

Echantillonnage

Cette étude concerne des adultes voiliers et non voiliers de *C. maculatus* qui ont déjà été échantillonnés au Sénégal, mis en élevage au laboratoire et conservés dans de l'éthanol après leur émergence. Tous provenant du Sénégal, nous avons défini comme population, tous les individus ayant la même capacité voilière. Tout compte fait, nous avons utilisé CmV pour coder les individus de la forme voilière et CmNV pour la forme non voilière.

II.2. Extraction de l'ADN

Première étape des manipulations au laboratoire, l'extraction de l'ADN des individus de *C. maculatus* a permis d'isoler l'ADN sous forme de solution aqueuse qui sera utiliser dans les étapes suivantes. Cette extraction a été réalisée en utilisant la trousse d'extraction QIAGEN DNeasy tissue et en suivant le protocole d'extraction standard des insectes (voir annexe 1). La tête, le thorax et les pattes de l'insecte sont seulement utilisés pour l'extraction de l'ADN. A la fin, pour vérifier la qualité totale de l'ADN extraite, une migration électrophorétique sur gel d'agarose 1,5% d'un mélange de 7 µl d'ADN et de 3 µl de bleu de charge est réalisée. Pour cela, le mélange est déposé dans les puits creusés avec des peignes lors du moulage dans le gel d'agarose. Ce dernier se trouvant dans la cuve remplie de solution tampon est exposé à une tension de 100 volts pendant 35 minutes. La révélation est faite après avoir laissé tremper le gel d'agarose dans du bromure d'éthidium pendant 30 minutes et rincer à l'eau par la suite avant la fluorescence sous lumière ultraviolet dans l'e-box V2®.

II.3. PCR et séquençage de l'ADN

La réaction de polymérisation en chaîne ou PCR permet de générer un nombre considérable de fragments identiques à la séquence cible facilitant ainsi le séquençage. Cette technique d'amplification *in vitro* de l'ADN repose sur la répétition de trois étapes : la dénaturation de l'ADN, l'hybridation des amorces sur la matrice et enfin l'élongation 3' du brin néosynthétisé par la Taq polymérase (une enzyme thermophile). Cette réaction se fait en présence de l'ADN extrait, des désoxyribonucléotides (dNTP), d'ion Mg^{2+} et des amorces Forward et Reverse du gène d'intérêt.

Pour notre travail, le gène d'intérêt est le cytochrome B, un gène mitochondrial. Gène à transmission maternel, le cytochrome B présente beaucoup d'intérêt dans les études visant à retracer l'histoire et la diversité des populations d'insecte. Il a été séquencé pour la première

fois chez *Sitophilus spp*, ravageur des graines de maïs et c'est également révélé polymorphe et discriminant chez les insectes. Pour ce gène, les amorces utilisées sont le mtD26 (5'- TATGT ACTACCATGAGGACAAATATC-3') et mtD28 (5'-ATTACACCTCCTAATTTATTAGGA AT-3') (Barry *et al.*, 2017) (Voir Annexe 2 et 3).

Le séquençage de l'ADN constitue une méthode dont le but est de déterminer la succession linéaire des bases A, C, G et T prenant part à la structure de l'ADN. Le séquençage se fait en suivant la technique de Sanger. Cette technique se fait en présence des composés habituellement utilisés lors de la synthèse du brin d'ADN lors d'une PCR en plus de didésoxynucléotides. Ces derniers sont des dNTPS dont le groupement OH a été remplacé par un groupement H et ont la particularité d'être couplés à des fluorochromes : ddATP-vert, ddTTP-rouge, ddCTP-bleu et ddGTP-jaune. Lorsqu'un ddNTP est incorporé à la place d'un dNTP, l'ADN polymérase ne peut plus continuer sa polymérisation. La réaction d'extension s'arrête car le ddNTP ne possède pas le groupe 3'-hydroxyle indispensable à la réaction de polymérisation. Statistiquement, au cours de la réaction, pour chaque « base » de l'ADN cible, au moins une fois, un ddNTP complémentaire sera incorporé à la place d'un dNTP. Par conséquent, à la fin de réaction, nous obtiendrons des fragments de tailles différentes. Lors de la migration, chaque fragment (contenant un ddNTP marqué par un fluorophore) sera excité par un laser et le signal obtenu sera analysé par un logiciel spécifique. L'analyse informatique des signaux permet d'obtenir la séquence étudiée, sous forme d'électrophorégramme de lecture manuelle aisée.

Pour nos travaux, le séquençage a été réalisé en externe pour un volume réactionnel de 50 µl de produit PCR par individu.

II.4. Analyses génétiques

II.4.1. Le nettoyage et alignement des séquences

Les séquences obtenues sont nettoyées et alignées à l'aide de la fonction ClustalW multiple alignement (Thompson *et al.*, 1994) du logiciel BioEdit version 7.2.5 (Hall, 1999). C'est une étape importante de l'analyse permettant de vérifier les éventuels gaps et mutations, de relever et corriger les erreurs éventuelles des séquences de chaque individus grâce aux chromatogrammes puis aligner l'ensemble des séquences. Cet alignement permet de ressortir les similitudes, les délétions, les insertions et/ou les substitutions.

II.4.2. Diversité génétique

A la suite de l'alignement, les paramètres basiques de la diversité génétique des séquences sont ressortis. Ces paramètres permettent de faire une description brute du jeu de

données global, de localiser les modifications pointues du cytochrome b de chaque forme et d'obtenir une estimation de leur polymorphisme. Ainsi grâce au logiciel DnaSP version 6.12.01 (Rozas *et al.*, 2017), le nombre de séquences (**n**), le nombre total de sites (**N**), le nombre de sites invariables (**i**) et également le nombre de sites variables (**v**) ont été déterminés. Parmi ces derniers, sont distingués les singletons variables et les informatifs en parcimonie. Le nombre total de mutation (**Eta**) correspondant au nombre total de sites synonymes et non synonymes a été également déterminé avec DnaSP 6.12.01 (Rozas *et al.*, 2017). Ce dernier a permis déterminer le nombre d'haplotype (**h**), le nombre moyen de différences nucléotidiques (**k**) de même que les diversités haplotypique (**Hd**) et nucléotidique (**Pi**) pour chaque population. Les taux de transversion (**v**), de transition (**s**) et de mutation (**R**), ainsi que les taux de mutations synonymes (**Ks**) et non synonymes (**Kns**) ont été calculés avec le logiciel MEGA version X (Kumar *et al.*, 2018). Les changements nucléotidiques et d'acide animés sont ressortis avec le logiciel MEGA version X (Kumar *et al.*, 2018).

II.4.3. Différenciation et structuration génétiques

L'indice de différenciation (**F_{ST}**), est un indice de différenciation génétique entre subdivisions d'une population. Le **F_{ST}** représente la corrélation entre allèles à l'intérieur d'une sous population par rapport à l'ensemble des sous populations. Il est déterminé par le logiciel Arlequin version 3.1 (Excoffier *et al.*, 2005).

La distance génétique de Nei (**D**) est un indice permettant de quantifier la divergence génétique existant entre les populations prises deux à deux. Cette distance présente des propriétés spécifiques et reste appropriée pour un modèle particulier d'évolution. Elle considère un modèle mutation-dérive et elle est destinée à mesurer le nombre de différences de codons entre une populations après leur divergence. Cette distance est la plus utilisée dans les recherches sur la diversité génétique. Pour ce travail, les distances génétiques inter et intra population sont obtenues en utilisant le modèle Kimura 2-parameter (1980) pour les deux formes de *C. maculatus* grâce au logiciel MEGA version X (Kumar *et al.*, 2018). Le choix du paramètre étant fait après avoir testé les modèles compatibles à ce jeu de données avec le logiciel MEGA. Au terme de ce test, le meilleur paramètre est celui qui a la plus petite valeur d'AIC (Akaike Information Criterion).

La structuration génétique représente la part de la différenciation génétique entre les entités de chaque niveau populationnel (groupes, populations, individus) dans la variabilité génétique global. Elle est obtenue en faisant une analyse de la variance moléculaire (AMOVA).

Le calcul de l'indice de structuration est réalisé en utilisant les informations sur le contenu haplotypique du gène de même que sur leurs fréquences. Cette analyse est faite en utilisant le logiciel Arlequin version 3.1 (Excoffier *et al.*, 2005).

II.4.4. Evolution démographique

Au cours de nos travaux, les indices démo-génétiques tels que le D de Tajima et le F_S de Fu sont calculés dans le but d'estimer l'évolution démographique de chaque forme de *C. maculatus* au Sénégal. Ces indices sont obtenus grâce au logiciel Arlequin version 3.1 (Excoffier *et al.*, 2005) et le niveau de significativité a été évalué après 10000 simulations coalescentes (Simulation consistant à retracer la généalogie jusqu'à trouver l'ancêtre commun le plus récent).

Les valeurs de la distribution du nombre observé de différences nucléotidiques entre paires d'haplotypes « mismatch distribution » sont obtenues avec le logiciel Arlequin version 3.1 (Excoffier *et al.*, 2005) et les graphiques sont obtenus par la suite grâce au logiciel Excel. Pour tester la qualité d'ajustement de la distribution, la somme des carrés des déviations (SSD) entre les distributions observées et attendues de même que l'indice d'irrégularité (Rg) ont été calculés et déterminés sous l'hypothèse d'une population en expansion en utilisant ARLEQUIN V3.1 (Excoffier *et al.*, 2005). La statistique R2 (Ramos-Onsin & Rozas, 2002), basée sur les différences entre le nombre de mutations de type singleton et la moyenne des différences nucléotidiques, est également calculée de façon à compléter les analyses précédentes avec DnaSP version 6.12.01 (Rozas *et al.*, 2017).

II.4.5. Reconstructions phylogénétiques

Afin de visualiser les relations existantes entre les individus, des arbres phylogénétiques ont été générés. Ces arbres sont réalisés selon les méthodes du Neighbor-Joining, du maximum de vraisemblance avec le logiciel MEGA version X (Kumar *et al.*, 2018) et enfin celle de l'inférence bayésienne avec le logiciel Mrbayes version 3.2.6 (Ronquist *et al.*, 2011). Le modèle Kimura-2-parameter est associé à 10000 randomisations par bootsrap pour l'obtention des deux premiers arbres. La méthode du Neighbor-Joining permet une construction rapide basée sur les distances génétiques. Cependant cette méthode peut induire une perte d'informations. La méthode de vraisemblance permet de tester toutes les histoires ayant pu engendrer le jeu de données actuelles analysés. La vraisemblance de chaque arbre est ainsi calculée en suivant le modèle choisi, et la reconstruction phylogénétique choisie est celle pour laquelle la vraisemblance est maximale. L'arbre de l'approche bayésienne a été réalisé après 1 000 000 de

génération pour chacune des chaînes en échantillonnant les différents paramètres toutes les 1000 générations. La distribution des probabilités postérieures des arbres a été estimée par MC3 en utilisant quatre chaînes simultanément (dont trois ont été « chauffées » de façon graduelle). Les générations réalisées pendant la période d'allumage sont éliminées des analyses. De manière conservative, les 250000 premières générations ont été éliminées (25%) et les inférences sont alors réalisées sur les 750000 générations suivantes. Les reconstructions ont été enracinées avec une séquence de *Caryedon serratus*.