESPECES DE <i>S. G. GREGARIA</i>	28
a.	<i>Intégrer dans l'analyse des marqueurs génétiques qui informent sur des temps anciens</i>	28
b.	<i>Ajouter une population plus divergente dans l'analyse.....</i>	29

Liste des annexes :

Annexe 1 : Marqueurs génomiques (a) et transcriptomiques (b) développés par l'équipe du CIRAD (les statistiques résumées ont été mesurées à partir de 84 individus répartis dans trois populations de Mauritanie).	II
Annexe 2 : Protocole de la PCR utilisée dans notre étude :	IV
Annexe 3 : Conditions d'élevage isolées du criquet pèlerin d'où sont issus les larves utilisées dans l'estimation du taux de mutation	V
Annexe 4 : Grandes lignes du protocole d'extraction ADN en plaque (©MAX).....	VI
Annexe 5 : Simulations de l'analyse pour tester la présence de migration (Chaîne de Monte-Carlo du logiciel 2MOD).....	VII
Annexe 6 : Distribution des fréquences alléliques pour le marqueur diEST12 entre les deux sous-espèces.....	VIII
Annexe 7 : Distribution des fréquences alléliques des marqueurs comparées entre les deux sous-espèces (en rouge S.g.gregaria et en vert S.g.flaviventris)	VIII
Annexe 8 : Statistiques résumées pour les marqueurs et les populations étudiées de S. g. flaviventris	X
Annexe 9 : Hétérozygotie attendue pour les marqueurs et les populations étudiées de S. g. gregaria	XI
Annexe 10 : Richesse allélique pour les marqueurs et les populations étudiées de S. g. gregaria (pour 9 individus)	XII
Annexe 11 : Matrice des Fst observés par paires pour nos 28 populations (en bas) et des intervalles de confiance (en haut). Une étoile correspond à un résultat significatif.	XIII
Annexe 12 : Posterior des paramètres en sortie de DIYABC pour le modèle retenu avec variation de la taille efficace suite à la divergence	XIV
Annexe 13 : Profil GeneMapper des « pools » de 4 individus où l'on suspecte une mutation sérieuse (les deux premières lignes correspondent aux profils parentaux et la troisième au profil du « pool »)	XVII

Tableau 1 : Taxonomie du criquet pèlerin, *Schistocerca gregaria* (FAO Guidelines 2003)

CLASSE	INSECTES	
ORDRE	ORTHOPTÈRES	Criquets et sauterelles (environ 20 000 espèces dans le monde)
SOUS-ORDRE	CAELIFÈRES	Criquets, locustes et sauteriaux (antennes courtes) (environ 10 000 espèces dans le monde)
SUPER-FAMILLE	ACRIDOIDEA	
FAMILLE	ACRIDIDAE	Acridiens ou criquets
SOUS-FAMILLE	CYRTACANTHACRIDINAE	
GENRE	<i>Schistocerca</i>	
ESPECE	<i>gregaria</i>	



Figure 1 : Photo d'un adulte grégaire mature du criquet pèlerin *Schistocerca gregaria* (A. Monard)



Figure 2 : Carte de répartition de l'espèce *Schistocerca gregaria* (Waloff, 1972 ; CPR, 1982)

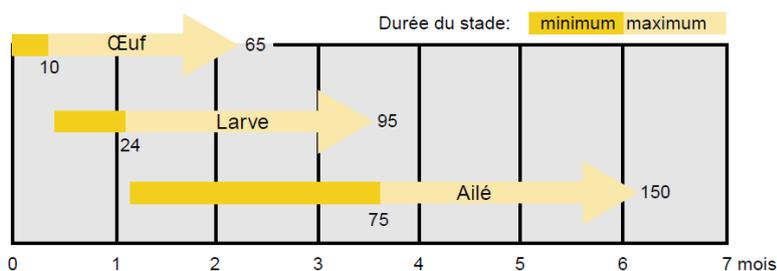


Figure 3 : Table de vie du criquet pèlerin (*Schistocerca gregaria gregaria*) (en nombre de jours) (FAO Guidelines, 2003)

I. Introduction

1. Contexte de l'étude

Parmi les ravageurs des cultures, les criquets ont toujours représenté une menace importante pour la production agricole mondiale. A l'heure actuelle, sur tous les continents, la gestion des populations de criquets est une préoccupation de tous les jours. Le Criquet pèlerin, en particulier, requiert une surveillance permanente puisqu'il peut présenter des explosions de population, ce qui le rend assez redoutable.

a. Biologie du Criquet pèlerin, une espèce ravageuse

Le Criquet pèlerin, *Schistocerca gregaria* (Forskål, 1775)(Fig.1), est un insecte de l'ordre des Orthoptères (Tableau 1) dont l'aire de répartition principale (en période de rémission, c'est-à-dire hors pullulation) s'étend dans des zones arides et semi-arides se situant en l'Afrique de l'Ouest (Sénégal, Mauritanie, Mali, Niger, Tchad), au Maghreb (Maroc, Algérie), en Afrique du Nord (Libye, Egypte), en Afrique de l'Est (Soudan, Ethiopie, Somalie, Erythrée, Kenya) mais aussi au Proche-Orient (Arabie saoudite, Yémen principalement), jusqu'en Asie du Sud-Ouest (Pakistan, Inde) (Fig.2 ; (Waloff 1972; (COPR) 1982). Par ailleurs, en Afrique australe, il existe des populations géographiquement isolées qui ne posent pas de problème en agriculture (Botha 1969).

La durée de développement du criquet peut s'étendre de 2 à 6 mois (Fig.3) et passe par trois stades : l'œuf, la larve (ou nymphe) et l'imago. Seul l'adulte après la mue imaginale possède des ailes fonctionnelles. Entre temps, les larves passent par cinq ou six stades. Au stade adulte, les individus sont d'abord sexuellement immatures et la période de maturation peut aller de 3 semaines à 9 mois. C'est sous cette forme que l'espèce est capable de rentrer en quiescence afin de supporter les environnements défavorables (milieu très aride ou température basse). La femelle peut pondre de 80 à 160 œufs par oothèque et en moyenne chacune d'entre elles pond deux oothèques en milieu naturel (Symmons et Cressman 2001). Dans les conditions optimales de température et d'habitat, il a été mesuré qu'une femelle peut produire de 16 à 20 criquets par génération. De ce fait, même si la prédation (insectes prédateurs et parasites, champignons entomophages, etc...) est assez efficace dans la régulation démographique, il est possible d'observer une augmentation très rapide, voire une explosion de la population sous des conditions écologiques favorables (végétation abondante et pluies bien réparties).

Des groupes peuvent se former, pulluler et causer des dégâts important aux cultures. Par exemple, les pertes agricoles pour les communautés affligées par la dernière invasion du Criquet pèlerin de 2003-2005 en Afrique de l'Ouest ont été évaluées à 2,5 milliards de dollars (Brader 2006). La grande capacité de dispersion du Criquet pèlerin le rend encore plus ravageur. Dans les périodes de grandes pullulations (densité supérieure à 300 individus/ha), la surface qu'il occupe est très importante (29 millions de km² sur une soixantaine de pays) comparée à la surface de 16 millions de km² (sur une trentaine de pays) lors de sa période de rémission (Fig.2). Les essaims de criquets grégaires se déplacent sous le vent et sont capables de parcourir au moins 100 km par jour (Symmons et Cressman 2001). Face à cette situation, les dégâts peuvent être considérables et la lutte est difficile à organiser, puisqu'elle est transfrontalière. La communauté internationale par le biais de l'Organisation des Nations Unies pour l'alimentation et l'agriculture (FAO) a ainsi investi, depuis les

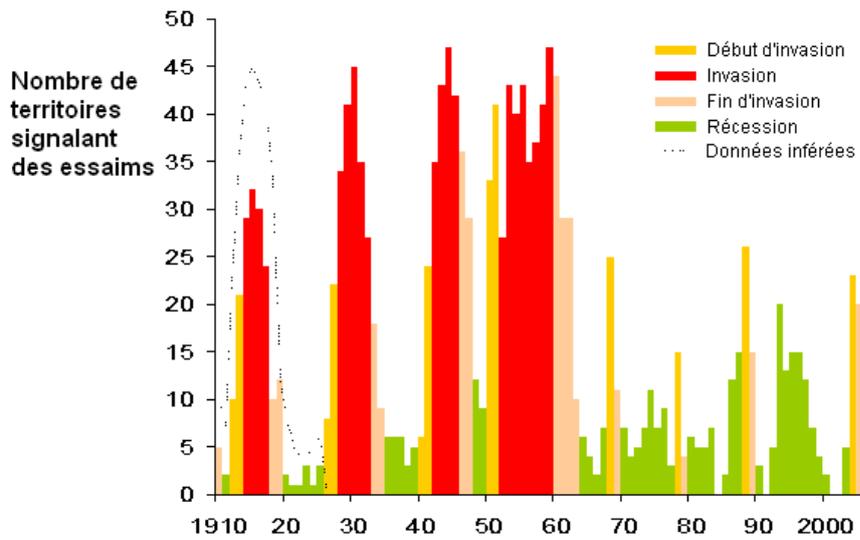


Figure 4 : Dynamique cyclique des populations de criquet pèlerin entre 1910 et 2005 avec en rouge les populations en périodes d'invasion et en vert les populations en périodes de récession (Lecoq, 2005)



Figure 5 : Différence de morphologie et de couleur entre individus en phase « grégaire » (à gauche) et individus en phase « solitaire » (à droite)

années 1960, beaucoup de moyens humains et matériels dans la lutte contre le Criquet pèlerin et dans sa surveillance, notamment au Maghreb, Afrique de l'Ouest et sur le pourtour de la Mer rouge. Historiquement, les périodes d'invasion et de récession sont apparues de manière cyclique (Fig.4). Cependant, à partir des années 1960, les nouvelles invasions ont pu être maîtrisées, notamment grâce à la lutte insecticide préventive et à une coordination supranationale de la surveillance. Aujourd'hui encore, la surveillance est nécessaire afin d'éviter de nouvelles tendances à la pullulation. A la suite de pluies exceptionnelles début 2012 au Sud-Ouest de la Libye et au Sud-Est de l'Algérie, les cultures céréalières de l'Ouest de l'Afrique sont à nouveau menacées par les essaims du Criquet pèlerin. En conséquence, la FAO a lancé un plan d'action d'urgence en juin 2012 pour limiter l'insécurité alimentaire.

b. Le polyphénisme de phase : un trait conditionnant la nuisibilité de l'espèce

Les augmentations numériques de population au début des pullulations sont associées à de fortes pluies, procurant une végétation abondante et un sol humide permettant le développement de plusieurs générations. A une échelle plus locale, il a été montré que la distribution agrégée des ressources (Despland et Simpson 2000) et une forte variation du microclimat (Farrow 1975) qui peut induire une raréfaction de la végétation, concentrent les individus, et de ce fait facilitent les départs de pullulation. L'influence environnementale sur la densité des individus est d'autant plus importante chez le Criquet pèlerin qu'elle interagit avec l'expression du polyphénisme de phase.

Le polyphénisme de phase peut être défini comme la réponse plastique du criquet face à la densité de population notamment sur les traits de comportement de migration et d'agrégation. Il peut alors passer d'une phase ou comportement « solitaire » à une phase ou comportement « grégaire ». Le comportement des individus grégaires a pour conséquence de concentrer les individus, et de cette manière, d'affecter directement la dynamique des populations. La grégarisation est un mécanisme clé qui entraîne une rétroaction positive conduisant toujours à de plus grands groupements des individus, et donc à une plus forte agrégation.

Au niveau du groupe, le comportement grégaire est très caractéristique. Les larves se suivent dès l'éclosion pour former une bande larvaire qui se déplace au sol. Plusieurs bandes peuvent alors se former et ensuite se rejoindre pour en former de plus grandes, puisque les individus grégaires s'efforcent de rester toujours le plus nombreux possible. Dans ces conditions, les criquets se développent plus rapidement et sont beaucoup plus actifs. Ils sont alors à la recherche de toutes les sources de nourriture qu'ils peuvent trouver. Au stade adulte, les individus regroupés forment des essaims et se déplacent alors tous ensemble en volant, comme un « tapis roulant », où les premiers individus du nuage se posent tandis que les derniers s'envolent. Ils peuvent couvrir des surfaces gigantesques : de 1 km² à plusieurs centaines de km², contenant de 40 millions à 80 millions de criquets par km² (Symmons et Cressman 1994). Ces essaims vont ensuite pondre, à de fortes densités, des descendants déjà grégaires (Uvarov 1966), ce qui aura tendance à entretenir, à long terme, le phénomène de grégarisation. Dans ce contexte, les végétaux présents sur le passage des essaims seront alors totalement mangés, sans distinction d'espèce, ce qui en fait un ravageur redoutable des cultures. Le changement de phase du Criquet pèlerin rend sa gestion problématique en agriculture.

D'autre part, le changement de comportement s'accompagne d'un changement de morphologie (Uvarov 1966; Pener et Simpson 2009). Les traits, classiquement utilisés en acridologie pour distinguer les phases grégaire et solitaire sont des rapports de mesures morphométriques

(Thomas 1941; Dirsh 1953; Uvarov 1956). Les traits les plus couramment mesurés sont le ratio de la longueur du fémur arrière sur la largeur de la tête) et le ratio de la longueur de l'élytre sur la longueur du fémur arrière, qui diminue et augmente respectivement avec la grégarisation. Un changement frappant associé à la grégarisation est le changement de couleur (Fig.5) : les larves grégaires présentent de nombreuses tâches noires sur leur cuticule et les adultes grégaires sont roses lorsqu'ils sont immatures et jaunes lorsqu'ils sont sexuellement matures, contrairement aux solitaires qui ont une couleur verte ou brune. Au niveau de l'oothèque, les femelles grégaires pondent des œufs moins gros et en plus petite quantité à chaque fois.

c. Deux sous-espèces décrites comme ayant des traits de phase différents

Au sein de l'espèce *Schistocerca gregaria*, deux sous-espèces géographiquement séparées ont été décrites (Botha 1967) : *Schistocerca gregaria gregaria* s'étendant dans la moitié Nord de l'Afrique jusqu'en Asie et *Schistocerca gregaria flaviventris* (Burmeister 1838) en Afrique australe, au Nord-Ouest de l'Afrique du Sud, en Namibie et au Sud de l'Angola (Faure 1932; Waloff 1966) (Fig.3). Ces deux sous-espèces sont interfertiles (Nolte 1965) mais ont été décrites comme des populations isolées (Uvarov 1966). *S. g. flaviventris* et *S. g. gregaria* sont morphologiquement très similaires et Dirsh a déclassé leur statut d'espèces à celui de sous-espèces en 1956. *S.g. flaviventris* diffère de l'espèce nominale par des cerques mâles plus larges (Dirsh 1974) et une taille des femelles plus élevée (Botha 1967). Ces caractères morphologiques décrits sur un petit nombre de spécimens illustrent probablement une variation continue au sein de chaque sous-espèce plutôt que des différences discrètes. Ainsi, la taxonomie du Criquet pèlerin repose premièrement sur la séparation géographique, et les facteurs qui la façonnent restent à ce jour largement incompris.

Ces deux sous-espèces ont particulièrement été étudiées du fait de leur différence de comportement et de nuisibilité en agriculture. Contrairement à *S. g. gregaria*, *S. g. flaviventris* n'est qu'un ravageur de très faible importance (Botha 1969; Price 2000). En effet, les populations de *S. g. flaviventris* pullulent rarement (Waloff et Pedgley 1986), avec seulement deux événements de pullulation, en 1934-1935 et en 1948-1949, localisés dans les dunes du désert du Kalahari, le long de la frontière entre la Namibie et l'Afrique du Sud. Uvarov (1966, 1977) a aussi rapporté l'existence d'essaims d'ailés et de bandes de larves de faibles niveaux de densité, d'agrégation et d'activité.

De plus, *S. g. flaviventris* semble ne présenter qu'un faible polyphénisme de phase (Uvarov 1966). En effet, en conditions environnementales contrôlées, et notamment de groupement, il a été montré que les deux sous-espèces présentent une différence de couleur. Les mâles groupés de *S. g. flaviventris* ne développent pas de coloration rose ou jaune (Faure 1932; Nolte 1965). Il existe aussi une différence dans les ratios morphométriques. Le groupement continu d'insectes de *S. g. flaviventris* solitaires change peu leur morphométrie, même après 6 générations (Botha 1967). De manière similaire, Ellis a noté un comportement d'agrégation moins fort chez *S. g. flaviventris* (Ellis 1962).

2. Apport de la génétique des populations et démarche suivie lors de ce stage

a. Une approche comparative des mécanismes de la grégarisation

La variation géographique de la capacité à pulluler chez les populations du Criquet pèlerin peut être liée à la variation géographique des facteurs climatiques entre l'Afrique du Nord et l'Afrique

de marqueurs microsatellites permet ainsi d'accéder à des processus évolutifs neutres et récents, telles que la migration et la dérive, et ainsi de renseigner sur des paramètres d'intérêt des populations, telles que les tailles de populations, les temps de divergence récents et les mouvements efficaces des individus entre populations.

c. Questions posées dans le cadre du stage et démarche

Explorer l'histoire évolutive des deux sous-espèces du Criquet pèlerin nous amène à devoir explorer les effets de la divergence, de la dérive et des migrations sur la variabilité génétique des populations du Criquet pèlerin. Nous comptons répondre à plusieurs questions : à quand remonte la dernière fois où les deux sous-espèces étaient rassemblées en une seule espèce (temps de divergence) ? A quel point les deux sous-espèces sont-elles séparées (présence de migration) ? Quel est l'influence de la dérive sur la variabilité génétique de nos populations (taille efficace) ? A cette fin, nous utiliserons une méthode d'inférence de paramètres démographiques basée sur les théories de la Coalescence et de Bayes (le logiciel DIYABC 2.0) (Cornuet *et al.* 2008). Nous avons choisi ce logiciel car il propose un modèle de mutation complexe qui représente des événements de mutations de manière réaliste pour nos données. L'inconvénient est que cet outil est construit sur l'hypothèse de divergence pure (c'est-à-dire en absence de migration). Il nous faudra alors vérifier que l'hypothèse de non migration est réaliste.

L'étude porte à une échelle globale, celle du continent africain, et nous souhaitons aussi comprendre la structuration au sein de chaque sous-espèce. L'information de différenciation au sein de chaque sous-espèce sera, en effet, indispensable pour mieux cibler les échantillons de populations à étudier par la suite et pour mieux appréhender les possibilités de migrations. Par ailleurs, mieux comprendre la structure et la dynamique de chacune des sous-espèces a d'autres intérêts, qui seront abordés de manière très mince dans ce stage. A cette fin, nous utiliserons des indices standards de différenciation, tel que le F_{st} (Weir et Cockerham 1984).

Pour répondre au mieux à ces objectifs, dans la suite de ce rapport, nous présentons :

- Une synthèse bibliographique pour établir (i) les différentes hypothèses démographiques associées à la divergence entre les deux sous-espèces et (ii) les informations préalables sur les paramètres de mutation nécessaires à une estimation juste des paramètres démographiques que nous cherchons à informer (temps de divergence, migration et tailles de populations).
- La méthodologie utilisée pour génotyper les populations du Criquet pèlerin, décrire la diversité et la structuration génétique, tester l'hypothèse démographique d'absence/présence de migration et enfin estimer les valeurs des paramètres démographiques.
- Les résultats de ce travail concernant en particulier la structuration génétique au sein de l'espèce et le scénario de divergence entre les deux sous-espèces supporté par nos données.
- Une discussion détaillant à la fois nos conclusions quant à ces deux résultats principaux, les implications pour les études comparatives futures et les perspectives méthodologiques.

australe. On peut, cependant, poser la question de la présence d'une variation génétique au sein des populations responsable de la variation de la capacité à pulluler. Cette hypothèse est plausible puisque des différences de traits de phase, qui participent à la formation ou au maintien des pullulations, existent entre les deux sous-espèces.

L'observation de différences de changements de phase densité-dépendants entre les populations de *S.g.gregaria* et *S.g.flaviventris* a conduit l'équipe du CIRAD à se tourner vers la compréhension des mécanismes intrinsèques, en centrant sa recherche sur l'échelle de la population et sur les variations phénotypiques et génétiques. Une étude comparative de ces deux sous-espèces paraît alors intéressante. Certains groupes de recherche se concentrent déjà sur l'évolution du polyphénisme de phase mais à une échelle interspécifique. Ainsi, des études sont en cours sur la phylogénie des caractères liés à la phase au sein de la sous-famille des Cyrtacanthacridinae (Song et Wenzel 2008).

L'équipe CIRAD souhaite s'intéresser à la source de variabilité des traits au sein des différentes populations de *Schistocerca gregaria* afin de mieux en comprendre les mécanismes de la grégarisation. Cette approche intra-spécifique évitera les difficultés liées à la divergence interspécifique au niveau génétique (variation trop importante de séquences de certains gènes par exemple) ou écologique (adaptation pour d'autres caractères que le polyphénisme de phase par exemple). Elle permettra également de procéder à des croisements entre sous-espèces divergentes afin de mieux déterminer les bases génétiques des traits étudiés.

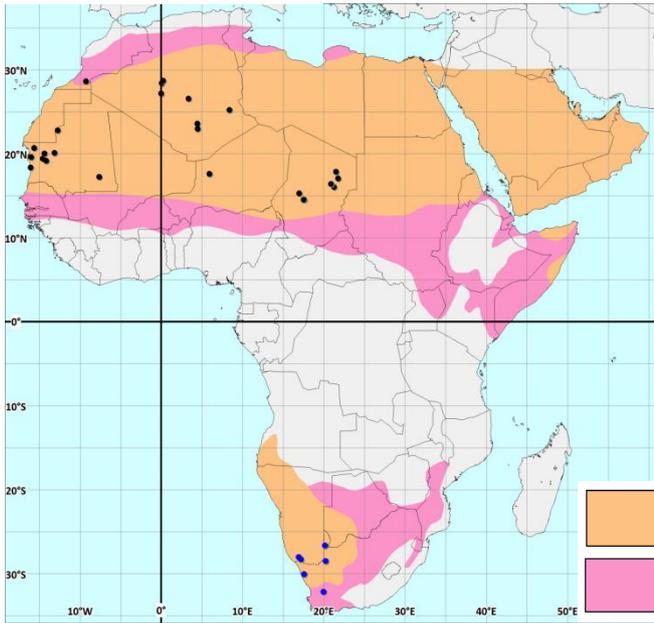
Cette nouvelle voie de recherche pourrait permettre de développer de nouvelles méthodes de lutte par l'altération des traits de population impliqués dans la capacité à pulluler. En effet, la nuisibilité du Criquet pèlerin provient de sa capacité à changer de « phase » et des traits associés. Ces traits pourraient être manipulés, par exemple, en identifiant des gènes candidats liés au changement de comportement. Il serait alors possible de perturber la régulation de ces gènes par l'utilisation de la technique d'ARN interférence (Price et Gatehouse 2008).

b. Etude de l'histoire évolutive et des processus neutres

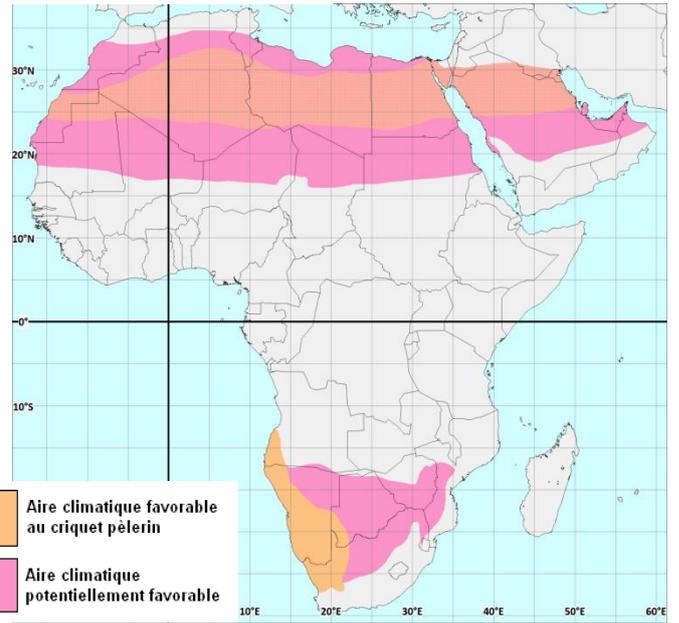
Les sources potentielles de la variation génétique observée entre les populations du Criquet pèlerin pour les traits liés à la phase proviennent à la fois des forces de sélection et de processus neutres. Il est nécessaire d'étudier séparément les deux types de processus afin de mieux comprendre leur interaction. La théorie neutraliste (ou synthétique) de Kimura soutient ainsi que l'évolution biologique est non seulement déterminée par la sélection naturelle, mais aussi par l'équilibre entre la mutation, qui crée de la variabilité, et la dérive génétique, qui cause la fixation des allèles. La théorie inclut également les effets, en termes de différenciation génétique, de la sous-division d'une population en groupes qui ne se croisent pas librement par des flux de gènes ou des migrations.

Ainsi, nous désirons ici appliquer les outils de la génétique des populations pour analyser l'origine évolutive neutre de la variation génétique observée entre les deux sous-espèces *S.g.gregaria* et *S.g.flaviventris*. A cette fin, nous utiliserons des locus microsatellites qui sont de bons marqueurs candidats. Les marqueurs microsatellites sont des marqueurs nucléaires qui présentent une variabilité intra-spécifique élevée. Ils sont constitués de la répétition de courts motifs de nucléotides (de 1 à 4 généralement), ce qui leur procure un taux d'évolution élevé par la haute probabilité d'erreur de réplication (ajout ou suppression de x motifs) (Oliveira *et al.* 2006). L'analyse

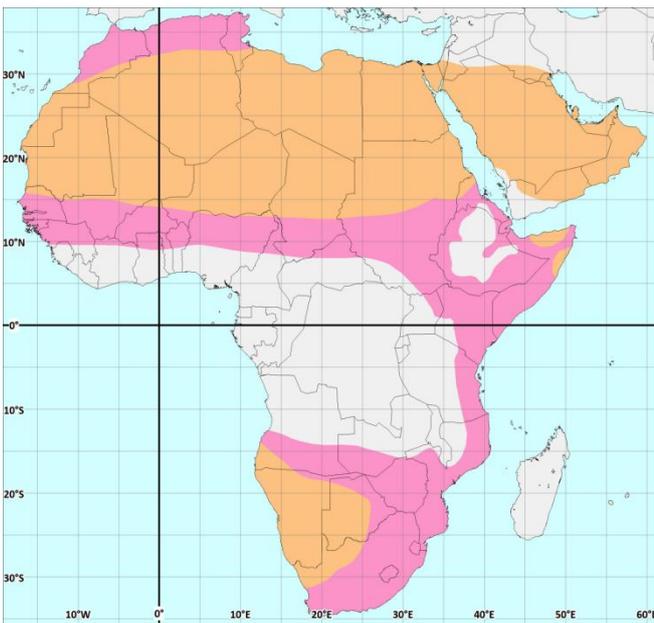
(a) Aujourd'hui



(b) -9 000 ans



(d) -11 000 ans



(c) -13 000-18 000 ans

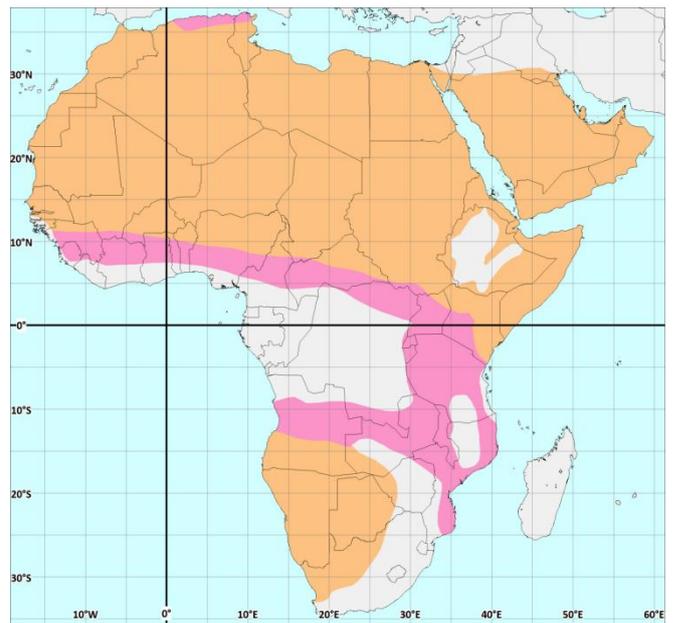


Figure 6 : Cartes de l'évolution du climat et des zones favorables au criquet pèlerin depuis aujourd'hui jusqu'au dernier âge glaciaire (Adams, J, 2001). Sur (a) ont été reportés la localisation des populations échantillonnées.

II. Bibliographie et hypothèses

1. Hypothèses démographiques pour notre étude

Des éléments de réponse aux questions sur l'histoire évolutive du Criquet pèlerin peuvent provenir des connaissances sur l'histoire africaine, la phylogéographie et les études de génétiques des populations.

a. Histoire paléoclimatique africaine

Les évolutions du climat africain et notamment l'évolution des zones arides et semi-arides, environnements les plus favorables au Criquet pèlerin, nous ont permis de poser des hypothèses biogéographiques. La figure 6 représente les évolutions des milieux arides et semi-arides depuis le dernier maximum glaciaire jusqu'à nos jours (Adams 2001). Nous ferons l'hypothèse que l'évolution de ces milieux favorables est un bon indicateur de l'évolution de l'aire de répartition de l'insecte. Il est d'abord possible de constater que l'aire de répartition actuelle est approximativement 10 fois plus petite pour *S. g. flaviventris* que pour *S. g. gregaria* (Fig. 6a). Par rapport à l'aire de répartition contemporaine, l'aire potentielle était extrêmement réduite il y a 9000 ans (Fig. 6b). Au Nord de l'Afrique, *S. g. gregaria* se retrouvait probablement sur une mosaïque de milieux désertiques, sur une aire morcelée environ 4 à 5 fois plus petite et *S. g. flaviventris* sur une aire approximativement 2 fois plus petite que l'aire actuelle. Si l'on remonte jusqu'au dernier maximum glaciaire, l'aire de répartition potentielle était beaucoup plus grande (Fig. 6c et 6d).

D'autre part, il existe déjà des études réalisées sur des espèces ayant une répartition proche de celle du Criquet pèlerin aujourd'hui, c'est-à-dire dans deux zones séparées : à la fois au Sahara et au Kalahari. De nombreuses espèces de plantes, appelées espèces des zones sèches à aire disjointe, sont concernées par cette répartition bipolaire (Lebrun 2001). Par ailleurs, on peut citer l'étude de l'histoire biogéographique des écureuils du genre *Xerus* (Herron *et al.* 2005) qui confirme l'existence d'un couloir de milieu aride qui s'étendait de la corne de l'Afrique au Cap de Bonne-Espérance. Ils n'ont cependant pas réussi à estimer de temps de divergence pour leurs espèces. Cependant, que ce soit pour les espèces végétales ou animales, il semble que c'est l'hypothèse du recul du milieu aride à la suite du dernier maximum glaciaire en Afrique qui expliquerait l'origine de la divergence ou de la répartition disjointe des populations.

Nous nous appuyons sur ces études, notamment paléoclimatiques, pour construire les hypothèses démographiques que nous souhaitons tester par la suite. Par exemple, il est possible que les tailles efficaces de population aient pu varier en fonction de l'aire de répartition des milieux favorables et que la date de divergence corresponde au moment où les aires devenaient très réduites. Nous émettons en particulier l'hypothèse que, pendant le dernier maximum glaciaire, les deux sous-espèces ne devaient probablement en former qu'une, et en ce sens nous rejoignons Uvarov (1977).

b. Etude de la structuration locale du Criquet pèlerin

Considérer à l'heure actuelle que les deux sous-espèces de *Schistocerca gregaria* sont parfaitement isolées peut être remis en cause. Face aux capacités de dispersion importantes de cet insecte en période d'invasion, il est probable que des migrations d'individus d'une sous-espèce vers l'autre aient pu avoir lieu. En effet, le Criquet pèlerin a été observé plusieurs fois en dehors de son

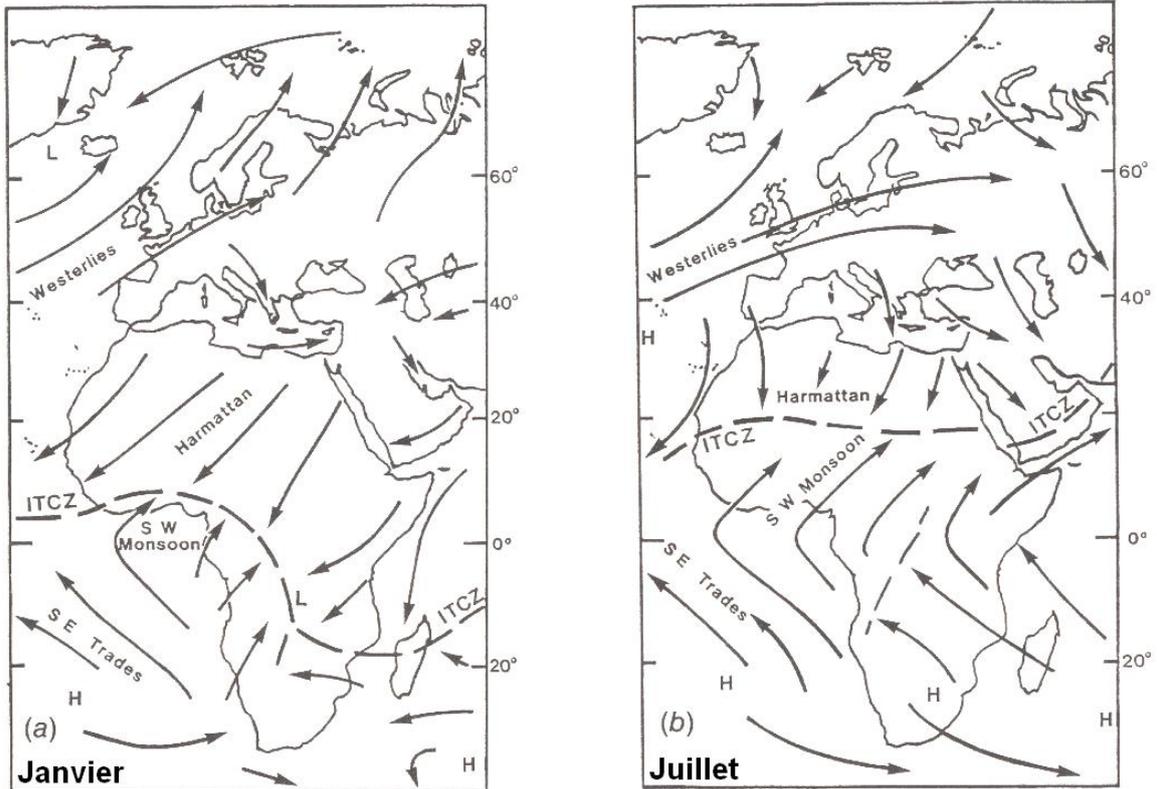


Figure 7 : Carte des vents dominants en Europe et en Afrique en janvier (a) et en juillet (b). Les lignes pointillées représentent les zones de convergence (Pedgley, 1995)

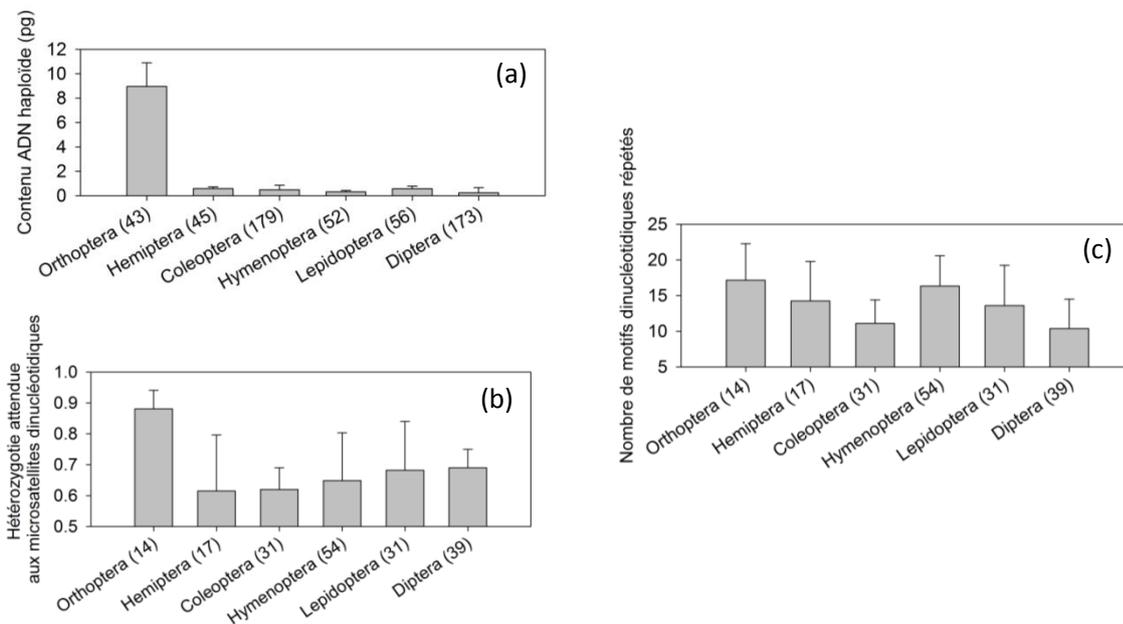


Figure 8 : Particularités du génome des Orthoptères par rapport aux principaux autres ordres d'insectes : (a) par rapport à la taille du génome, (b) par rapport à l'hétérozygotie attendue et (c) par rapport à la longueur des microsatellites dinucléotidiques (Gregory 2012 ; Chapuis *et al*, 2012)

habitat : *S. g. gregaria* en Europe méditerranéenne (Grèce, Balkans), au Mozambique à Tete en janvier 1945 (Waloff 1966), plusieurs fois en Amérique centrale, comme par exemple en 1988 dans les Caraïbes (Ritchie et Pedgley 1989), et dans les îles Ascension ou du Cap Vert (Uvarov 1923). Ces mouvements d'insectes peuvent être expliqués en partie par la direction des vents, comme la direction Nord/Sud en Afrique de l'Est menant au Mozambique ou la direction Sud-Ouest/Nord-Ouest menant à l'Océan Atlantique (Fig.7) (Pedgley 1981). La présence d'individus isolés de *S. g. gregaria* au Mozambique peut ainsi pousser à émettre l'hypothèse qu'une migration est possible. En effet, il ne reste qu'à parcourir la vallée du Limpopo pour rejoindre l'aire de *S. g. flaviventris*. Cependant, cette capacité de déplacement n'implique pas forcément la survie et la reproduction des individus immigrés et donc une migration efficace et significative. Nous savons que le Criquet pèlerin ne s'est jamais installé en Amérique centrale ou en Europe méditerranéenne par exemple.

Un travail de génétique des populations portant sur la diversité intra-spécifique du Criquet pèlerin (Ibrahim *et al.* 2000) s'est intéressé à la structuration des populations de la zone grégarigène de l'Erythrée pendant les années 1993-1995. Les auteurs concluent qu'il y a une structuration à petite échelle entre les populations de la côte et celles de l'intérieur des terres. Ce résultat est surprenant si on le confronte à l'intensité des mouvements d'individus grégaires lors des invasions. Les auteurs considèrent alors que la structuration observée peut être due à différents facteurs. Leur première hypothèse est qu'en période de rémission, période très défavorable, les individus grégaires immigrants sont mal adaptés et ne sont pas capables de se réinstaller assez vite dans les habitats où des individus solitaires autochtones existent. Leur deuxième hypothèse est que les populations subissent des événements de fondation-extinction dans des habitats fluctuants. La colonisation des milieux par une partie restreinte de quelques individus (équilibré ou non par des migrations) serait à l'origine d'un pool génétique différent du précédent, induisant alors une structuration. Cette étude suggère que les mouvements efficaces sont faibles, au moins lors des périodes de fin d'invasion.

En conclusion, ces données sont assez fragmentaires mais suggèrent qu'il est prudent de vérifier l'hypothèse d'isolement des deux sous-espèces.

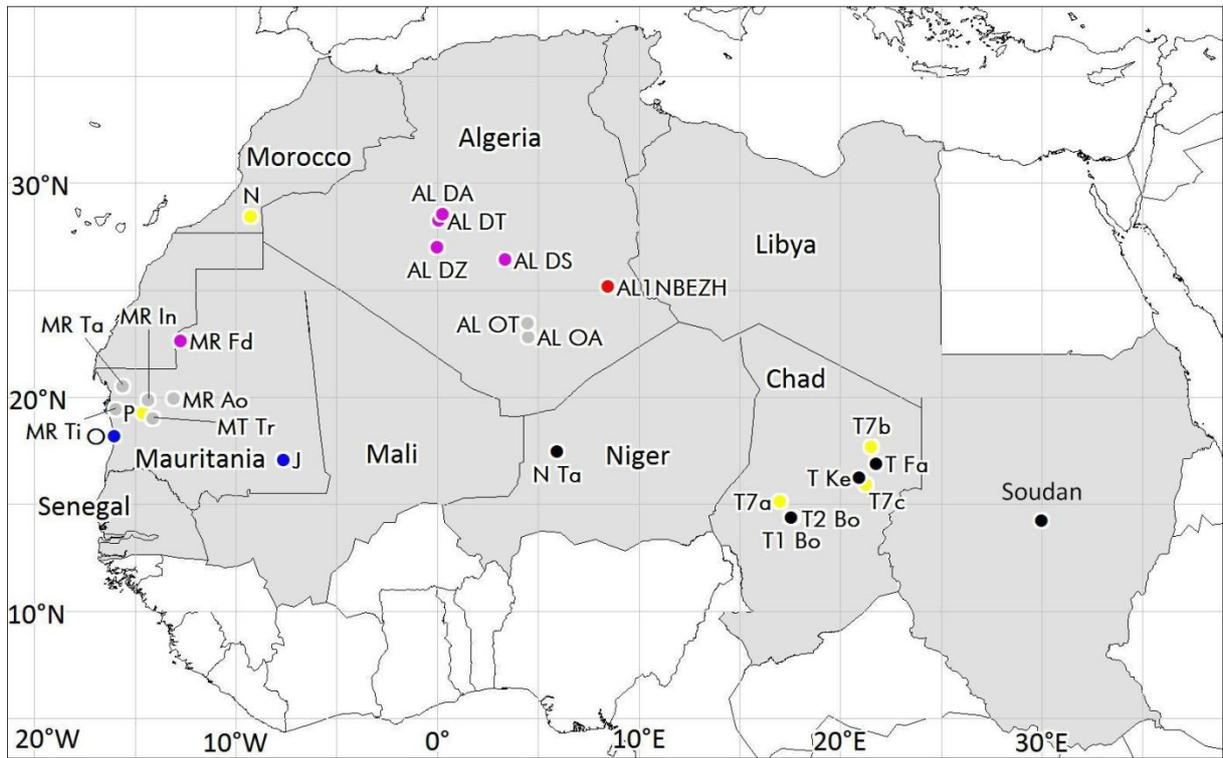
3. Spécificités des marqueurs microsatellites utilisés

a. Caractéristiques des marqueurs microsatellites des Orthoptères

Le génome des Orthoptères a un fonctionnement complexe. Il est assez différent des autres principaux ordres d'insectes : il est beaucoup plus gros (environ 9 pg au lieu d'environ 0,8 pg) (Gregory 2012)(Fig.8a), il possède des microsatellites plus longs et des hétérozygoties attendues plus élevée (Chapuis *et al.* 2011) (Fig.8b). On s'attend alors à avoir des mécanismes de régulation du génome différents de ceux des autres insectes. Ainsi, et puisque l'utilisation des marqueurs microsatellites est inédite sur le Criquet pèlerin, nous souhaitons aussi, dans le cadre de ce travail, mieux préciser le modèle mutationnel de ces marqueurs. Cette information est très importante afin de pouvoir accéder au taux d'évolution de ces marqueurs et d'avoir une estimation de l'histoire évolutive des deux sous-espèces du Criquet pèlerin. Une donnée essentielle est le taux de mutation des marqueurs.

Les études qui s'intéressent au taux de mutation des microsatellites concernent pour la plupart les grands animaux (oiseaux, mammifères, notamment l'Homme). Chez les insectes, peu d'études se sont attachées à estimer le taux de mutation des microsatellites : une étude sur la fourmi

(a)



(b)

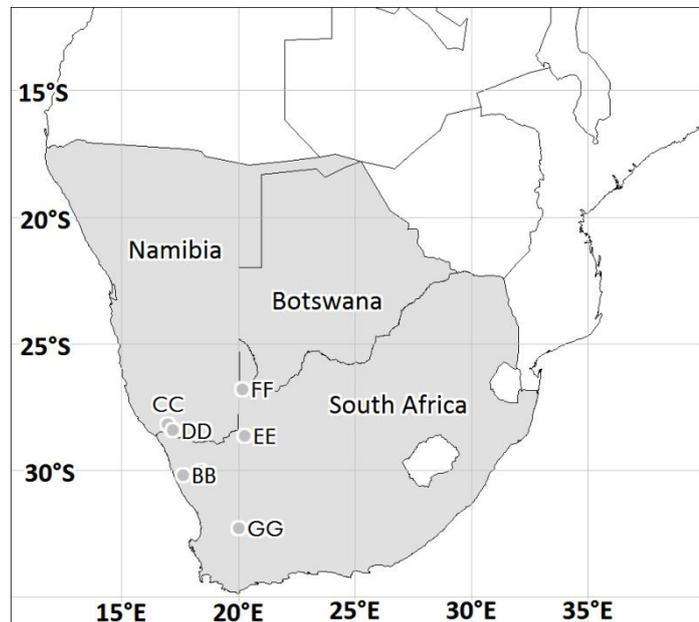


Figure 9 : Localisation géographique des populations échantillonnées: 26 populations de *S. g. gregaria* (a) et 6 populations de *S. g. flaviventris* (b)

Camponotus consobrinus (Crozier *et al.* 1999) et quatre études sur *Drosophila melanogaster* (Schug *et al.* 1997; Schlotterer *et al.* 1998; Schug *et al.* 1998; Vazquez *et al.* 2000). Pour la plupart des loci étudiés, il n'y a en fait aucune mutation qui a été observée avec le taux de détection disponible. Lorsque c'était possible, le taux de mutation était de $5,9.10^{-3}$ et 7.10^{-4} pour la fourmi et de $1,5.10^{-4}$ et $3,6.10^{-5}$ pour la drosophile. Dans notre cas, nous nous attendons à ce que le taux de mutation des microsatellites du Criquet pèlerin soit plus élevé que chez les autres insectes. En effet, le génome possède des microsatellites en moyenne plus longs que chez les autres insectes et il a été montré que plus un microsatellite est long et plus il a de chance de muter (Primmer 1998).

Ainsi, nous proposons de mesurer expérimentalement le taux de mutations d'une partie marqueurs microsatellites utilisés afin de vérifier que les valeurs utilisées pour informer les distributions *a priori* de DIYABC et informées à partir de la littérature sont réalistes pour notre espèce.

b. Spécificités des marqueurs microsatellites dérivés de séquences transcrites

Il est possible que les particularités du génome des Orthoptères que nous avons décrites ci-dessus expliquent le faible nombre d'études de génétique des populations sur les criquets (Zhang *et al.* 2003). En effet, chez les criquets, ces marqueurs souffrent de la présence très fréquente d'allèles nuls (Chapuis *et al.* 2008). Les allèles nuls sont des allèles qui ne sont pas amplifiés par PCR (Réaction en Chaîne par Polymerase). La PCR est une méthode de multiplication du fragment d'ADN d'intérêt, ce qui nous permet ensuite d'avoir suffisamment de quantité d'ADN pour le détecter. Dans le cas des allèles nuls, à cause de mutations dans les régions flanquantes, il y a une mauvaise fixation des amorces et donc une absence d'amplification. L'information partielle que l'on retire de tels marqueurs peut alors être biaisée (Chapuis et Estoup 2007).

Cette difficulté peut être en grande partie levée par la caractérisation d'amorces microsatellites à partir de séquences transcrites, issues par exemple de banques EST (Expressed Sequence Tags). En effet, les microsatellites issus de séquences transcrites présentent souvent moins d'allèles nuls que les microsatellites issus directement du génome (Coulibaly *et al.* 2005; Ellis et Burke 2007; Kim *et al.* 2008). Ils n'ont pas les mêmes contraintes évolutives et sont généralement mieux conservés. Cependant, du fait de leur place dans la région codante de l'ADN, il est possible qu'ils se retrouvent sous sélection et que cela induise un biais dans les analyses. Dans le cadre de ce travail, nous devons donc être particulièrement vigilants aux loci sous sélection pour les marqueurs « transcriptomiques ». Plusieurs études ont comparé les marqueurs d'origines génomique et transcriptomique et confirment l'intérêt d'utiliser les deux types en génétique des populations. Mais il faut tout de même noter que peu d'études les ont comparés chez les animaux. Seuls Coulibaly *et al.* (2005) chez la truite arc-en-ciel et Kim *et al.* (2008) sur la chrysome du maïs s'y sont intéressés et démontrent leur fiabilité.

Par ailleurs, les marqueurs transcriptomiques possèdent vraisemblablement un taux de mutation par glissement sensiblement différent de celui des marqueurs génomiques. Une étude a montré chez le ver à soie (Prasad *et al.* 2005) que les mutations observées pour les marqueurs « transcriptomiques » sont différentes de celles des marqueurs « génomiques » : ils sont plus concernés par la tranversion-transition et moins par les insertions-délétions. **Ainsi, nous nous attendons à un taux de mutation pour les marqueurs microsatellites d'origine transcriptomique inférieur à celui des marqueurs d'origine génomique, et nous nous efforçons de le prendre en compte dans les analyses Bayésiennes.**

Tableau 2 : Informations sur les 32 populations échantillonnées

Sous-espèce	Nom de la population	Pays	Long DMS100	Lat DMS100	Date	Nombre
<i>S.g. flaviventris</i>	BB	Afrique du Sud	-17°33' 44.5"	29°59'12.1"	18/06/2011	32
<i>S.g. flaviventris</i>	CC	Afrique du Sud	-16°59'20.4"	28°08'01.0"	21/06/2011	32
<i>S.g. flaviventris</i>	DD	Afrique du Sud	-17°13'10.9"	28°21'57.00"	22/06/2011	24
<i>S.g. flaviventris</i>	EE	Afrique du Sud	-20°17'15.5"	28°35'05.7"	24/06/2011	35
<i>S.g. flaviventris</i>	FF	Afrique du Sud	-20°12'08.4"	26°44'93.6"	28/06/2011	33
<i>S.g. flaviventris</i>	GG	Afrique du Sud	-20°01' 45.0"	32°13'53.3"	05/07/2011	31
<i>S.g. gregaria</i>	AlSgOA	Algérie	-04.31.23	22.52.17	14/11/2010	30
<i>S.g. gregaria</i>	AlSgOT	Algérie	-04.29.23	23.30.10	02/12/2010	30
<i>S.g. gregaria</i>	AlSgDS	Algérie	-03.22.49.00	26.28.21.00	28/04/2009	6
<i>S.g. gregaria</i>	AlSgDT	Algérie	-00.04.57.00	28.19.04.00	29/03/2009	9
<i>S.g. gregaria</i>	AlSgDA	Algérie	-00.16.00.00	28.36.04.00	06/04/2009	8
<i>S.g. gregaria</i>	AlSgDZ	Algérie	-00.00.35.00	27.03.26.00	11/02/2009	8
<i>S.g. gregaria</i>	AlSg1NBEZH	Algérie	08.25.39.00	25.07.56.00	03/03/2012	30
<i>S.g. gregaria</i>	N	Maroc	-09.12.56.00	28.30.57.00	07/05/2010	30
<i>S.g. gregaria</i>	J	Mauritanie	-07.36.24.96	17.08.26.64	13/09/2009	33
<i>S.g. gregaria</i>	O	Mauritanie	-16.01.27.09	18.15.00.59	25/11/2009	80
<i>S.g. gregaria</i>	P	Mauritanie	-14.33.56.00	19.19.03.00	02/02/2010	40
<i>S.g. gregaria</i>	MRSgTa	Mauritanie	15°36'35"	20°34'20"	14/03/2011	30
<i>S.g. gregaria</i>	MRSgTi	Mauritanie	15°57'04"	19°30'09"	22/01/2011	30
<i>S.g. gregaria</i>	MRSgTr	Mauritanie	14°06'59.8"	19°04'04.7"	04/01/2011	30
<i>S.g. gregaria</i>	MRSgIn	Mauritanie	14°19'47"	19°55'01"	05/01/2011	27
<i>S.g. gregaria</i>	MRSgAo	Mauritanie	13.05.11	19.59.35	17/11/2010	30
<i>S.g. gregaria</i>	MRSgFd	Mauritanie	-12.42.44.96	22.41.43.00	26/01/2009	21
<i>S.g. gregaria</i>	NSgTa	Niger	-17°30'26"	5°57'25"	40850	30
<i>S.g. gregaria</i>	Soud	Soudan	NA	NA	NA	11
<i>S.g. gregaria</i>	7a	Tchad	-17.01.28	15.11.29	06/11/2010	26
<i>S.g. gregaria</i>	7b	Tchad	-21.33.45,6	17.44.48,4	19/10/2010	30
<i>S.g. gregaria</i>	7c	Tchad	-21.17.31	15.57.44	21/10/2010	30
<i>S.g. gregaria</i>	TSgKe	Tchad	-20°55'57"	16°18'35'	15/11/2011	18
<i>S.g. gregaria</i>	TSgFa	Tchad	-21°47'54.5"	16°57'20"	28/11/2011	11
<i>S.g. gregaria</i>	TSg2BO	Tchad	-17°38'33"	14°25'26"	16/11/2011	15
<i>S.g. gregaria</i>	TSg1BO	Tchad	-17°33'36"	14°25'59"	16/11/2011	10

III. Matériels et Méthodes

1. Présentation des marqueurs utilisés et génotypage

a. Populations étudiées

Les populations disponibles de la sous-espèce *S. g. gregaria* sont issues de campagnes d'échantillonnage menées par les correspondants locaux de chaque pays qui travaillent sur le suivi et la gestion du Criquet pèlerin. 549 individus de vingt-six populations (6 pays) ont été obtenus de cette manière et sont représentatifs de 4 années différentes (2009 à 2012) (Fig. 9a et Tableau 2). Par ailleurs, six populations (186 individus) de *S. g. flaviventris* proviennent de l'échantillonnage d'Antoine Foucart et de Marie-Pierre Chapuis en 2011 en Afrique du Sud (Fig.9b et Tableau 2). Quatre populations (ALDA, ALDT, ALDS et ALDZ) ne seront finalement pas étudiées car trop faiblement représentées en raison du mauvais état de conservation des tissus ne permettant pas une extraction ADN de qualité. Pour les analyses de biologie moléculaire, un morceau du fémur de la patte arrière de chaque individu a été utilisé. L'ADN des 549 individus a été extrait suivant le protocole d'extraction en colonne QIAGEN (Protocole Spin-Column pour tissus animaux du DNeasy Blood & Tissue Handbook).

b. Génotypage des allèles

Dans cette étude, nous utilisons 28 marqueurs microsatellites dinucléotidiques qui sont issus de deux banques de données de séquences différentes :

- 15 marqueurs « génomiques » dont 6 ont été décrits par d'autres équipes (Yassin *et al.* 2006; Kaatz *et al.* 2007) et 9 ont été développés au CIRAD par Laurence Blondin et Marie-Pierre Chapuis.
- 13 marqueurs « transcriptomiques », développés par l'équipe du CIRAD et à partir des régions microsatellites formées de 8 à 12 motifs de séquences transcrites publiées par (Badisco *et al.* 2011).

Les caractéristiques (type de motif, longueur et range de taille) de ces marqueurs sont présentées en tableau 3 et en annexe 1.

Le protocole d'amplification ADN par PCR qui a été suivi est décrit en annexe 2. Afin de diminuer le nombre de PCRs, les 28 marqueurs microsatellites ont été amplifiés par la méthode de multiplexage QIAGEN (voir Protocole avec solution Q du QIAGEN Multiplex PCR Handbook). La figure 10 présente les 5 jeux de marqueurs génomiques et les 4 jeux de marqueurs transcriptomiques. Nous avons, en plus, « poolé » les produits PCR issus de 6 jeux de marqueurs deux à deux pour l'étape d'électrophorèse capillaire sur séquenceur. Nous avons alors génotypé en deux étapes : 14 populations avec 6 jeux de loci (28 marqueurs microsatellites) et 14 autres populations avec 5 jeux de loci (24 marqueurs). Nous avons considéré que 14 populations était un nombre suffisant pour représenter nos sous-espèces pour l'étude de divergence (6 populations de *S. g. flaviventris* et 8 populations de *S. g. gregaria*). Les autres populations seront utiles pour l'étude de la structuration spatiale au sein des sous-espèces.

Les produits PCR sont finalement dilués au 300^{ème} dans une solution de formamide/marqueur de taille. Les amplicons sont alors séparés et détectés sur un séquenceur ABI130 (Applied

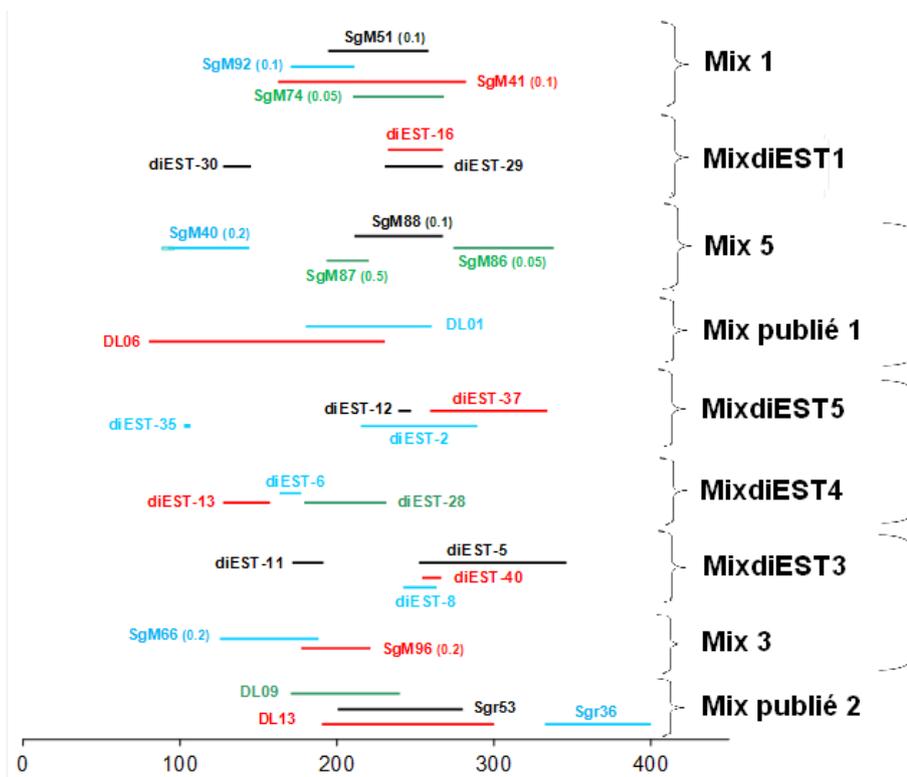


Figure 10 : Les 9 jeux de loci multiplexés (28 marqueurs) utilisés dans notre étude et développés par l'équipe CIRAD. Les jeux de loci « Mix » correspondent aux loci d'origine génomique et les « MixdiEST » ceux d'origine transcriptomique. Six d'entre eux sont « poolés » deux par deux (accolade sur la droite).

Tableau 3 : Caractéristiques des marqueurs microsatellites utilisées dans l'étude

Type de marqueur	Locus	Type de motif	Nombre moyen de motifs en Mauritanie	Intervalle de taille observé (en nombre de motifs)	Contrainte de taille estimée (en nombre de motifs)
Génomique	Sgr36	GT*	20	24	48
Génomique	SgM96	CT	26	29	37
Génomique	SgM88	GT	20	29	38
Génomique	DL13	CT	28	34	49
Génomique	Sgr53	GT*	25	34	103
Génomique	SgM86	GT	28	35	71
Génomique	SgM92	GT	12	36	51
Génomique	SgM74	GT*	10	37	112
Génomique	DL09	GT	19	39	48
Génomique	DL01	CT	33	39	79
Génomique	SgM66	GT*	23	44	74
Génomique	DL06	GT	29	48	62
Génomique	SgM87	GT	8	48	144
Génomique	SgM41	GT	29	68	86
Génomique	SgM51	GT*	29	84	108
	Moyenne		23	42	74
Transcriptomique	diEST-12	AT	7	8	16
Transcriptomique	diEST-6	CT*	12	9	11
Transcriptomique	diEST-35	GT	8	10	18
Transcriptomique	diEST-30	GT	11	17	22
Transcriptomique	diEST-40	TA	9	18	24
Transcriptomique	diEST-29	GT	10	20	60
Transcriptomique	diEST-11	AT*	11	23	29
Transcriptomique	diEST-16	GT	12	29	38
Transcriptomique	diEST-28	GT*	13	31	63
Transcriptomique	diEST-13	AT	10	38	50
Transcriptomique	diEST-2	GT	21	39	52
Transcriptomique	diEST-8	AT*	11	40	50
Transcriptomique	diEST-37	GT	22	49	75
	Moyenne		13	25	39

Biosystems). Les résultats sont ensuite lus sur GeneMapper v4.1, qui permet d'observer facilement la longueur des allèles pour chaque individu.

c. Détermination du taux de mutation des marqueurs génomiques

i. Estimation du taux de mutation pour des individus d'élevage

Dans notre étude, pour s'intéresser aux mutations qui peuvent apparaître sur les marqueurs microsatellites, nous nous concentrons sur les mutations germinales. Ainsi, nous avons génotypé environ 2 000 descendants (larves de premier stade, Fig. 2) issus d'un élevage maintenu en laboratoire au CIRAD. Ceci nous permet d'avoir un suivi précis des parents de nos individus. L'élevage de ces criquets est proche des conditions naturelles (voir annexe 3) : 34°C avec une photopériode de 12h-12h, en boîtes individuelles, élevés à partir de populations naturelles mauritaniennes depuis 4 générations. Tous les parents ont ainsi subi le même traitement, ce qui devrait normalement limiter l'effet du stress sur le taux de mutation (Fonville *et al.* 2011).

Le génotypage des individus nous permet de repérer les mutations éventuelles en comparant leur génotype à celui de leurs parents. L'utilisation de 2 000 descendants donne ainsi accès à l'observation de 4 000 événements de méiose par marqueurs. Nous avons alors un seuil de détection des taux de mutations par marqueurs de $2,5 \cdot 10^{-4}$ au maximum. Pour cette raison, il ne paraît pas possible d'estimer le taux de mutation des marqueurs « transcriptomiques » et des marqueurs qui sont courts (Lai et Sun 2003). Nous nous concentrons alors sur les 15 marqueurs « génomiques ». Cependant, trois marqueurs « transcriptomiques » sont génotypés en même temps grâce au multiplexage mais ne seront pas analysés dans le cadre de ce stage. Ceci correspondra alors à la lecture de 60 000 méioses parents-enfants, tous loci confondus.

L'extraction ADN des larves, en très grand nombre, a été faite à partir d'une patte arrière de chaque individu, directement en plaque d'extraction de 96 puits suivant un protocole proche de l'extraction en colonne QIAGEN (voir annexe 4). La concentration de l'extrait ADN de chaque individu est mesurée par Nanodrop afin de pouvoir ajuster les concentrations. Cependant, afin de diminuer le coût lié à la PCR et au génotypage de tant d'individus, nous avons optimisé la méthode en « poolant » plusieurs individus dans un même puits. Nous avons choisi de les « pooler » par 4 tout en ajustant les concentrations à plus de $1/8^{\text{ème}}$. Ce résultat provient d'une expérience préliminaire de « mutation artificielle » que nous avons menée préalablement (voir paragraphe suivant et IV.2.c.i). Nous avons ensuite suivi le même protocole de génotypage que décrit précédemment en III.1.b pour les 15 marqueurs génomiques dont nous souhaitons estimer le taux de mutation. Deux lecteurs (Laurence Blondin et Christophe Plantamp) ont lu indépendamment les profils sur GeneMapper.

ii. Développement d'un protocole d'amplification en pool

Le principe de cette expérience est d'identifier le nombre d'individus « frères » que nous pouvons « pooler » sans risquer de rater l'allèle mutant lorsqu'il est présent. Nous avons alors poolé des extraits ADN de larves dont le génotype est connu et simulé la présence d'un allèle mutant en choisissant un individu qui introduit un unique allèle « étranger » comparé aux autres allèles présents dans le mélange.

Nous avons, tout d'abord, testé 10 combinaisons différentes de 3, 4 et 5 individus poolés ensemble, que nous avons répétées deux fois et analysées pour trois marqueurs SgM86, SgM87 et SgM88. Les deux observations étaient répétables mais nous avons raté sur GeneMapper 9 mutations

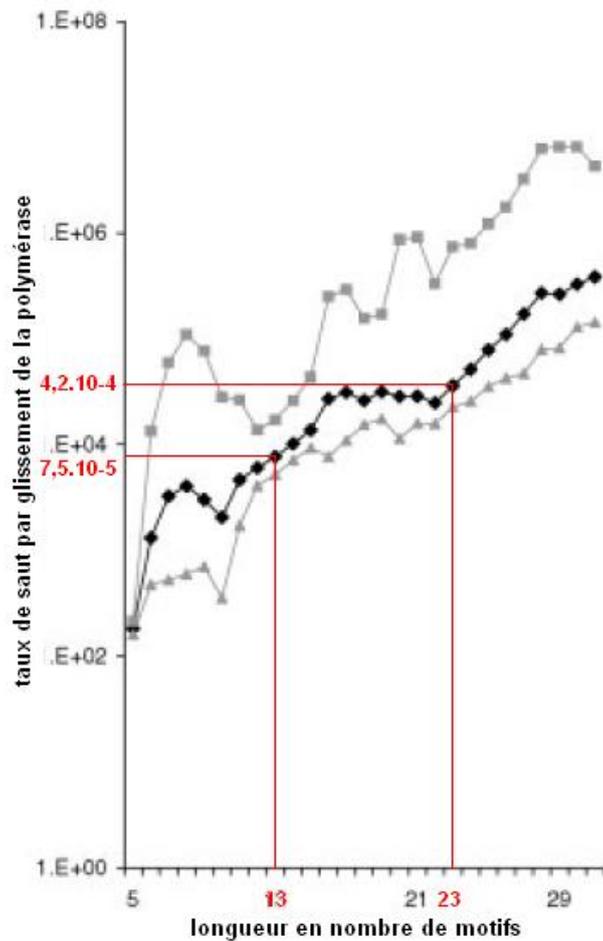


Figure 11 : Estimation chez l'Homme de la relation entre la longueur des microsatellites dinucléotidiques et le taux de mutation par glissement de la polymérase (Lai et Sun, 2003). Le taux de mutation est représenté en échelle logarithmique comparé au taux de mutation ponctuel de 10^{-8} . Le taux de mutation s'obtient donc à partir de l'ordonnée A de la manière suivante $\mu=A \cdot 10^{-8}$. (Les points noirs correspondent à la moyenne ; les points gris à l'enveloppe de 95% des valeurs).

sur 38 cas pour 5 individus poolés dans un même puits, 5 sur 38 cas pour 4 individus et 3 sur 38 cas pour 3 individus. Vu cette tendance, nous nous sommes concentrés sur les pools de 4 individus par la suite. Nous avons alors testé les 10 combinaisons de 4 individus (répété 2 fois et lu 2 fois indépendamment) pour les 15 marqueurs dont nous souhaitons estimer le taux de mutation.

Nous avons ensuite analysé les résultats de ces 228 cas d'allèle « mutant » par un GLM. Nous avons cherché à expliquer quels facteurs influencent la lecture du pic « mutant », mesurée par l'aire du pic « mutant » relativement à l'aire totale. Les facteurs testés sont la distance (en paires de bases) à l'allèle plus long le plus proche, le nombre d'allèles plus courts, la concentration mesurée de l'allèle « mutant ».

Les résultats de cette expérience servent à optimiser le protocole de génotypage pour l'étude d'estimation du taux de mutation afin de diminuer le nombre de PCRs.

2. Description de l'« état » génétique de l'espèce : signatures de déséquilibre, diversité génétique et structure génétique

A partir des 28 populations décrites en III.1.a. et des 28 marqueurs microsatellites décrits en III.1.b, nous comptons tester plusieurs hypothèses démographiques (divergence, migration, structuration) pour l'espèce *Schistocerca gregaria* et de ses deux sous-espèces. Dans cette optique, il est nécessaire de contrôler les biais ou déséquilibres aux hypothèses de génétique des populations. Les marqueurs et les populations qui démontreraient un biais sont retirés du jeu de données pour la suite des analyses.

a. Conformité aux équilibres d'Hardy-Weinberg et de liaison

Il est important de vérifier qu'au sein de notre jeu de données, il n'y a pas de trop de biais par rapport à la situation d'équilibre d'Hardy-Weinberg. Nous avons utilisé GENEPOP 4.1.4 (Raymond et Rousset 1995) pour tester la validité de l'équilibre d'Hardy-Weinberg dans nos données. Un déséquilibre qui apparaîtrait pour une population pourrait aussi être expliqué par une consanguinité ou un apparentement des individus au sein de la population. Un marqueur qui montrerait un écart avec Hardy-Weinberg pourrait, en fait, présenter des allèles nuls. Nous ne suspectons pas une présence forte d'allèles nuls dans nos données, puisqu'un effort important de sélection des marqueurs a été fait sur ce critère (les 28 marqueurs ont été choisis parmi 93 génotypés pour 6 populations d'Afrique de l'Ouest). Cependant comme ce travail précédent a été réalisé sur *S.g.gregaria*, il est possible que des allèles nuls apparaissent pour les populations de *S. g. flaviventris*. Ainsi, pour tous nos marqueurs, nous calculons aussi la proportion d'allèles nuls par population qu'ils peuvent avoir avec FreeNA (Chapuis et Estoup 2007).

D'un autre côté, nous avons testé, avec GENEPOP 4.1.4, le déséquilibre de liaison de nos marqueurs qui pourrait ajouter un autre biais dans nos données, en associant préférentiellement certains allèles de plusieurs loci entre eux.

Pour toutes ces analyses, comme ce sont des cas de multi-tests nous faisons une correction FDR (False Discovery Rate) (Benjamini et Hochberg 1995). Pour ce faire, nous utilisons le package R *qvalue* afin de corriger nos p-values.

b. Test de marqueurs sous sélection

Les tests de sélection, que nous utilisons, comparent les statistiques observées pour nos marqueurs et celles obtenues par simulation dans une situation de neutralité. Les simulations se font sur la base de la coalescence, pour des modèles neutres dont les valeurs de paramètres génèrent des statistiques résumées les plus proches possible de notre jeu de données. Les marqueurs qui possèdent des statistiques observées en dehors de l'enveloppe de 95% des statistiques obtenues en situation de neutralité sont alors considérés comme sous sélection. Nous utiliserons F_{dist} (Beaumont et Nichols 1996), qui s'appuie sur la relation entre F_{st} (estimation de Weir and Cockerham 1984) et H_e (Nei 1987) pour des populations interconnectées, dans le cas des tests intra sous-espèces.

Sur la base de deux sous-espèces divergentes à l'équilibre, nous testerons la sélection avec DETSEL (Vitalis *et al.* 2001), qui est construit à partir des statistiques multi-locus F_i pour décrire les marqueurs, avec $F_i = \exp(-t/N_i)$ avec t le temps de divergence et N_i la taille efficace de la population i . Pour l'instant, nous n'avons pas réussi à tester l'hypothèse de sélection sous ce scénario de divergence des populations. En effet, nous échouons à paramétrer θ (égal à quatre fois le produit de la taille efficace par le taux de mutation) de manière réaliste, ce qui semble être lié à une limite maximale sur ce paramètre, bien inférieure à la valeur compatible avec les diversités génétiques élevées observées chez le Criquet pèlerin. Nous utiliserons donc également $F_{dist} 2$ pour tester les marqueurs sous sélection entre les deux sous-espèces.

c. Diversité au sein des populations et différenciation entre populations

A partir du jeu de données considéré sans « biais », il paraît intéressant de décrire les populations présentes dans notre échantillonnage au moyen de plusieurs statistiques descriptives : le nombre d'allèles, la richesse allélique A_r (pour une taille minimale de 9 individus) et l'hétérozygotie attendue ou diversité génétique H_e (Nei 1987). Ces statistiques sont calculées avec Fstat 2.9.3.2 (Goudet 2002). Le test de comparaison de moyenne entre les sous-espèces est un test non paramétrique de Wilcoxon.

Il est aussi possible de caractériser la « proximité » des populations entre elles, c'est-à-dire d'avoir une idée de l'intensité des flux de gènes ou de divergence entre ces populations. Cette analyse de la structuration génétique est basée sur le F_{st} calculé par GENEPOP 4.1.4 (Weir and Cockerham 1984). Chaque sous-espèce est caractérisée par un F_{st} global et un F_{st} entre sous-espèces et leurs populations par des F_{st} par paires (matrices de F_{st}). La significativité des F_{st} par paire est mesurée à partir de l'intervalle de confiance calculé par bootstrap de 1 000 répétitions sur FreeNA. Etant donné que la précision du calcul de F_{st} est de l'ordre de 0,008 (valeurs négatives minimales des intervalles) et que des valeurs aussi petites de différenciation n'ont pas beaucoup de sens, nous considérons qu'il y a de la différenciation lorsque l'intervalle de confiance est strictement positif et inclut 0,01.

d. Analyse d'excès d'hétérozygotie, signature de goulot d'étranglement

A partir de l'hétérozygotie attendue H_e (Nei 1987), il est possible d'avoir une idée de la présence d'un effondrement de population ou « goulot d'étranglement » (grosse réduction de taille efficace) ou d'une expansion démographique (augmentation de taille efficace) au sein des populations que nous étudions. La méthode a été développée par Cornuet et Luikart (Cornuet et Luikart 1996). Elle s'appuie sur le fait que, lors d'un goulot d'étranglement ou « bottleneck » en anglais, le nombre d'allèles diminue beaucoup plus que l'hétérozygotie attendue H_e . Ainsi, en

comparant l'hétérozygotie attendue H_e (calculée à partir des fréquences alléliques) avec l'hétérozygotie attendue à l'équilibre H_{eq} (calculée à partir du nombre d'allèles), il est possible d'observer un excès d'hétérozygotie en cas de « bottleneck » et un déficit d'hétérozygotie en cas d'expansion (Cornuet et Luikart 1996). Les résultats par locus sont analysés par un test de Wilcoxon. Nous utiliserons ce principe avec le logiciel Bottleneck (Piry *et al.* 1999). Nous paramétrons le logiciel avec un modèle mutationnel TPM (Two Phase Model) autorisant des sauts multiples de motifs (0% de SMM) dont la variance suit une loi géométrique estimée à 0,36.

3. Test d'isolement démographique des deux sous-espèces

Pour avoir des précisions sur la présence de migration ou non, nous utiliserons un outil qui compare la vraisemblance des deux modèles possibles (dérive ou équilibre immigration/dérive) sachant nos données (Ciofi *et al.* 1999). Chacun des 2 modèles explore différentes valeurs de t/N et de $N*m$ (avec t le temps de divergence, N la taille efficace de la population et m le taux de migrants). Le calcul est fondé sur la théorie de la coalescence et la simulation de chaîne de Monte Carlo. Cette approche est implémentée dans le logiciel 2MOD (Ciofi *et al.* 1999);(Streiff *et al.* 2005). La mutation n'est pas simulée dans cette analyse. Au vu des temps de divergence et des tailles efficaces du Criquet pèlerin, cette hypothèse n'est pas très réaliste et cette analyse devra être complétée par d'autres (IMA par exemple).

Nous choisissons de lancer plusieurs fois les calculs pour vérifier la convergence de chacune des chaînes de Monte-Carlo. Chaque calcul est constitué d'un « burn-in » de 10 000 (valeurs retirées du calcul) parmi 100 000 simulations pour s'assurer d'avoir une bonne mixité (fluctuations des valeurs de vraisemblance). Cependant, l'analyse n'a pas abouti. Les valeurs de vraisemblance de six simulations (en annexe 5) illustrent le fait qu'il n'a pas été possible d'avoir une bonne mixité de la chaîne de Monte-Carlo. Il n'y a aucun moyen de changer les paramètres de calcul de 2MOD pour améliorer la mixité des simulations. Il n'est alors pas possible de tester l'isolement démographique de nos deux sous-espèces.

4. Estimation de l'histoire évolutive neutre des sous-espèces : temps de divergence et tailles efficaces

a. Présentation de l'analyse utilisée

L'analyse que nous avons utilisé pour étudier la divergence des deux sous-espèces s'appuie sur la théorie de la coalescence et sur de l'ABC (Approximate Bayesian Computation) afin d'inférer de nombreux paramètres de nuisance et les caractéristiques démographiques. Cette méthode permet d'obtenir une distribution des paramètres d'intérêts sans expliciter les calculs de vraisemblance. Suivant le principe de l'ABC, 500 000 données simulées sont produites à partir de lois de répartition *a priori* peu informées des paramètres que l'on souhaite estimer (voir les sections suivantes b et c). Ensuite, pour chaque simulation, nous avons calculé 3 statistiques résumées caractérisant la variation génétique intra-populationnelle et moyennées sur les loci : le nombre d'allèles, l'hétérozygotie attendue (Nei 1987), et la variance de taille allélique (en paires de bases) ainsi que deux estimateurs de la différenciation génétique : le F_{st} (Weir and Cockerham, 1984) et la distance de Golstein (Goldstein *et al.* 1995). Pour chacune de ces simulations, les valeurs simulées des statistiques résumées sont comparées aux valeurs réellement observées par un algorithme de rejet/sélection et seules les simulations les plus proches sont conservées (1%). Ces simulations sont alors sélectionnées et utilisées afin de tracer, après régression, une loi de répartition *a posteriori* des

Tableau 4 : Modèle mutationnel choisi en « prior » pour DIYABC

		transcriptomique	génomique	transcriptomique	génomique
	Loi	Minimum		Maximum	
Taux de mutation moyen	Uniforme	10^{-5}	10^{-4}	10^{-4}	10^{-3}
Taux de mutation par locus	Gamma (2)*	10^{-6}	10^{-5}	10^{-3}	10^{-2}
Paramètre P de la loi géométrique moyen	Uniforme	0,1		0,5	
Paramètre P de la loi géométrique par locus	Gamma (2)*	0,1		0,9	
Taux moyen d'insertion de base dans la région flanquante	Log-uniforme	10^{-8}		10^{-5}	
Taux d'insertion de base dans la région flanquante par locus	Gamma (2)*	10^{-9}		10^{-4}	

Le paramètre P détermine la variance de la loi géométrique utilisé pour modéliser le changement du nombre d'unités répétées (sous un GSM).

* Le paramètre de la loi Gamma correspond au paramètre de densité de probabilité.

paramètres. Indirectement, il est alors possible d'avoir une estimation plus ou moins précise des paramètres démographiques qui correspondent à notre jeu de données observées.

Pour cette étude, nous avons ainsi utilisé le logiciel DIYABC 2.0 dont l'interface permet de simuler des données complexes (Cornuet *et al.* 2008). DIYABC présente, en fait, plusieurs atouts. Entre autres, il propose l'utilisation d'un modèle GSM (Generalized Stepwise Mutation Model ; (Zhivotovsky *et al.* 1997)) symétrique, et la possibilité d'avoir des événements d'insertion-délétion dans la région flanquante, ce qui est soupçonné pour beaucoup de nos marqueurs (Pascual *et al.* 2007). Ce type de modèle, plus complexe, correspond bien mieux aux marqueurs microsatellites des Orthoptères, et rend l'analyse plus réaliste. À notre connaissance, parmi les approches pour estimer la divergence entre les populations ou les tailles effectives de populations, le traitement de DIYABC est le seul à modéliser un GSM (contrairement à IMA).

Un autre avantage, nouveau dans la version 2.0 de DIYABC, encore non disponible mais pour laquelle nous avons été beta-testeurs au cours de ce stage, est qu'il est possible d'avoir plusieurs catégories de marqueurs caractérisés par des modèles mutationnels différents (classiquement pour analyser conjointement des données de génotypes microsatellites et de variation de séquences nucléiques). Dans notre cas, il est particulièrement intéressant de pouvoir définir un modèle mutationnel différent pour chacun des types de marqueurs « transcriptomiques » et « génomiques » dans l'analyse, puisqu'il est hautement probable qu'ils ne partagent pas les mêmes mécanismes mutationnels (Prasad, 2005).

Nous avons utilisé comme jeu de donnée les 13 populations génotypées sur 6 jeux de marqueurs (Tch7c retirée pour problème de consanguinité). En plus des génotypes, le logiciel DIYABC nécessite l'implémentation de deux modèles *a priori* : le modèle mutationnel et le modèle populationnel. Il nous faut informer la distribution *a priori* de chacun des paramètres de ces modèles.

b. Le modèle mutationnel : information « réaliste » des paramètres de nuisance

Pour déterminer les priors décrivant les processus mutationnels, nous avons conservé les valeurs par défaut de DIYABC à l'exception du taux de mutation que nous avons tenté d'informer de manière réaliste pour chacun des types de marqueurs. Cependant, pour un même scénario populationnel, nous avons fait 3 analyses différentes : deux avec les marqueurs « transcriptomiques » « génomiques » séparés et une avec les deux types de marqueurs en même temps. Les résultats sont similaires avec une estimation des paramètres démographiques qui est plus précise avec les deux types de marqueurs ensemble (écart-type plus petit). Du coup, nous ne présentons que les résultats de cette analyse avec les deux types de marqueurs.

Dans un premier temps, nous avons utilisé l'information disponible sur les processus mutationnels des microsatellites chez les humains (Lai et Sun 2003). En utilisant la relation qui existe entre longueur des microsatellites et taux de mutation, nous avons cherché à estimer quel devrait être le taux de mutation de nos marqueurs. Le taux de mutation estimé est environ $4,2 \cdot 10^{-4}$ [$4,2 \cdot 10^{-3}$; $2 \cdot 10^{-4}$] pour les marqueurs génomiques (longueur moyenne de 23 motifs) et $7,5 \cdot 10^{-5}$ [$3 \cdot 10^{-4}$; $4 \cdot 10^{-5}$] pour les marqueurs transcriptomiques (longueur moyenne de 13 motifs) (Fig.11 et Tableau 3). Nous avons alors décidé d'informer le taux de mutation avec une fenêtre de prior de [10^{-4} ; 10^{-3}] pour les microsatellites génomiques et [10^{-5} ; 10^{-4}] pour les microsatellites d'origine transcriptomiques. Par la suite, pour chaque modèle populationnel que nous testerons, nous garderons le même modèle mutationnel (Tableau 4). Dans un second temps, nous espérons informer les distributions *a priori* des

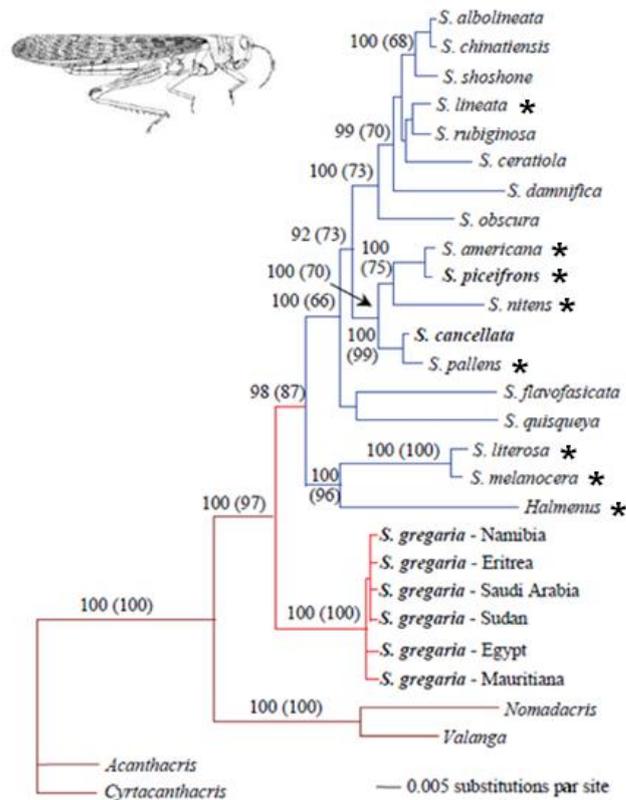


Figure 12 : Arbre phylogénétique des espèces du genre *Schistocerca*, basé sur l'ADN mitochondrial (ND1-ARnt-12S-16S) (Lovejoy 2006). Les étoiles correspondent aux espèces génotypées par le laboratoire pour informer la contrainte de taille de nos marqueurs.

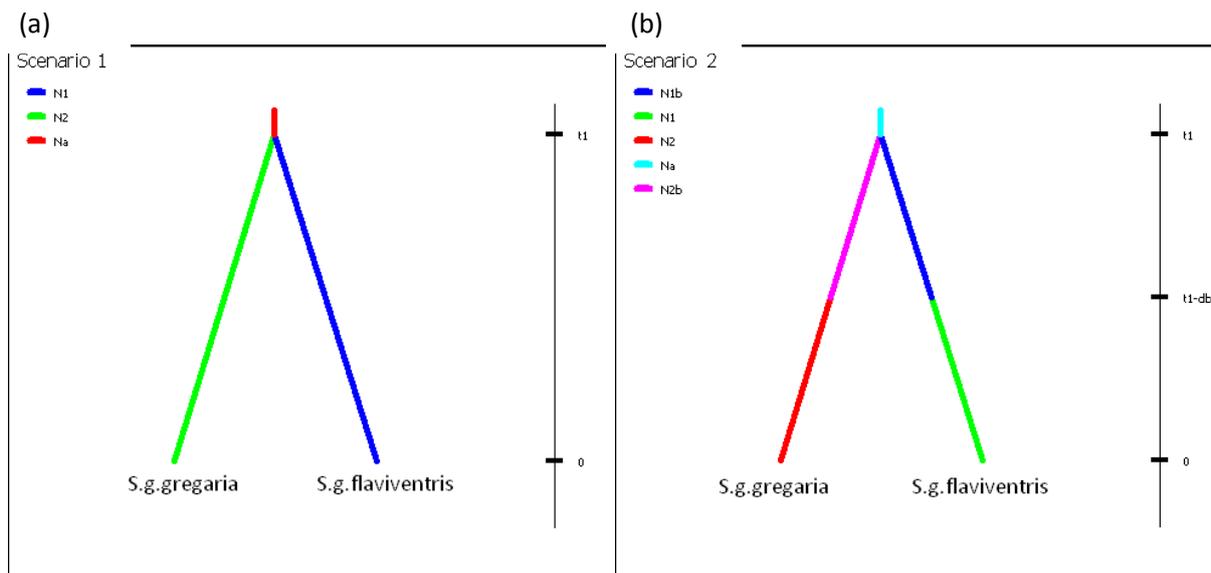


Figure 13 : Les deux modèles populationnels testés sur DIYABC : (a) un scénario de divergence avec démographie constante : les deux sous-espèces de taille $N1$ (*S.g. flaviventris*) et $N2$ (*S.g. gregaria*) divergent il y a un temps $t1$ d'une population ancestrale de taille Na ; (b) scénario similaire à (a) avec démographie variable : les deux sous-espèces augmentent de taille, db générations après la divergence, passant des tailles $N1b$ et $N2b$ à $N1$ et $N2$.

taux de mutations par la mesure expérimentale du taux de mutation. De cette manière, en spécifiant un modèle mutationnel propre à nos marqueurs et à notre espèce, nous cherchons à avoir une estimation qui correspond mieux à la réalité. DIYABC requiert une information *a priori* sur le taux moyen de mutation mais permet aussi qu'une variation par locus soit possible.

D'autre part, dans la paramétrisation de DIYABC, il est nécessaire de fournir une estimation de la contrainte de taille propre à chaque marqueur utilisé, obtenue à partir de l'écart maximum de taille allélique (en nombre de motifs). La valeur par défaut est de 40 motifs et des valeurs de 20 à 60 états continus sont traditionnellement choisies pour les études de simulation et d'inférence basées sur les données microsatellites (Verdu *et al.* 2009). Cependant, comme les microsatellites acridiens sont longs (Chapuis *et al.* 2012), l'équipe a obtenu une estimation précise en génotypant quelques individus de 8 espèces proches de *Schistocerca gregaria* (Fig.12). Avoir une idée des allèles, qui existent chez d'autres espèces, c'est-à-dire dans un contexte de divergence assez grande, fournit des informations sur la contrainte de taille maximale que les marqueurs peuvent présenter. En fait, à partir des allèles observés chez les différentes espèces, il est possible d'informer la contrainte de taille de chaque marqueur grâce à une relation approchée par Pollock *et al.* (Pollock *et al.* 1998) selon la formule $(W_k * (K+1)) / (K-1) + 1$, avec K le nombre de populations divergentes et $W_k = \text{allèle max} - \text{allèle min}$ (l'intervalle de taille en nombre de motif) (Tableau 3).

c. Le modèle populationnel : un modèle de divergence des deux sous-espèces

L'analyse issue de DIYABC fournit une estimation des paramètres composés, c'est-à-dire des paramètres originaux liés à la mutation : $NS * \mu$ et $t * \mu$. A partir de ces paramètres qui devraient être relativement bien informés grâce aux données observées, il est envisageable d'obtenir une estimation des paramètres démographiques comme le temps de divergence t et les tailles efficaces N_1 de *S. g. flaviventris* et N_2 de *S. g. gregaria*.

Le modèle populationnel que nous souhaitons tester est assez simple (Fig. 13a) : les deux sous-espèces de taille efficace N_1 et N_2 ont divergé à un temps t_1 . Avant ce temps, les deux sous-espèces formaient une population ancestrale de taille efficace N_a . Nous avons choisi des lois de répartition de valeurs *a priori* non informatives pour ces paramètres : N_a , N_1 et N_2 en Log-uniforme s'étendant sur [1000 ; 10 000 000] et t en Log-uniforme sur [100 ; 1 000 000]. Le temps de divergence s'exprime en nombre de générations, en considérant 3 générations par an pour le Criquet pèlerin. Cette distribution *a priori* inclut le scénario suggéré par l'évolution du climat africain, décrite en II.1.a et en figure 6. En effet, sous ce scénario, nous nous attendons à une divergence aux alentours de 9 000 à 21 000 ans avant notre ère, avec la fin du dernier âge glaciaire (Uvarov, 1966).

Pour évaluer la qualité du modèle, nous utilisons des analyses incluses dans DIYABC. Nous nous intéressons à l'ACP calculée à partir des 500 000 simulations et sur laquelle est projetée également la valeur correspondant aux données observées, ainsi que les 5000 simulations retenues pour le calcul du posterior des paramètres. Si la valeur des données observées se retrouve au milieu du nuage de point, le modèle est considéré comme correct. Par ailleurs, nous vérifions aussi que les statistiques résumées simulées retenues dans le posterior soient proches des statistiques observées.

d. Analyse de sensibilité du modèle à la variation de taille efficace des populations

A partir des données bioclimatiques africaines, il est facile d'imaginer que la divergence des deux sous-espèces, dont la cause est la réduction de leurs aires de répartition, puisse s'accompagner

d'une réduction des tailles efficaces (Fig.6b), suivi d'une augmentation de tailles efficaces jusqu'à nos jours (Fig. 6a). Notre objectif est, ici, de mesurer la sensibilité de l'estimation des paramètres démographiques à la possibilité d'une variation de taille : réduction de taille suivie d'une expansion. Nous pouvons nous attendre à un temps de divergence plus petit dans ce cas, puisque la différenciation génétique actuelle sera expliquée par un temps de coalescence moins long du fait de la présence d'un « bottleneck ». De ce fait, nous avons souhaité ajouter au modèle précédent un paramètre de taille supplémentaire pour chacune des deux sous-espèces. Le nouveau modèle populationnel à tester est présenté en figure 13b. Nous avons choisi d'utiliser des priors informés par l'histoire biogéographique africaine : la réduction de tailles des deux sous-espèces est concomitante à la divergence et dure db générations. La distribution a priori de la période db est une Log-uniforme [3 000 ; 12 000] . Les nouveaux paramètres de tailles efficaces $N1b$ et $N2b$ sont informés a priori par une Log-uniforme sur [10 ; 100 000] et contraints à être inférieurs aux valeurs de tailles $N1$ et $N2$ du modèle populationnel originel. Par le choix des valeurs de db , nous considérons que la réduction de taille a une durée égale pour les deux sous-espèces qui se situe entre 1 000 et 4 000 ans après l'évènement de divergence

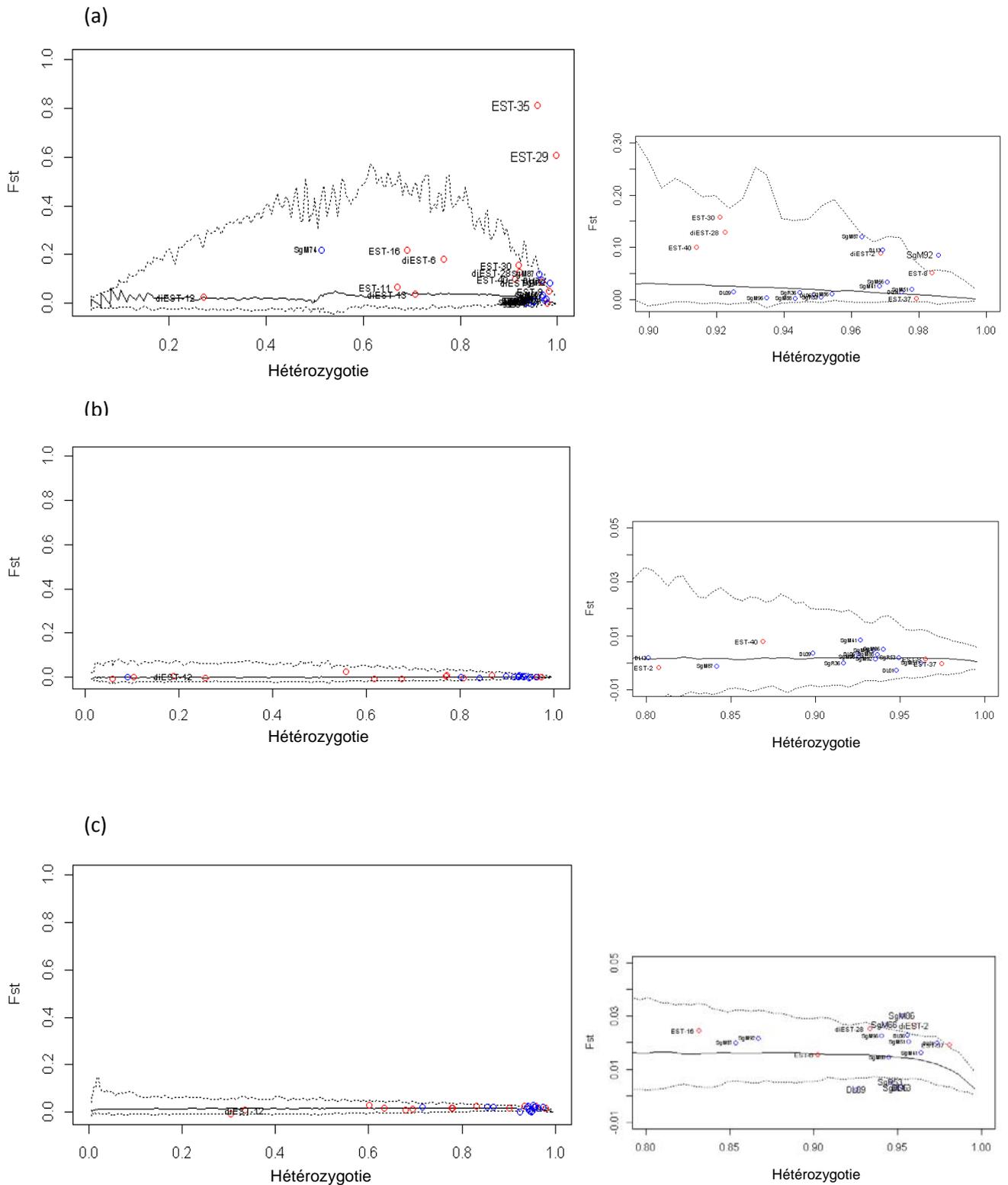


Figure 14 : Sorties graphiques en sortie de l'analyse Fdist 2 (avec un zoom à droite). (a) le test entre *S. g. gregaria* et *S. g. flaviventris*, (b) le test entre les populations de *S. g. flaviventris* et (c) le test entre populations de *S. g. gregaria*. Les traits pleins représentent la médiane et les pointillés l'enveloppe à 95% des valeurs simulées pour des marqueurs qui ne sont pas sous sélection.

IV. Résultats

1. Description de la variation génétique des populations et construction du jeu de données pour l'étude de la divergence

a. Conformité aux équilibres d'Hardy-Weinberg et de liaison

Nous avons vérifié les écarts aux hypothèses d'Hardy-Weinberg. Après correction FDR, 46 tests apparaissent significatifs parmi 723 (6,4% des tests). Nous constatons que, parmi ces 46 cas, la grande majorité (30) n'est pas associée préférentiellement à une population ou à un marqueur. Cependant, 10 tests parmi 28 sont significatifs pour la population Tch7c et 6 tests parmi 14 sont significatifs pour le locus Sgr36. De plus, le locus Sgr36 ne présente des allèles nuls que sur la sous-espèce *S. g. flaviventris* (moyenne de 16%). Si l'on retire la population Tch7c et le locus Sgr36 de l'analyse, le nombre de tests significatifs est alors inférieur à 5% (4,4%).

D'autre part, nous nous sommes intéressés au déséquilibre de liaison. Après correction FDR, 150 tests sont significatifs parmi 7233 (2,1%). Cette faible proportion ne paraît pas très inquiétante d'autant plus que ces tests ne sont pas associés préférentiellement à un locus ou une paire de loci. Cependant, 60 tests parmi 378 (15,9 %) pour la population Tch7c et 52 tests parmi 378 (13,8%) pour la population GG sont significatifs.

b. Test de marqueurs sous sélection

Nous avons testé la présence de marqueurs sous sélection au sein des populations échantillonnées. Ces tests se font sous l'hypothèse de populations reliées par des flux de gènes (équilibre migration/dérive) au travers de $F_{dist} 2$ (Beaumont et Nichols 1996), puisque DETSEL ne fonctionne pas. Les sorties de $F_{dist} 2$ sont présentées en Figure 14. Pour le test entre les deux « sous-espèces », il est possible de constater que diEST29 et diEST35 sont clairement en dehors de l'enveloppe des 95% (Fig. 14a). Ils sont sous sélection divergente et présentent un niveau de différenciation entre les sous-espèces de 0,61 et 0,81 respectivement. Nous pouvons remarquer que ce sont deux marqueurs « transcriptomiques » qui sont observés sous sélection, confirmant la littérature (Ellis, 2007).

Pour ce test et les deux tests au sein des sous-espèces, quelques autres loci, tous avec des hétérozygoties H_e très élevées, sont à l'extérieur mais très proches de l'enveloppe de confiance qui se rétrécit considérablement pour H_e supérieur à 0,9: SgM92 (Fig.14a), SgM66, SgM86, diEST2, Sgr53, Sgr36, DL09, DL13 (Fig.14c). Nous considérons que ces marqueurs ne sont pas problématiques.

D'autre part, il est curieux que le test de $F_{dist} 2$ ne détecte pas le marqueur diEST12 sous sélection stabilisante, bien qu'il soit peu diversifié et que les distributions de fréquences alléliques soient similaires entre les 2 sous-espèces (Annexe 6). Nous suspectons un problème de puissance de détection de la sélection stabilisante mais en attendant les résultats des analyses DETSEL, nous conservons ce marqueur dans notre jeu de données.

En conclusion des analyses de déséquilibre, le locus Sgr36 (écart à Hardy-Weinberg, allèles nuls) et les loci diEST29 et diEST35 (sous sélection divergente) sont retirés du jeu de données pour les analyses suivantes. La population Tch7c (écart à Hardy-Weinberg et déséquilibre de liaison) présente probablement de la consanguinité ou des individus apparentés et est également retirée

Tableau 5 : Statistiques descriptives des deux sous-espèces *S. g. gregaria* et *S. g. flaviventris* (en gras les populations qui sortent significatives au test de Wilcoxon)

		Hétérozygotie attendue		Richesse allélique			Hétérozygotie attendue He	Richesse allélique (pour 9 individus)	
		He		(pour 9 individus)					
		Moyenne <i>S.g.flaviventris</i>	Moyenne <i>S.g.gregaria</i>	Moyenne <i>S.g.flaviventris</i>	Moyenne <i>S.g.gregaria</i>				
Génomiques	SgM51	0,962	0,939	13,45	11,88	SgBB	0,765	8,96	
	SgM92	0,934	0,850	11,06	8,59	SgCC	0,781	8,90	
	SgM41	0,919	0,949	10,75	12,85	SgDD	0,783	9,09	
	SgM74	0,091	0,701	1,72	7,05	SgEE	0,776	8,65	
	SgM66	0,929	0,917	11,32	11,06	SgFF	0,777	8,84	
	SgM96	0,924	0,921	10,28	10,97	SgGG	0,779	8,80	
	SgM87	0,842	0,837	8,36	8,25	Moyenne <i>S.g.flaviventris</i>	0,777	8,87	
	SgM88	0,933	0,931	11,06	11,40	AlSg1NBEZH	0,852	9,70	
	SgM86	0,936	0,925	11,34	12,14	AlSgOA	0,850	9,77	
	DL01	0,950	0,955	12,23	13,76	AlSgOT	0,511	3,10	
	DL06	0,922	0,935	10,73	11,91	SgN	0,856	10,33	
	DL09	0,896	0,924	11,71	13,41	MRSgAO	0,840	9,60	
	SgR53	0,948	0,942	16,29	16,30	MRSgFd	0,783	7,82	
	DL13	0,800	0,949	8,10	17,49	MRSgln	0,850	9,44	
	Transcriptomiques	diEST-30	0,767	0,765	5,63	5,38	SgJ	0,847	10,04
		diEST-16	0,257	0,811	2,29	7,56	SgO	0,857	10,42
diEST-11		0,623	0,624	3,62	4,60	SgP	0,863	10,48	
diEST-8		0,964	0,889	13,55	9,12	MRSgTa	0,835	9,63	
diEST-40		0,862	0,767	7,97	5,20	MRSgTi	0,842	9,35	
diEST-6		0,544	0,686	4,08	4,22	MRSgTr	0,845	9,27	
diEST-28		0,681	0,913	4,01	10,73	Soud	0,846	9,51	
diEST-13		0,764	0,583	6,75	4,26	Tch7a	0,858	10,26	
diEST-12		0,189	0,332	2,11	2,70	Tch7b	0,847	10,09	
diEST-2		0,809	0,936	8,70	11,98	Tch7c	0,840	9,45	
diEST-37		0,976	0,962	14,71	14,39	TSg1BO	0,834	9,14	
Moyenne		0,777	0,838	8,873	9,888	TSg2BO	0,843	9,39	
						TSgFa	0,841	9,24	
					TSgKe	0,831	9,47		
					NSgTa	0,851	10,53		
					Moyenne <i>S.g.gregaria</i>	0,838	9,89		

Tableau 6 : Indice de différenciation F_{st} observé pour les sous-espèces et pour trois populations particulières. *S.g.gregaria** correspond à la sous-espèce toute entière incluant les individus des trois populations particulières (AlSgOT, MRSgFd, SgTch7c). *S.g.gregaria* est constituée des individus de 19 populations sans les trois populations. Les valeurs de F_{st} en gras avec un astérisque sont significatives.

	Fst global	Intervalle de confiance
<i>S.g. flaviventris</i>	0,0029	[0,0001;0,0051]
<i>S.g. gregaria</i>	0,0013	[0,0004;0,0022]
<i>S.g. gregaria</i> *	0,019*	[0,0152;0,0226]
	Fst par pair	Intervalle de confiance
<i>S.g. gregaria</i> - <i>S.g. flaviventris</i>	0,066*	[0,0391;0,0926]
<i>S.g. gregaria</i> *- <i>S.g. flaviventris</i>	0,064*	[0,041;0,0869]
<i>S.g. gregaria</i> -SgTch7c	0,0063	[0,0045;0,0083]
<i>S.g. gregaria</i> -MRSgFd	0,0191*	[0,0122;0,0273]
<i>S.g. gregaria</i> -AlSgOT	0,144*	[0,1255;0,1622]

du jeu de données. Pour la suite, le jeu de données sera donc constitué de 25 marqueurs et de 27 populations.

c. Diversité génétique des populations

L'hétérozygotie attendue H_e , le nombre d'allèles échantillonnés et la richesse allélique A_r calculés par F_{stat} sont en Tableau 5 (tableau complet en annexes 8, 9 et 10). Comme nous pouvons le constater, toutes ces statistiques descriptives sont assez équivalentes entre nos populations.

Cependant, la population AISgOT apparaît significativement différente de toutes les autres populations ($p < 0,0008$ pour H_e et $p < 0,0001$ pour A_r). Elle possède une diversité génétique ($H_e = 0,511$ et $A_r = 3,1$) plus petite que les autres ($H_e = 0,829$ et $A_r = 9,5$). A partir du test de déséquilibre démographique calculé à partir de l'hétérozygotie, cette même population possède un excès d'hétérozygotie significatif ($p = 0,002$), ce qui suggère un goulot d'étranglement récent subi par la population. La population MRSgFd semble avoir une diversité génétique plus faible que les autres populations ($H_e = 0,783$ pour une moyenne de 0.838 sur les autres populations) mais le test est non-significatif.

Si l'on s'intéresse à la comparaison des deux sous-espèces (Tableau 5), les statistiques résumées H_e et A_r ne montrent pas de différence significative ($p = 0,66$ pour H_e et $p = 0,33$ pour A_r).

d. Structuration génétique entre populations

Les résultats de l'indice de différenciation F_{st} sont présentés en tableau 6 et annexe 11. Deux populations de *S. g. gregaria* présentent une différenciation significative par rapport à la sous-espèce (Tableau 6) : AISgOT (F_{st} par pair = 0,144), MRSgFd (F_{st} par pair = 0,0191). Ces deux populations ont subi des événements démographiques qui sont alors différents de ceux de l'espèce. Comme nous l'avons vu précédemment, AISgOT présente notamment des signes d'effondrement de taille de population.

Au niveau des sous-espèces, il n'y a quasiment aucune différenciation ($F_{st\ global}(flaviventris) = 0,0029$ et $F_{st\ global}(gregaria) = 0,0013$) (Tableau 6). Par contre, il est possible d'observer une structuration assez importante entre les deux sous-espèces (F_{st} par pair = 0,066). Cette différenciation traduit un isolement relatif de *S. g. flaviventris* par rapport à *S. g. gregaria*.

En conclusion de ces résultats sur la diversité et la structuration génétique, nous décidons de retirer de nos données les deux populations AISgOT et MRSgFd pour la suite des analyses afin d'éviter de biaiser nos résultats avec ces populations qui ont une histoire démographique particulière. Pour la suite, le jeu de données sera donc constitué de 25 marqueurs et de 25 populations.

2. Estimation de l'histoire évolutive neutre des sous-espèces : temps de divergence et tailles efficaces

a. Comparaison des modèles démographiques

La qualité d'estimation du modèle de divergence simple des deux sous-espèces est comparée à celle du modèle introduisant une variation de taille efficace après la divergence des deux sous-espèces. En figure 15 se trouvent les ACPs de la projection des valeurs simulées par rapport aux données observées. Comme nous pouvons le constater pour le modèle de divergence simple (Fig.15a), le point correspondant à nos données (observées) se situe en marge du nuage de points tirés de la distribution a posteriori. En figure 15b, nous constatons que le point correspondant à nos

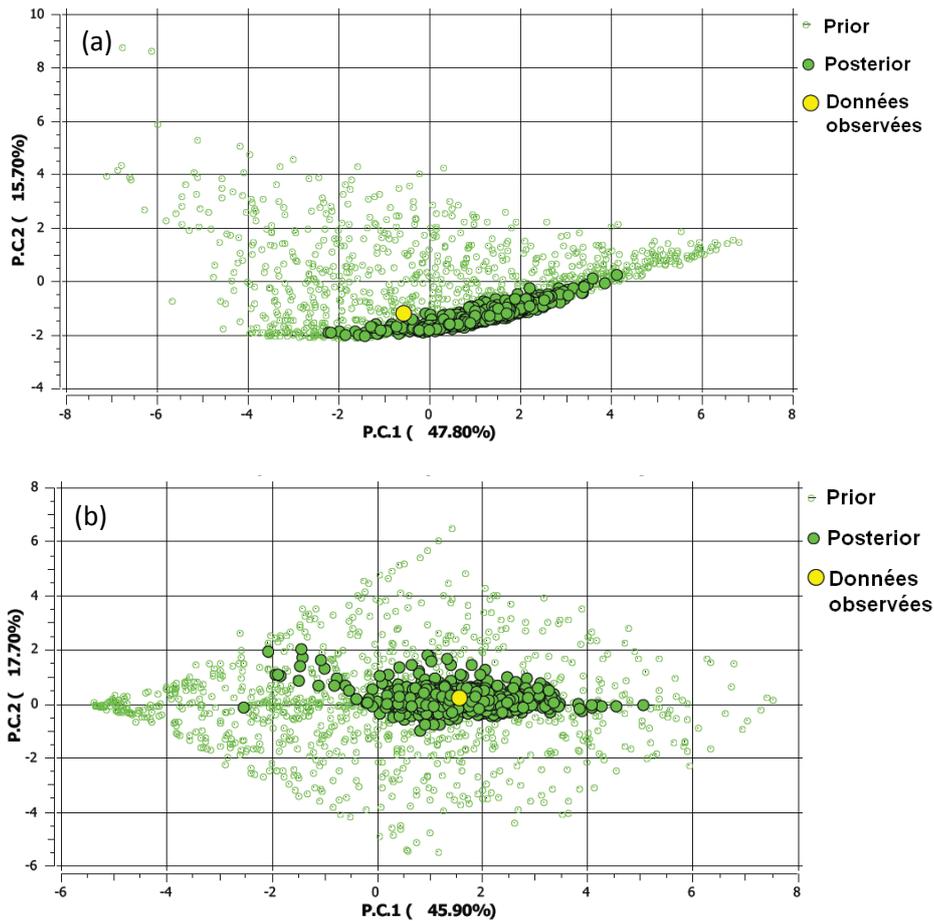


Figure 15 : Analyse de qualité du modèle par l'intermédiaire d'une ACP (logiciel DIYABC) : (a) modèle de divergence avec un ancêtre commun, (b) modèle avec l'ajout d'un goulot d'étranglement. Un modèle est satisfaisant lorsque le point des données observées est projeté dans le nuage des points des données simulées.

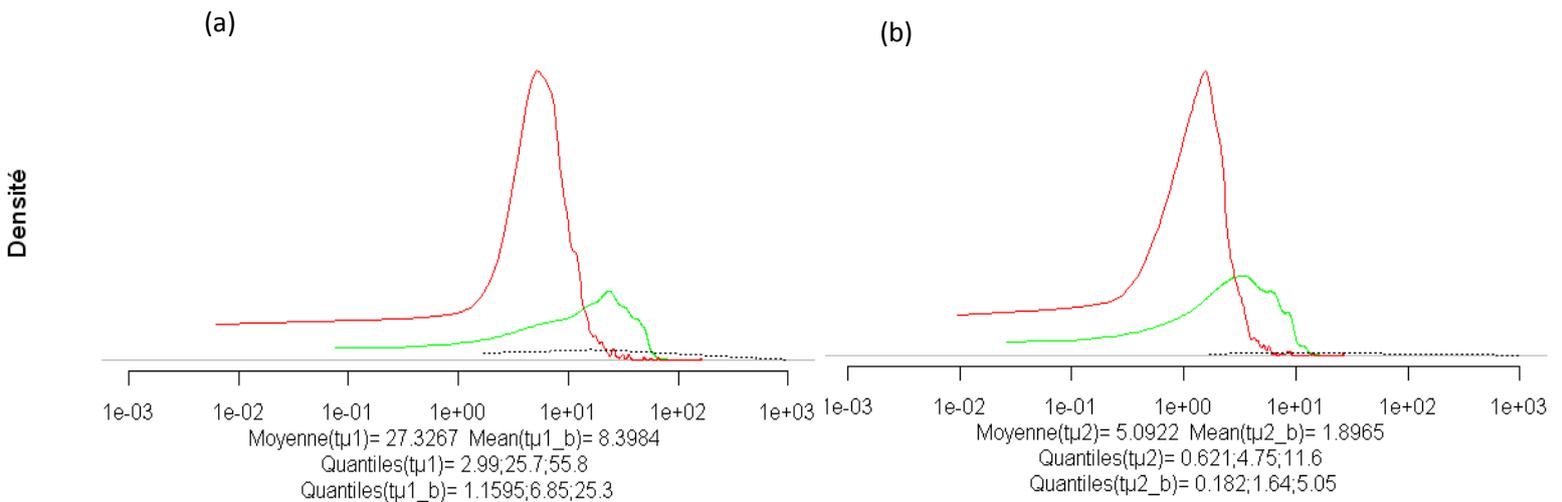


Figure 16 : Posterior des paramètres τ_1 (a :génomique) et τ_2 (b :transcriptomique) du modèle (vert) de divergence le plus simple et du modèle (rouge) avec un goulot d'étranglement (logiciel DIYABC). (Les quantiles présentés sont ceux à 0,025 et 0,975 des valeurs ; τ_{1_b} et τ_{2_b} sont les paramètres du modèle avec goulot d'étranglement).

données observées est quand à lui bien centré, à la fois au milieu des simulations explorées à partir des « priors » mais surtout au milieu des simulations « posteriors ». Le modèle de divergence avec un changement de taille efficace simultané à la divergence semble donc mieux expliquer nos données que le modèle simple.

La comparaison des valeurs observées et simulées des statistiques une à une et pour le modèle de divergence simple nous informent sur la cause de sa mauvaise adéquation avec nos données. En effet, les hétérozygoties attendues H_e observées de *S.g. flaviventris* sont plus petites que celles retenues dans les simulations ($p=0,008$ pour les marqueurs génomiques et $p=0,02$ pour les transcriptomiques). Cette tendance existe pour *S.g. gregaria* mais n'est pas significative ($p=0,1025$ pour les génomiques et $p=0,0565$ pour les transcriptomiques). De la même manière, la valeur de F_{st} observée entre les deux sous-espèces est plus grande que les F_{st} simulés ($p=0,9945$ et $p=0,9955$). Il semble ainsi qu'un modèle incluant une variation de tailles des populations soit plus réaliste pour nos données. Ce résultat de DIYABC est confirmée par les résultats des tests de présence d'un effondrement ou d'une expansion de population (logiciel Bottleneck) qui mettent en évidence un déficit d'hétérozygotie de façon significative pour les deux sous-espèces ($p=0,0005$ pour *S.g.flaviventris* et $p=0,0007$ pour *S.g.gregaria*). Sur les 23 marqueurs testés, 8 marqueurs présentent un déficit d'hétérozygotie de façon significative pour *S.g. flaviventris* et 9 marqueurs pour *S.g. gregaria* ($p<0,05$). Ce résultat suggère qu'il existe un signal d'expansion au sein des deux sous-espèces.

Le choix de ce modèle démographique n'est pas anodin pour l'estimation du temps de divergence. En effet, si nous nous intéressons aux paramètres composés $t\mu_1$ (génomique) et $t\mu_2$ (transcriptomique) (Fig.16) qui sont comparables entre nos deux modèles, nous observons que les deux paramètres de divergence sont plus petits dans le modèle de variation des tailles des populations. De plus, sous ce même modèle, leur estimation est plus précise. Les temps de divergence correspondant exprimés en nombre de générations sont [3 630 ; 80 100] dans le modèle à démographie variable et [7 610 ; 330 650] dans le modèle à démographie constante.

b. Estimation des valeurs de paramètres pour le modèle démographique retenu

Etant donné que le modèle présentant une variation de taille efficace semble avoir plutôt une bonne qualité d'estimation, nous nous intéressons aux posteriors des paramètres évalués associés (valeurs en tableau 7 et graphiques en annexe 12). Le temps de divergence est situé à 7830 ans [1210 ; 26 700], en considérant 3 générations par an. Les paramètres composés $Ns*\mu_1$ et $Ns*\mu_2$ sont respectivement 2,6 et 1,7 fois plus grand pour *S.g.gregaria* que pour *S.g.flaviventris*. De ce fait, la taille efficace de *S.g.gregaria* est plus grande que celle de *S.g.flaviventris* : médiane et intervalles de confiance respectivement à 1 885 000 [312 000 ; 9 330 000] et à 1 200 000 [175 475 ; 8 610 000] individus. Ce rapport est accentué lors du goulot d'étranglement pour les paramètres composés $N1b*\mu_1$ et $N2b*\mu_2$ et originaux $N1b$ et $N2b$, puisque la taille efficace basse de *S.g.gregaria* est dix fois plus grande que celle de *S.g.flaviventris*.

D'autre part, il est étonnant de voir que même en permettant une variation de taille efficace lors de la divergence, les paramètres composés $Na*\mu_1$ et $Na*\mu_2$ et la taille efficace N_a sont très petits. Les données semblent avoir suffisamment d'information pour estimer à 5500 individus [1000 ; 59 312] la taille efficace de la population ancestrale.

Tableau 7 : Présentation des estimations « postérieurs » de nos paramètres à partir du modèle de divergence avec variation de la taille efficace (en sortie de DIYABC)

Paramètres composés		Moyenne	Médiane	ModE	Quantile à 2,5%	Quantile à 97,5%
Na* μ 1 ancêtre	<i>Na*μ1</i>	3,19	1,55	0,39	0,21	18,3
NS* μ 1 Sg flaviventris	<i>N1*μ1</i>	564,58	242	82,7	33,1	3560
NS* μ 1 Sg gregaria	<i>N2*μ1</i>	1160,27	628	240	79	5390
Nb* μ 1 Sg flaviventris	<i>N1b*μ1</i>	1,9	0,35	0,00519	0	15,15
Nb* μ 1 Sg gregaria	<i>N2b*μ1</i>	9,44	3,19	0,0249	0,02	54,7
	<i>t*μ1</i>	8,4	6,85	4,7	1,16	25,3
	<i>db*μ1</i>	2,47	2,01	1,41	0,56	7,26
Na* μ 2 ancêtre	<i>Na*μ2</i>	0,73	0,34	0,139	0,04	3,66
NS* μ 2 Sg flaviventris	<i>N1*μ2</i>	118,11	59,2	43,6	8,46	696
NS* μ 2 Sg gregaria	<i>N2*μ2</i>	182,94	101	50,5	16,7	758
Nb* μ 2 Sg flaviventris	<i>N1b*μ2</i>	0,43	0,09	0,000817	0	3,42
Nb* μ 2 Sg gregaria	<i>N2b*μ2</i>	1,73	0,76	0,0144	0	7,53
	<i>t*μ2</i>	1,9	1,64	1,64	0,182	5,05
	<i>db*μ2</i>	0,51	0,47	0,421	0,138	1,07
Paramètres démographiques						
temps de divergence	<i>t</i>	29595	23500	22100	3620	80100
durée du bottleneck	<i>db</i>	7140	6810	4320	3240	11700
taille efficace de la population ancestrale	<i>Na</i>	11052	5500	1000	1000	59312
taille efficace de Sg flaviventris	<i>N1</i>	1961828	1200000	781000	175475	8610000
taille efficace de Sg gregaria	<i>N2</i>	2808129	1885000	826000	312000	9330000
taille efficace réduite de Sg flaviventris	<i>N1b</i>	6298	1310	NA	14	49700
taille efficace réduite de Sg gregaria	<i>N2b</i>	24594	12300	NA	49	93200
Paramètres mutationnels						
taux de mutation des marqueurs génomiques	<i>μ1</i>	3,30E-04	2,90E-04	1,92E-04	1,12E-04	8,12E-04
taux de mutation des marqueurs transcriptomiques	<i>μ2</i>	6,80E-05	7,10E-05	8,65E-05	2,50E-05	9,80E-05
Paramètre de la loi géométrique modélisant le changement du nombre de motifs pour les marqueurs génomiques	<i>P1</i>	3,17E-01	3,19E-01	3,14E-01	1,23E-01	4,93E-01
Paramètre de la loi géométrique modélisant le changement du nombre de motifs pour les marqueurs transcriptomiques	<i>P2</i>	4,44E-01	4,63E-01	4,88E-01	2,49E-01	5,17E-01
Taux d'insertion dans la région flanquante	<i>SN1</i>	2,70E-06	2,00E-06	2,62E-07	7,00E-08	9,20E-06
Taux d'insertion dans la région flanquante	<i>SN2</i>	2,50E-06	1,40E-06	5,88E-08	4,00E-08	9,60E-06

c. Information des paramètres mutationnels : confirmation expérimentale des taux de mutation des marqueurs microsatellites utilisés

i. Résultat des tests préliminaires pour le développement de la méthode

Lors des tests préliminaires, nous avons observé que parmi les 15 marqueurs microsatellites « génomiques », deux marqueurs, DL01 et DL09, présentent des profils GeneMapper rendant la détection de l'allèle « mutant » introduit difficile (environ 20% d'erreur). Pour ces deux marqueurs, nous risquons de sous-estimer le taux de mutation et nous ne les considérerons donc pas par la suite. Par ailleurs, pour les autres marqueurs, il est possible de lire les allèles mutants la plupart du temps, y compris dans des situations difficiles (mutant long ou dans le stutter d'un allèle proche, voir des exemples Fig.17). Mais nous n'avons pas vu l'allèle « mutant » dans 11 cas. Ces 11 erreurs peuvent s'expliquer par la concentration trop faible de l'allèle mutant lors du « pool » des quatre extraits d'ADN. En effet, pour ces 11 cas, l'allèle mutant était dans une concentration bien inférieure à la concentration théorique de 1/16 (Fig.18). La concentration est en fait le seul facteur ayant un impact significatif sur l'aire relative du pic mutant sur GeneMapper (GLM, $p=0,005$) (Fig. 18). Pour notre expérience d'estimation du taux de mutation, nous avons été particulièrement vigilants à l'ajustement systématique de la concentration au-dessus de $1/8^{\text{ème}}$.

ii. Taux de mutation observé par marqueur

Parmi les 13 marqueurs « génomiques » dont nous souhaitons estimer le taux de mutation, nous avons observé une cinquantaine de profils « suspects » sur GeneMapper dont 10 très sérieux (Fig. 19 et annexe 13). Si nous considérons ces 10 mutations comme sûres, il est alors possible d'estimer que les marqueurs DL06 et SgM74 ont un taux de mutation de $5 \cdot 10^{-4}$ (2 mutations) et Sgr53, Sgr36, SgM51, SgM66, SgM96 et DL13 un taux de mutation de $2,5 \cdot 10^{-4}$ (1 mutation). Ces taux de mutation peuvent être utiles pour affiner le modèle mutationnel des analyses mais doivent, avant tout, être confirmés (avec les autres profils « suspects ») par un génotypage individuel qui sera fait prochainement.

Le taux de mutation global, calculé comme le rapport entre le nombre de mutations et le nombre d'événements de méioses multiplié par le nombre de locus testés ($10/(4000 \times 13)$), est de ($2 \cdot 10^{-4}$). Ce taux est très proche du taux de mutation médian estimé par DIYABC ($2,9 \cdot 10^{-4}$) et du taux estimé à partir de la longueur en motifs ($4,2 \cdot 10^{-4}$) (Lai et Sun, 2003). Ainsi, il n'y a pas nécessité, à ce jour, de tester de nouveaux priors du taux de mutation dans DIYABC.

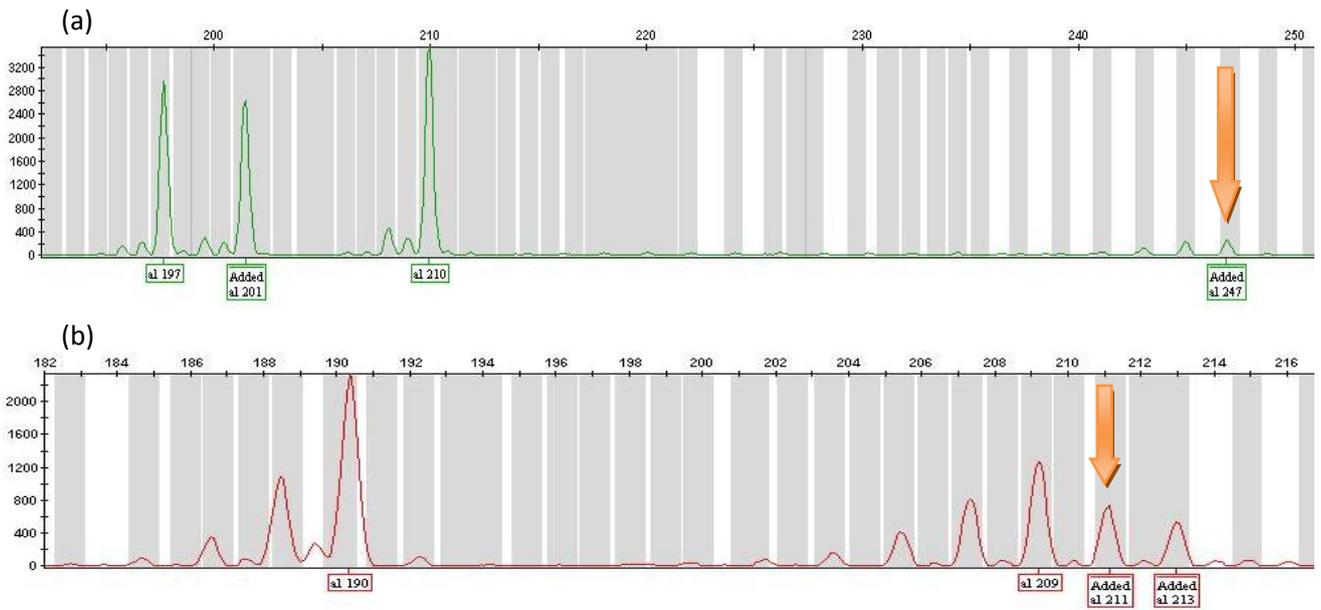


Figure 17 : Profil GeneMapper des tests de détection de l'allèle « mutant » (indiqué par la flèche) dans deux difficiles : (a) un allèle « mutant » haut (locus SgM87) et (b) allele mutant dans un « stutter » (locus SgM96) lors des tests préliminaires de « pool » de quatre individus entre eux (8 copies de gènes).

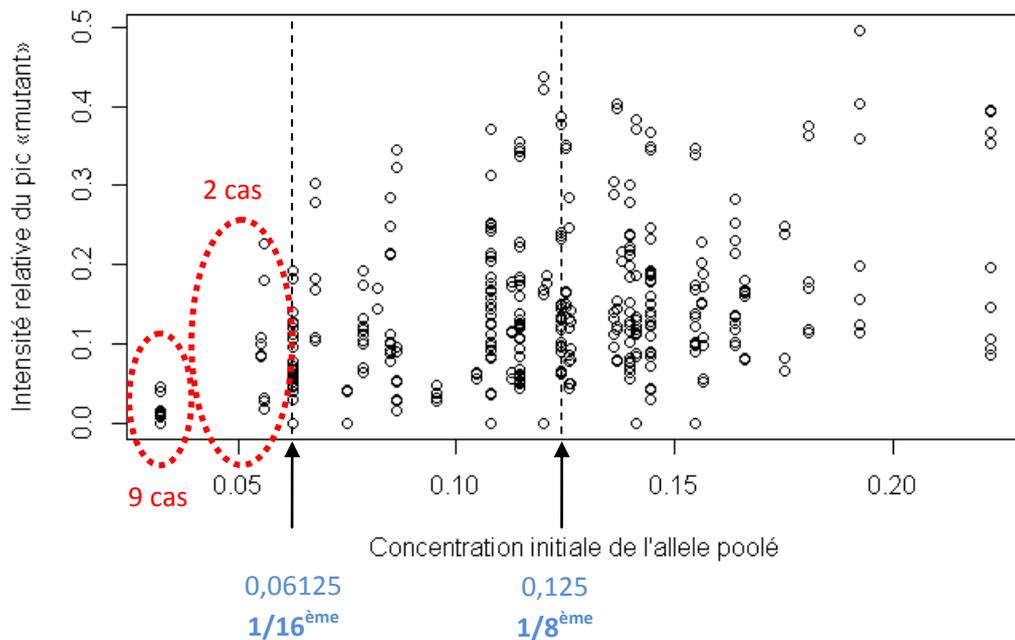


Figure 18 : Relation entre l'intensité relative du pic « mutant » sur GeneMapper et la concentration de l'allèle mutant dans l'échantillon. Les 11 cas correspondant à des mutants non détectés non vus sont sous la concentration de $1/16^{\text{ème}}$.

V. Discussion

1. Conclusions sur la structuration génétique

a. Absence de structuration au sein des sous-espèces

A partir de 25 marqueurs microsatellites et de 28 populations, notre étude a montré que la structuration au sein de chaque sous-espèce du Criquet pèlerin est faible, voire inexistante. Ce résultat n'a rien d'étonnant face à la capacité de dispersion de cet insecte, notamment lors des grandes périodes de pullulations (centaines de kms par jour)(Fig. 2). Nous savons que les populations se déplacent en fonction des fluctuations saisonnières pour trouver des habitats. En période d'invasion, ces mouvements peuvent s'étendre de l'Afrique de l'Ouest à l'Afrique de l'Est (Waloff, 1966 ; suivi satellitaire FAO 2012). Pourtant, deux populations de la sous-espèce *S. g. gregaria* présentent une différenciation génétique significative. AISgOT et MRSgFd présentent une signature génétique de fondation ou de goulot d'étranglement (faible diversité et test de Bottleneck significatif pour AISgOT). Les événements de fondation-extinction de populations dans des environnements fluctuants tels que les milieux semi-arides peuvent induire rapidement une différenciation génétique. Nos résultats font écho à la présence de structuration génétique observée en Erythrée par Ibrahim *et al* (2000).

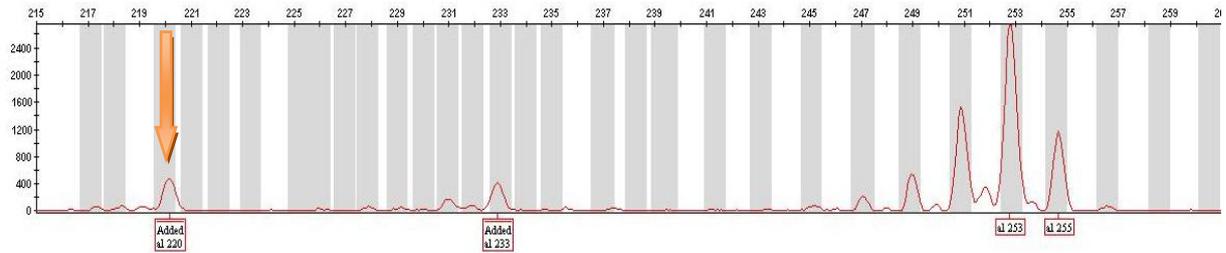
b. Une différenciation entre sous-espèces probablement sous-estimée

Le niveau de différenciation génétique observée entre sous-espèces est significatif (6,6%). Etudier la divergence entre les deux sous-espèces est alors cohérent. Cette valeur de différenciation peut paraître peu élevée. Cependant, Hedrick montre que le niveau de différenciation génétique entre population estimé par le F_{st} est maximisé par le niveau d'hétérozygotie(Hedrick 1999). Afin d'avoir une meilleure estimation du niveau de différenciation pour nos marqueurs qui ont des hétérozygoties élevées, il est possible d'utiliser un F_{st} « corrigé ». La mesure standardisée du F_{st} proposée par Meirmans (Meirmans 2006) est défini comme le rapport entre le F_{st} de Weir and Cockerham (1984) et le F_{st} maximal qui peut être lu si tous les allèles de nos données étaient différents entre les deux sous-espèces. Après correction, nous mettons en évidence un niveau de différenciation génétique de 0.31, donc élevé, entre nos sous-espèces.

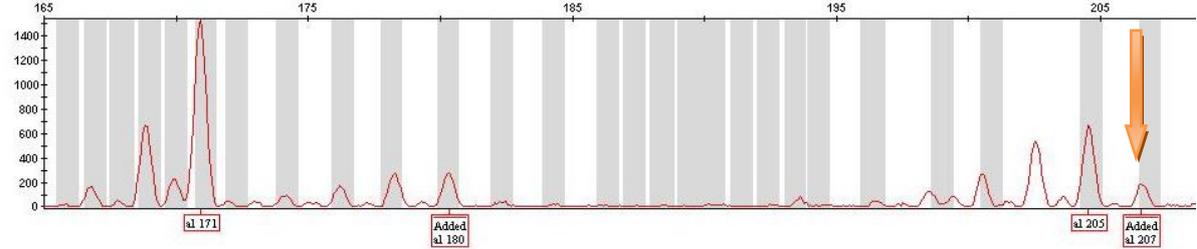
2. Conclusions sur le modèle et les paramètres démographiques

Les analyses sous DIYABC montrent que le scénario démographique de divergence qui explique le mieux nos données est celui qui permet une variation de taille efficace après la divergence. Cette variation est en fait une expansion démographique de chaque sous-espèce peu après la divergence et confirme les hypothèses paléoclimatiques. Par la suite, il sera possible de tester ces différents scénarii en les comparant directement dans DIYABC. Ce serait plus élégant qu'une simple comparaison de la qualité de l'estimation. Ce résultat est également confirmé par l'analyse utilisée dans Bottleneck (Cornuet et Luikart 1996). Cette analyse permet de visualiser un événement récent qui remonte à $2N-4N$ générations avec N représentant la taille efficace de la population (Piry *et al* 1999). Il est ainsi cohérent de visualiser une variation de taille efficace sur la longue période qui nous sépare de la divergence, vu que les tailles efficaces sont de plusieurs centaines de milliers d'individus pour l'espèce.

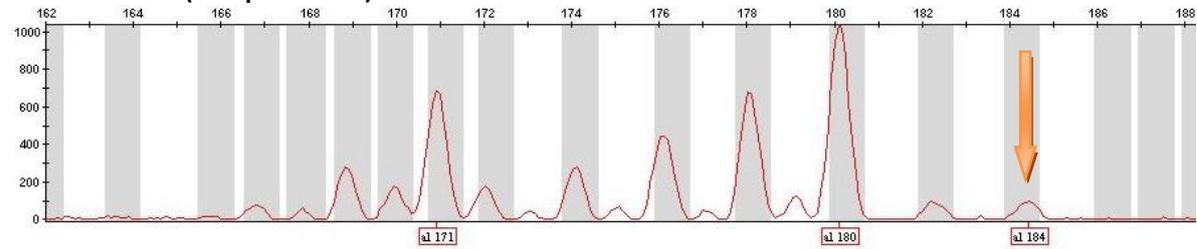
Allèle mutant 220 (marqueur DL13)



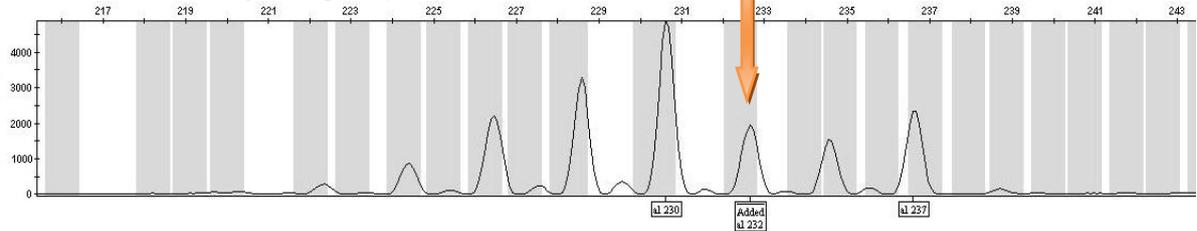
Allèle mutant 207 (marqueur DL06)



Allèle mutant 184 (marqueur DL06)



Allèle mutant 232 (marqueur SgM51)



Allèle mutant 360 (marqueur Sgr36)

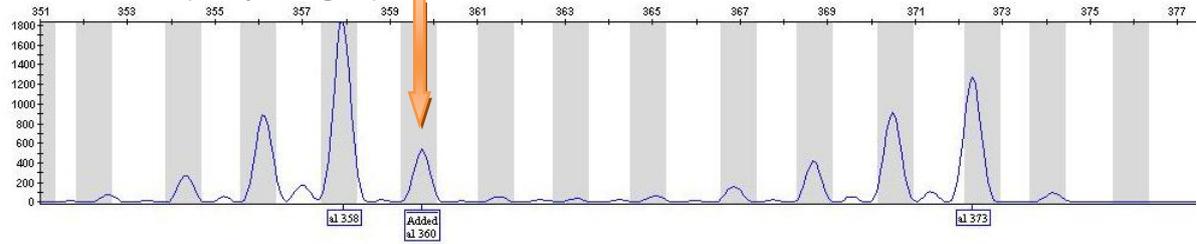


Figure 19 : Cinq premiers profils GeneMapper où l'on suspecte la présence de mutations de façon très sérieuse

En effet, sous ce scénario, les valeurs estimées des tailles efficaces actuelles sont élevées, bien qu'imprécises pour les deux sous-espèces : 1 200 000 [175 475 ; 8 610 000] pour *S. g. flaviventris* et 1 885 000 [312 000 ; 9 330 000] pour *S. g. gregaria*. La taille de *S. g. gregaria* est plus élevée que celle de *S. g. flaviventris*, en accord avec les indices génétiques observés. Les tailles efficaces pendant le goulot d'étranglement sont mal estimées mais celle de *S. g. gregaria* est 10 fois plus grande. Ceci est un peu étonnant vu que l'aire de répartition de *S. g. flaviventris* est censée être plus stable que l'aire de *S. g. gregaria* du point de vue de l'évolution du climat africain. Cependant, l'aire de *S. g. flaviventris* est également bien plus petite dès l'origine et les populations peuvent ainsi être plus fortement affectées même par une réduction d'aire d'habitat moins importante.

Le temps de divergence associé à ce scénario est centré sur 7 800 ans et est inférieur à 27 000 ans (intervalle : [1210 ; 26 700]). Cela confirme plutôt bien que l'événement de divergence a eu lieu entre le dernier âge glaciaire et notre époque. La médiane du temps de divergence place de façon cohérente la divergence lorsque l'aire potentielle de répartition du Criquet pèlerin était la plus réduite (Fig. 6b). On peut noter qu'une source d'imprécision de l'estimation du temps de divergence provient de la valeur utilisée pour le nombre annuel de générations de Criquet pèlerin. Nous avons postulé une moyenne de trois générations par an en se référant à la biologie actuelle. Mais ce nombre varie sûrement entre les deux sous-espèces et surtout, il a certainement évolué au cours des périodes paléoclimatiques.

3. Deux sources de biais potentielles dans nos résultats

a. La non prise en compte de la migration

Le fait de ne pas avoir pris en compte la possibilité de migration dans l'analyse DIYABC a pu insérer un biais dans nos résultats. Les signaux qui nous ont poussés à considérer une variation de taille efficace dans notre modèle pourrait provenir de l'absence de migration dans le modèle. Etudier la présence de migration, en absence de mutation, comme sur 2MOD, est imparfait puisque le temps de divergence et les tailles efficaces du Criquet pèlerin sont élevés, ainsi que probablement le taux de mutation aux marqueurs microsatellites. Pour tenir compte de la migration, l'utilisation du logiciel IMA (Hey et Nielsen 2007), fondé sur un modèle de deux populations en équilibre migration-mutation-dérive est pertinente. L'estimation qui en ressortirait serait intéressante. Cependant, cette analyse s'appuie sur un SMM, modèle mutationnel très différent de celui de nos marqueurs (Kimura et Ohta 1978). Il est peu vraisemblable qu'un modèle SMM représente bien la variation microsatellite observée chez le Criquet pèlerin. D'ailleurs, d'autres analyses utilisées au cours de ce stage, telle que Fdist par exemple, sont fondés sur ce modèle mutationnel irréaliste.

b. Des résultats sensibles au modèle mutationnel

Dans notre analyse DIYABC, nous nous attendons aussi à ce que notre choix de modèle mutationnel influence nos résultats. Alors que la valeur du paramètre de la loi géométrique du GSM, P , est suspectée être de 0,22 pour les microsatellites humains, nous avons choisi une fenêtre *a priori* pour ce paramètre large [0,1 ; 0,5]. Ce choix a été conditionné par les distributions *a posteriori* de ce paramètre pour les marqueurs transcriptomiques qui butait sur la droite lorsque le prior était limité par l'intervalle par défaut [0,1 ; 0,3]. Nous redoutons que les valeurs de P et nos choix sur les priors influencent les statistiques résumées de la même manière que pour la différence SMM/GSM (voir résultats (Chapuis *et al.* 2010) (Fig.20). En effet, dans ce dernier cas, on observe des réponses différentes du nombre d'allèles et des hétérozygoties aux modèles mutationnels, qui créent un

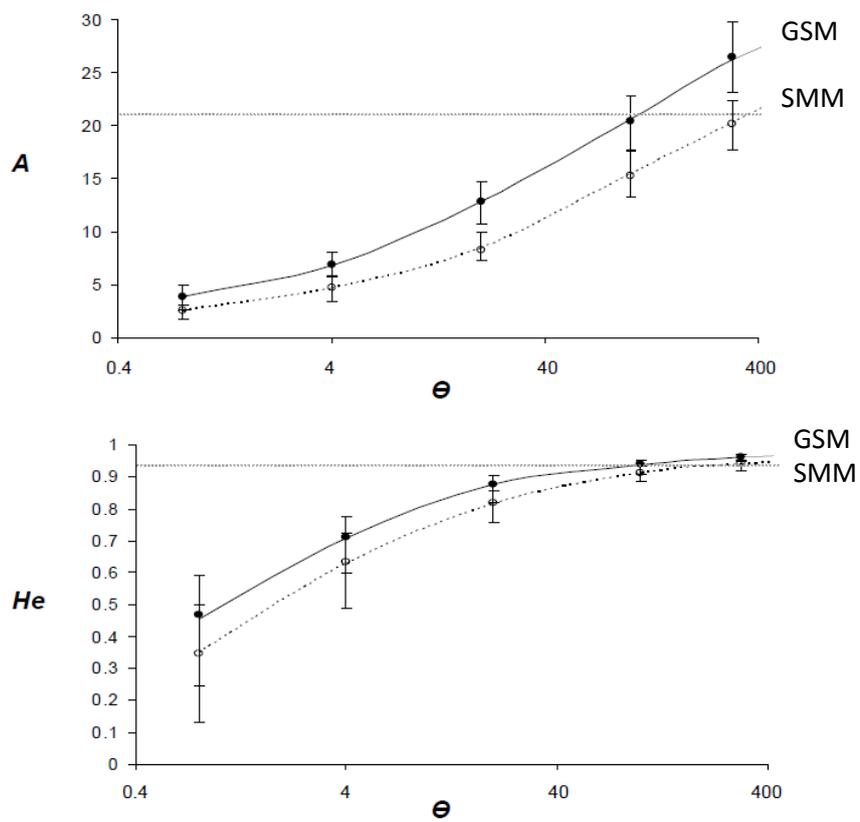


Figure 20 : Effets du modèle mutationnel sur les statistiques résumées (les moyennes (ligne) sont calculées à partir de 50 valeurs et les barres d'erreur correspondent au quantile 2,5 et 97,5%) (Chapuis *et al*, 2010)

d'hétérozygotie' tout comme une expansion démographique. . L'influence du paramètre P sur les statistiques résumées mérite une attention particulière pour faire face à ce problème. Des simulations additionnelles sur DIYABC viennent d'invalider cette hypothèse puisqu'un prior standard pour le paramètre P ([0,1 ; 0,3]) donne des résultats similaires à ceux trouvés au cours du stage.

4. Une estimation cohérente du taux de mutation de nos marqueurs

Il est important d'avoir une estimation des paramètres de notre modèle mutationnel qui corresponde à notre espèce pour que les résultats soient plus réalistes. C'est ce que nous avons commencé à faire dans ce travail avec le taux de mutation. Le taux global estimé de 2.10^{-4} paraît cohérent avec les résultats de DIYABC et les estimations de la littérature. Ce qui est rassurant pour notre étude. Mais le seuil de détection de notre expérience est encore trop faible pour informer tous nos marqueurs (surtout pour les marqueurs courts et les marqueurs transcriptomiques). Il faut alors par la suite augmenter le nombre d'individus génotypés. C'est en partie possible rapidement car nous avons encore en réserve 1500 œufs frères et sœurs des individus d'élevage génotypés. Cependant, il faut être particulièrement vigilant à l'influence des conditions d'élevage et de stress sur le taux de mutation (Fonville, 2011). Il ne sera peut-être pas facile de reproduire exactement les mêmes conditions, ce qui pourrait changer un peu le taux de mutation.

5. Une meilleure connaissance des processus neutres pour calibrer les études sur les traits liés à la phase

L'échelle spatiale de la structuration génétique au sein de chaque sous-espèce permet de définir les unités génétiques qui guideront, avec les données de l'environnement, le choix des populations à échantillonner pour des études expérimentales. Nos résultats montrent l'impossibilité d'utiliser des réplicats génétiques dans le cas présent, puisque la grande majorité des populations de chaque sous-espèce partagent la même distribution de fréquences alléliques (non différenciées).

Grâce aux résultats de l'analyse de divergence des deux sous-espèces de *S. gregaria*, nous avons une idée de l'intensité de la force de dérive qui s'exerce sur chaque sous-espèce. Cette force de dérive n'est pas très grande puisque le temps de divergence est petit (médiane à 7800 ans) et que les tailles efficaces des sous-espèces sont très grandes (médiane à plus d'un millions d'individus). Il est maintenant possible grâce à ces estimations d'informer les futures analyses sur la force de dérive et de mieux discerner les forces évolutives. En particulier, les valeurs de paramètres démographiques des deux population (temps de divergence et tailles efficaces) seront précieuses pour modéliser de manière réaliste les processus évolutifs neutres agissant dans les populations dans le cadre (1) du développement de modèles démo-génétiques pour évaluer l'influence d'un trait divergent (ex comportement grégaire) sur la dynamique des populations et (2) de la paramétrisation de tests de détection de gènes sous sélection (ex sur données génomiques).

6. Perspectives méthodologiques pour mieux informer l'histoire évolutive des deux sous-espèces de *S. g. gregaria*

a. Intégrer dans l'analyse des marqueurs génétiques qui informent sur des temps anciens

L'utilisation de séquences ADN peut être complémentaire de l'utilisation de marqueurs microsatellites. Grâce à la version 2 de DIYABC, il est possible de reproduire nos analyses de

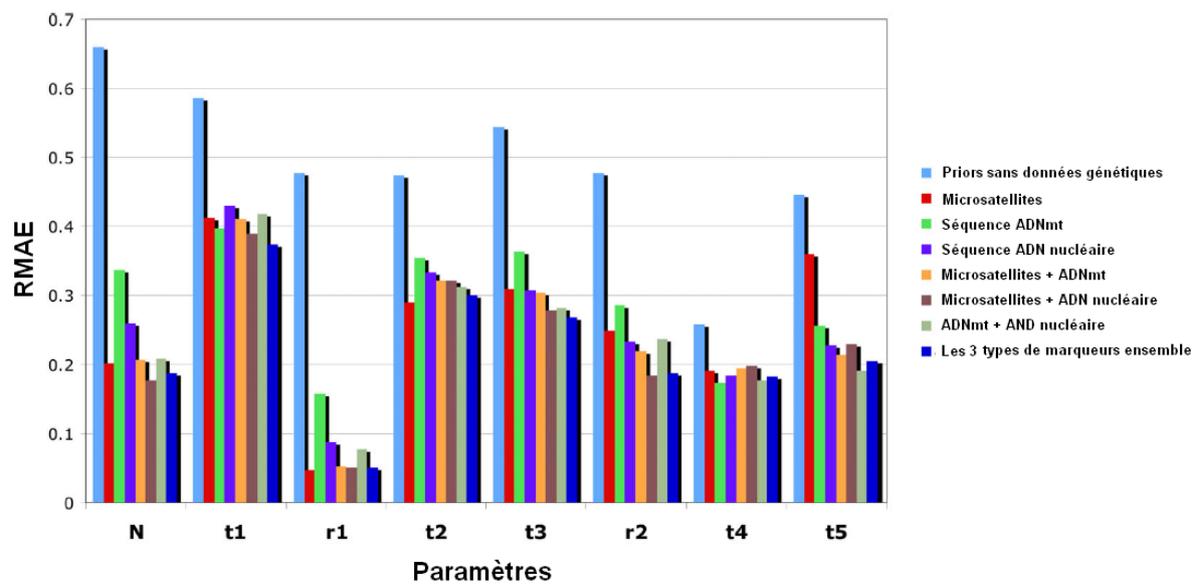


Figure 21 : Influence des différents marqueurs sur l'erreur médiane absolue (RMAE) de l'estimation des paramètres démographiques (testé dans DIYABC) (Cornuet *et al*, 2010)

divergence en utilisant ces deux types de marqueurs en même temps. Les séquences sont un meilleur indicateur des temps de divergence très grand (Fig.21) (Cornuet *et al.* 2010).

Une possibilité pourrait être de développer des séquences mitochondriales sur le Criquet pèlerin. Certaines séquences utilisées classiquement telles que les séquences correspondant aux sous-unités ribosomales (28S 5.8S 18S par exemple) ou au COI (Cytochrome oxydase I) peuvent être très intéressantes car elles sont bien connues (métagénomique, barcoding, etc...). Cependant, ce type de séquence ne paraît pas le plus adapté dans le cas présent. En effet, la variabilité intraspécifique observée est insuffisante. Des analyses préliminaires récentes (Lovejoy 2006 ; équipe CIRAD) ont montré que les différences de séquences entre individus de *S. g. gregaria* et de *S. g. flaviventris* sont infimes. De plus, de nombreux Orthoptères, et notamment le Criquet pèlerin, présentent des introgressions de séquences mitochondriales dans le génome nucléaire Numts (Nuclear Mitochondrial Pseudogenes) (Bensasson *et al.* 2000);(Zhang et Hewitt 1996). Il est alors difficile lors de l'amplification ADN de véritablement distinguer la séquence mitochondriale parmi toutes les séquences observées, si ce n'est par un effort intense de clonage. Des tentatives pour trouver des amorces spécifiques des versions mitochondriales de Col (Moulton *et al.* 2010) ont échoué jusqu'à présent sur le Criquet pèlerin. Pour ces raisons, toutes les séquences mitochondriales sont peu utilisables.

Les séquences nucléaires constituent une autre source de marqueurs génétiques. Les ITS (Internal Transcribed Spacer), sont très étudiés et souvent assez variables. Dans le cas du Criquet pèlerin, les ITS ne sont pas utilisables car ils présentent des insertions-délétions dans les introns qui empêchent un bon alignement. Les séquences de gène (ou de leurs introns associés) sont moins bien connus, souvent propres à chaque espèce et étudiées pour leur lien avec des traits sous sélection. Un enjeu actuel au laboratoire est d'identifier des gènes nucléaires opérationnels chez les criquets et ayant de bonnes résolutions phylogéographiques. Des pistes sont à l'étude dans l'équipe : Facteur d'élongation 1, Actine et GAPDH. Par ailleurs, la séquence Homéobox utilisée par Ibrahim *et al* (2000) serait une piste à ce sujet. En parallèle, l'utilisation des techniques de séquençage de nouvelle génération serait une voie prometteuse pour cibler plus efficacement un grand nombre de gènes.

b. Ajouter une population plus divergente dans l'analyse

Pour préciser la divergence entre les deux sous-espèces, il pourrait être intéressant d'avoir une autre espèce du genre *Schistocerca* dans les analyses. Cela est possible car les marqueurs microsatellites possèdent, en fait, une assez bonne amplification croisée chez les espèces proches du genre *Schistocerca*. Treize d'entre eux amplifient sur une population de 31 individus de *Schistocerca americana* et sont polymorphes. Les marqueurs transcriptomiques se sont révélés être meilleurs à ce sujet que les génomiques, en accord avec la littérature (Ellis 2007, Kim, 2008).

Lovejoy *et al.* (2006) ont observé un taux de divergence d'au moins 3-4% entre les séquences de *S. gregaria* et les autres espèces du genre *Schistocerca* présentes en Amérique. La date de séparation des autres espèces remonterait ainsi à au moins 3 millions d'années en prenant un taux de divergence constant de 1,1% par million d'années (Brower 1994). La divergence entre *Schistocerca americana* et l'ancêtre des deux sous-espèces pourrait servir de référence dans le modèle. Il est possible que l'ajout de cet évènement calibré dans DIYABC, bien que beaucoup plus ancien, permette de mieux estimer le temps de divergence entre les deux sous-espèces de *Schistocerca gregaria*.

BIBLIOGRAPHIE

- (COPR), A., 1982. The locust and grasshopper agricultural manual. London, COPR: -690.
- Adams, J. (2001). "AFRICA DURING THE LAST 150,000 YEARS." 2012, from <http://geoecho.snu.ac.kr/nercAFRICA.html>.
- Badisco, L., J. Huybrechts, G. Simonet, H. Verlinden, E. Marchal, R. Huybrechts, L. Schoofs, A. De Loof et J. Vanden Broeck, 2011. Transcriptome Analysis of the Desert Locust Central Nervous System: Production and Annotation of a *Schistocerca gregaria* EST Database. *PLoS ONE* 6(3): e17274.
- Beaumont, M. A. et R. A. Nichols, 1996. Evaluating Loci for Use in the Genetic Analysis of Population Structure. **263**: 1619-1626.
- Benjamini, Y. et Y. Hochberg, 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57(1): 289-300.
- Bensasson, D., D.-X. Zhang et G. M. Hewitt, 2000. Frequent Assimilation of Mitochondrial DNA by Grasshopper Nuclear Genomes. **17**: 406-415.
- Botha, 1969. Locust and their control in South Africa. *Farming in South Africa* 45: 110-116.
- Botha, D. H., 1967. Some phase characteristics of the Southern African form of the Desert locust (*Schistocerca gregaria*, Forskål). *South African Journal of Agricultural Science* 10: 61-76.
- Brader, L., Djibo, H, Faye, FG, 2006. Towards a More Effective Response to Desert Locusts and their Impacts on Food Security, Livelihoods and Poverty. *Multilateral Evaluation of the 2003–05 Desert Locus*(FAO).
- Brower, A. V., 1994. Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. **91**: 6491-6495.
- Chapuis, M.-P., M. Lecoq, Y. Michalakis, A. Loiseau, G. A. Sword, S. Piry et A. Estoup, 2008. Do outbreaks affect genetic population structure? A worldwide survey in *Locusta migratoria*, a pest plagued by microsatellite null alleles. *Molecular Ecology* 17(16): 3640-3653.
- Chapuis, M.-P., J.-A. M. Popple, K. Berthier, S. J. Simpson, E. Deveson, P. Spurgin, M. J. Steinbauer et G. A. Sword, 2010. Challenges to assessing connectivity between massive populations of the Australian plague locust. **278**: 3152-3160.
- Chapuis, M. P. et A. Estoup, 2007. Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution* 24(3): 621-631.
- Chapuis, M. P., R. Streiff et G. A. Sword, 2011. Long microsatellites and unusually high levels of genetic diversity in the Orthoptera. *Insect Molecular Biology* 21(2): 181-186.
- Ciofi, C., M. A. Beaumontf, I. R. Swingland et M. W. Bruford, 1999. Genetic divergence and units for conservation in the Komodo dragon *Varanus komodoensis*. **266**: 2269-2274.
- COPR, 1982. The locust and grasshopper agricultural manual. London, COPR: -690.
- Cornuet, J.-M., V. Ravigne et A. Estoup, 2010. Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). **11**: 401.
- Cornuet, J. M. et G. Luikart, 1996. Description and Power Analysis of Two Tests for Detecting Recent Population Bottlenecks From Allele Frequency Data. **144**: 2001-2014.

- Cornuet, J. M., F. Santos, M. A. Beaumont, C. P. Robert, J.-M. Marin, D. J. Balding, T. Guillemaud et A. Estoup, 2008.** Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics* **2008**: 2713-2719.
- Coulibaly, I., K. Gharbi, R. G. Danzmann, J. Yao et C. E. Rexroad, 2005.** Characterization and comparison of microsatellites derived from repeat-enriched libraries and expressed sequence tags. *Animal Genetics* **36**(4): 309-315.
- Crozier, R. H., B. Kaufmann, M. E. Carew et Y. C. Crozier, 1999.** Mutability of microsatellites developed for the ant *Camponotus consobrinus*. *Molecular Ecology* **8**(2): 271-276.
- Despland, E. et S. J. Simpson, 2000.** The role of food distribution and nutritional quality in behavioural phase changes in the Desert locust. *Animal Behaviour* **59**: 643-652.
- Dirsh, V. M., 1953.** Morphometrical studies on phases of the Desert locust (*Schistocerca gregaria* Forskål). *Anti-Locust Bulletin* **16**: 1-34.
- Dirsh, V. M., 1974.** Genus *Schistocerca* (Acridomorpha, Insecta). Den Haag, Dr W. Junk BV: -240.
- Ellis, J. R. et J. M. Burke, 2007.** EST-SSRs as a resource for population genetic analyses. *Heredity* **99**(2): 125-132.
- Ellis, P. E., 1962.** Behavioural mechanisms in locust gregariousness. *Symposia genetica et biologica Italica* **X**: 225-234.
- Farrow, R. A., 1975.** The African Migratory locust in its main outbreak area of the middle Niger: quantitative studies of solitary populations in relation to environmental factors. *Locusta* **11**: 1-198.
- Faure, J. C., 1932.** The phases of Locusts in South Africa. *Bulletin of Entomological Research* **23**(3): 293-405.
- Fonville, N. C., R. M. Ward et D. Mittelman, 2011.** Stress-Induced Modulators of Repeat Instability and Genome Evolution. *Journal of Molecular Microbiology and Biotechnology* **21**(1-2): 36-44.
- Goldstein, D. B., A. Ruiz Linares, L. L. Cavalli-Sforza et M. W. Feldman, 1995.** An evaluation of genetic distances for use with microsatellite loci. **139**: 463-71.
- Goudet, J. (2002).** "Fstat 2.9.3.2." from <http://www2.unil.ch/popgen/softwares/fstat.htm>.
- Gregory, T. R. (2012).** "Animal Genome Size Database." from <http://www.genomesize.com>.
- Hedrick, P. W., 1999.** Perspective: Highly Variable Loci and Their Interpretation in Evolution and Conservation. *Evolution* **53**(2): 313-318.
- Herron, M. D., J. M. Waterman et C. L. Parkinson, 2005.** Phylogeny and historical biogeography of African ground squirrels: the role of climate change in the evolution of *Xerus*. *Molecular Ecology* **14**(9): 2773-2788.
- Hey, J. et R. Nielsen, 2007.** Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. **104**: 2785-2790.
- Ibrahim, K. M., P. Sourrouille et G. M. Hewitt, 2000.** Are recession populations of the Desert locust (*Schistocerca gregaria*) remnants of past swarms? *Molecular Ecology* **9**: 783-791.
- Kaatz, H. H., H. J. Ferenz, B. Langer et R. F. A. Moritz, 2007.** Isolation and characterization of nine polymorphic microsatellite loci from the Desert locust, *Schistocerca gregaria*. *Molecular Ecology Notes* **7**(4): -.
- Kim, K. S., S. T. Ratcliffe, B. W. French, L. Liu et T. W. Sappington, 2008.** Utility of EST-Derived SSRs as Population Genetics Markers in a Beetle. **99**: 112-124.

- Kimura, M. et T. Ohta, 1978.** Stepwise mutation model and distribution of allelic frequencies in a finite population. **75**: 2868-2872.
- Lai, Y. et F. Sun, 2003.** The Relationship Between Microsatellite Slippage Mutation Rate and the Number of Repeat Units. **20**: 2123-2131.
- Lebrun, J.-P., 2001.** Introduction à la flore d'Afrique, Seuil.
- Meirmans, P. G., 2006.** USING THE AMOVA FRAMEWORK TO ESTIMATE A STANDARDIZED GENETIC DIFFERENTIATION MEASURE. *Evolution* **60**(11): 2399-2402.
- Moulton, M. J., H. Song et M. F. Whiting, 2010.** Assessing the effects of primer specificity on eliminating numt coamplification in DNA barcoding: a case study from Orthoptera (Arthropoda: Insecta). *Molecular Ecology Resources* **10**(4): 615-627.
- Nei, M., 1987.** *Molecular Evolutionary Genetics*. (Columbia University Press).
- Nolte, D. J., 1965.** The pigmentation of locusts. *S. Afr. J. Sci.* **61**: 173-178.
- Oliveira, E., J. Pádua, M. Zucchi, R. Vencovsky et M. Carneiro-Vieira, 2006.** Origin, evolution and genome distribution of microsatellites. *Genetics and Molecular Biology* **29**(2): 294-307.
- Pascual, M., M. P. Chapuis, F. Mestres, J. Balanyà, R. B. Huey, G. W. Gilchrist, L. Serra et A. Estoup, 2007.** Introduction history of *Drosophila subobscura* in the New World: a microsatellite-based survey using ABC methods. *Molecular Ecology* **16**: 3069-3083.
- Pedgley, 1981.** Desert Locust Forecasting Manual.
- Pener, M. P. et S. J. Simpson, 2009.** Locust phase polyphenism: an update. *Advances in Insect Physiology* **36**: 1-272.
- Piry, S., G. Luikart et J.-M. Cornuet. (1999).** " BOTTLENECK : A program for detecting recent effective population size reductions from allele data frequencies." from <http://www.ensam.inra.fr/URLB/bottleneck/pub.html>.
- Pollock, D. D., A. Bergman, M. W. Feldman et D. B. Goldstein, 1998.** Microsatellite Behavior with Range Constraints: Parameter Estimation and Improved Distances for Use in Phylogenetic Reconstruction. *Theoretical Population Biology* **53**(3): 256-271.
- Prasad, M. D., M. Muthulakshmi, M. Madhu, S. Archak, K. Mita et J. Nagaraju, 2005.** Survey and Analysis of Microsatellites in the Silkworm, *Bombyx mori*. **169**: 197-214.
- Price, Brown, 2000.** A Century of Locust Control in South Africa. *Workshop on Research Priorities for Migrant Pests of Agriculture in Southern Africa*.
- Price, D. R. G. et J. A. Gatehouse, 2008.** RNAi-mediated crop protection against insects. *Trends in biotechnology* **26**(7): 393-400.
- Primmer, C. R. S., N. Moeller, A. P. Ellegren, H, 1998.** Unraveling the Processes of Microsatellite Evolution Through Analysis of Germ Line Mutations in Barn Swallows *Hirundo rustica*. *MOLECULAR BIOLOGY AND EVOLUTION* **15**: 1047-1054.
- Raymond, M. et F. Rousset, 1995.** An Exact Test for Population Differentiation. *Evolution* **49**(6): 1280-1283.
- Ritchie, M. E. et D. E. Pedgley, 1989.** Desert Locusts cross the Atlantic. *Antenna, Bulletin of the Royal Entomological Society of London* **13**(1): 10-12.
- Schlotterer, C., R. Ritter, B. Harr et G. Brem, 1998.** High mutation rate of a long microsatellite allele in *Drosophila melanogaster* provides evidence for allele-specific mutation rates. **15**: 1269-1274.

- Schug, M. D., C. M. Hutter, K. A. Wetterstrand, M. S. Gaudette, T. F. Mackay et C. F. Aquadro, 1998.** The mutation rates of di-, tri- and tetranucleotide repeats in *Drosophila melanogaster*. *Molecular Biology and Evolution* **15**: 1751-1760.
- Schug, M. D., T. F. Mackay et C. F. Aquadro, 1997.** Low mutation rates of microsatellite loci in *Drosophila melanogaster*. *Nat Genet* **15**(1): 99-102.
- Song, H. et J. W. Wenzel, 2008.** Phylogeny of bird-grasshopper subfamily Cyrtacanthacridinae (Orthoptera: Acrididae) and the evolution of locust phase polyphenism. *Cladistics* **24**(4): 515-542.
- Streiff, R., P. Audiot, A. Foucart, M. Lecoq et J.-Y. Rasplus, 2005.** Genetic survey of two endangered grasshopper subspecies, *Prionotropis hystrix rhodanica* and *Prionotropis hystrix azami* (Orthoptera, Pamphagidae): within- and between-population dynamics at the regional scale. *Conservation Genetics* **7**(3): 331-344.
- Symmons, P. et K. Cressman, 1994.** The Desert locust guidelines. I : Biology and behaviour. -27.
- Symmons, P. M. et K. Cressman, 2001.** Desert locust guidelines 1 : Biology and behaviour. Rome, FAO. **2e**: -43.
- Thomas, J. G., 1941.** The relative size of the eye as a phase character in the African Migratory locust. *Bulletin of Entomological Research* **31**(4): 431-433.
- Uvarov, B. P., 1923.** Records and descriptions of Orthoptera from North West Africa. *Novitates Zoologicae* **30**: 59-78.
- Uvarov, B. P., 1956.** The Locust and Grasshopper Problem in Relation to the Development of Arid Lands. Washington, DC, American Association for the Advancement of Science: 383-389.
- Uvarov, B. P., 1966.** Grasshoppers and locusts. A handbook of general Acridology. I : Anatomy, physiology, development, phase-polymorphism, introduction to taxonomy. Cambridge, University Press: -475.
- Vazquez, J. F., T. Perez , A. JESÚS et D. ANA, 2000.** Estimation of microsatellite mutation rates in *Drosophila melanogaster*. *Genetical Research*, **76**: 323-326.
- Verdu, P., F. Austerlitz, A. Estoup, R. Vitalis, M. Georges, S. Thery, A. Froment, S. Le Bomin, A. Gessain, J.-M. Hombert, L. Van der Veen, L. Quintana-Murci, S. Bahuchet et E. Heyer, 2009.** Origins and Genetic Diversity of Pygmy Hunter-Gatherers from Western Central Africa. *Current Biology* **19**(4): 312-318.
- Vitalis, R., K. Dawson et P. Boursot, 2001.** Interpretation of Variation Across Marker Loci as Evidence of Selection. **158**: 1811-1823.
- Waloff, Z., 1966.** The upsurges and recessions of the Desert locust plague : an historical survey. *Anti-Locust Memoir* **8**: 1-111.
- Waloff, Z., 1972.** The plague dynamics of the Desert locust, *Schistocerca gregaria* (Forsk.). C. F. Hemming et T. H. C. Taylor. Londres, COPR: 343-349.
- Waloff, Z. et D. E. Pedgley, 1986.** Comparative Biogeography and Biology of the South American Locust, *Schistocerca cancellata* (Serville) and the South African Desert locust, *S. gregaria flaviventris* (Burmeister) (Orthoptera : Acrididae) : A Review. *Bulletin of Entomological Research* **76**: 1-20.
- Weir, B. S. et C. C. Cockerham, 1984.** Estimating F-Statistics for the Analysis of Population Structure *Evolution* **38**(6): 1358-1370.

- Yassin, Y. A., E. J. Heist et K. M. Ibrahim, 2006.** PCR primers for polymorphic microsatellite loci in the Desert locust, *Schistocerca gregaria* (Orthoptera: Acrididae). *Molecular Ecology Notes* **6**(3): 784-786.
- Zhang, D. X. et G. M. Hewitt, 1996.** Highly conserved nuclear copies of the mitochondrial control region in the Desert locust *Schistocerca gregaria*: some implications for population studies. *Molecular Ecology* **5**: 295-300.
- Zhang, D. X., L. N. Yan, Y. J. Ji, L. Kang, G. M. Hewitt et Z. S. Huang, 2003.** Isolation, characterization and cross-species amplification of eight microsatellite DNA loci in the Migratory locust (*Locusta migratoria*). *Molecular Ecology Notes* **3**: 483-486.
- Zhivotovsky, L. A., M. W. Feldman et S. A. Grishechkin, 1997.** Biased mutations and microsatellite variation. **14**: 926-933.

Rapport-Gratuit.com

ANNEXES

Annexe 1 : Marqueurs génomiques (a) et transcriptomiques (b) développés par l'équipe du CIRAD (les statistiques résumées ont été mesurées à partir de 84 individus répartis dans trois populations de Mauritanie).

Locus	Core repeat (cloned allele)	Size (bp)	Size range (bp)	N_o	H_o	H_e	HW test	r_D	Primer (μM)	Primer sequence (5'-3')	GenBank Accession No.
<i>SgM51</i>	(TG) ₂₄ (GT) ₁₆	226	194-258	23	0.845	0.954	0/3	0.055	0.1	F: Ned-AAATCTGTGCTGCCGAGA R: AGGTGATGTTGTTCCCATGT	JX101522
<i>SgM92</i>	(AC) ₈	174	170-211	16	0.777	0.859	0/3	0.038	0.3	F: Fam-TTTTAAATGGTTGATTCACATT R: CTCTGTAAGGCAACTGTTCA	JX101523
<i>SgM41</i>	(TG) ₂₃	213	162-282	27	1.000	0.965	0/3	0.000	0.2	F: Pet-ACAAAGTGGCATTGACGGAAT R: GCAGGATGAAATACGTGGGTA	JX101524
<i>SgM23</i>	(AC) ₆	243	230-242	4	0.370	0.590	1/3	0.122	0.1	F: Fam-GAAGGAGCCGTAGGAGGTAGA R: AGCCCTTCGTCTGATGTC	JX101532
<i>SgM36</i>	(TG) ₈	153	152-161	5	0.458	0.673	1/3	0.125	0.1	F: Ned-CCATTTGTATFCGCAACCTCA R: GTGAGGGCGTCCTTAATCTG	JX101533
<i>SgM74</i>	(CA) ₁₀ CG (CA) ₈	245	210-268	14	0.702	0.727	0/3	0.043	0.1	F: Vic-CGTCTATGACGCGTGGTAGTT R: CCACAAATCGCTGAACAACA	JX101525
<i>SgM37</i>	(GT) ₂₂	236	215-268	23	0.549	0.955	3/3	0.263	0.2	F: Pet-TCTTCTGAATGGTTGCACCTT R: TTCATACCGTCTAAGTGTGGATGT	JX101535
<i>SgM66</i>	(GT) ₁₉ G(CGT) ₁₀	162	125-188	20	0.865	0.929	0/3	0.023	0.2	F: Fam-CCTAAATCAGATGGCTGGA R: TAGCCACCTAGCACCACTT	JX101530
<i>SgM35</i>	(CA) ₂₈	303	270-320	24	0.769	0.955	2/3	0.089	0.2	F: Fam-CGATGCATTCACAGACTAGAG R: GTTGTGCACCCCTTAATAITGG	JX101534
<i>SgM96</i>	(AG) ₂₂	191	177-221	19	0.859	0.940	0/3	0.027	0.2	F: Pet-ACTACCATTTGGCTTCGTGA R: CAGTCCAAACACCACCTCCAT	JX101531
<i>SgM40</i>	(TG) ₁₄	103	92-144	18	0.775	0.935	1/3	0.073	0.2	F: Fam-GGTGTAGCAAATTTAATGACG R: TGGAAAGCACCCACAAGAC	JX101526
<i>SgM67</i>	(CA) ₁₂	214	202-273	24	0.730	0.955	2/3	0.106	0.2	F: Fam-AAGGAGACAAGCCACTGGAA R: ACCCAATCCACCTGTTGAAA	JX101536
<i>SgM68</i>	(AC) ₁₂	207	199-226	12	0.522	0.818	2/3	0.218	0.2	F: Ned-ACCTCAGCACTCCGACATCT R: TCATCCTAAGCCCACTCACAC	JX101537
<i>SgM87</i>	(CA) ₇	201	193-220	15	0.885	0.871	0/3	0.000	0.1	F: Vic-TGCCACAGACTTCTTGAAA R: GTATTCGGCGACACATTGC	JX101528
<i>SgM88</i>	(AC) ₁₀	217	211-267	19	0.879	0.944	1/3	0.027	0.2	F: Ned-GGAAACACCAAGGCAATGAG R: TGCAAGAGCGTTTATCAGGA	JX101527
<i>SgM86</i>	(TG) ₂₃	298	274-338	24	0.910	0.958	0/3	0.025	0.2	F: Vic-ACTGCACGGACTTTGCTTT R: GACGGGACAAAATGTGGAGA	JX101529
Total				17.9	0.743	0.877	0.75/3	0.077			

HW test: number of significant probability Hardy-Weinberg exact tests (Rousset 2008) after false discovery rate correction (Storey and Tibshirani 2003; package QVALUE in statistical software R 2.12.0; R Development Core Team 2010), r_D : mean null allele frequency computed with the method of Dempster et al. (1977) (Freena software, Chapuis and Estoup 2007), *Primer*: primer concentration for both forward and reverse primers.

Locus	Core repeat (cloned allele)	Size (bp)	Size range (bp)	N_a	H_o	H_e	HW test	r_D	Primer (μM)	Primer sequence (5'-3')	Unigene ID
<i>diEST-29</i>	(GT) ₁₀	238	230-267	9	0.705	0.708	0/3	0.001	0.1	F: Vic-TGGATAGAAATAAATTTATGGGTGCTT R: CAGAGCTCACAATTAAGGCACA	LC01023A2F07.f1
<i>diEST-30</i>	(GT) ₁₁	133	127-145	7	0.788	0.771	0/3	0.011	0.2	F: Med-TCTTGGCTCAACAATAATCGAG R: CCACTTCCTATTGTGCGGT	LC.3018.C1.Contig3172
<i>diEST-16</i>	(CA) ₁₂	240	232-267	12	0.811	0.832	0/3	0.013	0.2	F: Pet-CCCGAATTTATCACTTCCA R: CAGCTGCCACAGCTCCTTA	LC01024A2A05.f1
<i>diEST-4</i>	(AC) ₈	271	267-318	20	0.827	0.940	1/3	0.095	0.1	F: Fam-GTCCCAGTTCGAGTCTCAC R: GCTTCTGCCACAATGTTGA	LC02007B1G07.f1
<i>diEST-47</i>	(TG) ₈	260	253-291	10	0.755	0.763	0/3	0.002	0.1	F: Fam-TGTGTCTGTGTCGAATCGT R: TGTCAAGAGAAATGCAGTT	LC.2018.C1.Contig2180
<i>diEST-20</i>	(CT) ₁₁	130	110-180	33	0.962	0.972	0/3	0.000	0.1	F: Vic-GTCTTGCCATATTTTCAGCC R: GATCACAGAAAGTACATGGGA	LC.3905.C1.Contig4033
<i>diEST-42</i>	(TC) ₁₀	279	271-321	11	0.725	0.723	0/3	0.000	0.1	F: Vic-GCTGCAGTGGAAACGGTAA R: CTTCTCTTTCACAGCGTCC	LC03017B2F09.f1
<i>diEST-11</i>	(AT) ₃ AA(AT) ₈	177	171-191	8	0.630	0.643	0/3	0.005	0.2	F: Med-AAGCAACTCAGAAAAGTGGCAA R: AAAATAGACCGTCTGTCG	LC.161.C1.Contig225
<i>diEST-5</i>	(AC) ₉ AT(AC) ₂	270	252-346	26	0.872	0.963	1/3	0.053	0.1	F: Med-GACAAGATGGATGGACGGAG R: ACAGGCATGGTCTACTGAG	LC.1901.C1.Contig2058
<i>diEST-40</i>	(TA) ₉	263	254-266	7	0.749	0.774	0/3	0.012	0.3	F: Pet-TTTCTTCAATGGGAAAAGCC R: AACATCTGTTTACAAGGTTGCAG	LC.568.C1.Contig667
<i>diEST-8</i>	(AT) ₁₀	252	242-263	14	0.855	0.901	0/3	0.015	0.2	F: Fam-TGCATACTCAACAGTAGCTT R: CAGCTGCAGTATGGTTGCCCT	LC01005A1A08.f1
<i>diEST-6</i>	(AG) ₈ (ACAG) ₄	170	167-175	5	0.641	0.709	0/3	0.019	0.1	F: Fam-ATCAGGTTCTGCATAGGGCT R: GTGGCTTCCCGTTAACTAGC	LC.382.C1.Contig464
<i>diEST-28</i>	GTAT(GT) ₁₀	189	179-231	21	0.895	0.938	1/3	0.025	0.1	F: Vic-TGGACAGTTTGTGACGGAA R: AGATTTCAFCCTTTCCTCC	LC.2033.C1.Contig2195
<i>diEST-13</i>	(AT) ₈	127	127-157	7	0.617	0.599	0/3	0.006	0.2	F: Pet-TCGGGCTTTGTGTAAAGTG R: GACAAATCTCAAATCTTTTCATGC	LC01067B1H12.f1
<i>diEST-37</i>	(GT) ₉	263	259-334	33	0.976	0.979	0/3	0.000	0.2	F: Pet-TTGTACGATAGTGCATGGC R: GCTGTAGTCCACCCAAATGGT	LC.1999.C1.Contig2161
<i>diEST-12</i>	(AT) ₈	247	239-247	4	0.401	0.384	0/3	0.000	0.1	F: Med-ATACGACAGAAAGCGCAGTT R: TGCTGTCTTGTGGGAGA	LC01057B1D11.f1
<i>diEST-2</i>	(AC) ₁₂	227	215-289	24	0.903	0.959	2/3	0.018	0.2	F: Fam-GTTGCTGACAAGCACAAATGG R: TGCAGAAAGCAGTTCTTCA	LC03005A2H07.f1
<i>diEST-35</i>	(GT) ₈	105	102-106	4	0.257	0.269	0/3	0.022	0.1	F: Fam-GCATAGGGCGAAAGACATTTG R: TCACACCACATCCTTAGAGTTCA	LC01018B2E05.f1
<i>Total</i>							0.28/3	0.015			

HW test: number of significant probability Hardy-Weinberg exact tests (Rousset 2008) after false discovery rate (Storey and Tibshirani 2003); package QVALUE in statistical software R 2.12.0; R Development Core Team 2010), r_D : mean null allele frequency computed with the method of Dempster et al. (1977) (FreeNA software, Chapuis and

Estoup 2007), *Primer*: primer concentration for both forward and reverse primers.

Annexe 2 : Protocole de la PCR utilisée dans notre étude :

Solution pour la PCR par Multiplexage QIAGEN (96 puits) :

- 500µL MasterMix QIAGEN (MgCl₂, HotStart Taq DNA polymerase, dNTPs, buffer)
- 100µL solution Q
- 100µL d'amorces (multiplexées) 10X
- 100µL d'eau
- Dépôt 8µL de cette solution et de 2µL d'ADN dans chaque puits

Programme PCR utilisé :

- L'amplification ADN par PCR est faite sur un thermocycler TC-512 (Techne)
- Un programme Touchdown qui consiste à baisser la température de 0,5°C par cycle pendant les 10 premiers cycles de 60°C à 55°C afin d'améliorer l'amorçage de l'amplification est utilisé pour l'amplification ADN.
- Une phase de dénaturation de 15 min à 94°C démarre la PCR à proprement parler
- Elle est suivie de 35 cycles constitués des étapes ci-dessous :
 - ◆ dénaturation pendant 30 s à 94°C
 - ◆ hybridation pendant 90 s
 - ◆ élongation pendant 75 s à 72°C
- La PCR s'achève ensuite par une étape d'élongation de 20 min à 72°C

Préparation pour génotypage :

- Dilution du produit PCR au 30^{ème} dans l'eau (2µL dans 68µL d'eau)
- Préparation d'une solution de 1,5mL de formamide auquel on ajoute 11µL de marqueur de taille GeneScan-500LIZ
- Dilution de l'ADN dilué au ~1/10^{ème} (Par puits : 1,5µL dans 15µL de solution formamide/marqueur de taille)
- Solution prête pour génotypage sur séquenceur ABI 3130

Annexe 3 : Conditions d'élevage isolées du criquet pèlerin d'où sont issus les larves utilisées dans l'estimation du taux de mutation

○ Structure d'élevage

Les structures d'élevage individuel mises au point par l'équipe sont modifiées d'après Hoste *et al.* (2003). Elles ont été conçues selon trois principes : la maniabilité, l'homogénéité des conditions et la limitation des stimuli de grégarisation entre cages. Roessingh *et al.* (1998) ont montré sur le criquet pèlerin que (i) la grégarisation est fortement induite par des stimuli tactiles, (ii) que la vue et l'olfaction combinées ou la vue à long terme permet une grégarisation notable, (iii) que la vue seule ne permet qu'une grégarisation très partielle et (iiii) que l'odorat ne permet pas de grégarisation (quatre composants de la phéromone d'agrégation mâle ont été testés). Cependant, deux composants de cette hormone ainsi que la phéromone d'agrégation larvaire n'ont pas été testés. Le système d'élevage proprement dit est composé de chariots sur roulettes à six plateaux contenant chacun 36 cages en plastique d'un litre (8cm x 8cm x 12cm) suspendues. Trois améliorations ont été apportées au protocole de Hoste *et al.*, 2003. Le couvercle des cages est percé d'une soixantaine de trous de deux millimètres de diamètre pour permettre l'aération de la cage. Afin d'homogénéiser les conditions de luminosité au sein d'un chariot, des néons froids sont installés sur les chariots toutes les deux rangées. Afin d'éliminer tout stimulus visuel entre cages individuelles, chacune d'elles est opacifiée par ponçage.

○ Maintenance

Les chambres ont été maintenues à une température de 34° C jour avec une humidité constante de 50%. Afin de minimiser les stimuli olfactifs entre cages, un renouvellement de l'air a été mis en place pour chaque chambre, par une ventilation de 20 volumes horaire. La totalité de l'expérience s'est déroulée sous un photo-régime de 12 heures de présence d'éclairage et 12 heures d'absence d'éclairage. Les adultes sont nourris tous les 2 ou 3 jours avec du blé germé et une ration de son en complément. Le blé est cultivé sur vermiculite en plaque de culture. Chaque portion introduite dans la cage est contenue dans une boîte d'emballage de pellicule photo pour garder la turgescence des pousses.

○ Reproduction

Les accouplements ont été déterminés pour satisfaire les besoins d'une expérience de génétique quantitative menée par Benjamin Pélissié. A maturité sexuelle (environ 30 jours après la mue imaginale), une femelle et un mâle sont placés ensemble 48 h dans une cage individuelle. A la séparation, le grillage du bas des cages femelles est enlevé afin d'insérer un pot de ponte (5cm x 10cm). Le pot de ponte est renouvelé une fois au bout de 2 jours et 7 jours plus tard, les mêmes couples sont formés à nouveau et on réitère ce processus jusqu'à la 4^{ème} ponte si possible. Les pots de pontes sont remplis de sable humidifié à 10% et sont percés à la base pour ajuster l'humidité. Le sable est du sable d'aquarium. Il est recyclé à partir des pontes précédentes par tamisage puis par stérilisation en étuve à 90 °C pendant 24 heures. Les pots de ponte sont vérifiés tous les 2 jours en les faisant tourner sur leur axe. Les femelles pondent généralement 1 à 2 jours après l'accouplement. La seconde ponte a lieu généralement 2 à 3 jours plus tard et une troisième quelques jours plus tard. Les pots de ponte sont alors recouverts de moustiquaire en tissu et réhumidifiés légèrement à l'aide d'une pissette pour compenser l'évaporation. Il faut environ 12 jours avant l'éclosion de la nouvelle génération. 24 heures après l'éclosion, les larves sont sexées à l'aide d'une binoculaire et mesurées. Elles sont ensuite conservées dans l'alcool absolu, disponibles pour les extractions ADN.

Annexe 4 : Grandes lignes du protocole d'extraction ADN en plaque (©MAX)

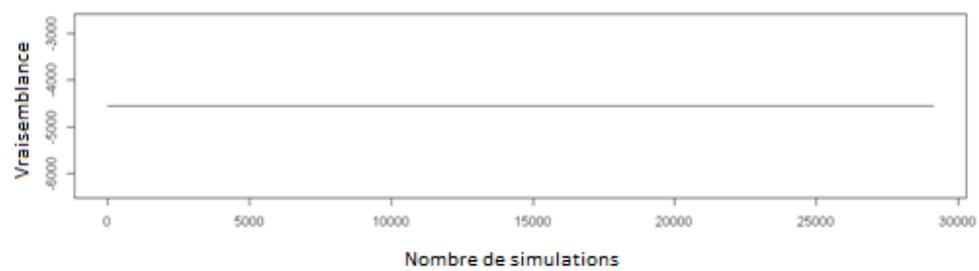
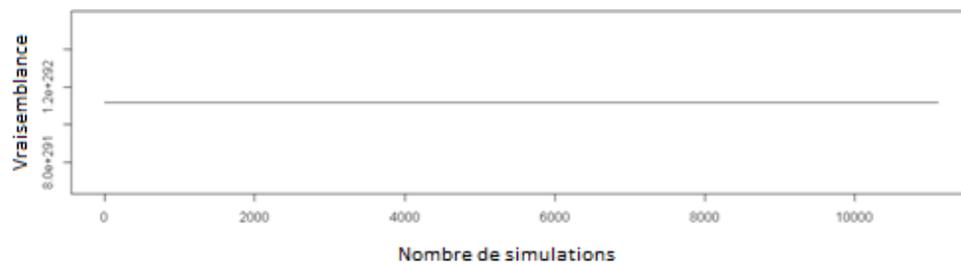
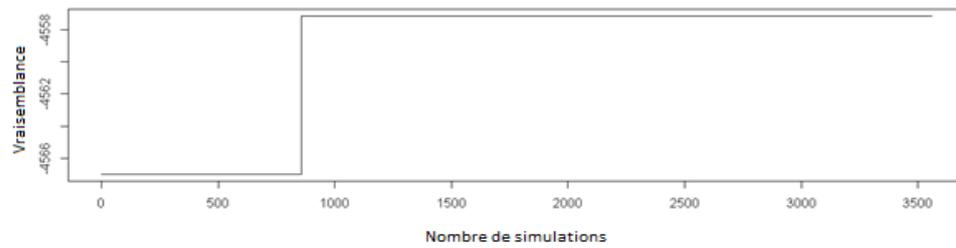
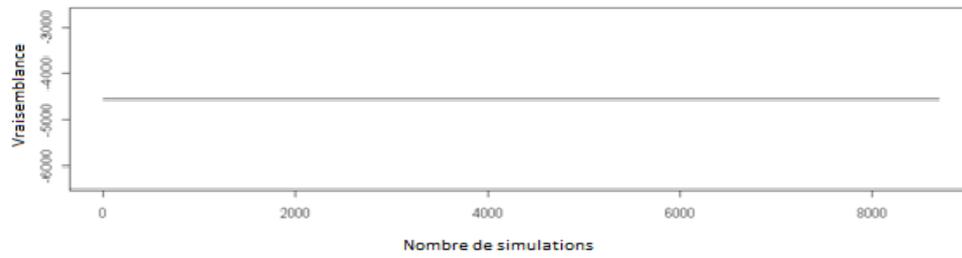
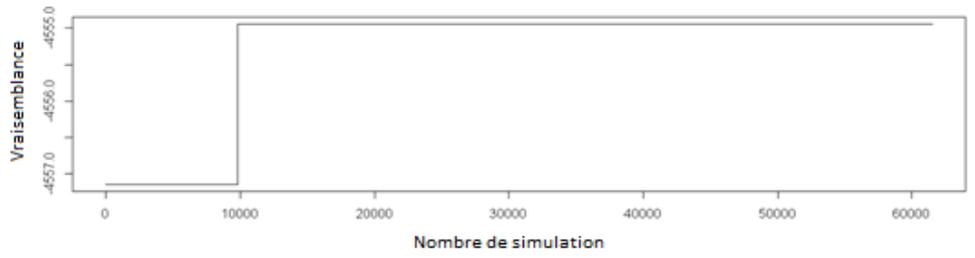
1^{ère} phase : dissection et digestion

- ◆ Préchauffer l'étuve à 55°C et mettre le tampon ACL au bain-marie à 37°C
- ◆ Pour le nettoyage du matériel de dissection, préparer une solution 2/3 eau – 1/3 d'eau de javel
- ◆ Dans une plaque QIAGEN bleue, mettre une patte par puits, et mettre le reste de l'individu dans une autre plaque dans le puits correspondant
- ◆ Mélanger **28 ml de tampon ACL** et **2 ml de protéinase K** dans un tube Falcon (trait de 30ml)
- ◆ Mettre **300 µl du mélange** dans chaque puits (combitips sur 300µl)
- ◆ Centrifuger pour que toutes les pattes soient bien en contact avec le milieu de lyse
- ◆ Fermer avec les bouchons pour éviter l'évaporation
- ◆ Mettre à **55°C** avec agitation pendant au moins 2-3 heures

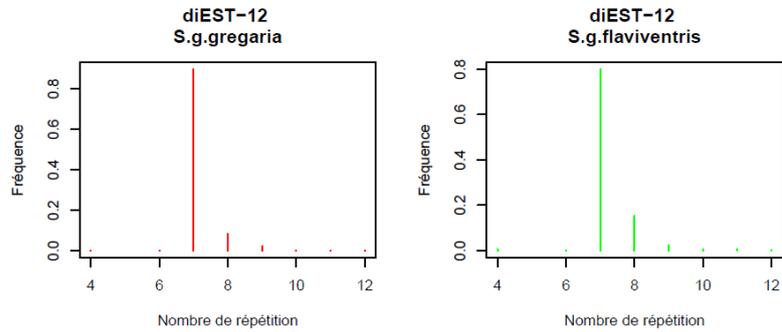
2^{ème} phase : extraction en colonne

- ◆ Mettre le tampon d'élution au bain marie à 60°C
- ◆ **Centrifuger** (1 pulse) avant d'ouvrir les bouchons des tubes de la plaque (sortie d'incubation)
- ◆ Enlever les bouchons en gardant l'ordre pour pouvoir les remettre
- ◆ Mettre 300 µl de solution AB dans chaque puits (multicanaux 3X300µl) et refermer (utilisation des pointes 1200µl BIOHIT)
- ◆ « Shaker » à la main 30 s les plaques et centrifuger 5 min à 6000 rpm
- ◆ Mettre les plaques avec colonnes BioBasic sur les portoirs QIAGEN, et **transférer 700 µl** de chaque puits sur les colonnes (700µl avec multicanaux en mode « pipette », vitesse 1) (plonger les cônes en deux étapes dans la barette pour éviter de faire déborder)
- ◆ Fermer les plaques avec un **film en aluminium** et **centrifuger 5 min à 6000 rpm**
- ◆ Ajouter **500 µl de « Wash solution »** (multicanaux 2X500µl, avec une 1^{ère} imprégnation)
- ◆ Refermer le **film alu** et **centrifuger 5 min à 6000 rpm**
- ◆ Ajouter **500 µl de « Wash solution »** (multicanaux 2X500µl). **Ne pas refermer avec un film alu !**
- ◆ **Centrifuger 15 min à 6000 rpm**
- ◆ Mettre les colonnes sur les plaques d'élution QIAGEN
- ◆ Ajouter sur chaque colonne **50 µl de tampon d'élution** (chauffé à 60°C) (multicanaux 12X50µl, avec une 1^{ère} imprégnation) et fermer avec un film aluminium
- ◆ Laisser incuber pendant 2 min
- ◆ **Centrifuger 2 min à 6000 rpm**
- ◆ **Répéter cette étape d'élution** une deuxième fois puis mettre des bouchons sur les plaques d'élution

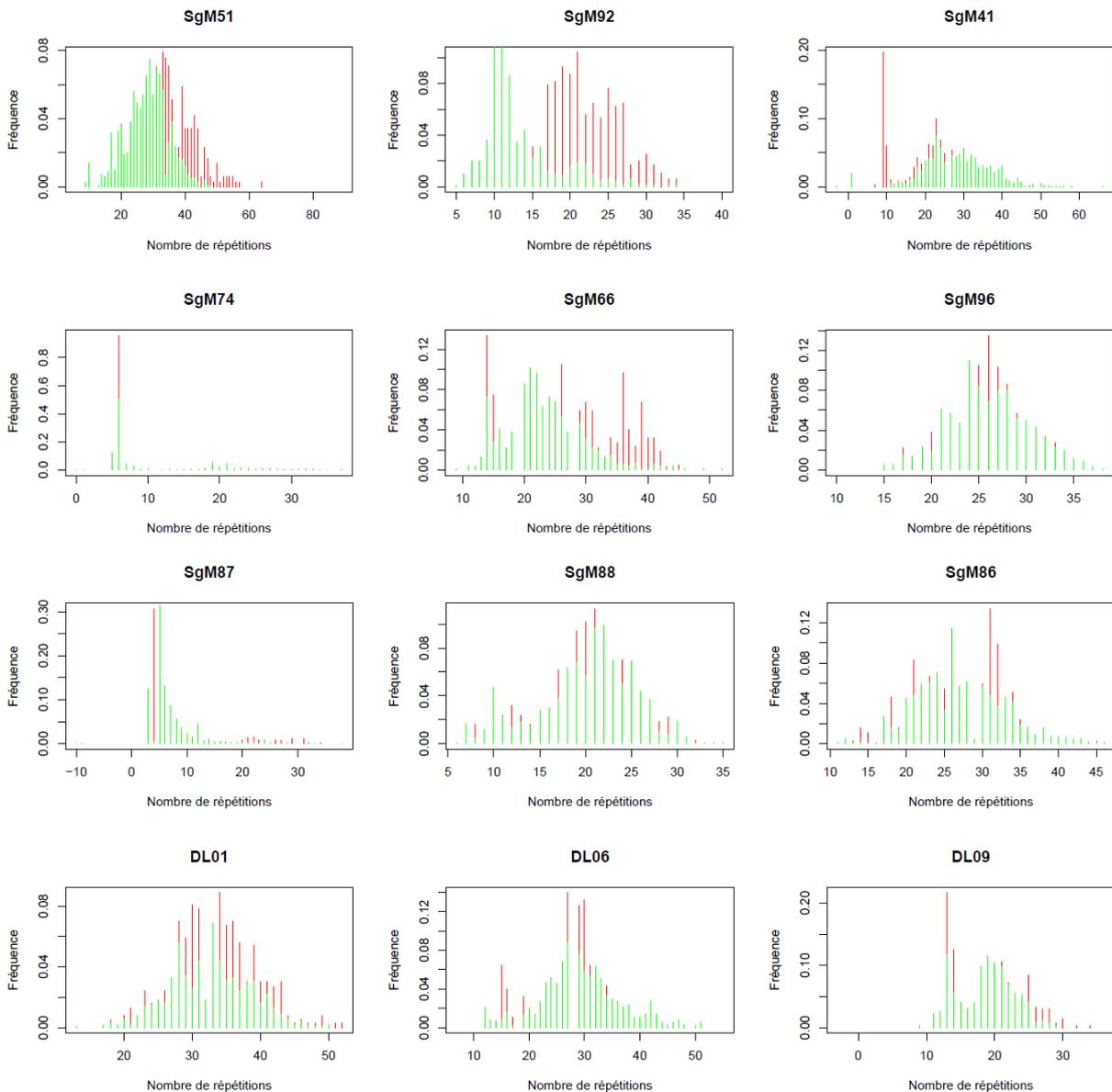
Annexe 5 : Simulations de l'analyse pour tester la présence de migration (Chaîne de Monte-Carlo du logiciel 2MOD)

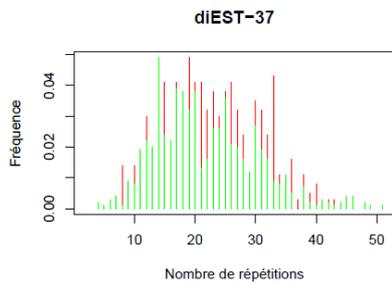
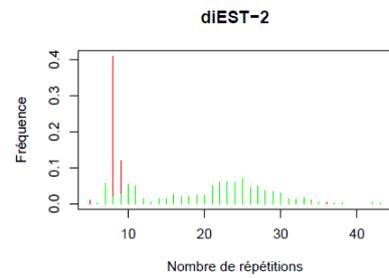
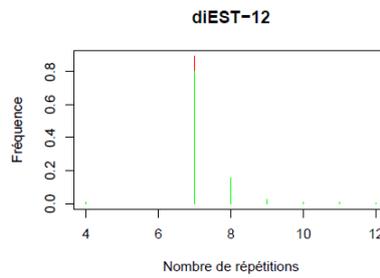
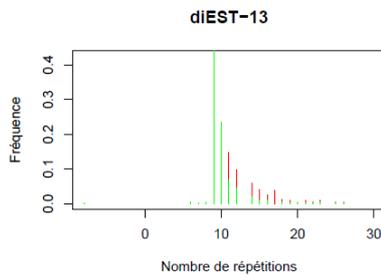
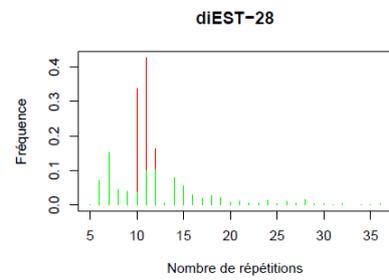
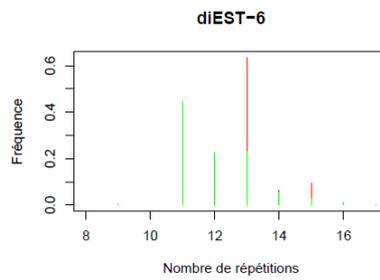
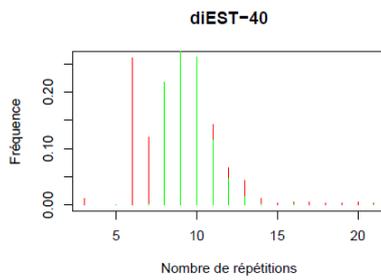
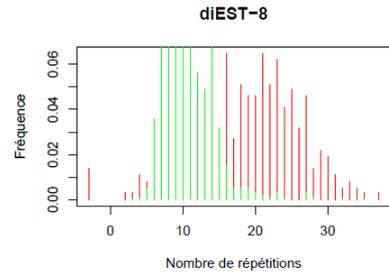
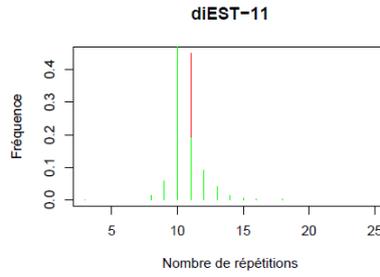
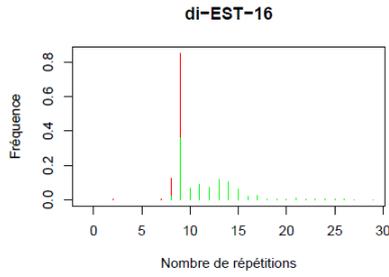
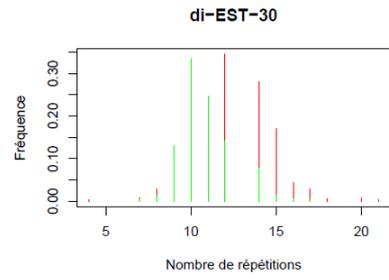
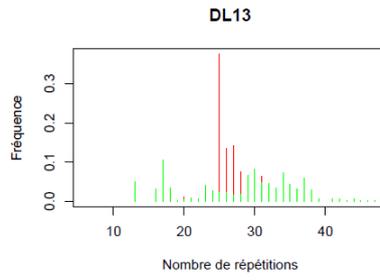
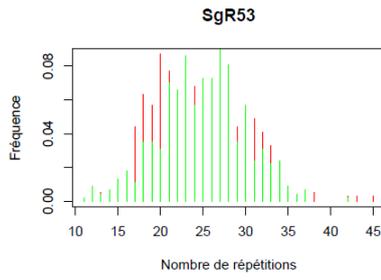


Annexe 6 : Distribution des fréquences alléliques pour le marqueur diEST12 entre les deux sous-espèces



Annexe 7 : Distribution des fréquences alléliques des marqueurs comparés entre les deux sous-espèces (en rouge *S.g.gregaria* et en vert *S.g.flaviventris*)





Annexe 8 : Statistiques résumées pour les marqueurs et les populations étudiées de *S. g. flaviventris*

Hétérozygotie attendue He

Aucune population n'est significativement différente des autres (pour He et Ar).

	SgBB	SgCC	SgDD	SgEE	SgFF	SgGG	Moyenne S,g, flaviventris
SgM51	0,958	0,965	0,961	0,968	0,959	0,963	0,962
SgM92	0,948	0,932	0,936	0,92	0,933	0,937	0,934
SgM41	0,883	0,916	0,938	0,898	0,941	0,938	0,919
SgM74	0,062	0,062	0,179	0,03	0,061	0,153	0,091
SgM66	0,934	0,93	0,9	0,944	0,932	0,936	0,929
SgM96	0,934	0,925	0,928	0,907	0,921	0,929	0,924
SgM87	0,83	0,875	0,816	0,824	0,856	0,852	0,842
SgM88	0,932	0,936	0,92	0,932	0,936	0,943	0,933
SgM86	0,951	0,946	0,954	0,902	0,939	0,922	0,936
diEST-30	0,709	0,765	0,778	0,722	0,81	0,815	0,767
diEST-16	0,32	0,222	0,268	0,26	0,321	0,151	0,257
diEST-11	0,656	0,593	0,617	0,625	0,604	0,641	0,623
diEST-8	0,953	0,97	0,976	0,957	0,962	0,966	0,964
diEST-40	0,826	0,888	0,926	0,813	0,863	0,858	0,862
diEST-6	0,448	0,61	0,583	0,683	0,551	0,39	0,544
diEST-28	0,662	0,72	0,705	0,671	0,624	0,704	0,681
diEST-13	0,664	0,733	0,76	0,761	0,796	0,872	0,764
diEST-12	0,175	0,273	0,156	0,287	0,117	0,125	0,189
diEST-2	0,831	0,816	0,803	0,759	0,778	0,867	0,809
diEST-37	0,975	0,982	0,97	0,975	0,98	0,975	0,976
DL01	0,95	0,954	0,946	0,956	0,937	0,958	0,950
DL06	0,926	0,908	0,918	0,928	0,939	0,912	0,922
DL09	0,903	0,905	0,878	0,898	0,899	0,892	0,896
SgR53	0,951	0,943	0,952	0,951	0,945	0,943	0,948
DL13	0,748	0,768	0,796	0,835	0,812	0,838	0,800
Moyenne	0,765	0,781	0,783	0,776	0,777	0,779	0,777

Richesse allélique (pour 9 individus)

	SgBB	SgCC	SgDD	SgEE	SgFF	SgGG	Moyenne S,g, flaviventris
SgM51	13,14	13,57	13,23	13,99	13,33	13,45	13,45
SgM92	11,77	10,80	11,05	10,64	10,75	11,38	11,06
SgM41	10,33	10,34	10,64	9,75	11,70	11,73	10,75
SgM74	1,56	1,56	2,38	1,27	1,49	2,05	1,72
SgM66	11,57	11,00	11,17	11,58	11,19	11,43	11,32
SgM96	10,93	10,15	10,50	9,59	10,18	10,32	10,28
SgM87	8,80	9,06	7,82	7,51	8,66	8,32	8,36
SgM88	11,11	11,21	10,03	11,15	11,14	11,74	11,06
SgM86	12,25	11,62	12,56	9,51	11,49	10,64	11,34
diEST-30	4,98	5,32	5,73	5,10	6,43	6,22	5,63
diEST-16	2,46	1,94	2,98	1,97	2,54	1,83	2,29
diEST-11	4,30	3,40	3,61	3,41	3,48	3,53	3,62
diEST-8	12,96	14,02	14,52	12,75	13,32	13,74	13,55
diEST-40	7,17	8,83	10,37	6,60	7,71	7,15	7,97
diEST-6	3,86	4,37	4,66	4,28	3,84	3,51	4,08
diEST-28	3,74	4,49	4,13	4,37	3,41	3,93	4,01
diEST-13	6,10	6,09	6,27	7,08	7,01	7,93	6,75
diEST-12	2,10	2,24	1,86	2,61	1,90	1,94	2,11
diEST-2	9,51	8,97	8,30	7,81	8,40	9,19	8,70
diEST-37	14,57	15,13	14,06	14,50	15,20	14,83	14,71
DL01	12,13	12,58	11,95	12,60	11,28	12,84	12,23
DL06	10,72	10,00	10,76	11,47	11,39	10,03	10,73
DL09	12,81	12,91	12,15	11,75	11,21	9,42	11,71
SgR53	16,99	15,08	18,38	16,45	16,33	14,52	16,29
DL13	8,14	7,92	8,11	8,50	7,71	8,22	8,10
Moyenne	8,96	8,90	9,09	8,65	8,84	8,80	8,87

Annexe 9 : Hétérozygotie attendue pour les marqueurs et les populations étudiées de *S. g. gregaria*

Seul AISgOT est significativement différent des autres (Wilcoxon $p < 0,0008$)

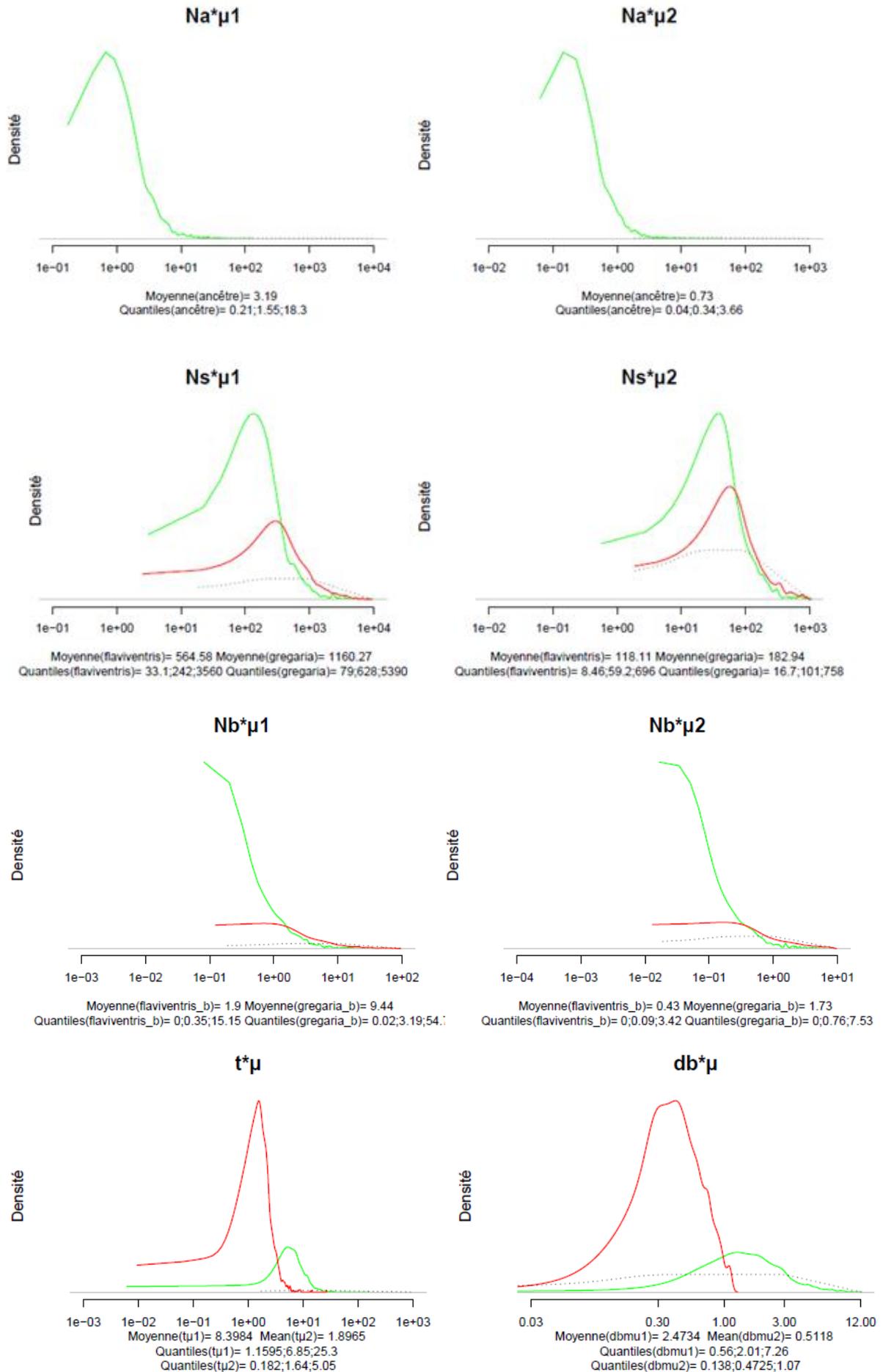
	AISg1NBEZH	AISgOA	AISgOT	Sg1N	MRSgAO	MRSgFd	MRSgIn	SgJ	SgO	SgP	MRSgTa	MRSgTi	MRSgTr	Soud	Tch7a	Tch7b	Tch7c	TSG1BO	TSG2BO	TSGFa	TSGKe	NSgTa	Moyenne Sg, gregaria
SgM61	0,959	0,948	0,689	0,951	0,959	0,93	0,938	0,954	0,952	0,96	0,952	0,941	0,963	0,977	0,941	0,933	0,943	0,95	0,948	0,968	0,962	0,947	0,939
SgM62	0,907	0,894	0,642	0,871	0,889	0,641	0,897	0,898	0,829	0,853	0,846	0,847	0,873	0,723	0,879	0,862	0,878	0,878	0,931	0,909	0,873	0,885	0,850
SgM41	0,973	0,964	0,701	0,964	0,966	0,936	0,956	0,963	0,967	0,96	0,964	0,957	0,956	0,964	0,974	0,967	0,935	0,972	0,96	0,945	0,965	0,961	0,949
SgM74	0,7	0,892	0,433	0,781	0,733	0,27	0,813	0,683	0,685	0,797	0,715	0,728	0,798	0,759	0,795	0,691	0,763	0,694	0,693	0,709	0,715	0,572	0,701
SgM66	0,939	0,949	0,506	0,958	0,939	0,924	0,94	0,9	0,94	0,947	0,952	0,955	0,938	0,905	0,94	0,945	0,913	0,922	0,942	0,932	0,938	0,958	0,917
SgM66	0,951	0,943	0,506	0,932	0,943	0,921	0,937	0,936	0,95	0,935	0,948	0,922	0,938	0,968	0,945	0,937	0,93	0,961	0,952	0,917	0,947	0,935	0,921
SgM87	0,871	0,851	0,582	0,872	0,859	0,767	0,887	0,857	0,866	0,89	0,849	0,789	0,856	0,882	0,864	0,843	0,794	0,894	0,893	0,855	0,768	0,815	0,837
SgM88	0,949	0,96	0,709	0,953	0,947	0,892	0,935	0,942	0,95	0,943	0,93	0,936	0,941	0,923	0,953	0,953	0,949	0,922	0,955	0,936	0,939	0,956	0,931
SgM86	0,948	0,951	0,321	0,956	0,957	0,92	0,941	0,964	0,958	0,956	0,961	0,957	0,94	0,968	0,945	0,954	0,935	0,961	0,938	0,977	0,965	0,966	0,925
dEST-30	0,75	0,76	0,794	0,772	0,786	0,704	0,714	0,76	0,762	0,787	0,847	0,73	0,75	0,755	0,735	0,73	0,814	0,789	0,745	0,777	0,761	0,806	0,765
dEST-16	0,823	0,874	0,126	0,845	0,827	0,837	0,855	0,834	0,845	0,824	0,808	0,879	0,848	0,886	0,782	0,847	0,764	0,867	0,931	0,832	0,849	0,863	0,811
dEST-11	0,681	0,707	0,097	0,689	0,59	0,637	0,705	0,595	0,645	0,687	0,576	0,742	0,601	0,755	0,638	0,613	0,652	0,556	0,631	0,625	0,656	0,649	0,624
dEST-8	0,91	0,897	0,672	0,912	0,933	0,857	0,916	0,889	0,906	0,905	0,895	0,867	0,899	0,882	0,88	0,917	0,901	0,944	0,893	0,895	0,879	0,91	0,889
dEST-40	0,775	0,842	0,507	0,809	0,737	0,772	0,801	0,769	0,76	0,788	0,775	0,811	0,785	0,736	0,796	0,788	0,733	0,767	0,757	0,759	0,814	0,793	0,767
dEST-6	0,684	0,663	0,608	0,734	0,69	0,558	0,729	0,706	0,714	0,718	0,72	0,701	0,732	0,459	0,715	0,712	0,689	0,711	0,731	0,673	0,772	0,672	0,686
dEST-28	0,931	0,92	0,551	0,932	0,941	0,886	0,917	0,914	0,929	0,957	0,93	0,935	0,92	0,968	0,938	0,936	0,917	0,967	0,938	0,868	0,949	0,933	0,913
dEST-13	0,628	0,512	0	0,564	0,604	0,717	0,551	0,653	0,631	0,552	0,603	0,676	0,606	0,755	0,633	0,65	0,541	0,606	0,564	0,755	0,487	0,538	0,583
dEST-12	0,52	0,313	0,234	0,21	0,319	0,297	0,422	0,305	0,455	0,378	0,216	0,28	0,392	0,577	0,408	0,249	0,372	0,189	0,295	0,332	0,169	0,381	0,332
dEST-2	0,964	0,952	0,563	0,953	0,966	0,938	0,957	0,958	0,954	0,966	0,963	0,963	0,958	0,964	0,956	0,959	0,93	0,939	0,931	0,955	0,95	0,951	0,936
dEST-37	0,978	0,982	0,668	0,983	0,98	0,947	0,964	0,98	0,974	0,981	0,98	0,97	0,975	0,986	0,983	0,98	0,975	0,956	0,981	0,986	0,978	0,962	
DL01	0,972	0,973	0,621	0,98	0,967	0,945	0,971	0,976	0,98	0,968	0,974	0,972	0,964	0,955	0,976	0,975	0,956	0,978	0,981	0,994	0,971	0,968	0,955
DL06	0,94	0,956	0,716	0,955	0,948	0,933	0,957	0,951	0,959	0,963	0,966	0,961	0,96	0,868	0,943	0,948	0,936	0,933	0,952	0,894	0,967	0,97	0,935
DL09				0,921				0,916	0,901	0,933					0,932	0,909	0,926					0,952	0,924
SgR53				0,944				0,936	0,945	0,946					0,948	0,933	0,932					0,953	0,942
DL13				0,95				0,945	0,964	0,969					0,947	0,944	0,924					0,952	0,949
Moyenne	0,852	0,850	0,511	0,856	0,840	0,783	0,850	0,847	0,857	0,863	0,835	0,842	0,845	0,846	0,858	0,847	0,840	0,834	0,843	0,841	0,831	0,851	0,838

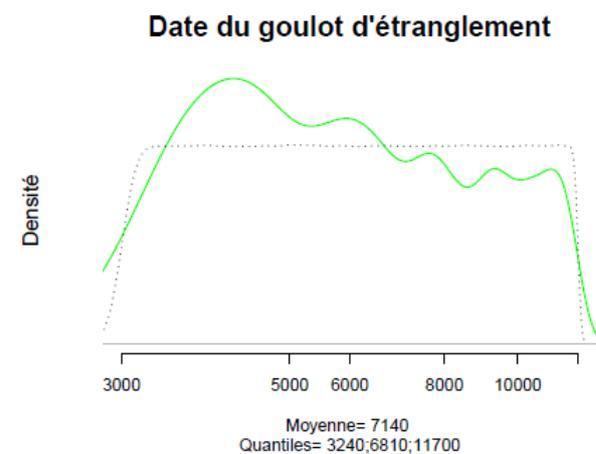
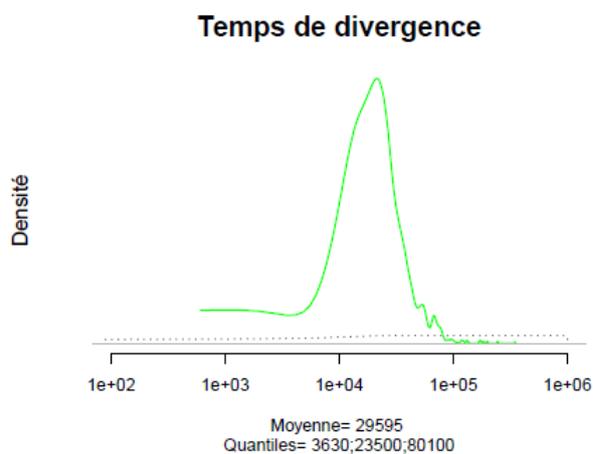
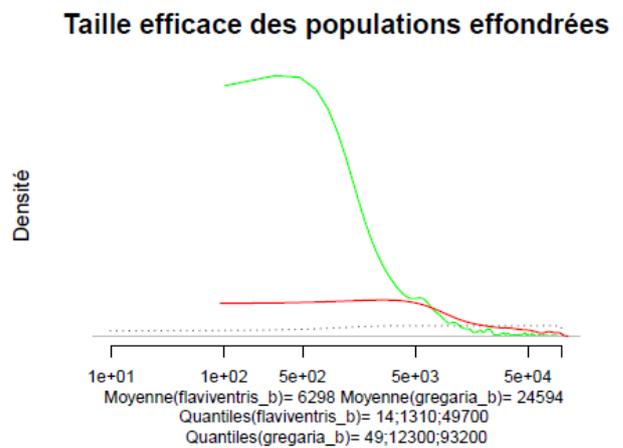
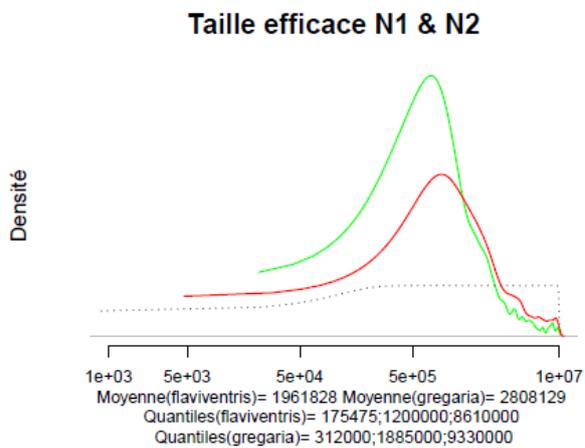
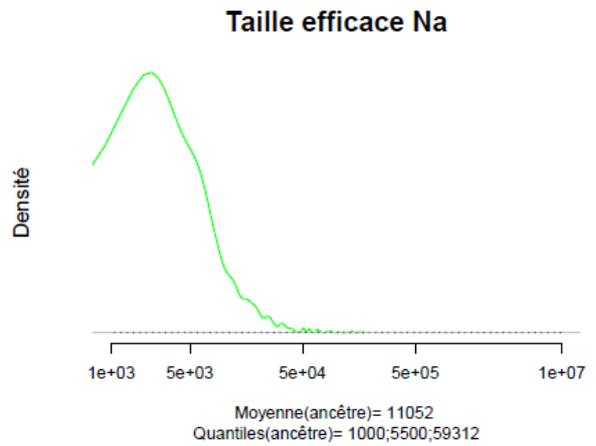
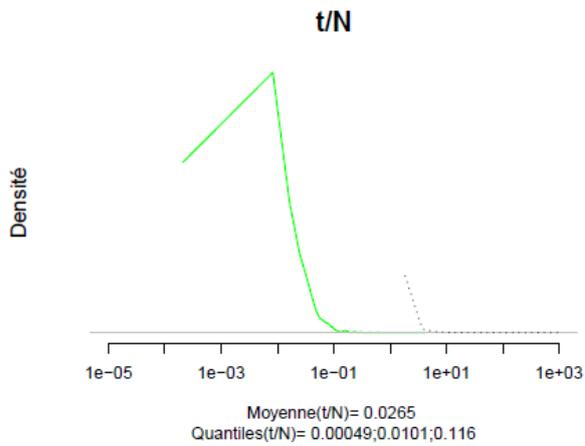
Annexe 10 : Richesse allélique pour les marqueurs et les populations étudiées de *S. g. gregaria* (pour 9 individus)

Seul AISgOT est significativement différent des autres (Wilcoxon $p < 0,00001$)

	AISgINBEZH	AISgOA	AISgOT	SgN	MRSgAO	MRSgFd	MRSgIh	SgJ	SgO	SgP	MRSgTa	MRSgTI	MRSgTr	Soud	Tch7a	Tch7b	Tch7c	TSg1BO	TSg2BO	TSgFa	TSgKe	NSgTa	Moyenne Sg. gregaria
SgM61	12,87	12,05	3,66	12,23	12,94	10,72	11,54	12,21	12,42	13,01	12,76	11,77	12,59	14,03	11,79	10,90	11,85	12,18	11,96	13,10	12,40	12,29	11,88
SgM62	10,35	9,21	3,59	9,02	9,43	5,76	10,04	9,54	8,82	8,56	9,11	7,44	9,00	5,61	8,97	8,59	8,84	8,68	10,45	9,85	9,04	9,19	8,59
SgM41	14,38	13,53	4,90	13,37	13,48	11,32	13,05	13,37	13,91	13,31	13,56	13,01	12,68	13,74	13,99	13,59	11,58	14,69	12,76	11,51	13,66	13,23	12,85
SgM74	7,35	9,77	2,00	8,24	6,20	2,97	8,78	7,34	6,46	7,50	7,15	7,43	8,85	7,07	8,14	7,67	6,51	7,49	6,80	7,73	8,11	5,58	7,05
SgM66	11,34	12,21	2,51	12,80	11,42	10,55	11,40	10,29	11,53	12,15	12,34	12,45	11,46	10,65	11,48	11,75	9,99	11,19	10,94	10,18	11,19	12,71	11,06
SgM66	12,33	11,50	2,51	11,22	11,36	10,03	10,99	11,04	12,04	11,08	12,12	10,13	11,61	12,44	11,56	10,84	10,96	12,36	11,61	11,19	11,43	10,99	10,97
SgM67	9,54	8,37	3,44	9,38	8,79	6,64	8,85	8,11	9,16	9,50	8,50	7,43	8,97	9,04	9,04	8,80	7,25	6,90	9,84	8,07	7,12	8,73	8,25
SgM88	12,20	13,08	4,73	12,77	11,79	8,95	11,47	11,58	12,08	11,74	11,09	11,22	11,57	10,03	12,33	12,56	12,23	11,19	12,32	11,49	11,84	12,55	11,40
SgM86	12,31	12,48	2,70	12,57	12,79	10,58	11,34	13,21	13,07	12,77	13,07	12,93	11,67	13,23	12,11	12,52	11,67	12,98	12,07	14,69	12,89	13,48	12,14
dEST-30	5,38	5,41	4,86	5,11	5,93	5,11	4,59	5,24	5,20	5,62	6,94	4,89	5,76	5,58	5,43	4,94	5,58	5,88	5,14	4,95	4,69	6,08	5,38
dEST-16	8,01	8,38	1,77	7,70	7,87	6,97	8,05	7,81	7,91	7,15	7,84	9,18	7,45	9,04	7,02	7,29	6,35	6,80	10,13	6,61	8,10	9,01	7,56
dEST-11	5,26	5,48	1,67	5,26	4,89	4,29	5,93	4,32	5,43	5,45	4,21	5,29	4,03	5,76	4,44	4,57	5,09	3,90	4,19	3,00	4,22	4,64	4,60
dEST-8	9,57	9,42	4,24	9,95	11,11	8,06	10,09	8,73	10,07	9,10	9,15	8,16	8,79	8,88	9,48	9,97	9,41	9,68	9,22	9,06	8,78	9,81	9,12
dEST-40	5,39	7,20	2,00	5,81	5,03	5,12	6,47	5,10	5,24	5,10	4,79	6,58	4,94	4,79	5,21	5,40	4,35	4,99	4,55	4,95	5,83	5,52	5,20
dEST-6	4,06	4,16	2,95	4,70	4,32	3,62	4,82	4,20	4,20	4,39	4,79	4,17	4,25	2,97	4,89	4,20	3,96	3,90	4,45	3,95	5,22	4,73	4,22
dEST-28	10,57	10,80	2,84	10,98	11,47	9,26	10,09	10,37	11,23	12,72	10,99	11,28	10,66	13,74	11,38	12,16	10,36	12,99	10,90	8,09	12,18	10,95	10,73
dEST-13	4,51	4,45	1,00	3,48	4,29	4,73	3,94	5,24	4,52	3,95	4,50	5,07	3,50	7,07	4,10	4,77	3,36	3,99	3,45	5,45	3,95	4,49	4,26
dEST-12	3,48	2,28	1,95	1,93	2,85	1,99	3,04	2,87	2,62	2,59	2,44	2,74	2,67	4,61	2,92	2,43	3,22	2,00	2,58	3,61	2,32	2,36	2,70
dEST-2	13,28	12,12	2,97	12,47	13,45	11,33	12,97	12,73	12,43	13,61	13,32	13,19	12,67	12,30	12,58	12,96	10,94	10,48	10,72	12,30	12,18	12,60	11,98
dEST-37	15,03	15,53	3,79	15,51	15,33	11,71	13,81	15,25	14,75	15,25	15,20	13,93	14,47	16,01	15,48	15,33	14,67	13,79	15,59	16,01	14,73	15,29	14,39
DL01	14,12	14,49	3,54	15,20	13,93	11,53	13,55	14,62	15,22	13,80	14,45	14,34	13,51	12,95	14,73	14,39	13,21	14,69	15,04	17,20	14,12	14,03	13,76
DL06	12,14	13,01	4,54	13,01	12,44	10,69	12,89	12,21	12,88	13,19	13,59	13,14	12,94	9,67	12,10	12,38	10,78	10,38	11,97	10,29	13,64	14,04	11,91
DL09				12,92				12,57	12,35	13,35					13,83	12,19	13,81					16,25	13,41
SgR63				15,66				16,34	17,14	16,23					16,17	15,44	16,22					17,24	16,30
DL13				17,06				16,61	19,93	20,91					17,29	16,50	14,16					17,50	17,49
Moyenne	9,70	9,77	3,10	10,33	9,60	7,82	9,44	10,04	10,42	10,48	9,63	9,35	9,27	9,51	10,26	10,09	9,45	9,14	9,39	9,24	9,47	10,53	9,89

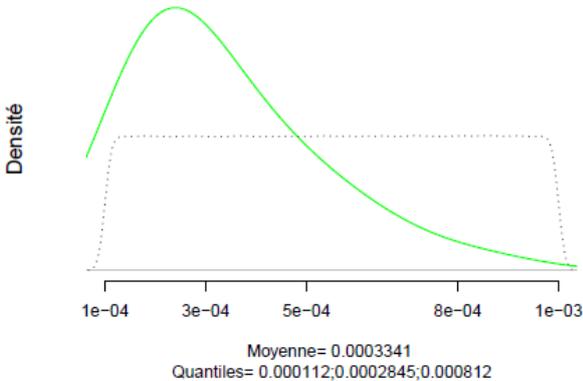
Annexe 12 : Posterior des paramètres en sortie de DIYABC pour le modèle retenu avec variation de la taille efficace suite à la divergence



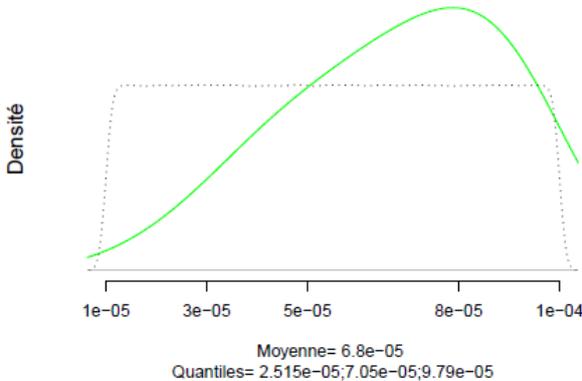


Paramètres mutationnels

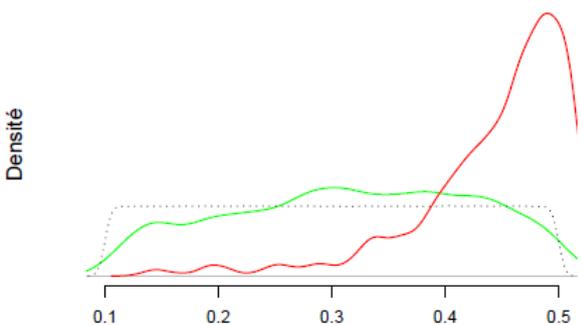
μ_1 taux de mutation moyen 1



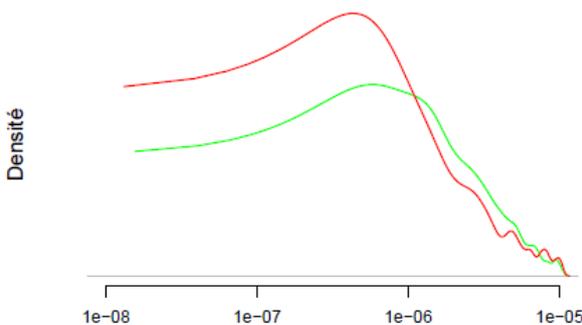
μ_2 taux de mutation moyen 2



Moyenne P

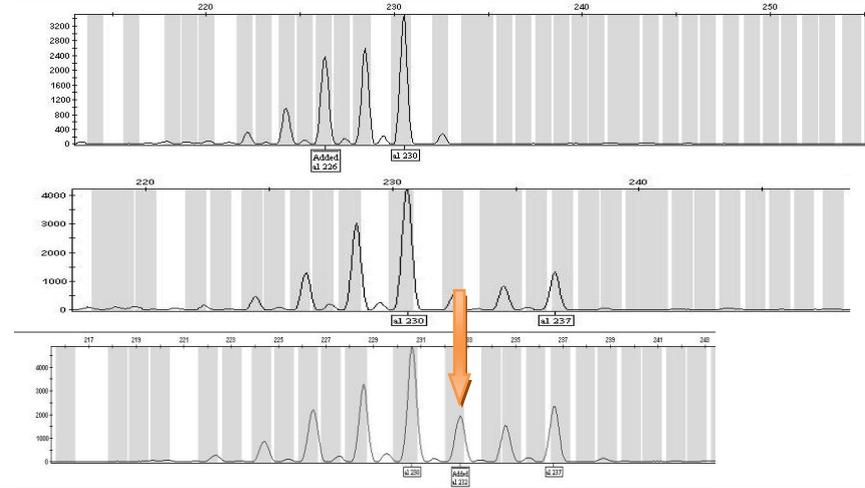


Moyenne SNI

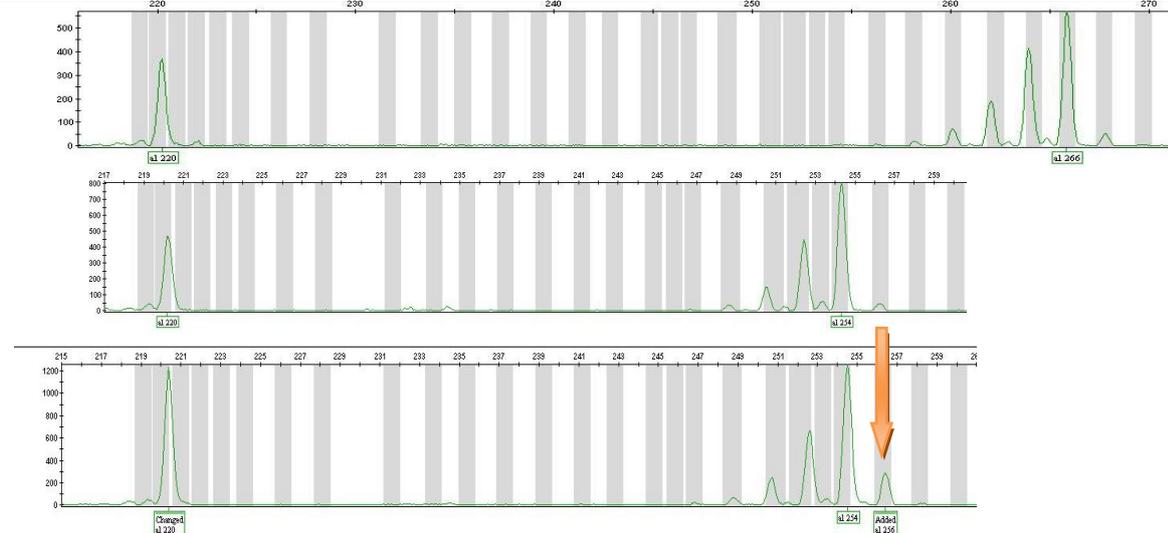


Annexe 13 : Profil GeneMapper des « pools » de 4 individus où l'on suspecte une mutation sérieuse (les deux premières lignes correspondent aux profils parentaux et la troisième au profil du « pool »)

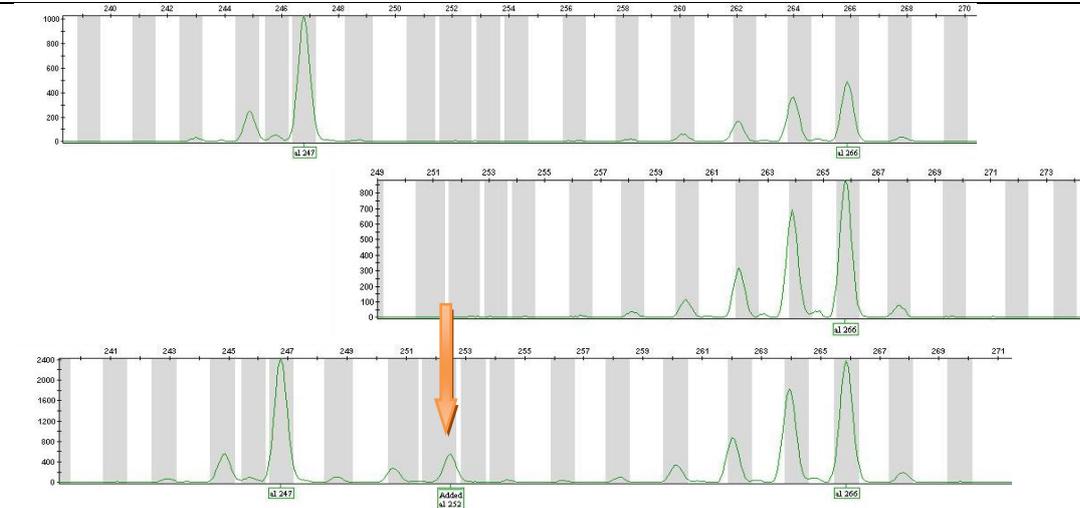
Allèle 232 mutant (marqueur SgM51)



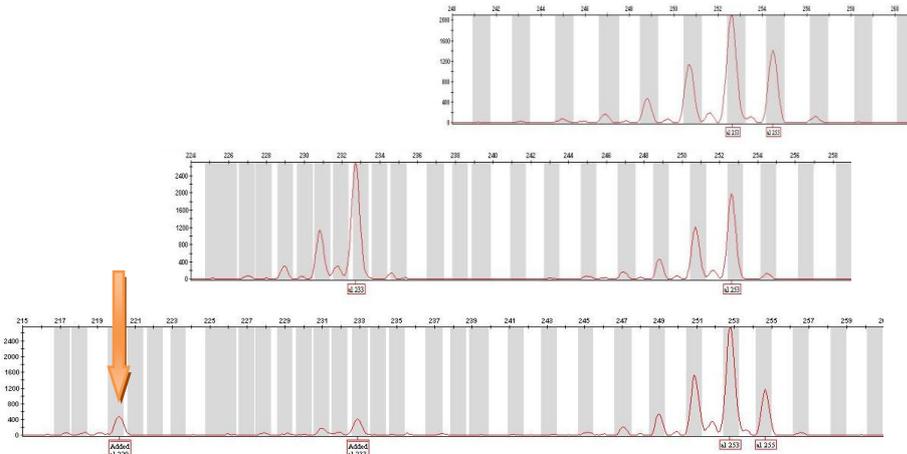
Allèle 256 mutant (marqueur SgM74)



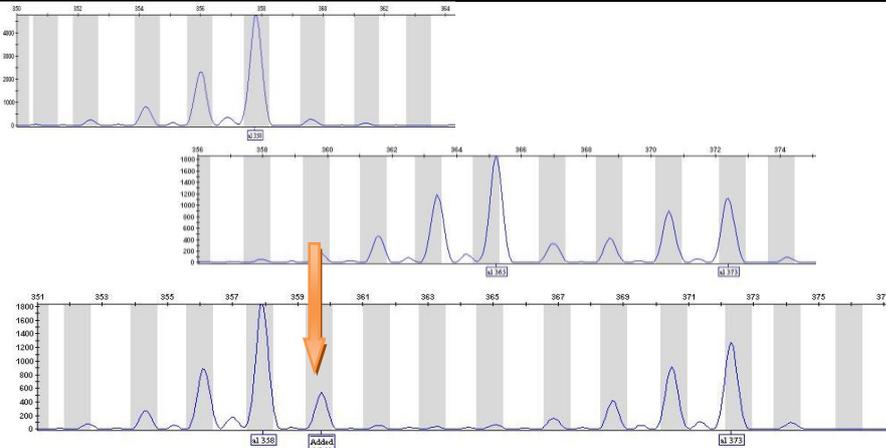
Allèle 252 mutant (marqueur SgM74)



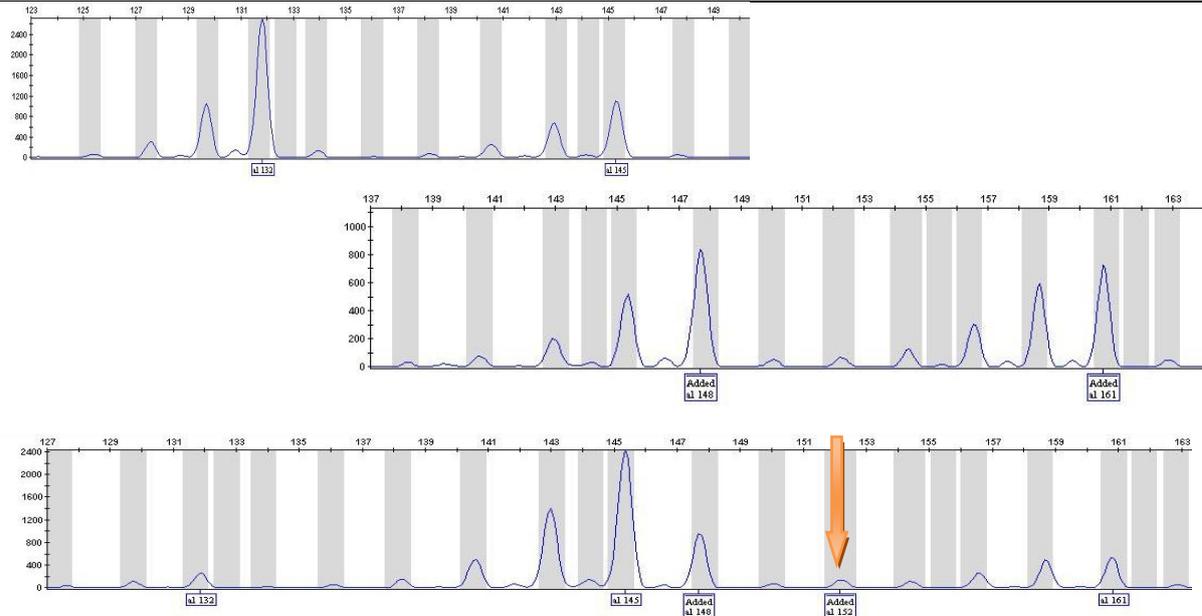
Allèle 220 mutant (marqueur DL13)



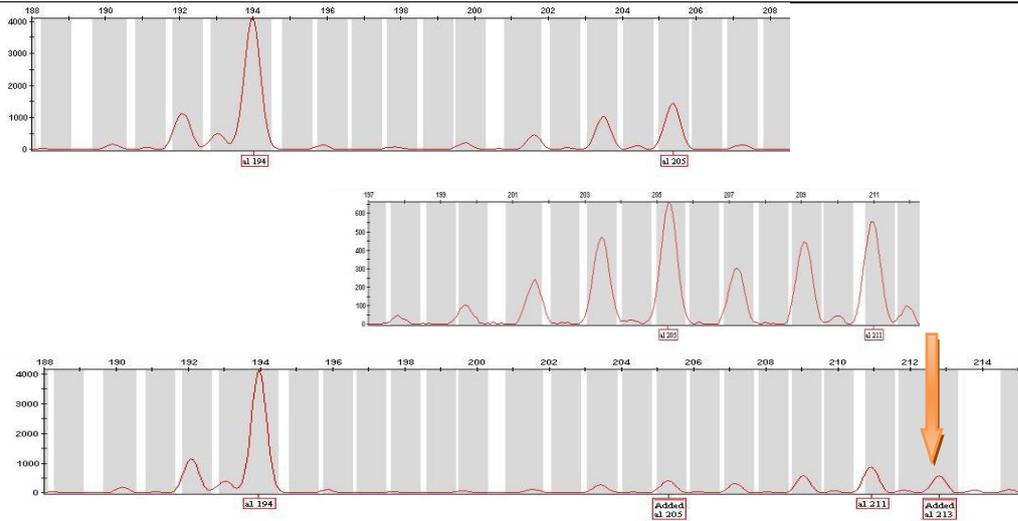
Allèle 360 mutant (marqueur Sgr36)



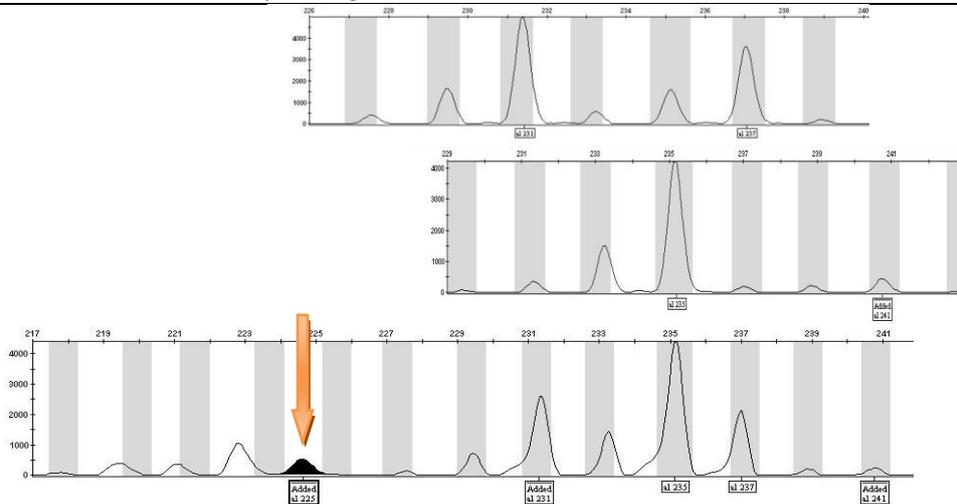
Allèle 152 mutant (marqueur SgM66)



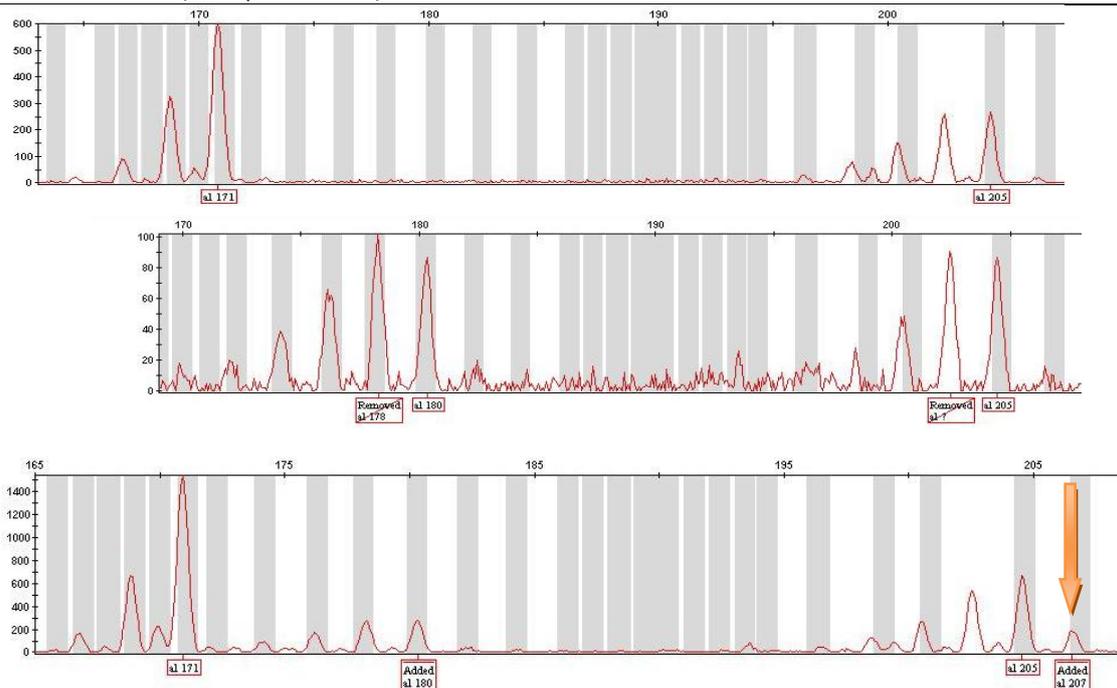
Allèle 213 mutant (marqueur SgM96)



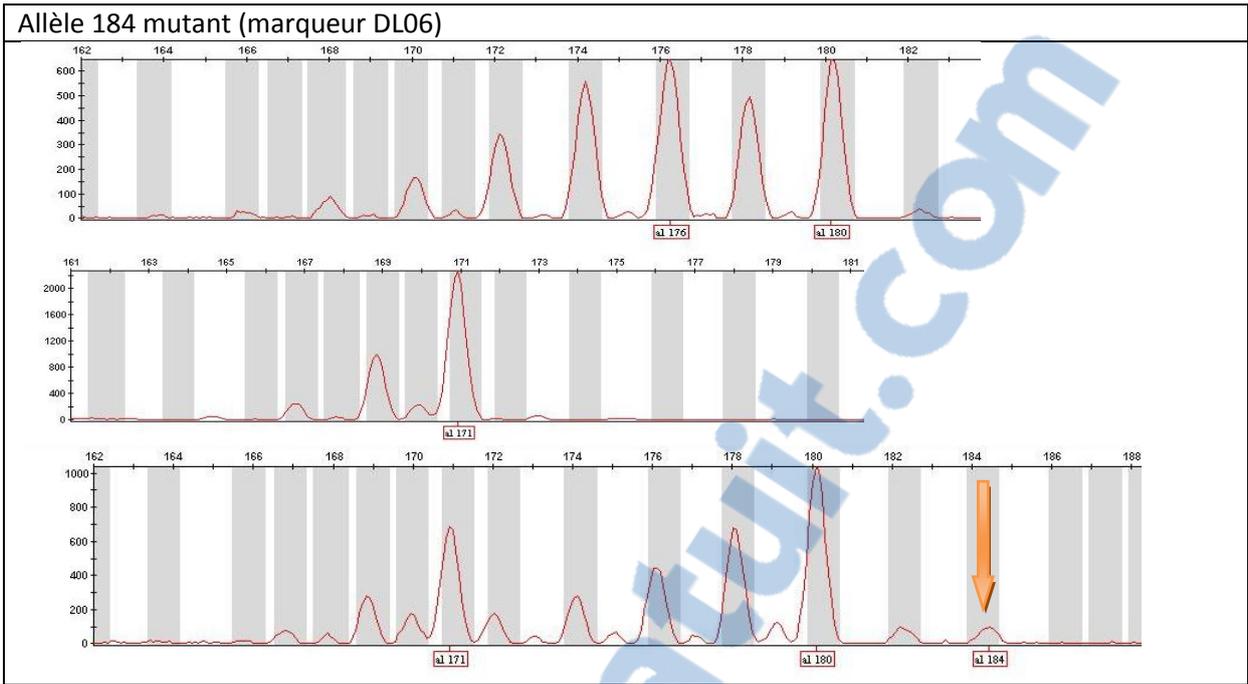
Allèle 225 mutant (marqueur Sgr53)



Allèle 207 mutant (marqueur DL06)



Allèle 184 mutant (marqueur DL06)



Résumé

Le Criquet pèlerin, *Schistocerca gregaria* est une espèce d'importance économique majeure au niveau mondial, qui sévit en Afrique du Nord, au Moyen-Orient et en Asie. Cet insecte ravageur est soumis à des augmentations énormes et imprévisibles de la taille de ses populations, associées à un changement d'une phase solitaire, inactive et dispersée, à une phase grégaire, très mobile et agrégative. Bien que *S. gregaria* soit l'espèce de locuste la plus redoutée et étudiée, son histoire évolutive et sa structuration génétique sont encore mal connues. L'objectif de ce stage est de décrire, à l'aide de 28 marqueurs microsatellites, les relations génétiques entre la sous-espèce nominale *S. g. gregaria* très redoutée et la sous-espèce *S. g. flaviventris*, anecdotique en Afrique australe et dont les populations présentent une faible propension à grégairiser. Nous nous intéressons à l'influence des processus neutres sur la différence observée entre ces deux sous-espèces. Nous avons utilisé une méthode d'inférence bayésienne basée sur la coalescence pour estimer le temps de divergence entre les deux sous-espèces et les tailles efficaces de chacune d'elles. L'histoire démographique des sous-espèces présentent en fait un événement d'expansion récent avec un temps de divergence de 7 800 ans environ. Les tailles efficaces des sous-espèces sont grandes et s'élèvent à plusieurs centaines de milliers d'individus. La force de dérive qui s'exerce sur nos sous-espèces n'est alors pas très élevée. Afin d'augmenter la précision d'inférence des paramètres démographiques par l'approche bayésienne utilisée, nous avons estimé expérimentalement le taux de mutations germinales aux marqueurs microsatellites. Nous avons observé un taux global de mutation de 2.10^{-4} ce qui est comparable à l'attendu.

Mot-clé : Criquet pèlerin, microsatellite, ABC, dérive, taux mutation germinale, biogéographie

Abstract

The desert locust, *Schistocerca gregaria*, is one of the major pests in the world, mostly in West Africa, in Middle East and in some Asian countries. This insect is able to increase drastically its population, while it changes from one solitary phase to a very active gregarious phase. Gregarious locusts change their behavior and aggregate a lot, forming swarms which can disperse really far. When huge outbreak occurs the agricultural production is at risk. Although it is a well known pest, its evolutionary history and its genetic structure has not been really studied. The aim of this work is to study the genetic link between two subspecies which slightly differ in the ability to gregarise: the infamous *S. g. gregaria* and *S. g. flaviventris* which swarms only rarely. We focus particularly on neutral processes that could have influenced this difference in the ability of gregarisation. Population genetics is a useful tool to infer population demographic parameters such time since divergence and occurrence of gene flow between *S. g. flaviventris* and *S. g. gregaria*, by applying Bayesian inference based on the coalescent to the genotyped populations. The demographic history that we identified includes an expansion event for both species just after the divergence around 7 800 years ago. The effective sizes of the subspecies are quite big because they can reach several hundred thousand. With these values, the drift is expected not to play a big influence on differentiation. To improve the mutation model accuracy in the Bayesian inference, 2000 nymphs were genotyped to estimate the microsatellite mutation rate. We observe a global mutation rate of 2.10^{-4} that is not far from literature estimation.

Key-words : Desert locust, microsatellite, ABC, drift, mutation rate, biogeography