

Développement de marqueurs
moléculaires polymorphes chez une espèce
nouvellement introduite en **E**urope :
Leptoglossus occidentalis



REMERCIEMENTS

Je tiens à remercier en premier lieu Monsieur **Alain ROQUES**, Directeur de l'Unité de Zoologie Forestière, qui m'a permis d'effectuer mon stage au sein de son unité.

Ensuite, je tiens à remercier tout particulièrement mon maître de stage, Monsieur **Vincent LESIEUR**, Doctorant, qui m'a proposé ce stage, et qui m'a permis de faire partie de ce projet. Je tiens à le remercier pour la confiance qu'il m'a accordée en me permettant d'être aussi autonome, pour sa disponibilité et toute l'aide qu'il a pu m'apporter durant toutes ces semaines, et j'espère sincèrement que mon travail lui aura été utile pour soutenir sa thèse de Doctorat.

Je remercie Madame **Emmanuelle MAGNOUX**, Assistante Ingénieur, pour m'avoir accordé du temps pour les résultats de séquençage et pour l'aide apportée lors de la rédaction de la partie « Matériel et Méthode » de ce rapport.

Puis, je remercie aussi Madame **Marie-Anne AUGER-ROZENBERG**, pour ses explications sur les microsatellites, ainsi que sur leurs détections.

Je remercie aussi Madame **Claudine COURTIN**, Technicienne de Recherche, qui m'a encadré lors de ma présence en laboratoire, pour sa gentillesse et pour son aide précieuse.

Enfin, je remercie toute l'Unité, pour sa bonne humeur, son accueil, et sa disponibilité lorsque j'en ai eu besoin, ce qui m'a fait apprécier ce stage.

ABREVIATIONS

% : Pourcent

μL : microlitre

A : Adénine (base azotée)

ADN : Acide Désoxyribonucléique

AERES : Agence d'Évaluation de la Recherche et le l'Enseignement Supérieur

AGPF : Amélioration, Génétique et Physiologie Forestières

BET : Bromure d'éthidium

C : Cytosine (base azotée)

dNTP : désoxynucléotide triphosphate

G : Guanine (base azotée)

GBFor : Génétique et Biomasse Forestières

GPS : Global Positioning System

H₂O : Eau

INRA : Institut National de la Recherche Agronomique

mL : millilitre

mM : millimolaire (mmol.L⁻¹)

mmol : millimole

pb : paire de base

PCR : Polymerase Chain Reaction, (Réaction d'amplification en chaine)

qsp : Quantité suffisante pour

T : Thymine (base azotée)

Taq : Thermus aquaticus polymérase

UE : Unité Expérimentale

UR : Unité de Recherche

URZF : Unité de Recherche en Zoologie Forestière

GLOSSAIRE

ADN polymérase : Enzyme servant à la réplication de l'ADN.

Allèles : Forme possible d'un même gène qui produit des effets phénotypiques reconnaissables.

Amorce : Fragment d'ADN monocaténaire complémentaire avec l'une des extrémités de la séquence cible d'ADN à amplifier et servant de point de départ à une ADN polymérase.

Gène : Séquence d'ADN qui code pour la synthèse d'une protéine ou d'un ARN fonctionnel.

Génome : Ensemble du matériel génétique d'un individu.

Hémiptère : Ordre taxonomique d'Insectes.

Hétérozygotie : Caractéristique d'un individu possédant 2 allèles différents sur un locus donné.

Homozygotie : Caractéristique d'un individu possédant 2 allèles identiques sur un locus donné.

Locus : Emplacement exact d'un gène sur un chromosome.

Microsatellites : Courtes régions d'ADN non codantes présentant un motif répété qui peut porter sur 2 à 5 nucléotides.

Mitochondrie : Organite intracellulaire caractéristique des Eucaryotes possédant son propre génome. Celui-ci se transmet de génération en génération uniquement par la mère.

Phylogéographie : Etude des lignées généalogiques, leur processus, leur distribution.

Qsp : « Quantité suffisante pour », quantité nécessaire pour atteindre un volume donné.

Séquence : Succession de bases nucléotidiques constitutives de l'ADN.

Taq polymérase : ADN polymérase thermorésistante intervenant dans la réplication de l'ADN et utilisée lors d'une intense chaleur. Cette enzyme a un faible taux d'erreur (1 sur un million de paires de bases), ce qui permet de l'utiliser pour le séquençage.

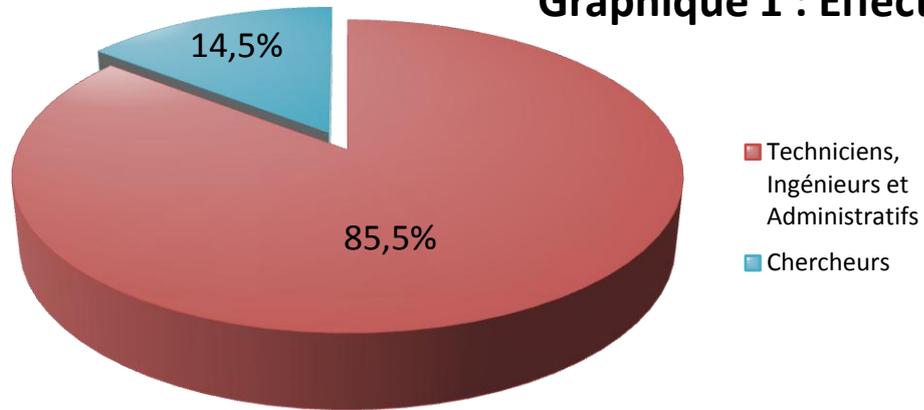
SOMMAIRE

Partie 1 - L'entreprise	1
I. L'INRA.....	2
A. Historique.....	2
B. Quelques chiffres	3
II. Le centre d'Orléans.....	3
A. Les effectifs du centre d'Orléans.....	2
B. Les moyens financiers	3
III. L'Unité de Recherche de Zoologie Forestière.....	3
Partie 2 - Le sujet de Recherche	4
I. Introduction	5
II. Généralités.....	6
1. Biologie Animale	6
2. Impacts de <i>Leptoglossus occidentalis</i>	6
III. Matériel et Méthodes.....	6
A. Microsatellites : Quelques informations.....	6
B. Obtention de données et études de variabilité	7
1. Echantillonnage	7
2. L'extraction d'ADN	7
3. Choix des amorces.....	8
4. L'amplification de l'ADN par PCR	9
5. Etape de préparation au séquençage.....	10
6. Séquençage et analyse	10
IV. Résultats expérimentaux	11
A. Analyse des microsatellites	11
1. Marqueurs développés par GenoScreen.....	11
2. Marqueurs développés à l'INRA.....	13
B. Analyse des données.....	14
V. Conclusions	15
A. Conclusion scientifique	15
B. Conclusion personnelle	15
Bibliographie	16
ANNEXES	17

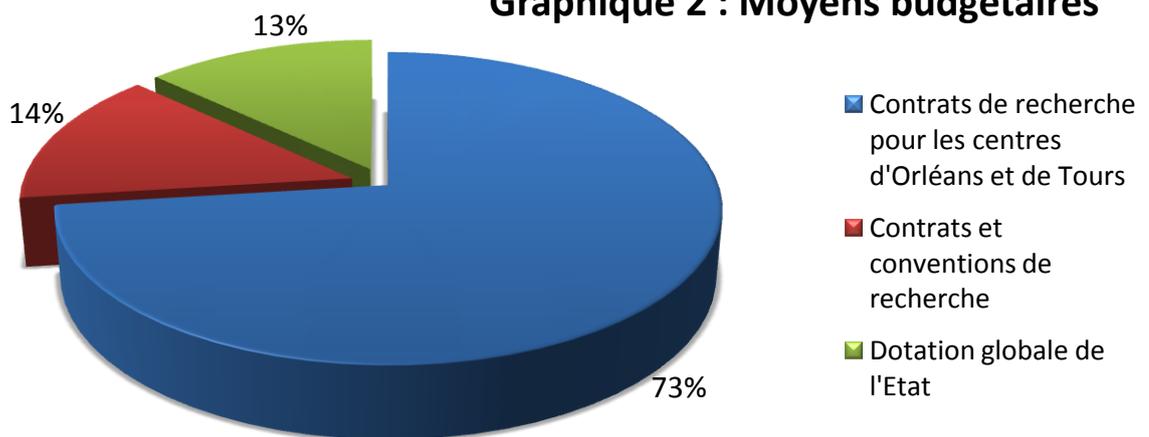
Partie 1

L'entreprise

Graphique 1 : Effectifs



Graphique 2 : Moyens budgétaires



I. L'INRA

A. Historique

1921 : Création de l'Institut de Recherches Agronomiques (IRA) organisé par disciplines scientifiques et par centre de recherche à vocation régionale.

1924 : L'institut s'installe à Versailles sous le nom de CNRA (Centre National de Recherche Agronomique).

Les centres suivants sont ceux de Versailles (centre national), Antibes, Bordeaux, Clermont-Ferrand, Colmar et Montpellier (centres régionaux).

1946 : Création de l'Institut National de la Recherche Agronomique d'après la loi du 18 mai.

Il s'agit d'un Etablissement Public Administratif placé sous l'autorité du Ministère chargé de l'Agriculture, qui organise, exécute, et publie les travaux de recherche relatifs à l'amélioration et au développement des productions végétales et animales. Il transforme et conserve les produits agricoles.

1984 : Transformation de l'INRA en Etablissement Public à caractère Scientifique et Technologique (EPST), géré par les Ministères de la Recherche et de l'Agriculture.

Aujourd'hui, l'INRA est la deuxième institution de recherche mondiale en agronomie, la première en Europe dans les domaines de l'agriculture, de l'alimentation et de l'environnement.

B. Quelques chiffres

L'INRA est composé de 19 centres répartis sur l'ensemble du territoire français engagés dans 21 pôles thématiques.

- 49 unités expérimentales, 200 unités de recherche (dont les 2/3 sont mixtes associant d'autres organismes, des universités, des écoles ou des instituts techniques).
- Le budget de l'INRA, en 2012, s'élève à près de 845 millions d'euros, dont 70% sont consacrés aux charges salariales.
- Les ressources propres en 2011 sont composées à :
 - 35% en dotation globale (Etat).
 - 65% en contrats et conventions de recherche

II. Le centre d'Orléans

Le centre de recherche d'Orléans est composé de 7 unités :

- **3 unités de recherche** (UR AGPF, UR ZF et UR Science du sol)
- **1 unité de service** (US Infosol)
- **2 unités expérimentales** (UE GBFor et UE Bourges)
- **1 unité expérimentale** (SDAR)

3 départements de recherche sont représentés :

- Ecologie, Prairie et Milieux Aquatiques
- Environnement et Agronomie
- Génétique Animale

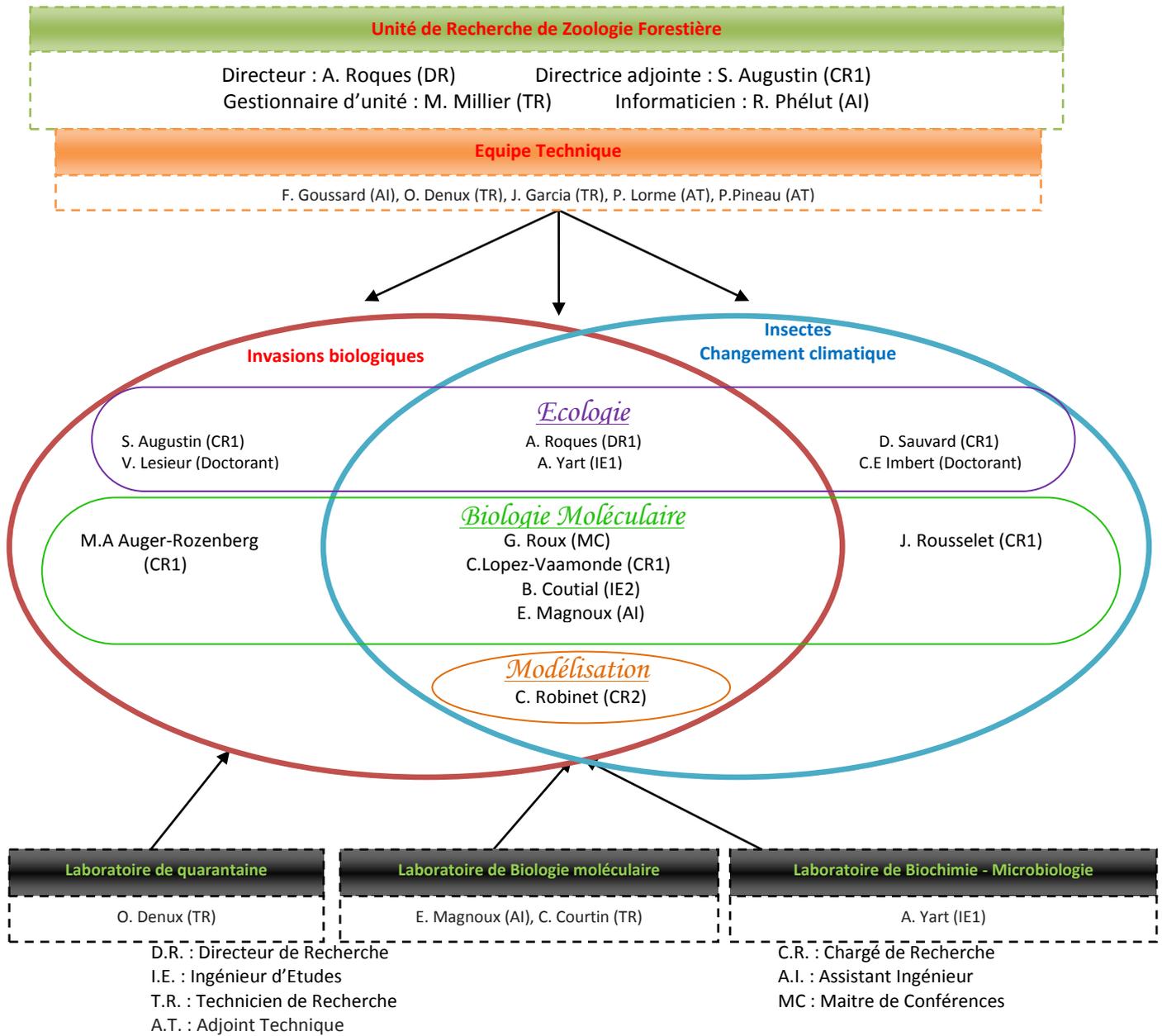
A. Les effectifs du centre d'Orléans (Cf. Graphique 1)

Les effectifs du Centre d'Orléans sont de **186 agents titulaires** et 144 agents non titulaires en 2011.

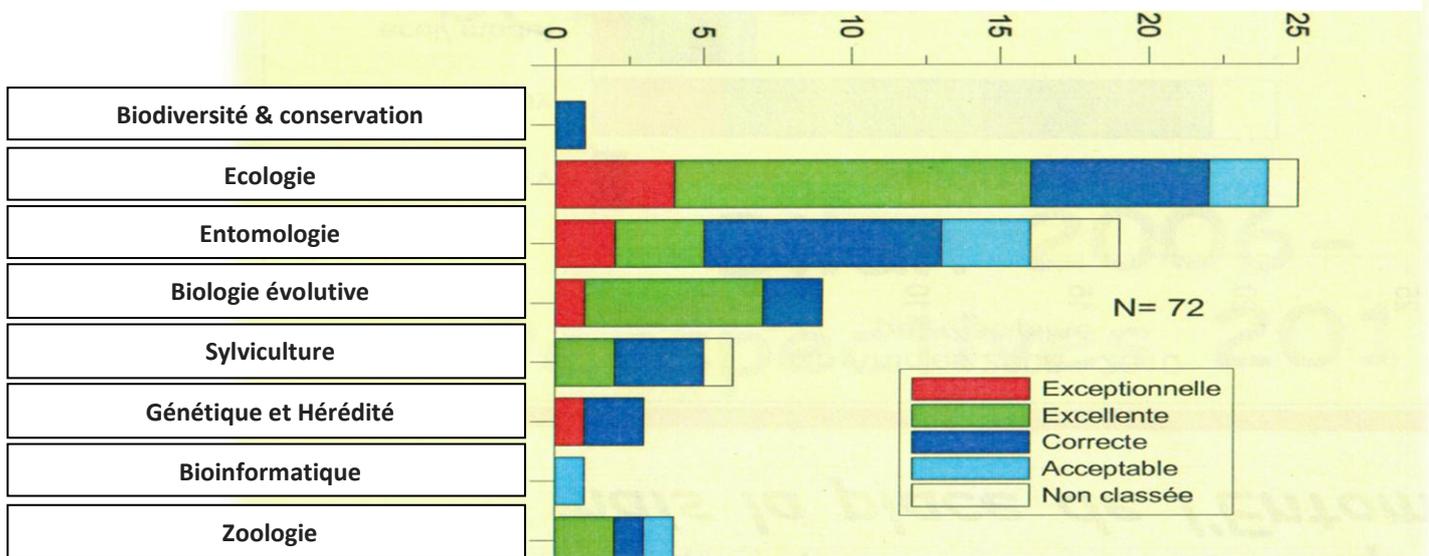
- | | |
|------------------------------|-------------------------------|
| ◆ 6 Directeurs de Recherche | ◆ 27 Assistants Ingénieur |
| ◆ 21 Chargés de Recherche | ◆ 55 Techniciens de Recherche |
| ◆ 18 Ingénieurs de Recherche | ◆ 38 Agents Technique |
| ◆ 21 Ingénieurs d'Etude | |

Document 1

Personnel de l'URZF



Graphique 3 : Nombre d'articles par thème en fonction de leur qualité



De plus, l'INRA d'Orléans accueille 144 agents non-titulaires en 2011, comme par exemple des étudiants, doctorants, post-doctorants, et chercheurs étrangers.

B. Les moyens financiers

La somme des moyens budgétaires de l'INRA d'Orléans, s'élevant à 12.3 millions d'euros en 2012, est issue de diverses sources. Ainsi, la répartition de ces ressources financières est disposée suivant le **Graphique 2**.

III. L'Unité de Recherche de Zoologie Forestière

L'unité de Zoologie Forestière a plus une vocation de pôle d'entomologie que de biologie animale générale. En effet, elle a pour but de comprendre les mécanismes qui motivent les populations d'insectes. Les deux principales thématiques de l'unité sont par ailleurs intitulées « Populations en expansion avec le changement climatique » et « Invasions biologiques ».

Les recherches au sein de l'unité ont tant une vocation de prévention envers certaines espèces, comme par exemple la Processionnaire du pin, que de protection, comme pour l'Isabelle de France, une espèce menacée d'extinction.

Par conséquent, plusieurs actions ont été effectuées. Un inventaire des espèces exotiques d'insectes présents en Europe a été créé, permettant d'identifier les risques et l'origine des espèces invasives. La phéromone sexuelle de l'Isabelle a pu être identifiée, ce qui pourra concourir à sa sauvegarde. Un modèle d'expansion a été mis en place en ce qui concerne la Processionnaire, en fonction des différents scénarii possibles du réchauffement climatique.

Outre ces projets d'envergure européenne, d'autres sont élevés au niveau international, comme par exemple un test programmé en Chine, où l'implantation d'arbres européens permet de définir les insectes potentiellement envahissants pouvant élire domicile en Europe.

Les partenariats sur les sujets d'études de l'URZF sont présents dans le monde entier, comme la Chine, la Suisse, l'Italie, le Maroc, le Canada,...

A cause d'un incendie accidentel, l'URZF a été en grande partie détruite. Sa seconde inauguration s'est faite en 2005.

Le personnel de l'unité, 20 personnes au total, est composé de 7 Chercheurs INRA, 5 ingénieurs INRA, 1 Maître de conférence et 7 techniciens. La totalité du personnel atteint une trentaine de personnes en comptant le personnel non-titulaire, tel que les stagiaires, les masters, les doctorants,... (**Document 1**).

De plus, l'unité a accueilli, depuis 2006, 6 doctorants, 7 Masters II d'Orléans, Rennes, ou encore Valencia (Espagne), 14 Masters I d'Orléans ou Metz et 12 BTS en Biotechnologies de Saint-Jean-de-Braye et Pontivy. Elle a également accueilli des post-doctorants étrangers, venus de Chine, de Russie ou du Mexique, ainsi qu'un chercheur d'Angleterre.

Entre 2006 et 2010, l'URZF a produit 72 publications scientifiques. Chaque publication est évaluée par l'INRA ou par l'AERES et classée en fonction de sa mention allant de l'article non classé à l'article exceptionnel. En plus de ce classement, les publications sont regroupées par domaines (**Graphique 3**).

Partie 2

Le sujet de Recherche

I. Introduction

Les invasions biologiques, considérées comme la 2e cause de perte de biodiversité, ont connu une augmentation exponentielle depuis les 200 dernières années, en nette relation avec les activités humaines et notamment les échanges commerciaux intercontinentaux. Parmi ces invasions biologiques, les insectes occupent une place prépondérante. *Leptoglossus occidentalis*, la punaise prédatrice des graines de conifère, en est un bon exemple.

Originaire de l'ouest du continent nord-américain, cette punaise a été introduite au cours de la deuxième moitié du XXème dans le nord-est américain avec une première observation dans l'Iowa en 1956. Elle a ensuite colonisé une bonne partie de cette région et atteint la côte dans les années 1990 (McPherson *et al.*, 1990). La première observation de cet insecte en dehors du continent nord-américain a eu lieu en 1999 dans le nord de l'Italie (Villa *et al.*, 2001). Elle a par la suite envahi en très peu de temps, une dizaine d'année seulement, l'Europe centrale et l'Europe de l'ouest. En effet, cette espèce est maintenant présente depuis la Sicile jusqu'en Norvège et du Portugal jusqu'au sud de la Russie (Fent & Kment, 2011 ; Putshkov *et al.*, 2012).

L'introduction en Europe par voie maritime semble l'hypothèse la plus probable. Plusieurs observations dans d'importantes zones portuaires laissent supposer que l'invasion européenne résulte de multiples introductions via le trafic maritime. L'exemple du contrôle d'un container, en 2006 au Havre, a révélé la présence de nombreux individus au sein de ce dernier. Ce container contenait des planches de Chêne blanc en provenance des Etats-Unis (Dusoulier *et al.*, 2007).

Longtemps sous-estimé, cet Hémiptère est considéré comme le principal prédateur de graines de conifères dans les vergers à graines commerciaux du continent nord-américain (Blatt *et al.*, 2000 ; Strong, 2006). L'espèce possédant une progression fulgurante constitue donc une menace tant pour les forêts européennes de conifères que pour la production de graines à but commercial. Un projet est actuellement mené par l'URZF, ce projet s'insère dans un vaste programme européen "ISEFOR", qui a pour but de préserver la pérennité des forêts européennes. Le projet développé par l'URZF vise à étudier *L. occidentalis*, notamment à définir quelles peuvent être les conséquences de son introduction en Europe (essences de conifères attaqués, relations avec les espèces européennes d'insectes, etc...) mais aussi à déterminer la zone d'origine des populations européennes. La découverte des zones d'origine apparait comme un facteur important pour identifier d'éventuels parasites issus des mêmes zones et pouvant à terme être utilisés en lutte biologique. La description des routes d'invasion et la mise en évidence d'introductions multiples joue un rôle important dans la mise en place de méthodes de luttes appropriées, ainsi que dans la compréhension du phénomène d'invasion.

Mon stage s'insère dans cette problématique de recherche des zones d'origine des populations européennes de punaises à l'aide de marqueurs génétiques appropriés. Une première approche phylogéographique basée sur un marqueur mitochondrial (fragment du gène cytochrome b) suggérait que les populations européennes proviennent de la zone d'introduction primaire, c'est-à-dire l'est américain (Lesieur *et al.*, 2012). Cependant, cette étude ne permettait pas de conclure sur l'origine précise des populations d'Amérique de l'est et l'origine exacte des populations européennes reste incertaine.

L'utilisation de marqueurs plus polymorphes, tels que des marqueurs microsatellites, permettrait d'obtenir une vision plus précise de la structuration génétique des populations de *L. occidentalis* dans la zone d'origine mais aussi dans les zones d'introduction : L'Europe et l'est américain. L'objectif de ce stage est donc de tester différents marqueurs microsatellites disponibles au laboratoire, de déterminer les marqueurs polymorphes et de comparer les génotypes des échantillons provenant à la fois de la zone d'origine et des zones d'invasion grâce à un important jeu de données.

II. Généralités

Le matériel biologique : *Leptoglossus occidentalis*

1. Biologie Animale

Le cycle biologique de *Leptoglossus occidentalis* est relativement bien connu. Cet insecte présente 5 stades larvaires. Dans un climat tempéré, il y a une génération par an. Mais dans des climats plus chauds comme celui du Mexique, 3 générations peuvent se succéder la même année, chaque génération comptant tous ses stades larvaires.

La femelle pond en moyenne 80 œufs. Les adultes sont très différenciables des larves au niveau morphologique. En automne, les individus se regroupent pour hiverner, dans des bâtiments, des cavités d'arbres, ou des nids. Ce phénomène est le résultat d'une sécrétion de phéromone d'agrégation produite par le mâle. Même si la punaise ne représente pas un danger pour l'Homme, les rassemblements peuvent compter de quelques uns à quelques milliers d'individus. Ils sont suffisamment impressionnants pour inquiéter les populations et créer des nuisances.

Adultes et nymphes se nourrissent des graines de conifères. Ils insèrent leur stylet buccal entre les tissus des cônes jusqu'aux graines dont ils digèreront l'intérieur. Dans sa zone native, l'insecte ne semble pas être dépendant d'un type de conifère. Cependant, les dégâts semblent varier d'une essence à une autre et d'un stade de développement à un autre.

On dispose de peu d'informations quant à ses préférences en Europe. La punaise s'observe autant sur des pins, que des épicéas, des cèdres, des sapins, ou encore des genévriers.

2. Impacts de *Leptoglossus occidentalis*

Leptoglossus occidentalis est considéré comme premier ravageur de graines de conifères commerciaux. Son aspect ravageur a été longtemps sous-estimé du fait de la discrétion avec laquelle il consomme les graines. En effet, à vue d'œil, il est très difficile de distinguer une quelconque intrusion du stylet buccal de la punaise dans la graine.

De plus, la consommation d'une graine par l'insecte laisse peu ou pas de trace. De récentes études ont démontré que la plupart des graines vides et avortées observées en vergers, dont l'origine était préalablement attribuée à des problèmes de pollinisation correspondent en réalité à la prédation effectuée par ces punaises.

L'impact de *Leptoglossus occidentalis* n'est donc pas immédiat mais n'est remarquable qu'une fois la baisse de production de conifères constatée.

L'introduction d'une telle espèce peut également avoir des répercussions sur la faune présente dans la zone d'introduction notamment sur les autres insectes partageant la même ressource. Les relations entre *L. occidentalis* et *Pissodes validirostris*, un coléoptère présent dans l'hémisphère nord et dépendant de pins sont abordés au sein de l'URZF.

III. Matériel et Méthodes

A. Microsatellites : Quelques informations

Un microsatellite est une courte région d'ADN présentant une séquence de 2 à 5 nucléotides répétée en moyenne de 10 à 30 fois. Cette particularité du génome peut être monomorphe (nombre de répétitions fixe et invariable) ou polymorphe (taux de mutation élevée, nombre de répétition variable). Ce détail structural est induit par un dérapage répliatif : l'ADN polymérase réplique l'ADN, et à force de

répliquer le même motif, finit par déraiper, ajoutant ou supprimant une ou plusieurs unités de répétitions de nucléotides du microsatellite.

De plus, les individus étudiés étant diploïdes, chaque allèle est présent sur les 2 chromosomes complémentaires. Il s'agira alors de définir si l'individu est homozygote ou hétérozygote pour un locus donné.

Les microsatellites sont des marqueurs génétiques nucléaires présents sur l'ensemble du génome principalement dans les zones non codantes de l'ADN.

Ces marqueurs suivent les lois de Mendel. Par conséquent, une mutation qui survient sur un locus microsatellite ne sera pas éliminée. Elle sera donc conservée pour les générations suivantes.

B. Obtention de données et études de variabilité

Le processus de détection et d'observation de la variabilité génétique est composé de 6 étapes :

A - Echantillonnage

B - Extraction d'ADN

C - Amplification de l'ADN par PCR

D - Choix des amorces

E - Etape de préparation au séquençage

F - Séquençage et analyse

Durant mon stage, j'aurais eu à extraire l'ADN, l'amplifier, le préparer au séquençage puis analyser les résultats de plus de 410 individus (**Annexe1**).

1. Echantillonnage

Les punaises sont prélevées et sont mises sur le terrain en tube de 10 mL d'alcool à 95°. Les coordonnées géographiques sont enregistrées par GPS et associées à un identifiant provisoire. L'échantillonnage des punaises est effectué aussi bien dans la zone d'origine que dans les zones d'introduction.

Avant l'extraction d'ADN au laboratoire, on prélève sur une punaise une des deux pattes métathoraciques (3^{ème} paire de pattes) car elle contient plus de tissus musculaires que les autres. On peut toutefois utiliser les autres pattes en cas d'absences de celles-ci. On attribuera un code d'identification pour la punaise afin de lier la punaise à son futur tube d'extraction d'ADN. Le code est composé des 4 premières lettres du lieu d'échantillonnage et d'un nombre à 2 chiffres. Le tube qui contiendra la patte après dissection sera utilisé en totalité.

2. L'extraction d'ADN

La première étape consiste à extraire l'ADN génomique (des compartiments nucléaire et mitochondrial). Il faudra que l'ADN soit pur, et sans contaminants, comme des protéines, pour assurer sa conservation (enzymes dégradant l'ADN) et éviter d'inhiber la polymérase utilisée dans l'étape suivante lors de la PCR. Lors de l'utilisation d'amorces spécifiquement créées pour l'espèce comme c'est le cas dans notre étude, il faut par ailleurs veiller à éviter la contamination par d'autres sources d'ADN que celle de notre échantillon.

Cette étape est effectuée grâce à différents kits d'extractions mis en place par différentes entreprises. Le kit utilisé sur la totalité du stage aura été NucleoSpin® Tissue XS de MACHEREY-NAGEL.

L'extraction d'ADN commence par une collecte de tissus d'un individu. Les tissus les plus riches en ADN sont ceux du tube digestif ou des muscles, et sont également les plus pauvres en pigments et autres composés susceptibles d'affecter la pureté de l'échantillon. On utilisera donc les muscles, car ce

Photo 1 : Le Nanodrop



sont les tissus les plus accessibles. Une petite quantité sera donc ôtée de l'individu, ce qui sera suivi par le protocole d'extraction de l'**Annexe 2**.

La quantité d'ADN présente pourra être contrôlée de deux façons :

- ♦ Par électrophorèse : 5 µL de produit d'extraction vont être mélangés à 5 µL de Bleu de charge, puis ces 10 µL seront ensuite déposés dans un puits de gel d'électrophorèse 3% (3g d'agarose pour 100 mL de Solution Tampon 1X) et le résultat pourra être comparé à un marqueur de taille. Cela permettra de vérifier l'intégrité de l'ADN (une seule bande de haut poids moléculaire) ou s'il est morcelé en petits fragments. La révélation des bandes se fait par BET (**Annexe 3**)
- ♦ Par Nanodrop (*Photo 1*) : Le Nanodrop est un spectrophotomètre. A la différence des spectrophotomètres habituels, il permet de n'utiliser que de très petits volumes (1 µL) pour effectuer des mesures. De plus, le logiciel utilisé avec le Nanodrop permet, en mesurant l'absorbance, de connaître la quantité et la qualité de l'ADN extrait (**Annexe 4**). Il calcule le rapport de l'absorbance de l'ADN (260 nm) par rapport à l'absorbance des protéines (280 nm), mais également par rapport à l'absorbance de l'ARN (230 nm).

Enfin, une fois l'extraction d'ADN effectuée, et l'ADN mis en solution dans un tube Eppendorf, on en collectera 25 µL sur les 50 µL obtenus, que l'on placera sur une plaque de 96 puits afin de faciliter les manipulations suivantes.

3. Choix des amorces

Pour pouvoir reproduire un microsatellite, il faut pouvoir le situer sur un génome et ne reproduire que cette séquence de nucléotides. Il est donc nécessaire de mettre au point des amorces capables de se fixer sur les zones flanquantes du microsatellite.

Dans cette thématique de recherche de marqueurs moléculaires, l'unité de Zoologie Forestière fait appel à une société française du nom de GenoScreen. Il s'agit d'une société spécialisée dans le domaine de la microbiologie, de la biologie moléculaire et des bio-informatiques. Pour notre recherche de marqueurs, elle peut réaliser des séquençages haut débit et donc trouver bien plus rapidement et plus facilement des séquences contenant des microsatellites.

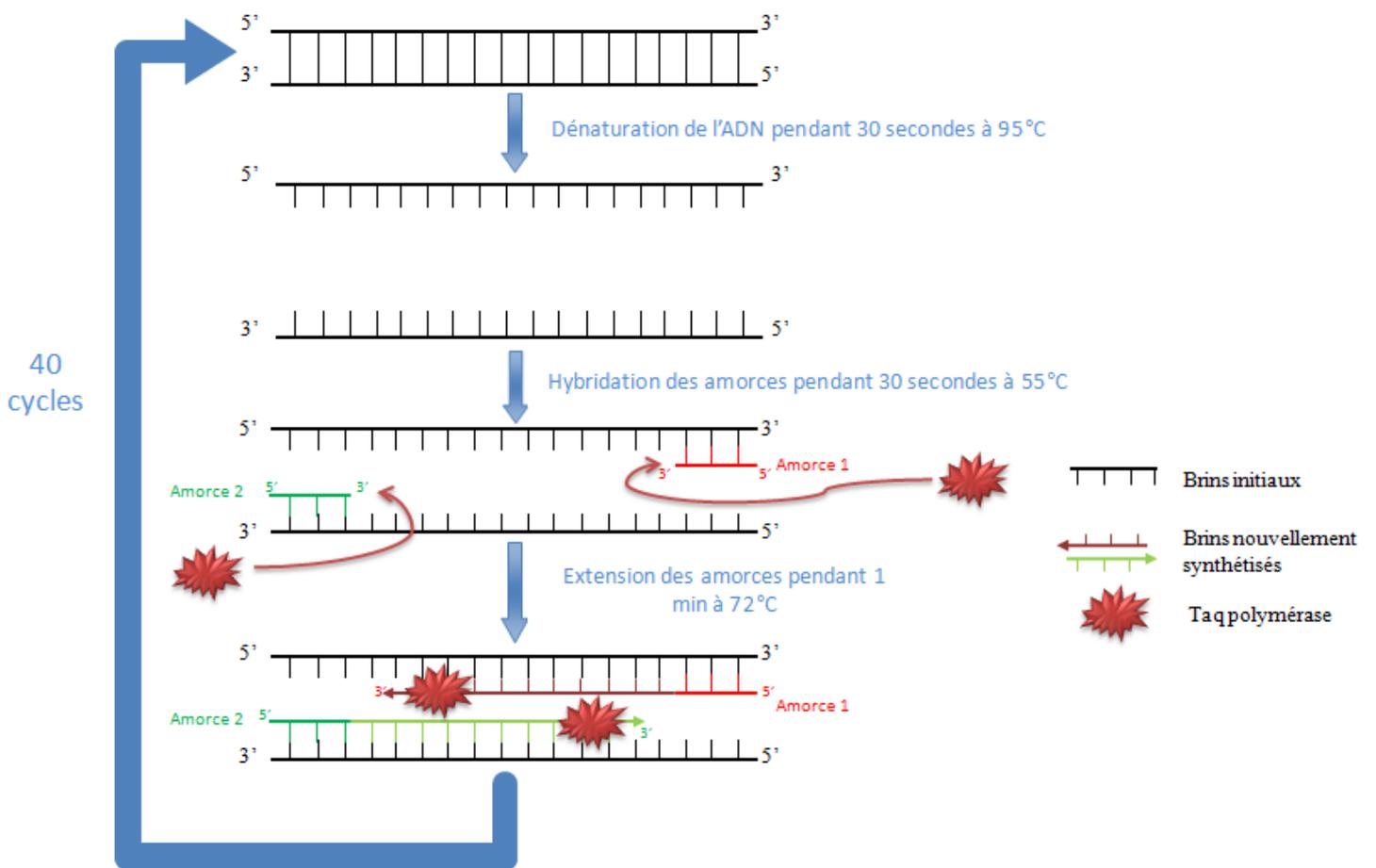
Pour obtenir des résultats de GenoScreen, il faut d'abord envoyer à l'entreprise des individus de l'espèce que l'on souhaite étudier. Il leur a donc été envoyé 7 *Leptoglossus occidentalis* couvrant l'ensemble de l'aire de répartition actuelle, c'est-à-dire l'Amérique du Nord et l'Europe. GenoScreen a par la suite décelé les séquences microsatellites, leurs tailles et leurs nombres de répétitions d'un motif donné. L'entreprise a envoyé un compte-rendu de 24 séquences obtenues et des amorces spécifiques qu'elle a mis en place afin de répliquer les microsatellites. Ces amorces correspondent à une série de nucléotides complémentaires du brin d'ADN matrice. L'URZF a choisi alors parmi ces données celles qui pourront être nécessaires et concluantes, et donc d'obtenir des marqueurs apparaissant comme les plus polymorphes.

Dans le cadre de ce projet, GenoScreen se devait de développer 12 couples d'amorces polymorphes.

Pour chaque locus, un couple d'amorces spécifiques est déterminé. Pour chaque couple, une amorce (l'amorce Forward notée F) est marquée par une molécule fluorescente, ou fluorochrome.

De plus, pour des manipulations simplifiées, les couples d'amorces sont regroupés en multiplexes de 4 couples. Le terme de « PCR multiplexe » désigne une mise au point de la technique PCR autorisant l'amplification, en une seule réaction, de plusieurs segments d'ADN distincts. Ces multiplexes ont donc été définis en fonction des paramètres de PCR des amorces qui doivent être proches. Les produits de PCR sont alors différenciés par leur taille et/ou un marquage fluorescent différent.

Document 2 Principe de la PCR



Par conséquent, pour un individu, plusieurs loci peuvent être déterminés en même temps sur une seule et même migration sur le séquenceur car plusieurs couleurs de marquage par fluorochrome sont possibles. De plus, 2 couples d'amorces peuvent avoir le même marquage car les gammes de tailles des allèles attendues ne se chevauchent pas.

Le premier avantage de cette adaptation technique est la réduction du coût et la diminution du temps de réalisation et, éventuellement, d'analyse des résultats. Le second avantage est la réduction de la quantité d'ADN nécessaire à ces analyses.

Nous disposons ainsi de 12 couples d'amorces répartis en 2 multiplexes (2x 4 couples d'amorces) et de 4 couples d'amorces non multiplexés (**Annexe 5**). Ces derniers correspondent à des marqueurs qui ne répondaient pas à tous les critères définis par GenoScreen (amplification pour tous les individus envoyés) mais tout de même fournis par la société.

En parallèle aux marqueurs développés par la société GenoScreen. L'URZF disposait de 4 marqueurs sélectionnés parmi une vingtaine selon différents critères. Cependant les amorces marquées n'avaient pas été testées, je devais donc tester ces amorces afin de définir si les marqueurs étaient bel et bien polymorphes et donc s'ils pouvaient être utilisables dans de futures analyses.

4. L'amplification de l'ADN par PCR

Cette étape permet d'amplifier sélectivement le fragment cible grâce à des amorces spécifiques. Ici, les fragments cibles sont des répétitions d'une série de nucléotides appelées microsatellites.

Les amorces ont une taille d'une vingtaine de paires de bases en moyenne. La spécificité des amorces est telle qu'il est plus que probable qu'elles ne se fixeront nulle part ailleurs que sur le site pour lequel elles ont été créées.

Par exemple, pour une amorce de 20 nucléotides, on peut trouver 4 possibilités aux 20 sites : A, T, C ou G.

Donc la probabilité de trouver un autre site d'accroche identique est de :

$$\frac{1}{4} \times \frac{1}{4} \times \frac{1}{4}$$

$$\text{Soit } \frac{1}{4} \text{ à la puissance } 20 = 0,00000000000090949470177$$

On a donc _____ chances de trouver un autre lieu d'accroche sur le génome, soit 1 sur 1 099 511 627 780 pour une seule amorce !

L'ADN présent après l'extraction n'est pas utilisé en entier. Une fraction (1 µL) sera utilisée pour réaliser l'étape d'amplification par PCR de façon à conserver le reste pour d'autres tests si nécessaire, et éviter ainsi de repasser par l'étape d'extraction.

Le microlitre d'ADN passera par l'étape de PCR. Cette étape permet de démultiplier l'ADN (**Document 2**).

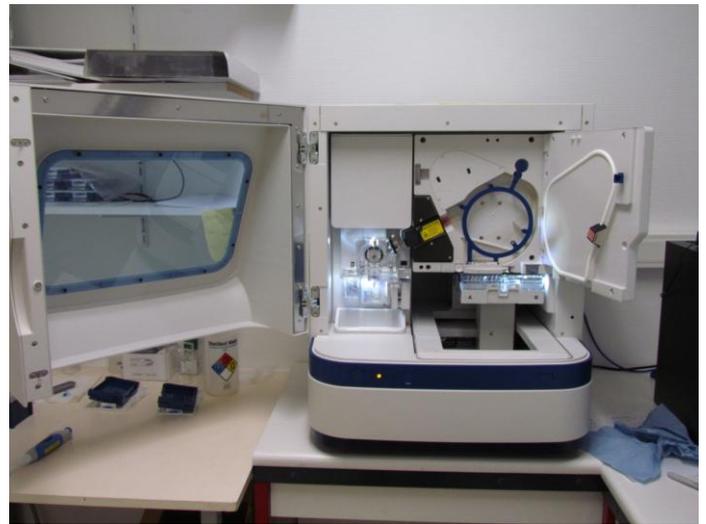
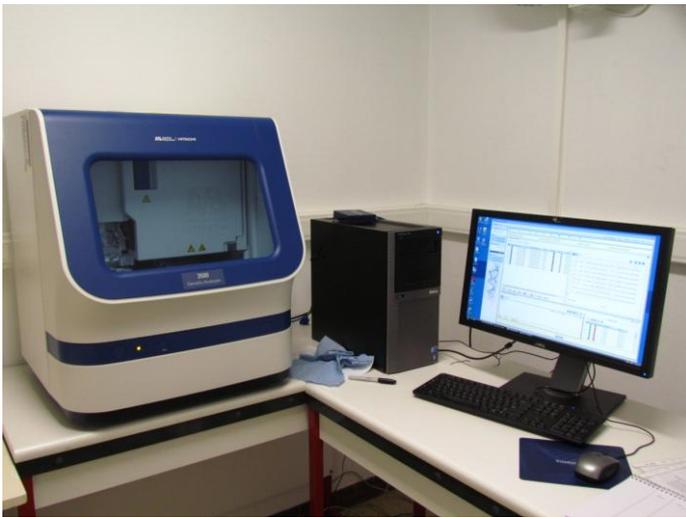
On effectue un mélange réactionnel comme indiqué en **Annexe 6**. Chaque réactif a une fonction bien spécifique :

- ◆ **Tampon Taq** : il maintient un pH constant et optimal pour l'activité de la Taq polymérase. Ce tampon contient du MgCl₂, un cofacteur essentiel de la Taq polymérase. Il a une action sur la spécificité des amorces.
- ◆ **dNTP** : ils permettent un apport de bases azotées (A, T, C, G) à la solution, bases qui seront utilisées à des fins de reproduction de l'ADN.

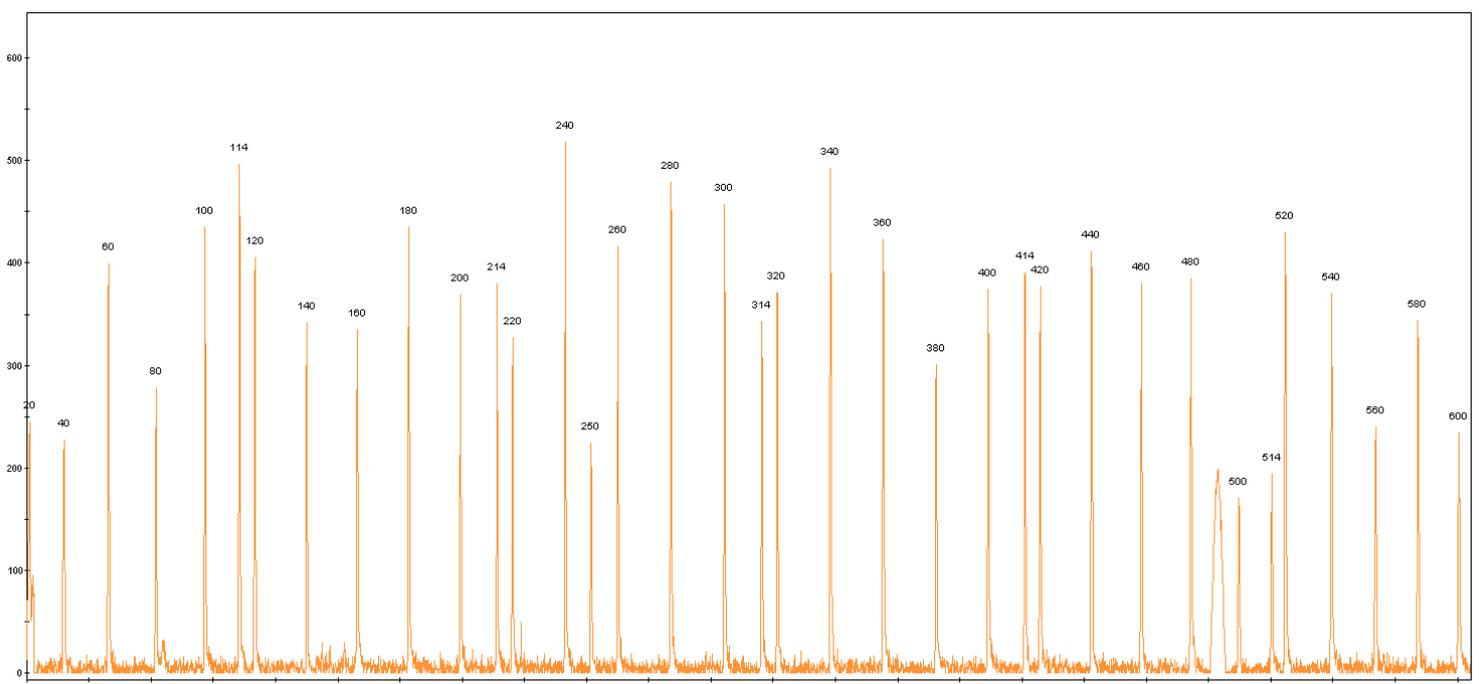
Photo 2 : Les thermocycleurs



Photo 3 et 4: Le séquenceur



Document 3
Signal du marqueur 600 LIZ



- ♦ Amorces : elles permettent de fixer la Taq polymérase sur le début de la séquence d'ADN à amplifier. On utilise des amorces Forward marqué (notée F*) et Reverse (notée R) qui ont des températures d'actions identiques.
- ♦ Bétaïne : elle permet de stabiliser la Taq polymérase et de faciliter l'amplification des structures riches en liaisons G-C, comme l'ADN.
- ♦ Dream Tag : Permet une grande rapidité de synthèse d'ADN à haute température. Ainsi, à 95°C, la Taq polymérase associe environ 60 bases/sec.
- ♦ H₂O ultrapure : utilisée comme solvant.

Au mélange réactionnel sera ajouté à l'ADN à amplifier et la solution passera au thermocycleur comme indiqué dans le protocole (*Photo 2*).

Après des premiers essais réalisés en utilisant les conditions PCR fournies par GenoScreen, les conditions, notamment les concentrations de chaque amorce, ont du être légèrement réajustées pour optimiser l'amplification. L'Annexe 5 inclut ces valeurs modifiées.

5. Etape de préparation au séquençage

L'étape de préparation au séquençage permet de préparer l'amplifiât de PCR tout juste obtenu, pour le passer au séquenceur. La finalité de cette étape est de pouvoir analyser chaque brin d'ADN. Pour cela, on effectue un mélange de formamide et de marqueur 600LIZ de l'entreprise GeneScan, comme indiqué en **Annexe 7**.

Le formamide est un dénaturant, il permet de rompre des liaisons hydrogène entre les bases azotées. Très toxique, il est à manipuler sous hotte avec des gants, afin de ne pas respirer les vapeurs pouvant s'en dégager, et d'éviter le contact avec le produit (**Annexe 3**).

Le marqueur 600 LIZ est une solution de brins d'ADN marqués, qui servira de marqueur référence de taille moléculaire.

Après ajout de la quantité d'ADN nécessaire, il faudra ensuite dénaturer l'ADN. Outre l'utilisation du formamide, l'augmentation de température jusqu'à 95°C pendant 3 minutes permettra de rompre plus facilement les liaisons hydrogènes solides entre les bases C et G (3 liaisons hydrogènes contre 2 entre A et T).

6. Séquençage et analyse

Le séquenceur Applied Biosystems® 3500 Genetic Analyzer est installé dans une salle climatisée et il n'y a que 3 personnes de l'INRA d'Orléans habilitées à l'utiliser (*Photo 3 et 4*).

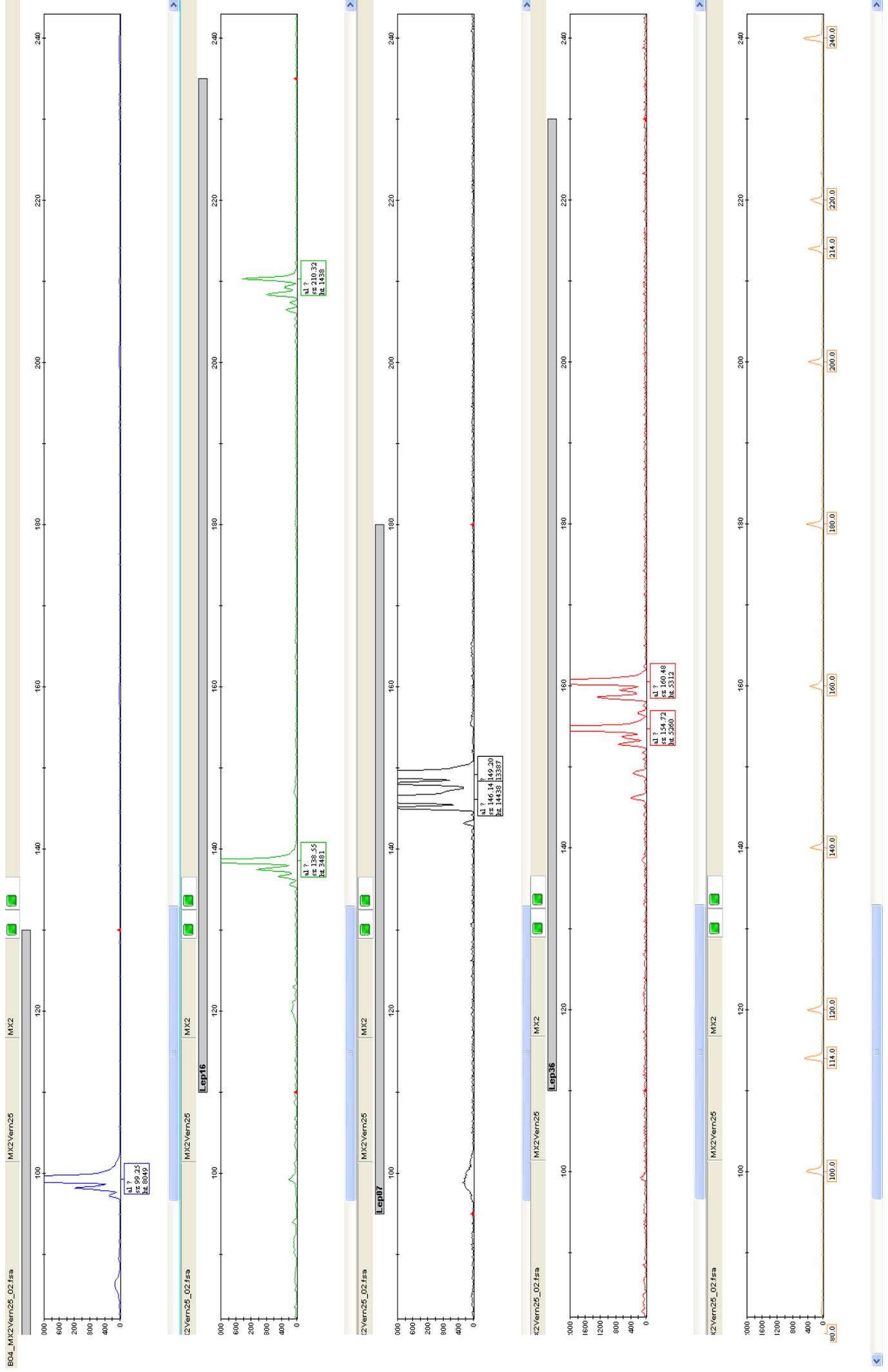
L'étude de séquençage s'effectue sur plusieurs heures, à raison de 8 puits toutes les deux heures environ, sous le principe de l'électrophorèse capillaire.

Le retour des données du séquenceur apparait sous forme d'un ensemble d'électrophorégrammes, avec un électrophorégramme pour chaque échantillon, correspondant au signal émis par chaque fluorochrome. Les deux multiplexes contiennent chacun 4 amorces marquées, il y aura donc 4 signaux, auquel le signal du marqueur 600 LIZ est ajouté (**Document 3**).

Le but de l'analyse est d'observer le signal émis et de le comparer aux données GenoScreen. Grâce aux échantillons envoyés à l'entreprise, la taille moyenne attendue a été déterminée et sert donc de référence.

Pour cela, on étudie les pics des électrophorégrammes. L'amplitude et la résolution sont des facteurs clés pour la détermination des allèles. Un pic peut révéler un allèle ou simplement un artefact (importante amplitude) ou un bruit de fond du séquenceur (faible amplitude). De même, lors de

Document 4 : Electrophorégrammes de l'individu Vern25 de Vernon, CANADA



l'analyse de la résolution d'un pic, un allèle est caractérisé par un pic d'une bonne amplitude et de faible largeur. Une largeur importante ne signifiera qu'un défaut du séquenceur ou de l'échantillon analysé.

Enfin, l'analyse aura deux buts. La première finalité, comme dit précédemment, sera de définir les allèles pour chaque locus étudié de chaque individu. L'étude permet de déterminer pour un microsatellite donné, si l'individu étudié est homozygote ou hétérozygote.

Le second but sera de valider ou non des marqueurs moléculaires dans l'optique de développer des marqueurs propres à l'INRA, d'en gérer les volumes utilisables pour une bonne lecture de séquence.

IV. Résultats expérimentaux

A. Analyse des microsatellites

L'analyse se fait marqueur par marqueur, indépendamment les uns des autres, via le logiciel GeneMapper.

1. Marqueurs développés par GenoScreen

Pour illustrer l'obtention de données, deux exemples sont présentés, l'un provenant de la zone d'origine, et l'autre provenant de la zone d'introduction de *Leptoglossus occidentalis*.

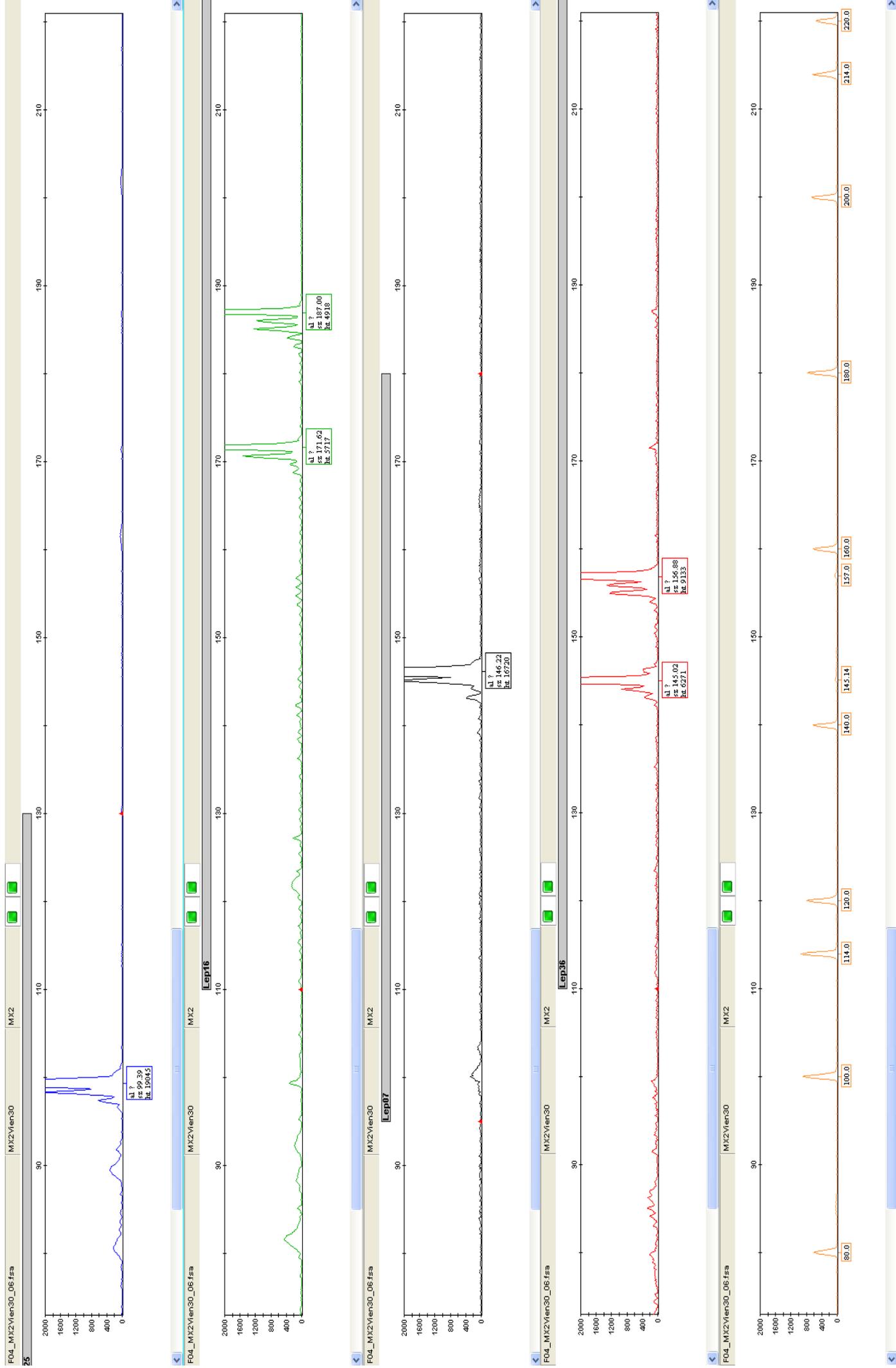
- Individu de Vernon, BC, CANADA : L'individu a été collecté dans la zone native de *Leptoglossus occidentalis* (**Document 4**).
 - Lep 25 : il s'agit de l'électrophorégramme bleu. Le motif microsatellite est (AG)₈
1 pic est observé.
La taille attendue définie par GenoScreen est de 99 pb.
La taille obtenue après marquage est de 99 pb.
L'amplitude ainsi que la résolution sont correctes.

→ L'individu est homozygote ; il possède donc 2 exemplaires de l'allèle.
 - Lep 16 : il s'agit de l'électrophorégramme vert. Le motif microsatellite est (AC)₉
2 pics sont observés.
La taille attendue définie par GenoScreen est de 176 pb.
Les tailles obtenues après marquage sont de 138 et 210 pb.
L'amplitude ainsi que la résolution sont correctes.

→ L'individu est hétérozygote ; il possède 2 allèles très différents l'un de l'autre.
 - Lep 07 : il s'agit de l'électrophorégramme bleu. Le motif microsatellite est (TCA)₇
2 pics sont observés.
La taille attendue définie par GenoScreen est de 146 pb.
Les tailles obtenues après marquage sont de 146 et 149 pb.
L'amplitude est correcte, la résolution est moyenne mais encore convenable.

→ L'individu est hétérozygote ; il possède 2 allèles différents l'un de l'autre.

Document 5 : Electrophorogrammes de l'individu Vien30 de Vienne, AUTRICHE



- Lep 36 : il s'agit de l'électrophorégramme bleu. Le motif microsatellite est (AC)₁₃
2 pics sont observés.
La taille attendue définie par GenoScreen est de 157 pb.
Les tailles obtenues après marquage sont de 155 et 161 pb.
L'amplitude ainsi que la résolution sont correctes.

→ L'individu est hétérozygote ; il possède 2 allèles différents l'un de l'autre.

- Individu de Vienne, AUTRICHE : L'individu a été collecté dans la zone d'invasion européenne de *Leptoglossus occidentalis* (**Document 5**).
 - Lep 25 : il s'agit de l'électrophorégramme bleu. Le motif microsatellite est (AG)₈
1 pic est observé.
La taille attendue définie par GenoScreen est de 99 pb.
La taille obtenue après marquage est de 99 pb.
L'amplitude ainsi que la résolution sont correctes.

→ L'individu est homozygote ; il possède donc 2 exemplaires de l'allèle.

 - Lep 16 : il s'agit de l'électrophorégramme vert. Le motif microsatellite est (AC)₉
2 pics sont observés.
La taille attendue définie par GenoScreen est de 176 pb.
Les tailles obtenues après marquage sont de 172 et 187 pb.
L'amplitude ainsi que la résolution sont correctes.

→ L'individu est hétérozygote ; il possède 2 allèles différents l'un de l'autre.

 - Lep 07 : il s'agit de l'électrophorégramme bleu. Le motif microsatellite est (TCA)₇
1 pic est observé.
La taille attendue définie par GenoScreen est de 146 pb.
La taille obtenue après marquage est de 146 pb.
L'amplitude est correcte, la résolution est moyenne mais encore convenable.

→ L'individu est homozygote ; il possède donc 2 exemplaires de l'allèle.

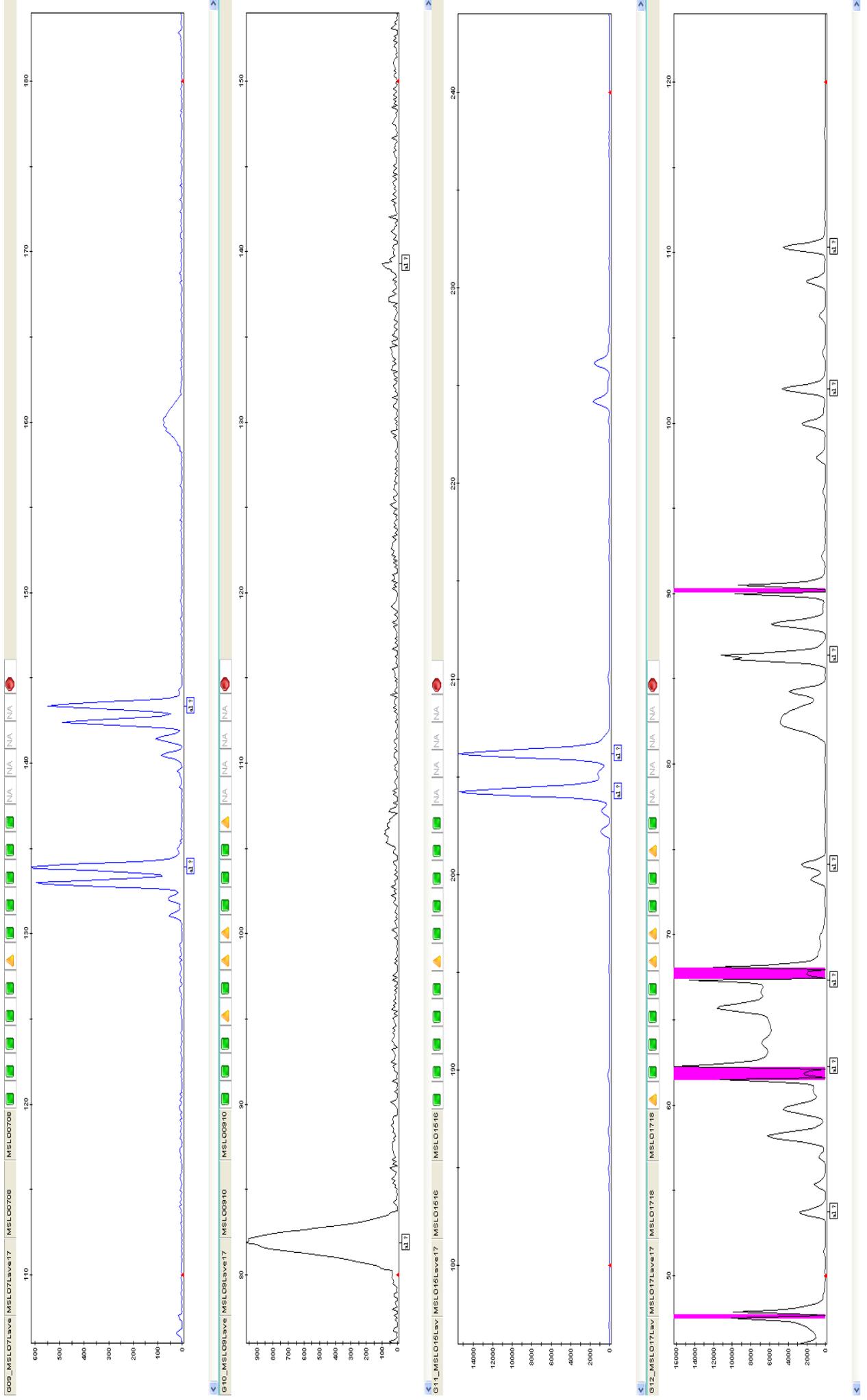
 - Lep 36 : il s'agit de l'électrophorégramme bleu. Le motif microsatellite est (AC)₁₃
2 pics sont observés.
La taille attendue définie par GenoScreen est de 157 pb.
Les tailles obtenues après marquage sont de 145 et 157 pb.
L'amplitude ainsi que la résolution sont correctes.

→ L'individu est hétérozygote ; il possède 2 allèles différents l'un de l'autre.

Les données ainsi produites sont ensuite regroupées dans un tableau qui comportera pour chaque individu, les données de chaque locus étudié et servira de base pour le chercheur.

Une fois ce travail effectué et le tableau complété pour chaque individu et chaque marqueur microsatellite, le jeu de données est transmis par le technicien de laboratoire au chercheur chargé de l'analyser. Celui-ci utilisera différents logiciels de génétique de populations pour évaluer la variabilité et

Document 6 : Electrophorogrammes de l'individu Lave17 testé avec les marqueurs en développement



la structuration génétique des populations introduites en comparaison de celles observées dans les zones d'origine.

L'**Annexe 8** représente le récapitulatif des données obtenues après passage au séquenceur. Ce tableau comporte des cases blanches fonctionnant en doublet. Une des deux cases contient obligatoirement une valeur définie par GeneMapper après analyse, qui correspond à la taille de l'allèle marqué. La seconde case peut contenir ou non une valeur. S'il y en a une, cela signifie que l'individu est hétérozygote et la valeur obtenue est celle de la taille du second allèle. Il y a, en plus de ces cases, des cases colorées en jaune. Ces cases correspondent aux analyses n'ayant pas abouti. Plusieurs possibilités peuvent avoir pour résultat ce défaut de données. Par exemple, si l'on n'obtient aucune donnée pour tous les marqueurs testés, on peut avoir le problème d'une mauvaise qualité de l'ADN (par une possible mauvaise conservation de l'individu), d'une extraction défectueuse ou d'une concentration en ADN très importante ce qui peut gêner l'ADN polymérase lors de la PCR. S'il ne manque qu'une donnée, cela peut provenir d'un défaut de PCR. Il faudra alors procéder à une seconde PCR. Enfin, s'il s'agit de la seconde PCR et qu'il n'y a toujours pas de résultat pour le même marquage, il y a donc un obstacle au niveau de l'amplification, comme une mutation sur les zones flanquantes utilisées pour fixer les amorces. Dans ce cas, l'amorce ne pouvant se fixer, aucune amplification n'aboutira.

2. Marqueurs développés à l'INRA

Les marqueurs dont l'INRA disposait ont été testés afin d'obtenir des marqueurs microsatellites supplémentaires. Pour avoir la certitude d'obtenir des résultats, un individu ayant obtenu des résultats concluants avec tous les autres marqueurs développés par GenoScreen sera utilisé. L'individu **Lave 17** sera donc le témoin positif de ce test (**Document 6**).

Quatre marqueurs seront utilisés : MSLO 07/08, MSLO 09/10, MSLO 15/16, MSLO 17/18.

- MSLO 07/08 : il s'agit du premier électrophorégramme bleu.
2 pics sont observés.
L'amplitude est correcte et la résolution est assez bonne.

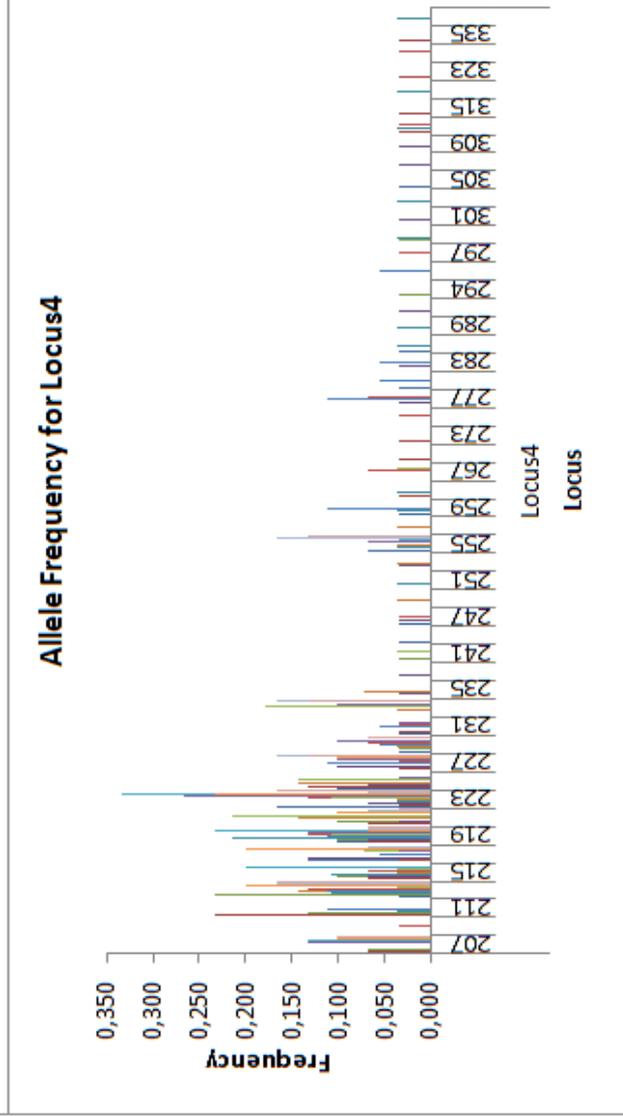
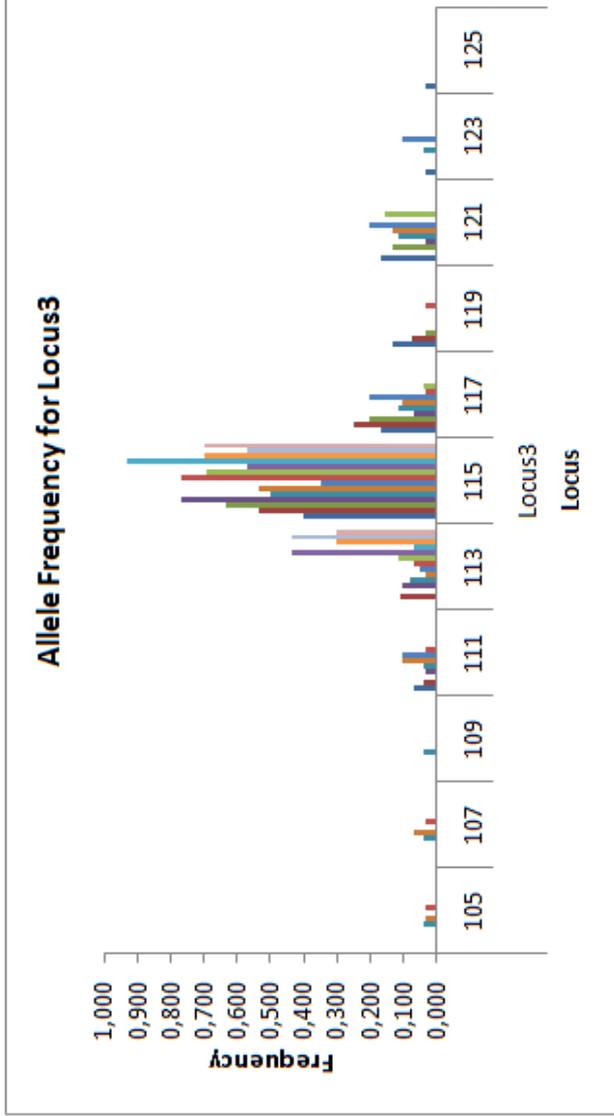
→ L'individu est hétérozygote ; il possède donc 2 allèles différents l'un de l'autre. Ce marqueur semble exploitable.
- MSLO 09/10 : il s'agit du premier électrophorégramme noir.
Aucun pic est observé, mis à part un artefact.

→ Le marqueur n'est pas exploitable.
- MSLO 15/16 : il s'agit du second électrophorégramme bleu.
2 pics sont observés.
L'amplitude est très élevée, la résolution est moyenne mais encore convenable.

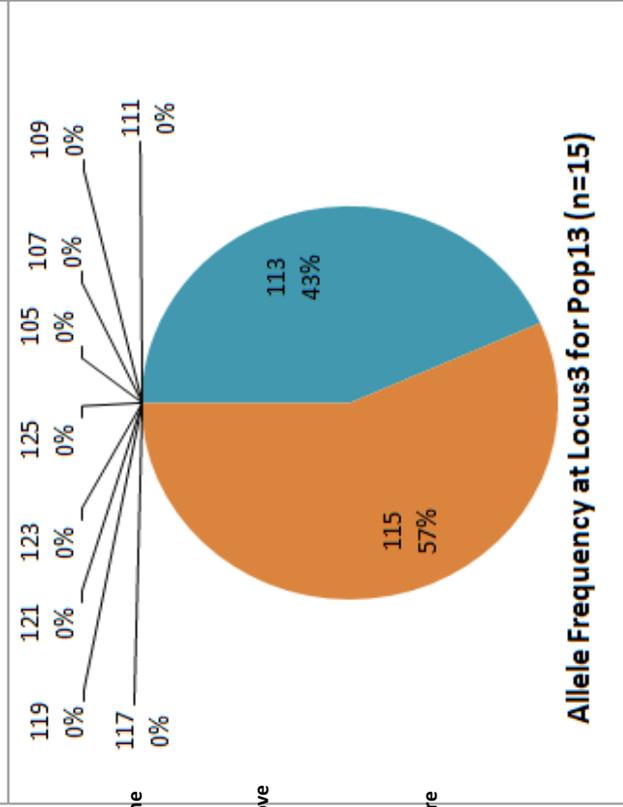
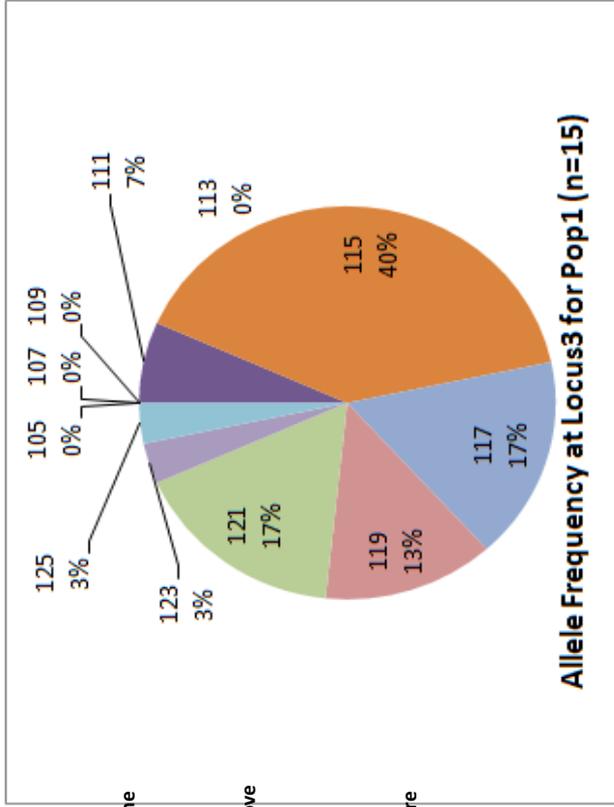
→ L'individu est hétérozygote ; il possède donc 2 allèles différents l'un de l'autre. Ce marqueur semble exploitable. Il sera cependant diminué pour obtenir des résultats plus fins.
- MSLO 17/18 : il s'agit du second électrophorégramme noir.
Beaucoup de pics sont observés, l'amplitude est très importante mais la résolution est peu fiable.

→ Il faut donc diminuer la concentration pour obtenir de meilleurs résultats.

Document 7 : Différence de polymorphisme du marqueur Lep 17 et Lep 43 pour les 14 populations



Document 8 : Fréquence d'allèle et taille des loci entre la population de Sechelt – Canada (a) et la population de Vienne –Autriche (b)



Comme énoncé précédemment, Lave 17 est un témoin positif. Le test ayant donné de bons résultats, un test peut être effectué avec une plaque entière de 96 puits, donc sur 96 individus, afin de tester ces marqueurs pour des populations entières. Cependant, au vu des résultats obtenus pour MSLO 09/10, il sera exclu du test. De plus, les trois autres marqueurs MSLO sont par la suite multiplexés.

B. Analyse des données

Les premières analyses comprennent 14 populations (de 15 individus) et ont été réalisées à l'aide de 8 marqueurs microsatellites correspondant aux marqueurs des multiplexes 1 et 2. Ces analyses ont été réalisées à l'aide du logiciel GenAlEx v.6.5.

Ces premières analyses montrent que le polymorphisme, soit le nombre d'allèles d'un même marqueur, varie en fonction des marqueurs. Lep 43 correspond au marqueur le plus variable avec 52 allèles détectés alors que Lep 04 et Lep 05 sont les moins variables et ne possèdent que 5 allèles (**Document 7**).

Le nombre moyen d'allèles par locus (N_a) montre clairement que les populations européennes sont moins riches que les populations de la zone d'origine. Ainsi les populations européennes présentent toutes un N_a inférieur à 5 ($3,875 < N_a < 4,750$) alors que le N_a des populations de la zone d'origine est toujours supérieur à 6 ($6,375 < N_a < 9,125$). On remarque également que tous les allèles retrouvés dans les zones d'introductions (Amérique de l'Est et Europe) ont été retrouvés dans la zone d'origine. Seules les populations d'Amérique de l'Ouest, donc de la zone native de *L. occidentalis*, possèdent des allèles « exclusifs ». Par exemple, l'allèle 107 du marqueur Lep 25 est présent dans la zone d'origine à des fréquences assez élevées, comprises entre 0,1 et 0,14, ne se retrouve dans aucune population européenne ou d'Amérique de l'Est (**Document 8**).

Cette réduction du nombre d'allèles dans la zone d'introduction est la conséquence d'un effet fondateur : seule une partie de la diversité retrouvée dans la zone d'origine a été introduite aussi bien en Europe qu'en Amérique de l'Est, on parle alors de goulot d'étranglement (appelé *bottleneck*) c'est-à-dire d'appauvrissement génétique de la population.

Les premières analyses de structuration des populations (basées sur des indices nommés F_{st} , qui mesurent le degré de différenciation génétique des sous-populations) montrent globalement une très faible différenciation au sein des populations de la zone d'origine (de la Colombie Britannique à la Californie). Ces résultats sont en accord avec les fortes capacités de dispersion de l'insecte. En effet, les études menées au laboratoire sur les capacités de vol des punaises montrent que ces dernières sont capables de parcourir jusqu'à 18km en une seule journée.

Ces analyses révèlent également peu ou pas de différenciation entre populations européennes à l'exception de la population barcelonaise ce qui pourrait traduire une double introduction, indépendante, en Europe. Cependant le taux important d'allèles nuls (insuffisance ou absence totale d'amplification par PCR résultant d'une mutation ponctuelle ou/et d'une insertion/délétion dans les régions flanquantes complémentaires de l'une des amorces nucléotidiques) pour le locus Lep 07 au sein de cette population pourrait avoir biaisé l'analyse.

V. Conclusions

A. Conclusion scientifique

Le travail réalisé au cours de ce stage a permis d'aboutir à l'estimation de premières données concernant la structuration des populations de *L. occidentalis* et son histoire invasive en Europe et en Amérique de l'Est.

Les premières analyses n'ont pas pu donner d'indications quant à l'origine des populations européennes. D'autres manipulations sont donc envisagées pour compléter le jeu de données. Il serait intéressant notamment d'intégrer de nouvelles populations mais aussi de prendre en compte les autres marqueurs testés au cours de mon stage tels que MSLO-07/08, MSLO-15/16, MSLO-17/18 ou encore les marqueurs du multiplexe 3.

B. Conclusion personnelle

Ce stage aura été une très bonne expérience dans le monde du travail. Il a nécessité une bonne organisation pour pouvoir faire toutes les manipulations demandées. La gestion du temps était nécessaire afin d'être coordonné dans les plannings pour des extractions, des passages au thermocycleur ou encore des passages de plaques au séquenceur, aussi bien avec l'Unité de Zoologie Forestière qu'avec les autres unités du site de l'INRA d'Orléans.

J'ai dû faire preuve d'autonomie afin de pouvoir effectuer des manipulations sans avoir à requérir de l'aide tous les jours, mais il a fallu également savoir s'organiser avec l'Unité pour utiliser le matériel à disposition. J'ai dû également être scrupuleux, de façon à exécuter les manipulations quotidiennes sans erreur.

De plus, ce stage m'a permis de garder un lien avec la Biologie Générale que j'avais découvert lors de mes études théoriques en Faculté de Biologie, et que j'ai préféré abandonner pour les travaux pratiques du BTS Biotechnologies, ce dernier étant centré sur la biologie moléculaire, la microbiologie, la biochimie ou la biologie cellulaire.

Grâce à ce stage, j'ai pu voir tous les aspects de la Biologie et je pense que c'est ce qui m'a le plus plu. De plus, la cohésion de l'équipe dans laquelle j'ai été admis a rendu ce stage agréable tout au long des huit semaines.

BIBLIOGRAPHIE

Ouvrages de référence :

- M. Harry,
Génétique moléculaire et évolutive, 2^{ème} édition, Collection *Sciences fondamentales*,
Edition MALOINE, 27 Rue de l'Ecole de Médecine, 75006 Paris, 2008
- F. Dusoulier, R. Lupoli, H-P. Aberlenc, J-C. Streito,
*L'invasion orientale de *Leptoglossus occidentalis* en France : bilan de son extension biogéographique en 2007*, L'entomologiste, Tome 63 n°6, 2008
- V. Lesieur, M.A. Auger-Rozenberg, A. Roques,
A bridgehead effect in the invasion of the Western conifer seed bug in Europe,
Poster, 2011
- Neil Campbell, Jane Reece,
Biologie, 7^{ème} édition
Pearson Education, Etats-Unis, 2007

Citations d'Introduction:

- J. E. McPherson, R. J. Packauskas, S. J. Taylor, M. F. O'Brien,
*Eastern Range Extension on *L. occidentalis* with a key to *Leptoglossus* species of America North of Mexico*,
The Great Lakes Entomologist, 1990
- M. Fent, P. Kment,
*First record of the invasive western conifer seed bug *Leptoglossus occidentalis* (Heteroptera: Coreidae) in Turkey*
North Western Journal of Zoology, 2011
- S. L. Bates, J. H. Borden, A. R. Kermode, R. G. Bennett,
*Impact of *Leptoglossus occidentalis* (Hemiptera: Coreidae) on Douglas-Fir Seed Production*,
Entomological Society of America, Octobre 2000
- W. B. Strong,
*Seasonal changes in seed reduction in lodgepole pine cones caused by feeding of *L. occidentalis**,
Entomological Society of Canada, 2006
- P. V. Putshkov, A. I. Gubin, G. V. Popov, V. I. Kalesnik, V. V. Syzhko,
*The North American Intruder *Leptoglossus occidentalis* Heidemann (Heteroptera: Coreidae) Settled Down in Ukraine*
2012

Sources :

- Document 1 et 2 : Créations personnelles
- Document 3, 4, 5 et 6 : GeneMapper
- Document 7 et 8 : GenAIEX
- Graphique 1 et 2 : Créations personnelles
- Graphique 3 : Diagramme URZF
- Photos 1, 2, 3 et 4 : Photos personnelles

ANNEXES

<u>ANNEXE 1</u> – POPULATIONS ETUDIEES	I
<u>ANNEXE 2</u> - PROTOCOLE D'EXTRACTION D'ADN A PARTIR DE TISSUS CELLULAIRES.....	II
<u>ANNEXE 3</u> - RISQUES SANITAIRES LIES AUX PRODUITS UTILISES	III
<u>ANNEXE 4</u> - SEQUENCE DE MICROSATELLITES	IV
<u>ANNEXE 5</u> - COURBE D'ABSORBANCE ET TABLEAU RECAPITULATIF.....	V
<u>ANNEXE 6</u> - PROTOCOLE DE PCR A PARTIR D'UNE EXTRACTION D'ADN.....	VI
<u>ANNEXE 7</u> - PROTOCOLE DE PREPARATION AU SEQUENÇAGE	VI
<u>ANNEXE 8</u> - TABLEAU RECAPITULATIF DES DONNEES OBTENUES PAR GENEMAPPER	VII

ANNEXE 1

POPULATIONS ANALYSEES :

Ville	Pays/Etat	Nombre
Missoula	Montana ; USA	25
Big Fork	Montana ; USA	21
Cœur d'Alène	Idaho ; USA	32
Camino	Californie ; USA	16
Vernon	Canada	26
Donetsk city	Ukraine	27
Lavercantière	France	25
Pittston	Maine ; USA	3
Alessandria	Italie	21
Montréal	Canada	9
Fort Collins	Colorado ; USA	24
Sechelt	Canada	20
Vienne	Autriche	24
Barcelone	Espagne	20
Vienne	Autriche	12
Cottage Grove	Oregon ; USA	24
Yvoy-le-marron	France	24
Olympia	Washington ; USA	12
Pittston	Maine ; USA	22
Coruche	Portugal	4
Grenade	Espagne	4
Almeria	Espagne	5
Segovia	Espagne	6
Rostrov	Russie	5
Camino	Californie ; USA	2

Total	413
-------	-----

ANNEXE 2

PROTOCOLE D'EXTRACTION D'ADN A PARTIR DE TISSUS CELLULAIRES

UTILISATION DU KIT NUCLEOSPIN® TISSUE XS :

- ✓ Après avoir ouvert la chenille, extraire quelques tissus musculaires et des tubes de Malpighi, stockés en tube Eppendorf.
- ✓ Ajouter **80 µL** de tampon de lyse T1 du kit.
- ✓ Broyer à l'aide d'un pilon plastique.
- ✓ Ajouter **8 µL** de Protéinase K du kit.
- ✓ Vortexer.
- ✓ Mettre en Bain-marie à 56°C au minimum 1h30. Possibilité de maintien plusieurs heures.
- ✓ Vortexer.
- ✓ Ajouter **80 µL** de tampon B3 du kit.
- ✓ Vortexer.
- ✓ Incuber 10 min à 70°C au Bain-marie.
- ✓ Vortexer.
- ✓ Ajouter **80 µL** d'éthanol 96-100%.
- ✓ Vortexer.
- ✓ Préparer des tubes du Kit (Tube de récupération + Colonne filtrante), annotés suivant l'échantillon.
- ✓ Déposer la solution d'ADN dans la colonne avec une pipette.
- ✓ Centrifuger à 11 000 g pendant 1 min (cf. *Photo 5*)
- ✓ Placer la colonne sur un nouveau tube de récupération.
- ✓ Ajouter dans la colonne filtrante **50 µL** de Tampon B5 du kit.
- ✓ Centrifuger 11 000 g pendant 1 min.
- ✓ Ajouter dans la colonne filtrante **50 µL** de Tampon B5 du kit.
- ✓ Centrifuger à 11 000 g pendant 2 min.
- ✓ Eliminer le tube de récupération et placer la colonne sur un tube Eppendorf 1,5 mL.
- ✓ Ajouter **50 µL** de Tampon BE du kit
- ✓ Incuber 1min sur la paillasse et centrifuger à 11 000 g pendant 1 min.
- ✓ Ajouter **50 µL** de Tampon BE du kit.
- ✓ Incuber 1min sur la paillasse et centrifuger à 11 000 g pendant 1 min.
- ✓ Eliminer la colonne de filtration.
- ✓ Vortexer et centrifuger le tube Eppendorf.

L'ADN peut être conservé au congélateur à -20°C.

ANNEXE 3

RISQUES SANITAIRES LIES AUX PRODUITS UTILISES :

- BET** :  Très toxique
- R22** : Nocif en cas d'ingestion.
- R26** : Très toxique par inhalation.
- R36/37/38** : Irritant pour les yeux, les voies respiratoires et la peau.
- R68** : Possibilité d'effets irréversibles.
- S26** : En cas de contact avec les yeux, laver immédiatement puis consulter un ophtalmologiste.
- S28** : Après contact avec la peau, se laver immédiatement et abondamment avec... (Produits appropriés à indiquer par le fabricant).
- S36/37** : Porter un vêtement de protection et des gants appropriés.
- S45** : En cas d'accident ou de malaise consulter immédiatement un médecin et lui montrer l'emballage ou l'étiquette.

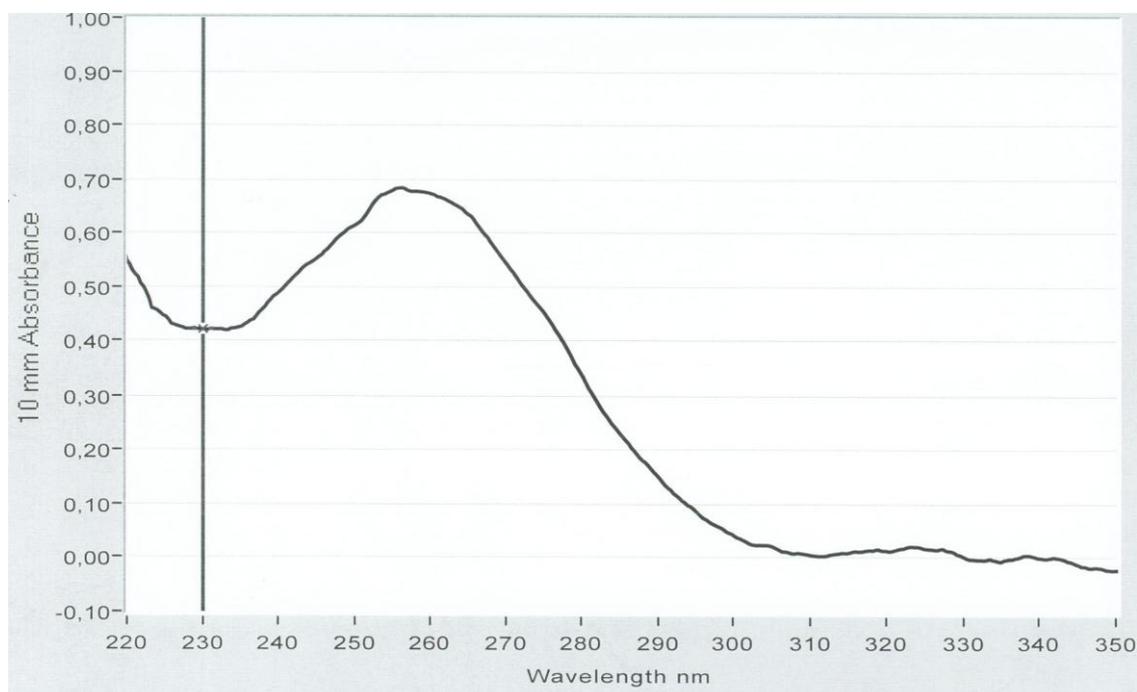
- Formamide** :  Toxique - Au-delà de 90°C, se modifie en acide cyanhydrique.
- R61** : Risque pendant la grossesse d'effets néfastes pour l'enfant.
- S45** : En cas d'accident ou de malaise consulter immédiatement un médecin et lui montrer l'emballage ou l'étiquette.
- S53** : Eviter l'exposition et se procurer des instructions spéciales avant l'utilisation.

Au-delà de 90°C, le formamide se modifie en acide cyanhydrique. Le protocole contenant une étape à 95°C, il est nécessaire de connaître les effets de l'acide cyanhydrique.

- Acide cyanhydrique** :  Très toxique  Dangereux pour l'environnement
- R26/27/28** : Très toxique (mortel) par inhalation, par contact avec la peau et par ingestion.
- R50/53** : Très toxique pour les organismes aquatiques, peut entraîner des effets néfastes à long terme pour l'environnement aquatique.
- S16** : Conserver à l'écart de toute flamme ou source d'étincelles - Ne pas fumer.
- S38** : En cas de ventilation insuffisante, porter un appareil respiratoire approprié.
- S45** : En cas d'accident ou de malaise, consulter immédiatement un médecin (si possible, lui montrer l'étiquette).
- S60** : Éliminer le produit et son récipient comme un déchet dangereux.
- S61** : Éviter le rejet dans l'environnement. Consulter les instructions spéciales/la fiche de données de sécurité.
- (S1/2)** : Conserver sous clef et hors de portée des enfants.
- S7/9** : Conserver le récipient bien fermé et dans un endroit bien ventilé.
- S36/37** : Porter un vêtement de protection et des gants appropriés.

ANNEXE 4

COURBE D'ABSORBANCE AU NANODROP D'UNE EXTRACTION ET TABLEAU RECAPITULATIF DE RESULTATS POUR 11 ECHANTILLONS :



Echantillon	Quantité d'ADN (ng/μL)	Quantité d'ADN rapportée au volume final (ng)*	Absorbance à 260nm	Absorbance à 280nm	A_{260}/A_{280}	A_{260}/A_{230}
Miss 15	12,72	636	0,254	0,143	1,78	0,81
Miss 30	38,47	1923,5	0,769	0,425	1,81	1,01
BiFo 04	90,08	4504	1,802	0,967	1,86	1,53
BiFo 15	80,54	4027	1,611	0,806	2,00	1,51
CoDA 05	38,71	1935,5	0,774	0,376	2,06	1,55
CoDA 23	32,76	1638	0,655	0,403	1,63	0,96
Cami 03	27,48	1374	0,55	0,314	1,75	1,23
Cami 15	33,31	1665,5	0,666	0,476	1,40	0,21
Vern 09	12,80	640	0,256	0,172	1,49	0,75
Vern 18	10,51	525,5	0,21	0,149	1,41	0,90
Vern 27	33,62	1681	0,672	0,333	2,02	1,60

* Volume final d'extraction : 50μL

ANNEXE 5

SEQUENCE DE MICROSATELLITES :

Nom de l'amorce	Séquence Forward	Séquence Reverse
Lep 04	GTGGCTTGCGCTGTGTATAG	TGACTCAGGAATAACAACAACACA
Lep 05	GGGACGAATTTCCCGTAGAT	GCGGGAGGTCTGACTTATGA
Lep 07	TCTTCCTCATCTTCATCAGAATCA	GGTGAAGTTAGCGCAGAGTCA
Lep 16	GGAGATGTTCTCTGCCGT	AGTATGATTTAAAAGGCTGCATAGTA
Lep 17	ACCCAGCTTCGGCTATTTAT	TGCGTAAAAACATACTCCACA
Lep 25	ACGAAAACGTTTGCTGTTTG	AACATTCTTTAATCGTCGGCT
Lep 36	TGTACATAACAGAATGAGACATGCAC	CATGAACACATCCTCTCGGA
Lep 43	CAATTTCAACAACCTCGGG	GTAGGATCCTGCGTGAGAGC
Lep 18	CAAAAGGAGCACTGTGAGTTTTT	AGCCAGTCTGCCATTACCAG
Lep 19	TCGAGCTTTATTGTCTAAAGTTG	CCAATTTTACCCCTACTACTGCC
Lep 24	ACCGAAATGAATAAATAGTTGAAGAATG	AAAAGGAAAGGCATCACTCC
Lep 31	TAAAAATGTTTTCTTTACTGCG	CAAATTTCTGTATGTTTGCTTG

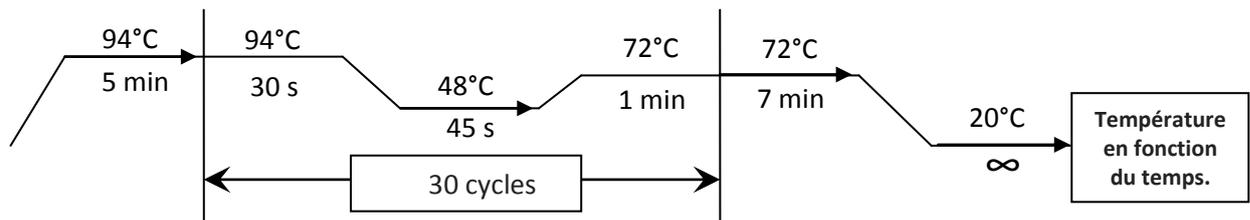
ANNEXE 6

PROTOCOLE DE PCR A PARTIR D'UNE EXTRACTION D'ADN :

25 μ L de Mix PCR par multiplex :

Multiplex 1	F*	R	Multiplex 2	F*	R
	Volume par tube (μ L)			Volume par tube (μ L)	
dNTP (10 mM)	0,6		dNTP (10 mM)	0,6	
Bétaïne	0,5		Bétaïne	0,5	
Tampon Dream Taq 10X	1,875		Tampon Dream Taq 10X	1,875	
Taq pol (5U/ μ L)	0,2		Taq pol (5U/ μ L)	0,2	
Lep 04 (20 pmol/ μ L)	0,25	0,25	Lep 07 (20 pmol/ μ L)	0,5	0,5
Lep 05 (20 pmol/ μ L)	0,1	0,1	Lep 16 (20 pmol/ μ L)	0,25	0,25
Lep 17 (20 pmol/ μ L)	0,25	0,25	Lep 25 (20 pmol/ μ L)	0,25	0,25
Lep 43 (20 pmol/ μ L)	0,1	0,1	Lep 36 (20 pmol/ μ L)	0,25	0,25
H ₂ O ultrapure	19,425		H ₂ O ultrapure	18,325	
ADN (20 à 80 ng)	1		ADN (20 à 80 ng)	1	

- ✓ Vortexer et centrifuger
- ✓ Mettre dans le thermocycleur et faire tourner le programme ainsi :



- ✓ Conserver les tubes au réfrigérateur

ANNEXE 7

PROTOCOLE DE PREPARATION AU SEQUENÇAGE :

- ✓ Mettre dans un tube Eppendorf : 960 μ L de formamide + 28.8 μ L de marqueur 600LIZ
- ✓ Bien mélanger au vortex
- ✓ Répartir dans une barrette environ 110 μ L par puits
- ✓ Répartir sur la plaque environ 9 μ L par puits
- ✓ Ajouter 1-2 μ L de chaque PCR
- ✓ Mettre un tapis blanc
- ✓ Dénaturer 3 min à 96°C puis placer directement sur la glace
- ✓ Remplacer le tapis blanc par un tapis gris percé sous hotte pour éviter les vapeurs toxiques
- ✓ Centrifuger
- ✓ Conserver à 4°C avant le passage au séquenceur

ANNEXE 8

TABLEAU RECAPITULATIF DES DONNEES OBTENUES PAR GENE-MAPPER :

1	A	B	C	D	E - Multiplex 1			H	I	J	K	L	M - Multiplex 2			O	P	Q
					Lep 04	Lep 05	Lep 17						Lep 43	Lep 07	Lep 16			
2	FoCo02	119.29	123.3	128.6	115.57	213.19	223.53	112.18	117.51	179.55	79.38	87.9	120.0					
3	FoCo03	119.16	128.41	128.41	115.57	215.29	221.37	146.26	165.7	167.72	99.25	107.55	162.62					
4	FoCo05	119.16	128.52	128.41	113.39	219.42	229.64	146.27	165.79	177.43	99.38	101.63	148.92					
5	FoCo06	119.17	128.44	128.44	115.57	231.49	298.88	146.27	162.35	194.96	99.38	107.48	156.87					
6	FoCo07	119.17	128.56	128.44	115.57	213.08	229.76	146.23	175.58	210.45	99.38	117.61	158.68					
7	FoCo08	119.18	128.44	128.44	113.51	227.85	277.95	146.22	179.42	179.3	99.38	109.58	164.49					
8	FoCo09	119.18	128.61	128.61	115.66	269.25	298.25	137.18	152.46	179.3	99.38	109.58	164.46					
9	FoCo10	119.17	128.6	130.48	115.57	234.03	336.3	146.27	149.28	198.84	99.38	109.58	150.96					
10	FoCo11	119.17	128.6	128.6	113.51	223.19	227.39	146.27	149.28	179.53	99.38	109.58	156.87					
11	FoCo12	119.17	128.6	128.6	106.93	219.54	227.39	146.27	149.28	150.48	103.54	107.55	187.55					
12	FoCo13	119.17	128.6	128.6	104.75	223.64	227.39	146.27	149.28	162.25	99.38	107.55	187.55					
13	FoCo14	119.17	128.55	128.55	115.57	209.04	249.07	146.27	149.28	175.56	99.38	107.55	187.55					
14	FoCo15	119.17	128.56	128.56	115.57	213.08	219.42	146.27	149.28	181.29	101.4	103.43	156.87					
15	FoCo16	119.17	128.56	128.56	115.45	213.08	219.42	149.34	187.28	187.28	99.38	101.51	148.98					
16	FoCo17	119.17	128.47	128.47	111.43	279.89	279.89	149.35	160.24	177.32	103.38	107.48	158.92					
17	FoCo18	119.29	121.3	128.71	106.92	238.29	254.85	143.49	179.53	177.32	99.38	105.67	156.87					
18	FoCo19	119.29	121.3	128.71	115.69	215.3	314.46	134.24	146.27	179.53	99.38	105.67	156.87					
19	FoCo20	119.17	128.59	128.59	115.57	221.37	227.78	149.28	179.3	179.3	99.38	103.41	166.38					
20	FoCo21	119.17	128.52	128.52	115.57	329.39	331.45	146.14	150.47	181.29	99.26	103.41	180.0					
21	FoCo22	119.16	128.52	128.52	115.57	221.49	227.34	146.35	149.34	181.3	99.38	105.3	168.37					
22	FoCo23	119.17	128.6	128.6	111.31	217.3	225.82	146.35	149.34	162.25	101.51	109.58	156.77					
23	FoCo24	119.07	121.17	128.57	115.54	213.08	215.18	149.34	152.34	191.14	101.38	125.28	154.97					
24	FoCo25	119.28	128.57	128.57	115.57	213.08	217.46	146.22	177.43	187.27	101.63	109.58	148.97					
25	FoCo26	119.16	128.65	128.65	115.57	290.4	337.87	146.35	149.35	162.23	99.38	107.6	170.2					
26	FoCo27	119.17	128.6	128.6	117.73	215.15	217.35	146.27	149.28	189.2	105.42	111.53	150.96					
27	FoCo28	119.17	128.6	128.6	115.57	221.49	221.49	137.18	146.27	189.2	103.41	111.53	160.83					
28	FoCo29	119.17	123.32	128.52	113.38	219.41	237.03	137.16	152.42	173.67	99.5	103.51	166.61					
29	FoCo30	119.17	128.6	128.6	117.73	211.12	261.48	146.39	149.4	148.56	101.39	103.41	151.93					
30	FoCo31	119.18	128.56	128.56	115.66	219.42	227.7	137.15	146.35	165.79	99.38	107.57	145.28					
31	FoCo32	119.17	121.29	130.55	115.57	221.59	284.28	137.2	146.23	165.76	99.38	103.5	154.85					
32	FoCo32	119.17	121.29	130.55	115.57	221.59	284.28	137.2	146.23	165.76	99.38	103.5	154.85					

Lay Audience Summary

I did my training session for a period of two months, in the National Institute of Agronomic Research of Orléans, more specifically in the unity of Forest Zoology directed by Alain ROQUES. I was supervised by Vincent LESIEUR, a Ph.D student.

I worked on an invasive insect species in Europe. An objective of the project is to characterize genetically the origin of the European populations. The goal of my training session was to compare the level of genetic variability in bug populations sampled across the entire native range in North America and in some European populations. I used microsatellite markers which are specific sequences on DNA.

First, I approached the plant and animal biology by studying bugs and their food, to conclude with a dissection.

Then I entered the field of molecular biology by extracting DNA from the cells, and amplifying. The extracting process is assured by a kit sold by a firm. Microsatellite markers were reproduced and amplified by PCR and then identified by a sequencer.

Finally, the work ends with bioinformatics and the comparison between microsatellites of different individuals coming from different populations.

This training session allowed me to discover the real working environment of a laboratory technician, and his relations with others members of a research team. I could see that the instructions being given, the technician had to manage his schedule to achieve all expectations. Each day brings its rigorous manipulations, as equipment maintenance, inventory management.

This training session has been very beneficial to me and comforted me in my choice of curriculum.