



Mémoire du Stage réalisé dans le cadre du module Initiation à la Recherche, M1-Mention Ecologie.

Analyse spatio-temporelle du régime alimentaire de l'Isard (*Rupicapra pyrenaica pyrenaica*) dans le Parc National des Pyrénées par une approche de metabarcoding ADN.

BUIREY Morgan



Isard dans le massif des Pyrénées © MaxPPP

Sous l'encadrement de Erwan QUEMERE
Stage effectué du 18 janvier au 15 avril 2016

INRA UR 035
Comportement et Ecologie de la Faune Sauvage CEFS,
Equipe Ecologie des populations,
Chemin de Borde Rouge, Auzeville, BP 52627, 31326 Castanet-Tolosan Cedex

Remerciements

Je remercie particulièrement Erwan Quéméré, pour m'avoir permis de réaliser ce stage dans les meilleures conditions, pour sa patience, sa gentillesse, sa disponibilité, son aide, ses conseils sur la démarche, sur le raisonnement à avoir devant un sujet de recherche, sur l'esprit scientifique, et sur la rédaction du rapport.

Je remercie aussi le directeur du laboratoire Mark Hewison qui m'a accueilli dans son laboratoire afin d'exercer mon stage.

Merci à Anaïs Denailhac et Laura Gervais pour leur encouragement et leurs motivations.

Je remercie mesdemoiselles Chloé Monestier et Ophélie Couriot de l'aide apportée lors des analyses statistiques, et de leur gentillesse.

Je tiens également à remercier l'ensemble des personnes du laboratoire C.E.F.S, pour leur accueil, leur convivialité et leurs connaissances qu'elles m'ont apportées.

Table des matières

Résumé - Abstract	2
Introduction	3
Matériels / Méthodes	6
Site d'étude.....	6
Stratégie d'échantillonnages	6
Procédure de laboratoire	6
Analyse des séquences et identification taxonomique.....	7
Etape 1 : Démultiplexage et indentification des séquences uniques.....	7
Etape 2 : Filtrage des données.....	7
Etape 3 : Identification taxonomique.....	8
Evaluation de l'effort d'échantillonnage.....	9
Utilisation Du Système d'Information Géographiques ArcGIS	9
Analyses statistiques des données.....	10
Résultats	12
Données spatiales	12
Détermination taxonomique	12
Régime alimentaire	13
Discussion	16
Régime alimentaire très diversifié	16
Les pouvoirs et les limites de l'approche metabarcoding ADN	17
Perspectives	18
Bibliographie.....	19
Annexes	21
Annexe 1 : Script dans terminal pour filtrer les données et identifier les taxons via Obitools.....	21
Annexe 2 : Toutes les séquences uniques (MOTUs) à la fin du traitement classées par ordre décroissant de Fo.....	23
Annexe 3 : Modèles linéaires généralisés pour l'analyse spatio-temporelle.....	25

Résumé - Abstract

Le changement climatique, la modification de l'usage des sols liés aux activités humaines (ex : pastoralisme, tourisme, sport d'hiver, infrastructures routières) et les changements saisonniers conditionnent la disponibilité des ressources alimentaires chez une espèce d'ongulé des Pyrénées : L'isard. Ainsi, Cette étude préliminaire permet de faire un point zéro du régime alimentaire de cette espèce via une approche de type metabarcoding ADN et des techniques de séquençage nouvelles générations pour évaluer la capacité de cette espèce à persister dans un environnement dominé par l'Homme, notamment lorsque celui ci réintroduit le bouquetin ibérique en 2014. L'analyse de l'ADN dans des pelottes fécales prélevées dans le domaine de Clot-Cayan à Cauterets relève que l'isard est un animal opportuniste ayant une stratégie alimentaire intermédiaire (entre sélectif et généraliste). Son régime alimentaire est très varié, et il peut s'adapter suivant les saisons. Nous constatons que son habitat varie au long de l'année (différences d'altitudes, types d'habitats, couvert végétale) car il est déterminé par les différentes conditions climatiques des saisons. Cette nouvelle approche est peu couteuse niveau temps et possède un fort pouvoir résolutif dans l'identification taxonomique. Cette étude préliminaire peut ouvrir les portes à d'autres études sur différentes espèces dans des projets de conservation de l'habitat et des écosystèmes.

Mots clés : Régime alimentaire, isard, metabarcoding ADN, conservation.

Climate change, modifications in land due to human activities (ex : pastoralism, tourism, winter sport, road infrastructure) and seasonal changes determine the availability of food resources in an ungulate species in the Pyrenees: The isard. Thus, this preliminary study allows for a zero point of the diet of this species via an approach like metabarcoding DNA and new generations sequencing techniques to evaluate the ability of this species to persist in an environment dominated by humans, including when he reintroduced the Iberian ibex in 2014. The analysis of the DNA in fecal pellets taken in the area of Clot-Cayan at Cauterets notes that the isard is an opportunistic animal with an intermediate food strategy (between selective and general). Its diet is diversified, and it can fit the seasons. We find that habitat varies throughout the year (altitude differences, habitat types, vegetation cover) because it is determined by the different climatic seasons. This new approach is an inexpensive time level and has a high resolving power in the taxonomic identification. This preliminary study may open doors to other studies on different species in conservation projects of habitat and ecosystems.

Key words : Diet, isard, metabarcoding DNA, conservation.

Introduction

L'alternance des saisons, le changement climatique (Post et al., 1999), et la modification de l'usage des sols liés aux activités humaines (ex : pastoralisme, tourisme, sport d'hiver, infrastructures routières) conditionnent la disponibilité des ressources alimentaires et donc la survie des espèces d'ongulés de hautes montagnes (Loison et al., 2003). Pour évaluer la capacité de ces espèces à persister dans des environnements dominés par l'Homme, il est important de connaître leur niveau de flexibilité notamment en termes de régime alimentaire. Or les techniques utilisées traditionnellement pour caractériser les espèces animales ou végétales consommées sont difficiles à mettre en œuvre à de larges échelles spatiales ou temporelles. L'observation directe de l'individu en train de se nourrir est une approche très coûteuse en temps et son efficacité est très variable en fonction du site et de l'espèce étudiée (habitat ouvert et accessibilité des individus) (Vaucher, 1988). L'identification des fragments de feuilles restant dans la panse des animaux morts (Couturier, 1938 ; Auvray, 1983 ; Babad, 1997) ne peut être faite que durant les périodes de chasse (Bertolino et al., 2009). Enfin l'identification des cuticules de plantes dans les pelotes fécales par l'observation microscopique (microhistologie) nécessite une formation considérable et une bonne expertise de l'observateur et des bases de références disponibles pour l'identification des produits alimentaires (Berducou, 1974 ; Garcia Gonzalez and Cuartas, 1995 ; Bertolino et al., 2009). Cette approche est également fortement influencée par la variation de la densité et la survie des cuticules végétales dans le tube digestif (Holechek et al., 1982). Il existe donc un risque de fournir des estimations biaisées des préférences pour certains fourrages (par exemple, les herbacées ont tendance à être sous-estimée (Anthony et Smith, 1974). De plus, son pouvoir résolutif au niveau taxonomique est limité avec souvent une détermination au niveau famille (La Morgia et Bassano, 2009). Il existe donc un besoin d'une nouvelle approche non-invasive permettant l'étude du régime alimentaire à un coût relativement faible et avec une résolution taxonomique élevée. L'émergence récente des technologies de séquençage de nouvelle génération (NGS) combinée au développement des méthodes d'identification taxonomique par code-barres ADN (Hebert et al., 2003, Hollingsworth et al., 2009) permet l'analyse du régime alimentaire d'un individu en identifiant l'ADN de plantes ou des animaux présent dans ses fèces et cela sans le besoin de mettre en place des étapes de clonage (Valentini et al., 2009, Pompanon et al., 2012, De Barba et al.,

2014). Grâce à ce nouveau type d'approche (dit de « métabarcoding alimentaire »), il est désormais possible d'obtenir le profil alimentaire de plusieurs centaines d'individus simultanément et donc d'évaluer les variations spatiales et temporelles de comportement alimentaire à l'échelle d'une population dans son ensemble (Quéméré et al. 2013). Ce type d'approche offre un pouvoir résolutif au niveau taxonomique grandement supérieur à l'approche micro-histologique (détermination possible jusqu'au rang générique et rang de l'espèce) mais fournit pas de données sur la quantité et le type de tissu.

Dans le cadre de ce stage, j'ai testé l'utilisation de cette approche pour évaluer la diversité et la structuration temporelle et spatiale du régime alimentaire d'une population d'isard du Parc National des Pyrénées (domaine de Clot-Cayan à Cauterets, département Hautes - Pyrénées) sur un cycle annuel.

L'isard (*Rupicapra pyrenaica pyrenaica*) est une espèce d'ongulé de la sous-famille des caprins emblématique des Pyrénées. Il se différencie du chamois des Alpes (*Rupicapra rupicapra*) par sa plus petite taille (1m-1m10) ainsi que sa couleur beige rougeâtre en été (Weber, 2001). La protection de la population d'isards fut l'un des moteurs de la création du parc national en 1967. Menacé d'extinction dans les années 50 pour cause de chasses abusives, la population du Parc National des Pyrénées compte actuellement plus de 5 000 individus (Site Parc national des Pyrénées). La recherche de nourriture est conditionnée par la disponibilité de celle-ci. Cette nourriture varie en fonction de la flore, la saison, la densité de la population, le type d'habitat, l'individu et son état physiologique (Cazajous, 2002). L'isard va alors migrer de façon saisonnière dans une zone de basse altitude, principalement recouverte de forêt appelée zone de refuge l'hiver (Couturier, 1938 ; Weber, 2001). Ces auteurs indiquent qu'il se nourrit principalement de plantes herbacées, de jeunes pousses et l'hiver de bourgeons, d'aiguilles de conifères, de feuilles sèches et de lichens.

Un second enjeu de cette étude est lié à la réintroduction dans les Pyrénées en 2014, du bouquetin ibérique (*Capra pyrenaica*) une autre espèce d'ongulé avec une niche écologie susceptible de recouvrir en partie celle de l'isard.

L'isard et le bouquetin ibérique occupent le même habitat (relief, altitude élevée en été, mais plus basse l'hiver, substrat rocheux) (Plan de restauration du bouquetin, 2012), et sont tous les deux herbivores avec un type d'alimentation intermédiaire (entre spécialiste et généraliste) (Hofmann, 1989). La concurrence liée à la nourriture est un mécanisme central de l'interaction entre les grands herbivores, qui agit quand il y a un chevauchement élevé d'utilisation de l'alimentation des habitats et quand les ressources partagées sont limitées (Belovski, 1986 ;

Latham, 1999). Autrement dit, pour la bonne gestion de ces deux espèces, il est nécessaire de déterminer quelles sont les espèces et familles de plantes communes dans leur alimentation. Cette étude préliminaire a pour objectif de faire l'état initial du régime alimentaire de l'isard avant une possible compétition avec le bouquetin. Les données de cette étude pourront donc être comparées à celles issues d'une prochaine étude sur le régime alimentaire du Bouquetin sur le même site pour déterminer s'il y a compétition ou non entre ces deux espèces.

Mon premier objectif a été d'évaluer la diversité taxonomique de plantes consommées par l'Isard en analysant, grâce à une suite d'outils bioinformatiques, les séquences code-barres obtenues à partir de l'ADN extrait de matières fécales. J'ai étudié les variations saisonnières la composition des taxons consommés à partir des fréquences d'occurrences dans les fèces au Printemps, en Été et en Hiver.

Puis, d'autre part, de m'intéresser aux variations spatiales du régime alimentaire de l'isard en utilisant un logiciel de Système d'Information Géographique : ArcGIS version 10.1 ESRI Software, Redlands, CA, USA (coordonnées GPS (Global Positioning System) XY des fèces couplées à deux cartes des habitats : une qui représente l'Unité Elémentaires de Paysage du Parc National des Pyrénées et l'autre, plus précise, la typologie française de "Corine Biotopes". En effet, d'après Couturier, en 1938, cette recherche de nourriture "conditionne sa dissémination géographique et l'altitude à laquelle il vit".

Avant de commencer l'étude, plusieurs hypothèses ressortent. Tout d'abord, on s'attend à ce qu'il y est une grande diversité dans l'alimentation de l'Isard car c'est une espèce avec un régime alimentaire plutôt généraliste (Hofmann, 1989). De plus, on devrait observer des différences importantes en termes de familles et d'espèces consommées l'hiver, l'été et le printemps car il n'y a pas la même disponibilité végétale. On s'attend à ce qu'il y est plus de Rosacées, Poacées et plantes à fleurs mangées au printemps et en été contrairement à l'hiver avec des espèces ligneuses sous couvert arborescent en hiver (Couturier, 1938 ; Weber, 2001). On s'attend à ce que ces différences entre saisons soit liée en partie aux différences d'habitat: l'hiver l'isard est plutôt dans un habitat de type arborescent, arbustif de faible altitude tandis que l'été un couvert herbacé, rocheux de haute altitude et donc on devrait mettre en évidence une migration hivernale pour trouver refuge à plus basse altitude pendant les périodes de fort enneigement et de froid toujours dans un but de recherche de nourriture (Couturier, 1938 ; Weber, 2001; Cazajous, 2002).

Matériels / Méthodes

Site d'étude

La zone d'échantillonnage pour la collecte de fèces est située dans la commune de Cauterets (département Hautes-Pyrénées de la région Languedoc Roussillon – Midi – Pyrénées en France, coordonnées : 42° 53' 20" Nord 0° 06' 52" Ouest). Le domaine étudié est Clot- Cayan d'une surface de 125 ha situé entre 1520m (Plateau du Clot) à 2326m d'altitude (Pic de Leytugouse) faisant partie du Parc National des Pyrénées (**Fig. 1**).

Le choix de la zone d'échantillonnage des fèces s'appuie sur la connaissance des domaines vitaux d'un groupe de femelles obtenue par le suivi des individus marqués de 1993 à 2013. Une particularité de la zone choisie est de présenter des faciès d'habitats différents (Voir Unités Elémentaires de Paysage du Parc National des Pyrénées de la **Fig. 1**).

Stratégie d'échantillonnages

Les techniciens de terrain du Parc National des Pyrénées dont Jean-Paul Crampe ont récolté 59 échantillons de fèces fraîches (20 échantillons le 23/05/13, 11 échantillons le 05/08/13, 21 échantillons le 05/12/13 et 7 échantillons le 11/03/14) (**Fig. 1**) de façon opportuniste sur une base visuelle le long d'un itinéraire. Chaque échantillon est décrit par la date, les conditions météorologiques du jour de la récolte et ses coordonnées GPS en coordonnées Lambert III.

Les fèces ont été asséchées avec du silica gel pour limiter la dégradation de l'ADN par les micro-organismes bactériens (Rayé et al., 2010).

Procédure de laboratoire

L'analyse de l'alimentation a été réalisée en utilisant une approche de métabarcoding ADN basée sur l'amplification du gène chloroplastique trnL (Pompanon et al 2012). En amont de mon travail, l'extraction d'ADN a été effectuée en utilisant environ 50 mg de la couche externe de pelotes fécales séchées en utilisant le kit d'extraction (ZR-96 Fecal DNA Kit - Zymo). Ensuite, la boucle de région P6 de l'intron chloroplastique du trnL (UAA) a été amplifiée au laboratoire EDB (UPS/CNRS Toulouse) en utilisant les amorces universelles g (5'- GGGCAATCCTGAGCCAA-3') et h (5'-CCATTGAGTCTCTGCACCTATC-3') (Taberlet, 2007) modifiées par l'addition de Tag spécifiques sur l'extrémité 5' pour permettre l'attribution de la séquence lue à l'échantillon concerné (Valentini et al. 2009a, b). Les produits

de la PCR (Polymerase Chain Reaction = technique de biologie moléculaire qui permet d'augmenter en laboratoire le nombre des copies d'une séquence d'ADN) ont été quantifiés en utilisant le spectrophotomètre NanoDrop ND-1000 UV-Vis (NanoDrop Technologies, Rockland, DE). L'ensemble des produits PCR ont ensuite été séquencés à l'aide d'un séquenceur haut débit Miseq (Illumina) à la plateforme génomique INRA de Toulouse - geT-PlaGe (Quéméré et al., 2013).

Analyse des séquences et identification taxonomique

La première étape de mon travail a été d'analyser les séquences de trnL issues du séquenceur Miseq afin d'assigner chaque séquence à un échantillon, caractériser les séquences uniques, filtrer les séquences artefactuelles et assigner à chaque séquence unique un nom de taxon en utilisant une base de code-barres de référence. J'ai utilisé pour cela une suite d'outils programmés en python (OBIttools, Boyer et al 2014) (<http://metabarcoding.org/obitools>). Je détaille ici les grandes étapes de mon analyse. Le script complet est détaillé en annexe 1. L'analyse a été effectuée sur le serveur informatique de la plateforme génomique GenoToul.

Etape 1 : Démultiplexage et identification des séquences uniques

Tout d'abord nous avons procédé à l'alignement des lectures avec l'outil « illumina-pairedend » puis, à l'aide de « obigrep » nous avons supprimé les lectures non alignées. Ensuite, afin de réduire à la fois la taille du fichier et de calculs de temps, et pour obtenir des résultats interprétables plus facilement, il est commode de travailler avec des séquences uniques au lieu de lectures. Pour déreplier ces lectures en séquence unique, nous utilisons la commande « obiuniq ». Le logiciel va tout d'abord comparer toutes les lectures, les unes aux autres dans un jeu de données, puis grouper les lectures strictement identiques ensemble et enfin sortir la séquence pour chaque groupe et son nombre dans le jeu de données d'origine (de cette manière, la duplication de toutes les lectures est retirée) (Seguritan V, Rohwer F., 2001).

Etape 2 : Filtrage des données

Les séquences obtenues ne sont pas toutes biologiquement utiles, certaines de ces séquences peuvent contenir des erreurs de séquençage et/ou de PCR, ou des chimères. Pour supprimer ces séquences, autant que possible, nous écartons premièrement les séquences rares puis des variantes de séquences qui correspondent probablement à des artefacts. Ainsi, nous avons gardé les séquences ayant un nombre de répétition supérieurs ou égales à 100 et un nombre de paire de base inférieur ou égale à 80pb avec l'outil « obigrep ». Ensuite j'ai utilisée

l'algorithme implémenté dans le programme « obiclean » afin d'identifier les séquences chimériques/artefactuelles issues d'erreurs d'amplification et/ou de séquençage. Ce programme groupe ensemble les séquences distantes au maximum d'une base dans des « clusters » et assigne à chaque occurrence, pour chaque échantillon, un statut : le statut « single » (S) signifie que la séquence en question est la seule du cluster présent dans l'échantillon, le statut « head » (H) correspond aux cas où il existe au moins une autre séquence du cluster dans l'échantillon mais cette autre séquence est minoritaire (en nombre de lectures). Enfin le statut « internal » (I) signifie au contraire qu'il existe une autre séquence majoritaire du même cluster dans l'échantillon. J'ai ensuite éliminé les séquences avec une proportion de « H » ou de « S » dans les échantillons supérieurs à 95%. Plusieurs témoins négatifs et des reliquats de PCRs ont été réalisés afin de nous assurer de la fiabilité des résultats et éliminer les contaminations potentielles.

J'ai choisi d'analyser les données en « présence/absence » plutôt qu'en « quantitatif » (nombre de lectures) car le nombre de lectures ne reflète que peu la quantité d'ADN matrice dans l'échantillon (Pompanon et al 2014). Cela est lié notamment au fait que les amorces utilisées, malgré leur caractère universel, n'ont pas la même efficacité pour toutes les familles de plantes. J'ai considéré une séquence comme présente dans un échantillon que lorsque le nombre de lecture est supérieur à 1% du nombre totale de lectures dans cet échantillon.

Etape 3 : Identification taxonomique

L'outil « *ecotag* » permet d'assigner à chaque séquence unique un nom de taxon en utilisant une base de référence de code-barres ADN. Puisque nous ne disposons pas de base locale de références (i.e. liste des code-barres des plantes présentes sur le site d'étude), nous avons utilisé la base complète des séquences de trnL présent dans GENBANK. Lorsqu'une séquence correspond à deux taxons de même rang avec la même homologie, le programme assigne la séquence au rang supérieur. Le problème est que la base de données « GENBANK » contient des codes-barres du gène trnL de plantes d'origine géographique diverse. J'ai précisé dans un second temps les identifications taxonomiques (familles, genre et espèces) en utilisant un relevé botanique très précis (919 plantes différentes) fourni par le Parc national des Pyrénées. Pour cela, j'ai utilisé l'outil « BLASTn » (=Basic Local Alignment Search Tool nucleotide) (<http://blast.ncbi.nlm.nih.gov/>) afin de lister les taxons dont les séquences sont identiques à 100% avec la séquence étudiée et je n'ai conservé que les noms de taxons présents localement. L'idéal aurait été de construire une base locale de référence de code-barres à partir de la liste de plantes disponibles fournir par le PNP mais cela n'a pas pu être fait faute de temps.

Lorsqu'une séquence est identifiée seulement au rang « familial » ou « générique », j'ai assigné comme nom, le nom du genre ou de la famille puis la dénomination « sp. » (ex : *Rosaceae sp1* pour la première espèce de Rosaceae non identifiée). J'ai fait l'hypothèse ici que chaque séquence correspond à une espèce unique. Mais il est probable que certaines séquences correspondent à des groupes d'espèces proches phylogénétiquement notamment dans les familles où le marqueur utilisé a évolué lentement (cf. discussion).

Evaluation de l'effort d'échantillonnage

Afin d'évaluer notre effort d'échantillonnage taxonomique, nous avons réalisé une courbe d'accumulation du nombre de familles ou d'espèces en utilisant la fonction *specaccum* du package R « vegan » (version 2.3-5). Cette courbe est réalisée en permutant au hasard la matrice de diversité (100 permutations effectuées) et en calculant pour chaque permutation le nombre accumulé de plantes différentes consommées pour un nombre d'échantillons donnés. Nous avons également calculé à l'aide de la fonction *specpool* du package « vegan » l'indice chao2 qui fournit une estimation de la diversité obtenue si l'effort d'échantillonnage était illimité et donc du nombre de plantes consommées non identifiées.

Utilisation Du Système d'Information Géographiques ArcGIS

Afin de pouvoir définir le régime alimentaire suivant la répartition spatiale et temporelle de l'isard, j'utilise un logiciel de Système d'Information Géographique, ArcGIS 10.1, j'ai à disposition les coordonnées GPS (XY Lambert III) des 59 échantillons, la couche d'habitat de la région de Péguyères, le relevé de végétation de la région Péguyères et les cartes IGN de Caudebec (Disponible sur www.geoportail.gouv.fr). Les plantes du relevé de végétation ont aucunes coordonnées GPS mais elles ont un code qui est le numéro de la station ou parcelle. Ainsi j'ai ajouté les coordonnées des centroïdes de chaque parcelle de la couche Habitat Péguyères, puis j'ai joint les deux couches (relevé de végétation + Habitat Péguyères) par le code, et j'obtiens donc des coordonnées GPS pour mes plantes observées. Ensuite, j'ai tracé (via l'outil "zone tampon") trois aires de répartition des isards qui correspondent à une zone de 200m de rayon autour de chaque échantillon en fonction de la saison. Puis j'ai "intersecté" ces zones de répartition avec la couche de relevé de végétation. J'obtiens ainsi les parcelles pour chaque aire de répartition avec les formules de la typologie "Corine Biotopes" et la végétation associée. Attention un triangle jaune correspond à pleins d'espèces de plantes car elles ont le même nom de station et donc les mêmes centroïdes (**Fig. 1**). Afin de déterminer la surface en pourcentage des types d'habitats, un tableau a été créé avec toutes les parcelles et les formules pour une

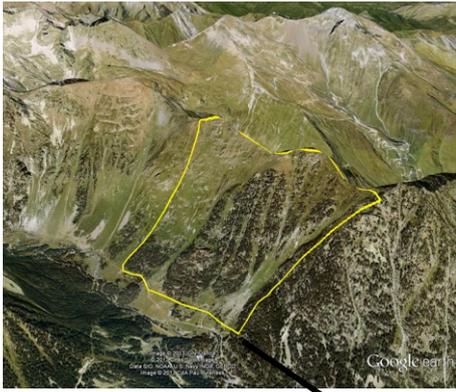
saison. J'ai fait l'hypothèse que pour chaque parcelle, les types d'habitats découpaient la parcelle en parties égales.

En géoréférençant les cartes IGN de Cauterets sur la couche Habitat Péguyères, j'ai pu déterminer les altitudes de chaque échantillon et ainsi analyser le déplacement des individus suivant un gradient altitudinal (**Fig. 2**).

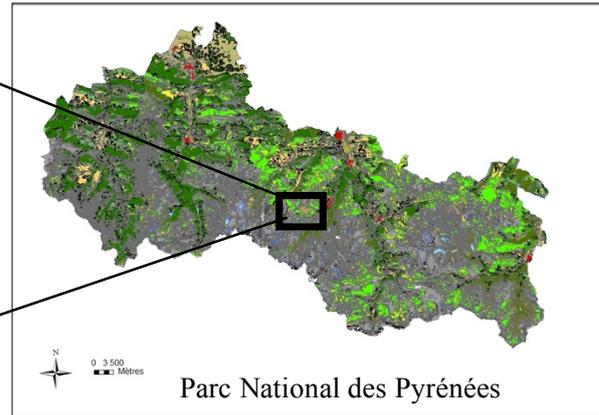
Analyses statistiques des données

Le régime alimentaire en fonction des saisons est décrit à l'aide de quelques statistiques. Nous créons à l'aide d'une boucle sous R studio et à partir de deux tableaux (un représentant les présences/absences des familles suivant les saisons et un représentant les présences/absences des espèces en fonction des saisons), un nouveau tableau afin de pouvoir leur appliquer un modèle linéaire généralisé (*Glm*) (Bertolino et al., 2008) avec la variable dépendante qui est la présence/absence d'une famille ou d'une espèce et les variables explicatives (espèces (ou familles) et saison en interaction. (Script de quelques lignes Annexe 3). Nous avons comparé l'AIC (Critère d'Information d'Akaike = critère de comparaison de modèle) du modèle avec le modèle nul. L'AIC le plus faible déterminera le meilleur modèle. Ensuite une ANOVA (test=chisq) est réalisé afin d'obtenir la validité du modèle.

Afin de déterminer la richesse spécifique par échantillons suivant les saisons ainsi que l'altitude, un test de rang non paramétrique sur k échantillons indépendants (*Kruskal.test*) a été réalisé. Puis pour, savoir entre quelles saisons il y a une différence significative, je fais un *kruskalmc*. Enfin pour avoir la p.valeur des comparaisons deux à deux de chaque saison je fais le test non paramétrique (avec un nombre d'échantillon inférieur à 30) de *Wilcoxon.test* (*paired=FALSE*).



Domaine de Clot-Cayan
(Cauterets, 65) (vue 3D)



Parc National des Pyrénées

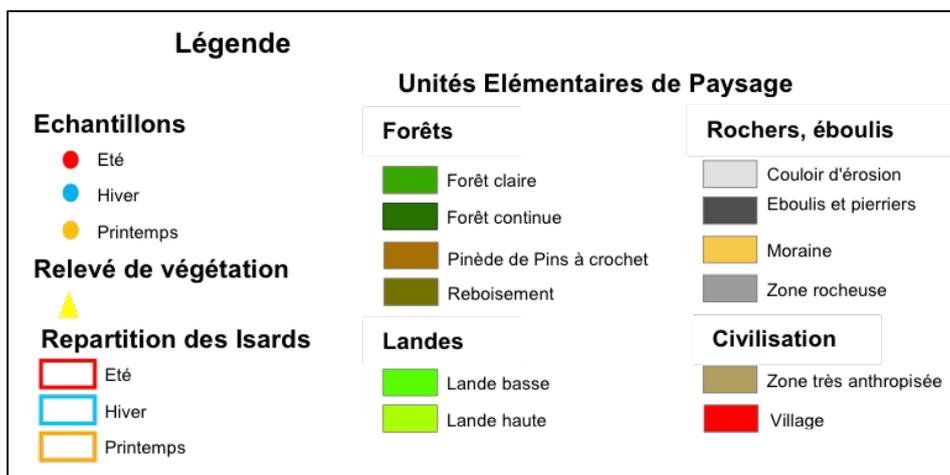
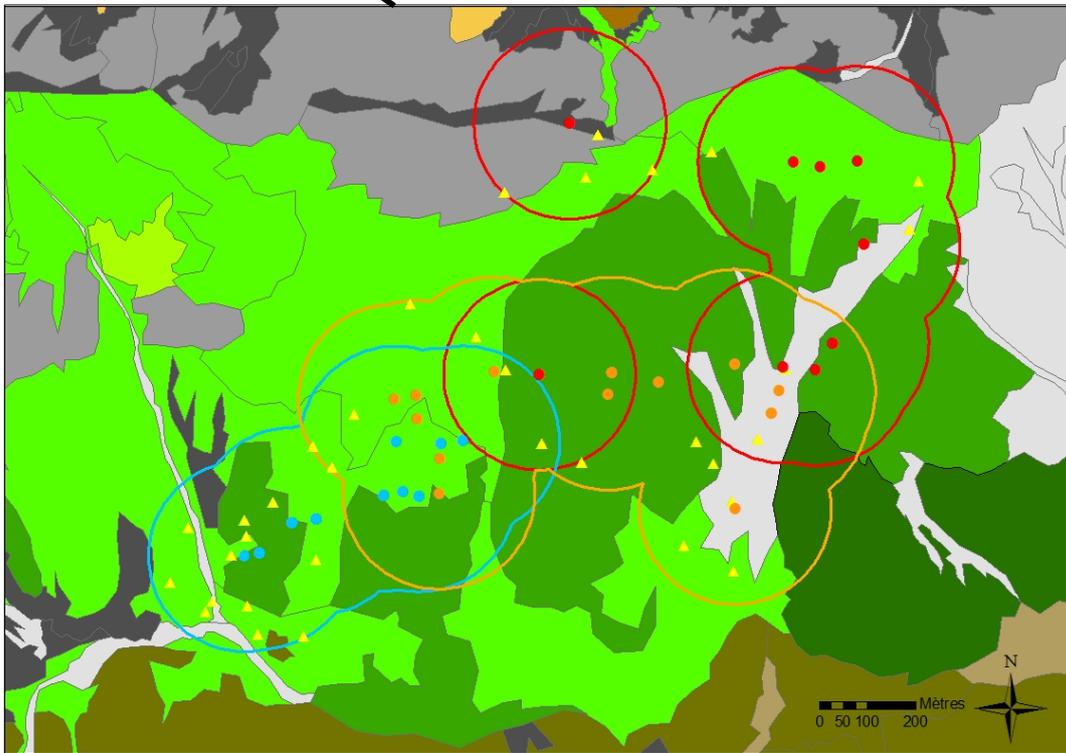


Fig.1 Site d'échantillonnage des fèces sur le domaine de Clot-Cayan dans le Parc National des Pyrénées

Résultats

Données spatiales

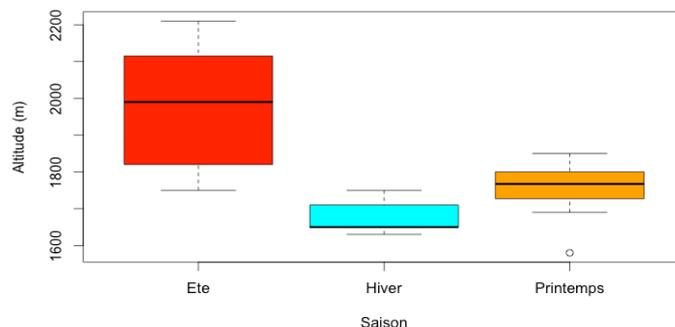


Fig. 2 : Altitudes des échantillons de fèces en fonction des saisons

L'altitude moyenne de l'isard en été est de 1760m (± 165.17 sd) (min : 1750m, max : 2210m). L'hiver l'altitude moyenne est de 1676m ($\pm 38,71$ sd) (min : 1630m, max : 1750m. Au printemps, l'altitude moyenne est de 1760m (± 63.37 sd) (min : 1580, max : 1850). Le test non paramétrique de rang de Wilcoxon-Mann-whitney nous indique qu'il y a une différence significative entre été-hiver (p-value = 1.206e-06), été-printemps (p-value = 0.0006304) et hiver-printemps (p-value = 4.225e-06).

L'été la surface totale de l'aire de répartition des isards est de 62.46 ha. Elle est de 35.91 ha en hiver et de 62,23 ha au printemps.

Détermination taxonomique

En comparant, l'identification taxonomique faite à partir de GENBANK et l'identification faite à partir du relevé de végétation de péguères Nous constatons que nous avons plus que doublé le nombre de MOTUs identifiés au niveau de l'espèce.

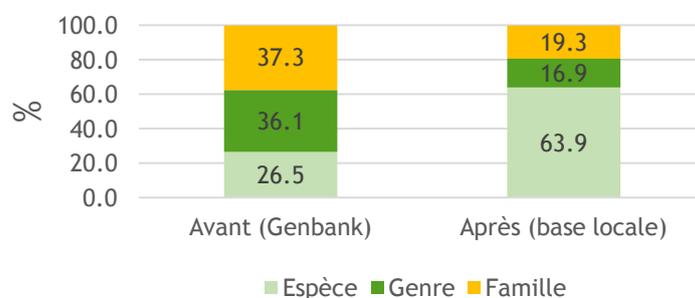


Fig. 2 : Détermination des taxons avant (avec GENBANK) et après (relevé Péguères de végétation).

La courbe d'échantillonnage des familles (courbe a) atteint une asymptote (Chao2 = 40 [35-45] (sd)) qui montre que nous avons bien échantillonné les familles. Cependant, la courbe d'échantillonnage des espèces (courbe b) n'atteint pas une asymptote (Chao2 = 108 [95-121] (sd)), qui montre que nous avons manqué l'échantillonnage de 12 à 38 espèces.

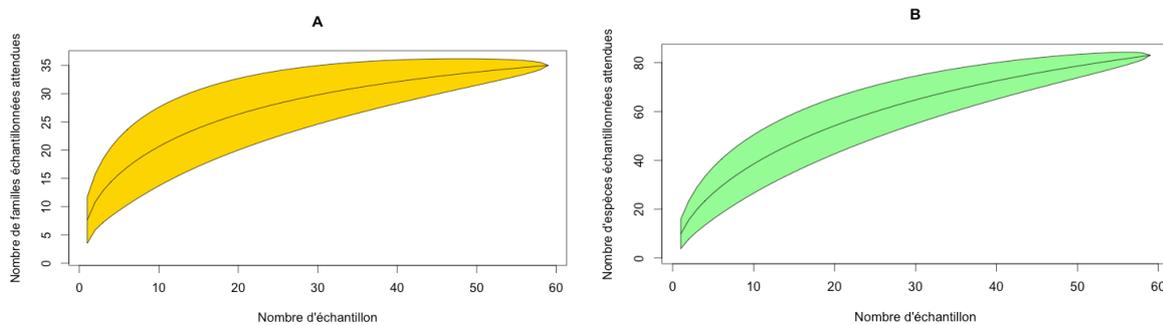


Fig. 3 : Courbe d'accumulation des familles (courbe a) et des espèces (courbe b), les bornes extérieures délimitent l'intervalle de confiance à 95%

Régime alimentaire

Le démultiplexage et l'identification des séquences uniques m'a permis d'obtenir 361 séquences uniques. Après le filtrage des données nous avons gardé 83 séquences uniques.

Après identification taxonomique, l'ensemble des espèces présentes dans le régime alimentaire de *Rupicapra pyrenaica pyrenaica* sont détaillées en Annexe 2.

Les échantillons de décembre 2013 et de Mars 2014 ont la même géolocalisation et les mêmes conditions météo (beau et froid avec 1m de neige). Ces échantillons décrivent donc la saison Hiver (**Fig. 1**).

Chaque échantillon de fèces contient une moyenne de 9.898 (± 3.06 sd) séquences uniques. Au printemps, la moyenne de séquences uniques par échantillon est de 9.7 (± 3.3 sd), en été 13.1 (± 2.0 sd) séquences uniques par échantillon et l'hiver la moyenne est de 8.8 (± 2.3 sd). D'après le test de Wilcoxon-Mann-whitney il y a une différence significative du nombre de plantes par échantillon entre l'été et l'hiver (p-value = 6.412e-05) et entre l'été et le printemps (p-value = 0.002576). Cependant, la différence est non significative entre le printemps et l'hiver (p-value = 0.48).

En utilisant la fréquence d'occurrence par saison (nombre d'échantillons qui possède une espèce sur l'ensemble des échantillons de cette même saison) (Rayé et al., 2010), un tableau a été réalisé représentant les familles (**Table. 1**) et les espèces (**Table. 2**) présentes au moins une fois dans le top 10 d'une saison.

Les 14 familles représentées ci-dessous représentent en terme d'abondance (Nombre de séquences) 95.3 % de l'ensemble des 35 familles.

Famille	Fo (%)	Fo Eté (%)	Fo Hiver (%)	Fo Printemps (%)
Poaceae	89.8	63.6	92.9	100.0
Cistaceae	88.1	100.0	89.3	80.0
Rosaceae	79.7	100.0	75.0	75.0
Ericaceae	78.0	81.8	64.3	95.0
Pinaceae	76.3	27.3	96.4	75.0
Betulaceae	64.4	54.5	82.1	45.0
Crassulaceae	52.5	54.5	64.3	35.0
Lamiaceae	39.0	0.0	46.4	50.0
Plantaginaceae	25.4	36.4	25.0	20.0
Fabaceae	23.7	100.0	3.6	10.0
Asteraceae	18.6	54.5	14.3	5.0
Cupressaceae	15.3	0.0	32.1	0.0
Dryopteridaceae	8.5	45.5	0.0	0.0
Cyperaceae	8.5	0.0	0.0	25.0

Table 1 : Familles présentes au moins une fois dans le top 10 d'une saison.

Les familles Poaceae et Cistaceae sont les deux familles les plus occurrentes, les plus représentées au sein des échantillons. Sur les 14 familles représentées dans cette table, 50 % sont présentes dans plus de 50 % des échantillons. Les Cistaceae, Rosaceae, Ericaceae, et Poaceae sont présentes dans plus de 50% d'échantillons toute l'année. Cependant, les Cypéraceae, Dryopteridaceae, Cupressaceae (représentée par une seule espèce : le genévrier commun ou *Juniperus communis*) sont mangées qu'une seule fois dans l'année (respectivement au printemps, en été et en hiver). La famille Pinaceae est majoritairement mangée l'hiver (Fo = 96.4%).

La présence/absence de la famille Pinaceae en été est significativement différente du printemps et de l'hiver (respectivement p.value = 0.000378, p.value = 2.05e-05 avec AIC = 633.32, p.value modèle <2e-16). De même pour Ericaceae, il y a une différence significative entre le printemps et l'été (p.value = 0.009293 AIC 633.32, p.value modèle <2e-16) mais aussi entre l'hiver et le printemps (p.value = 0.028052 AIC 633.32, p.value modèle <2e-16). Les autres résultats du modèle sont en **Annexe 3**.

Voici maintenant les 18 espèces qui représentent en fréquence d'abondance (Fs) 86.1 % des 83 espèces :

Espèce	Famille	Fo (%)	Fo Eté (%)	Fo Hiver (%)	Fo Printemps (%)
<i>Helianthemum nummularium</i>	Cistaceae	88.1	100.0	89.3	80.0
<i>Rosaceae sp1</i>	Rosaceae	78.0	100.0	71.4	75.0
<i>Pinus sylvestris</i>	Pinaceae	74.6	27.3	96.4	70.0
<i>Poaceae sp1</i>	Poaceae	71.2	9.1	89.3	80.0
<i>Corylus avellana</i>	Betulaceae	64.4	54.5	82.1	45.0
<i>Arctostaphylos uva-ursi</i>	Ericaceae	59.3	72.7	28.6	95.0
<i>Brachypodium pinnatum</i>	Poaceae	42.4	45.5	7.1	90.0
<i>Poaceae sp2</i>	Poaceae	42.4	9.1	60.7	35.0
<i>Teucrium scorodonia</i>	Lamiaceae	39.0	0.0	46.4	50.0
<i>Calluna vulgaris</i>	Ericaceae	25.4	27.3	42.9	0.0
<i>Lotus corniculatus</i>	Fabaceae	23.7	100.0	3.6	10.0
<i>Crassulaceae sp1</i>	Crassulaceae	23.7	18.2	35.7	10.0
<i>Sempervivum tectorum</i>	Crassulaceae	22.0	9.1	32.1	15.0
<i>Rubus idaeus</i>	Rosaceae	18.6	54.5	7.1	15.0
<i>Vaccinium myrtillus</i>	Ericaceae	15.3	18.2	0.0	35.0
<i>Asteraceae sp1</i>	Asteraceae	8.5	45.5	0.0	0.0
<i>Dryopteris sp1</i>	Dryopteridaceae	8.5	45.5	0.0	0.0
<i>Vicia orubus</i>	Fabaceae	8.5	45.5	0.0	0.0

Table 2 : Espèces présentes au moins une fois dans le top 10 d'une saison.

Parmi ces 18 MOTUs, 27.7% sont identifiées jusqu'à la famille, 5.5% jusqu'au genre et 66.6% sont identifiées jusqu'à l'espèce.

La famille Poaceae est la mieux représentée avec trois espèces (*Poaceae sp1*, *Brachypodium pinnatum* et *Poaceae sp2*). L'*Helianthemum nummularium* est l'espèce la plus présentes sur l'ensemble des 59 échantillons mais aussi l'espèce la plus abondante en terme de fréquence de séquences. Six espèces sont présentes dans plus de 50 % des échantillons. Seulement trois espèces sont présentes à plus de 50% sur l'ensemble des trois saisons : *Helianthemum nummularium*, *Rosaceae sp1* et *Corylus avellana*. *Lotus corniculatus* est majoritaire en hiver (> 30% de différence entre saison) tout comme *Rubus idaeus*, *Asteraceae sp1*, *Dryopteris sp1* et *Vicia orubus*.

La présence/absence en été et en hiver diffère pour sept espèces. Par exemple, *Pinus sylvestris* (p.value = 2.32e-05), *Poaceae sp1* (p.value = 1.34e-05) et *Corylus avellana* (p.value = 0.003657) avec une AIC pour ce modèle de 853 contre 1223 pour le modèle nulle et une p.value inférieur à 2e-16. Les autres exemples sont en **Annexe 3**.

Les fréquences de présence/absence dépendent des types d'habitats mais aussi des conditions climatiques ainsi pour chaque aire de répartition voici les différents types d'habitat

Types d'habitat	Été	Hiver	Printemps
Falaises silicieuses (lichens)	20.0	5.0	21.0
Falaises calcaires	6.0	0.0	0.0
Eboulis	7.0	7.8	17.4
Affleurement et rochers érodés	0.6	0.7	1.1
Forêt de Pins de montagne	16.9	0.4	7.2
Forêt de Pins sylvestre	12.3	28.2	22.8
Landes alpines (Rhododendron, Genévrier, <i>Arctostaphylos uva-ursis</i>)	13.4	18.6	10.2
Pelouses alpines acidiphiles (<i>Festuca</i>)	9.0	0.7	2.0
Pelouses calcaire semi-arides (<i>Brachypodium</i> , <i>Bromus erectus</i>)	9.0	23.7	11.0
Pelouses calcicoles alpines et subalpines	3.9	0.0	1.7
Prairie silicieuse sèches	1.2	5.6	2.2
Mégaphorbiaie alpines (grandes herbes)	0.7	6.5	3.4
Ruisselets	0.0	3.0	0.0

Table 3 : Types d'habitats (typologie "Corine Biotopes") en fonction des trois saisons

En été nous retrouvons l'isard majoritairement au niveau des falaises silicieuses, des forêts de Pins de montagnes (Pins à crochet), et des landes alpines (13.4%) (Rhododendron, Genévrier, *Arctostaphylos uva-ursis*) tandis qu'en hiver on le retrouve plutôt au niveau des forêts de Pins sylvestres, des pelouses calcaires (*Brachypodium*, *Bromus erectus*) mais aussi des landes alpines (18.6%). L' *Arctostaphylos uva-ursis* est présente dans 28.6% des échantillons l'hiver contre 72.7% en été. Au printemps l'habitat est plus homogène avec des falaises, des éboulis, de la forêt, des landes et des pelouses.

Discussion

Régime alimentaire très diversifié

L'étude du régime alimentaire de l'isard nous indique que sa nourriture est très diversifiée (83 espèces différentes mangées dans une année). Ensuite, l'étude de l'altitude des échantillons, et des aires de répartition nous montrent qu'une migration saisonnière s'effectue. Cette migration saisonnière est liée à la recherche de nourriture et donc à la recherche de zone de refuge ou zone d'hivernage (Weber, 2001). De plus, "La recherche de la nourriture joue un rôle considérable dans les mœurs de l'animal, conditionne sa dissémination géographique et l'altitude à laquelle il vit", Couturier en 1938. Le régime alimentaire est donc différent d'une saison à l'autre car les isards n'occupent pas les mêmes habitats, ni les mêmes altitudes. La différence entre les fréquences d'occurrence entre l'hiver et l'été ou le printemps chez certaines espèces comme *Arctostaphylos uva-ursis* est due au fait qu'en hiver les landes sont recouvertes de neige donc ce met est difficile d'accès. Pour l'obtenir l'isard doit gratter la neige

avec ses pattes (Weber, 2001), ce qui lui coûte de l'énergie. Encore faut-il que le manteau neigeux soit peu épais. L'hiver *Pinus sylvestris* couvre 28.3% de l'aire de répartition et est présent dans 96.4% des échantillons. Les isards se rabattent donc en forêt sur les arbrisseaux et arbustes (*Corylus avellana* (82% des échantillons) qui dépassent de la neige et sur les feuilles mortes, bourgeons, rameaux et aiguilles en se dressant sur leurs pattes arrières. (Weber, 2001).

Enfin, L'isard a une nette préférence pour les herbacées comme la famille des graminées (Poaceae) avec les fétuques, pâturins et avoines (Weber, 2001) , ou les légumineuses (Fabaceae) et les Rosaceae comme les framboisiers (Couturier, 1938) qui lui fournissent plus d'énergie ce qui est relié à sa petite taille et à l'énergie dépensée pour se déplacer dans un habitat extrême tel que les falaises, les rochers. Ses préférences changent en fonction des saisons et donc en fonction de la disponibilité ce qui montre qu'il possède une grande adaptation digestive du fait du changement net des plantes mangées suivant les saisons. C'est pourquoi il est opportuniste et entre dans la catégorie intermédiaire située entre spécialistes et généralistes (Hofmann, 1989). Cette adaptation peut être due à la remarquable richesse de la microflore et de la microfaune intestinale de leur appareil digestif (Weber, 2001). De plus, le volume de son ruminoreticulum augmente l'hiver par 150% (Drescher-Kaden, 1977) (300% pour Hofmann, 1989) lui facilitant l'assimilation et la digestion des fibres végétales.

Une diminution de la surface de répartition autour des fèces devrait avoir pour conséquence une baisse de la richesse spécifique par fèces car les individus sont plus concentrés et se cantonnent donc à des habitats moins différents. Cependant, ceci n'est pas le cas au printemps, la diversité par fèces est égale à l'hiver et la surface de répartition est égale à l'été. Il faut donc ajouter l'altitude à cette interprétation. En effet, l'été l'isard balaye des altitudes très différentes comparées à l'hiver et le printemps.

La comparaison entre saisons du régime alimentaire peut être biaisée par le peu d'échantillon par saison et par la différence du nombre de ces derniers entre les saisons.

Les pouvoirs et les limites de l'approche metabarcoding ADN

Dans cette étude, nous avons utilisé une technique de séquençage nouvelle génération à haut débit Miseq (Illumina) combinée à une approche novatrice, le metabarcoding ADN (Quéméré et al., 2013). Cette méthode est rentable car elle utilise moins de temps, de plus elle

permet d'évaluer le profil de l'alimentation d'un grand nombre d'individus en même temps. Il y a donc ici une possibilité d'effectuer des analyses à de grandes échelles (paysages par exemples). De plus cette méthode ne nécessite pas une grande connaissance taxonomique des plantes (Quéméré et al., 2013). utilisant l'intron trnL et sa boucle P6 a plusieurs avantages : les primers sont universelles et le système d'amplification est très robuste (Tabarlet et al., 2006), il a une petite longueur et peut être facilement amplifié (Quéméré et al., 2013). Cette boucle P6 peut être amplifiée en utilisant de l'ADN dégradé (ADN de matières fécales comme dans cette étude ou ADN gastriques) (Tabarlet et al., 2006) contrairement à la microhistologie qui est fortement influencée par la variation de la densité et la survie des cuticules dans les pelottes fécales (Holechek et al., 1982 ; Soininen et al. 2009). Cependant, sa capacité de résolution a été jugée particulièrement faible au sein de certaines familles majeures comme Poaceae et Asteraceae, le risque de sous-estimer la richesse de l'alimentation de cette espèce. Par conséquent, des efforts doivent être faits à l'avenir pour augmenter le pouvoir de résolution de la méthode, peut être à l'aide de codes -barres d'ADN supplémentaires avec une couverture taxonomique plus restreinte (par exemple en utilisant des amorces spécifiques à ces familles). Enfin, la probabilité que nos échantillons aient été contaminés par certaines conditions environnementales du sol, ou des débris soufflés est non nulle malgré les précautions d'échantillonnage.

De plus , la famille des Poacées peut être surreprésentée du fait de la résistance de leur épiderme aux processus digestifs (Neal, Pulkinen et Bartolome et al, 1995).

Perspectives

D'après les études, nous observons que l'isard est un "feeder" de type intermédiaire et que son alimentation change suivant les saisons. C'est pourquoi, nous pouvons penser que ce qu'il mange est le reflet du disponible. Ainsi, il serait intéressant d'avoir une liste d'espèces végétales présentes par saison sur ce domaine (en utilisant le relevé Péguyères par exemple) et d'ainsi observer si les plantes mangées sont bien le reflet des plantes disponibles.

Ensuite, il aurait été intéressant de construire notre propre base local de référence en s'appuyant sur les plantes disponibles à l'année sur le domaine étudié de Clot-Cayan. Ceci aurait permis une meilleure identification à la sortie du séquençage (comme l'ont fait Rayé et al., 2010 et Quéméré et al., en 2013).

L'étude du régime alimentaire d'une autre espèce d'ongulé tel que le bouquetin permettra la comparaison de ces deux régimes alimentaires et donc l'observation d'une éventuelle

compétition liée aux ressources alimentaires et aux zones de refuges. La conservation d'une espèce animale commence par la conservation des ressources et de l'habitat. Le régime alimentaire prend donc ici toute son importance.

Cette étude rentre dans le cadre d'une étude préliminaire nouvelle qui peut ouvrir les portes à d'autres études sur différentes espèces dans des projets de conservation de l'habitat et des écosystèmes. Cette technique est une prouesse, car à partir d'une seule pelote fécale nous pouvons savoir toute l'écologie de l'espèce sans l'approcher, sans la déranger dans son milieu de vie.

Bibliographie

Anthony RG, Smith NS (1974) Comparison of rumen and fecal analysis to describe deer diets. *J Wildl Manag* 38 :729–746

Auvray F. et al. (1983) Recherche sur l'e co-éthologie du mouflon (*Ovis musimon musimon* Schreber 1782) dans le massif de Caroux-Espi-nousse (Hé rault). PhD thesis, Montpellier University, France

Babad G. et al. (1997) Etude des relations entre un peuplement animal et la vé gé tation: impacts du chamois, du chevreuil et du mouflon sur les peuplements forestiers dans la Re sserve Nationale de Faune Sauvage et de Chasse des Bauges (Savoie). PhD thesis, Chambéry University, France

Bartolome, J., Franch, J., Gutman, M. & Seligan, N.A.G. (1995). Technical note : physical factors that influence faecal analysis estimates of herbivore diets. *J. Range Mgmt.* 48, 267–270.

Belovski, G.E. et al. (1986). Generalist herbivore foraging and its role in competitive interactions. *Am. Zool.* 26, 51–69.

Berducou C. et al. (1974) Contribution à l'étude d'un problème éco- physiologique pyrénéen: l'alimentation hivernale de l'isard. PhD thesis, Toulouse University, France.

Bertolino S., Di Montezemolo NC., Bassano B. (2008) Food-niche relationships within a guild of alpine ungulates including an introduced species. *J Zool* 277 :63–69

Cazajous T. et al (2002) Etude sociospatiale de l'isard et application à sa chasse. Thèse école vétérinaire de Toulouse. Pages 23,33-34

Couturier MAJ (1938) Le Chamois. Arthaud, Grenoble. 501-510

De Barba et al (2014) DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: application to omnivorous diet. *Molecular Ecology Resources* 14 (2) : 306–323

Drescher-Kaden V et al. (1977). Energy supply of chamois : microbial fermentations of natural food in rumen and cecum during summer and winter in the Alps. In : T.J. Peterle (Editor), XIIIth Congress of Game Biol. Wildlife Sot., Washington, pp. 114- 124.

Garcia-Gonzales R, Cuartas P (1996) Trophic utilisation of a mountain/subalpine forest by chamois (*Rupicapra pyrenaica*) in the central Pyrenees. *For Ecol Manage* 88 :15-23.

Hebert PDN, Cywinska A, Ball SL, De Waard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London Series B: Biological Sciences* 270: 313-321.

Hofmann R.R. et al. (1989) Evolutionary steps of ecophysiological adaptation and diversification of ruminants : a comparative view of their digestive system. *Oecologia* 78:443-57

Holechek J. et al. (1982) Botanical Composition Determination of Range Herbivore Diets: A Review. *Journal of range management* 35 (3): 309-315.

Hollingsworth PM, Forrest LL, Spouge JL, Hajibabaei M, Ratnasingham S, et al. (2009) A DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America* 106: 12794-12797.

La Morgia V & Bassano B. Feeding habits, forage selection, and diet overlap in Alpine chamois (*Rupicapra rupicapra* L.) and domestic sheep. *Ecol Res. Ecological Research* (Impact Factor: 1.3) 24(5):1043-1050.

Latham, J. et al. (1999). Interspecific interactions of ungulates in European forests: an overview. *For. Ecol. Mgmt.* 120, 13–21.

Loison A. et al. (2003) Large Herbivores in European Alpine Ecosystems : Current Status and Challenges for the Future. *Ecological Studies*, 167, 351-366

Neal, B.R., Pulkinen, D.A. & Owen, B.D. (1973). A comparison of faecal and stomach contents analysis in the meadow vole (*Microtus pennsylvanicus*). *Can. J. Zool.* 51, 715–721. Pompanon F, Deagle BE, Symondson WOC,

Brown DS, Jarman SN, et al. (2012) Who is eating what : diet assessment using next generation sequencing. *Molecular Ecology* 21 : 1931-1950.

Post, E. and N. C., Stenseth, N., C. (1999). Climatic variability, plant phenology, and northern ungulates. *Ecology*, 80 : 1322–1339.

Quéméré E, Hibert F, Miquel C, Lhuillier E, Rasolondraibe E, et al. (2013) A DNA Metabarcoding Study of a Primate Dietary Diversity and Plasticity across Its Entire Fragmented Range. *PLoS ONE* 8(3) : e58971.

Rayé G. et al. (2010) New insights on diet variability revealed by DNA barcoding and high-throughput pyrosequencing: chamois diet in autumn as a case study. *Ecol Res* (2011) 26: 265–276.

Seguritan V, Rohwer F (2001) FastGroup: a program to dereplicate libraries of 16S rDNA sequences. *BMC Microbiol Bioinformatics* 2

Soininen M et al. (2009) Analysing diet of small herbivores: the efficiency of DNA barcoding coupled with high-throughput pyrosequencing for deciphering the composition of complex plant mixtures. *BioMed central* 1-9.

Taberlet, P. et al. (2007) Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic Acids Res.* 35, e14.

Valentini, A. et al. New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the trnL approach. *Mol. Ecol. Res.* (in press)

Vaucher CA (1988) Contribution à l'étude éco-éthologique du chamois (*Rupicapra rupicapra*) au Mont Saleve (Haute Savoie). PhD thesis, Univ Nancy 1.

Weber E. (2001) Le chamois et l'isard. Edition delachaux et Niestlé Les sentiers du naturaliste.

Annexes

Annexe 1 : Script dans terminal pour filtrer les données et identifier les taxons via Obitools

```
ssh equemere@genotoul.toulouse.inra.fr
```

```
cd work
```

```
cd metabarcoding
```

```
cd morgan
```

```
cd Isard
```

####Récupération de la séquence assemblée, l'assemblage des lectures avant et arrière se fait avec l'utilitaire *illuminapairedend* qui aligne les deux lectures et retourne la séquence reconstruite.

```
illuminapairedend --score-min=40 -r Isards_ACTGAT_L001_R1.fastq Isards_ACTGAT_L001_R2.fastq > Isards.fastq
```

```
obihead --without-progress-bar -n 2 Isards.fastq
```

####Supprimer des enregistrements de séquences non alignées

Séquences non alignées ne peuvent pas être utilisées. *mode!* = "joined" signifie que si la valeur de l'attribut *mode* est différent de joint/lié, l'enregistrement de la séquence correspondante sera conservé. La commande suivante permet de les retirer de l'ensemble de données:

```
obigrep -p 'mode!="joined"' Isards.fastq > Isards.ali.fastq
```

```
obihead --without-progress-bar -n 2 Isards.ali.fastq
```

#####Attribuer chaque enregistrement de la séquence à la combinaison de l'échantillon/marqueur correspondant. Chaque enregistrement de la séquence est affecté à son échantillon et le marqueur en utilisant les données fournies dans un fichier texte correspondant (ici *Isards_diet_ngs filter.txt*). Ce fichier texte contient une ligne par échantillon, avec le nom de l'expérience (plusieurs expériences peuvent être inclus dans le même fichier), le nom des tags, la séquence de l'amorce sens, la séquence de l'amorce inverse

```
ngsfilter -t isard_ngs2.txt -u pasbon_2.fastq Isards.ali.fastq > Isards.ali.assigned_2.fastq
```

```
obihead --without-progress-bar -n 5 Isards.ali.assigned_2.fastq
```

#####Dérépliquer lectures en séquence unique : , il est commode de travailler avec des séquences uniques au lieu des lectures(reads). 1) Groupez les lectures strictement identiques

ensemble 2) Sortez la séquence pour chaque groupe et son nombre dans le jeu de données d'origine (de cette manière, la duplication de toutes les lectures est retirée)

```
obiuniq -m sample Isards.ali.assigned_2.fastq > Isards.ali.assigned.uniq_2.fasta
```

####Enlever les "Bruits/pollutions" des séquences du jeu de données (erreur de séquençage, chimères...)

```
obiannotate -k count -k merged_sample Isards.ali.assigned.uniq_2.fasta > $$ ; mv $$ Isards.ali.annotate.uniq_2.fasta
```

```
obihead --without-progress-bar -n 5 Isards.ali.annotate.uniq_2.fasta
```

####Obtenir les statistiques de comptage : nous donne le nombre de séquences qui se produisent une fois

```
obistat -c count Isards.ali.annotate.uniq_2.fasta | sort -nk1 | head -30
```

Conserver uniquement les séquences ayant un nombre >ou= à 100 et une longueur inférieure à 80 paires de bases.

```
obigrep -p 'count>=100' Isards.ali.annotate.uniq_2.fasta > Isards.uniq_2.100.fasta
```

```
obistat Isards.uniq_2.100.fasta
```

```
obihead --without-progress-bar -n 2 Isards.uniq_2.100.fasta
```

####Nettoyez les séquences pour les erreurs PCR/Echantillonnage

```
obiclean -s merged_sample -d 1 -r 0.05 -H Isards.uniq_2.100.fasta > Isards.uniq.clean_2.fasta
```

```
obihead --without-progress-bar -n 2 Isards.uniq.clean_2.fasta
```

####Assigner chaque séquence à un taxon via une base de référence (GENBANK)

```
ecotag -d /work/jchave/metabarcoding/db_creation/genbank_r209/genbank_r209 -R ./gh_genbank_r209_final.fasta --sort=count Isards.uniq.clean_2.fasta > Isards.ecotag_2.fasta
```

```
obihead --without-progress-bar -n 2 Isards.ecotag_2.fasta
```

####Générer le tableau de résultats final

```
obiannotate --delete-tag=scientific_name_by_db --delete-tag=obiclean_samplecount \  
--delete-tag=obiclean_count --delete-tag=obiclean_singletoncount \  
--delete-tag=obiclean_cluster --delete-tag=obiclean_internalcount \  
--delete-tag=obiclean_head --delete-tag=taxid_by_db --delete-tag=obiclean_headcount \  
--delete-tag=id_status --delete-tag=rank_by_db --delete-tag=order_name \  
--delete-tag=order Isards.ecotag_2.fasta > Isards.ecotag_clean_2.fasta
```

####Trier les séquences par ordre de nombre décroissant

```
obisort -k count -r Isards.ecotag_clean_2.fasta > Isards.ecotag_clean_sort_2.fasta
```

####Création du tableau final qui peut être lu par excel ou R studio

```
obitab -o Isards.ecotag_clean_sort_2.fasta > Isards.sort_2.
```

Annexe 2 : Toutes les séquences uniques (MOTUs) à la fin du traitement classées par ordre décroissant de Fo.

MOTU	Famille	Espec	Fs (%)	Fo (%)
MOTU4	Cistaceae	<i>Helianthemum nummularium</i>	14.09	88.14
MOTU2	Rosaceae	Rosaceae sp1	8.00	77.97
MOTU3	Pinaceae	<i>Pinus sylvestris</i>	15.18	74.58
MOTU9	Poaceae	Poaceae sp1	4.05	71.19
MOTU1	Betulaceae	<i>Corylus avellana</i>	20.64	64.41
MOTU14	Ericaceae	<i>Calluna vulgaris</i>	1.30	59.32
MOTU8	Poaceae	<i>Brachypodium pinnatum</i>	3.42	42.37
MOTU13	Poaceae	Poaceae sp2	1.72	42.37
MOTU15	Lamiaceae	<i>Teucrium scorodonia</i>	1.33	38.98
MOTU5	Ericaceae	<i>Arctostaphylos uva-ursi</i>	8.23	25.42
MOTU11	Crassulaceae	Crassulaceae sp1	2.42	23.73
MOTU12	Fabaceae	<i>Lotus corniculatus</i>	2.09	23.73
MOTU34	Crassulaceae	<i>Sempervivum tectorum</i>	0.56	22.03
MOTU16	Crassulaceae	<i>Sedum sp1</i>	1.25	18.64
MOTU24	Rosaceae	<i>Rubus ideaeus</i>	1.18	18.64
MOTU22	Cupressaceae	<i>Juniperus communis</i>	0.60	15.25
MOTU19	Ericaceae	<i>Vaccinium myrtillus</i>	0.58	15.25
MOTU41	Plantaginaceae	<i>Veronica sp1</i>	0.40	15.25
MOTU27	Cistaceae	<i>Cistus sp1</i>	0.51	10.17
MOTU60	Asteraceae	<i>Leontodon hispidus</i>	0.25	8.47
MOTU40	Dryopteridaceae	<i>Dryopteris sp1</i>	0.47	8.47
MOTU48	Fabaceae	<i>Vicia orubus</i>	0.46	8.47
MOTU10	Juncaceae	<i>Luzula sylvatica</i>	0.25	8.47
MOTU39	Poaceae	<i>Dactylis glomerata</i>	0.37	8.47
MOTU95	Poaceae	<i>Anthoxanthum odoratum</i>	0.08	8.47
MOTU51	Violaceae	<i>Viola sp1</i>	0.39	8.47
MOTU37	Asteraceae	Asteraceae sp2	0.52	6.78
MOTU53	Caryophyllaceae	<i>Cerastium sp1</i>	0.32	6.78
MOTU42	Crassulaceae	Crassulaceae sp2	0.50	6.78
MOTU32	Ericaceae	<i>Vaccinium ovalifolium</i>	0.29	6.78
MOTU54	Pinaceae	<i>Abies alba</i>	0.33	6.78
MOTU99	Adoxaceae	<i>Sambucus sp1</i>	0.14	5.08
MOTU149	Apiaceae	<i>Meum athamanticum</i>	0.04	5.08
MOTU58	Aspleniaceae	Aspleniaceae sp1	0.34	5.08
MOTU38	Asteraceae	Asteraceae sp3	0.51	5.08
MOTU29	Asteraceae	Asteraceae sp1	0.44	5.08
MOTU100	Euphorbiaceae	<i>Euphorbia hyberna</i>	0.14	5.08
MOTU75	Fabaceae	<i>Lathyrus sylvestris</i>	0.23	5.08

MOTU59	Orobanchaceae	<i>Melampyrum pratense</i>	0.34	5.08
MOTU	Famille	Espece	Fs (%)	Fo (%)
MOTU64	Plantaginaceae	<i>Veronica alpina</i>	0.27	5.08
MOTU84	Plantaginaceae	<i>Plantago lanceolata</i>	0.19	5.08
MOTU55	Rosaceae	<i>Potentilla micrantha</i>	0.26	5.08
MOTU86	Aspleniaceae	Aspleniaceae sp2	0.19	3.39
MOTU80	Asteraceae	<i>Hieracium pilosella</i>	0.20	3.39
MOTU62	Caryophyllaceae	<i>Silene nutans</i>	0.30	3.39
MOTU77	Caryophyllaceae	Caryophyllaceae sp2	0.16	3.39
MOTU106	Cyperaceae	<i>Carex sp1</i>	0.10	3.39
MOTU116	Cyperaceae	<i>Carex sp2</i>	0.08	3.39
MOTU28	Fagaceae	<i>Fagus sylvatica</i>	0.12	3.39
MOTU110	Grossulariaceae	<i>Ribes sp1</i>	0.09	3.39
MOTU7	Juncaceae	<i>Luzula nivea</i>	0.06	3.39
MOTU17	Poaceae	<i>Agrostis capillaris</i>	0.39	3.39
MOTU98	Polygonaceae	<i>Rumex sp1</i>	0.15	3.39
MOTU141	Ranunculaceae	<i>Ranunculus reptans</i>	0.08	3.39
MOTU36	Rosaceae	Rosaceae sp2	0.19	3.39
MOTU73	Rubiaceae	<i>Galium verum</i>	0.19	3.39
MOTU83	Apiaceae	<i>Bupleurum sp1</i>	0.20	1.69
MOTU103	Apiaceae	<i>Myrrhis odorata</i>	0.13	1.69
MOTU57	Asteraceae	Asteraceae sp4	0.25	1.69
MOTU109	Asteraceae	<i>Senecio latifolius</i>	0.13	1.69
MOTU67	Betulaceae	<i>Betula pendula</i>	0.26	1.69
MOTU89	Boraginaceae	<i>Pulmonaria officinalis</i>	0.17	1.69
MOTU170	Boraginaceae	Boraginaceae sp1	0.05	1.69
MOTU210	Boraginaceae	<i>Echium vulgare</i>	0.03	1.69
MOTU63	Campanulaceae	<i>Jasione montana</i>	0.29	1.69
MOTU120	Caryophyllaceae	<i>Atocion rupestre</i>	0.11	1.69
MOTU225	Caryophyllaceae	<i>Silene vulgaris</i>	0.03	1.69
MOTU78	Crassulaceae	Crassulaceae sp3	0.20	1.69
MOTU153	Cyperaceae	<i>Carex breviculmis</i>	0.06	1.69
MOTU186	Cyperaceae	<i>Carex sp3</i>	0.04	1.69
MOTU131	Dennstaedtiaceae	<i>Pteridium aquilinum</i>	0.07	1.69
MOTU132	Fabaceae	<i>Lathyrus heterophyllus</i>	0.09	1.69
MOTU68	Fabaceae	<i>Trifolium alpinum</i>	0.07	1.69
MOTU117	Gentianaceae	<i>Gentiana lutea</i>	0.07	1.69
MOTU96	Geraniaceae	<i>Geranium sylvaticum</i>	0.06	1.69
MOTU118	Iridaceae	<i>Crocus nudiflorus</i>	0.08	1.69
MOTU230	Malvaceae	<i>Malva moscheta</i>	0.03	1.69
MOTU206	Pinaceae	<i>Picea sp1</i>	0.02	1.69
MOTU292	Pinaceae	<i>Picea abies</i>	0.02	1.69
MOTU81	Plantaginaceae	<i>Globularia nudicaulis</i>	0.13	1.69
MOTU46	Poaceae	Poaceae sp3	0.29	1.69
MOTU148	Poaceae	<i>Phleum alpinum</i>	0.04	1.69

MOTU126	Saxifragaceae	<i>Saxifraga aizoides</i>	0.10	1.69
---------	---------------	---------------------------	------	------

Annexe 3 : Modèles linéaires généralisés pour l'analyse spatio-temporelle

➔ Comparaison des Familles présentes au moins une fois dans le top 10 d'une saison (Représentation des différences significatives seulement)

Model=glm(as.numeric(top10fam_ind_saison_hiver_ref[, "presenceabsence"])~top10fam_ind_saison_hiver_ref[, "famille"]*top10fam_ind_saison_hiver_ref[, "saison"], family="binomial")

➔ AIC: 633.32 / modele nul : AIC: 885.92

p.value <2e-16 ***

Hiver avec été

top10fam_ind_saison_hiver_ref[, "famille"]**Betulaceae**:top10fam_ind_saison_hiver_ref[, "saison"]ete 0.003221 **

top10fam_ind_saison_hiver_ref[, "famille"]**Crassulaceae**:top10fam_ind_saison_hiver_ref[, "saison"]ete 0.028522 *

top10fam_ind_saison_hiver_ref[, "famille"]**Pinaceae**:top10fam_ind_saison_hiver_ref[, "saison"]ete 2.05e-05 ***

top10fam_ind_saison_hiver_ref[, "famille"]**Poaceae**:top10fam_ind_saison_hiver_ref[, "saison"]ete 0.001598 **

Hiver avec printemps

top10fam_ind_saison_hiver_ref[, "famille"]**Ericaceae**:top10fam_ind_saison_hiver_ref[, "saison"]printemps 0.028052 *

Printemps avec été

top10fam_ind_saison_ref_printemps[, "famille"]**Ericaceae**:top10fam_ind_saison_ref_printemps[, "saison"]ete 0.009293 **

top10fam_ind_saison_ref_printemps[, "famille"]**Pinaceae**:top10fam_ind_saison_ref_printemps[, "saison"]ete 0.000378 ***

➔ Comparaison des espèces présentes au moins une fois dans le top 10 d'une saison (Représentation des différences significatives seulement)

Model=glm(as.numeric(top10esp_ind_saison[, "presenceabsence"])~top10esp_ind_saison[, "espece"]*top10esp_ind_saison[, "saison"], family="binomial")

➔ AIC: 853.48 / model nul : 1222.8

p.value = <2e-16 ***

Hiver avec été

top10esp_ind_saison[, "espece"]**Calluna.vulgaris**:top10esp_ind_saison[, "saison"]hiver 0.019887 *

top10esp_ind_saison[, "espece"]**Corylus.avellana**:top10esp_ind_saison[, "saison"]hiver 0.003657 **

top10esp_ind_saison[, "espece"]**Crassulaceae.sp1**:top10esp_ind_saison[, "saison"] 0.017418 *

top10esp_ind_saison[, "espece"]**Pinus.sylvestris**:top10esp_ind_saison[, "saison"] 2.32e-05 ***

top10esp_ind_saison[, "espece"]**Poaceae.sp1**:top10esp_ind_saison[, "saison"]hiver 1.34e-05 ***

top10esp_ind_saison[, "espece"]**Poaceae.sp2**:top10esp_ind_saison[, "saison"]hiver 0.000731 ***
top10esp_ind_saison[, "espece"]**Sempervivum.tectorum**:top10esp_ind_saison[, "saison"]hiver 0.012189 *

Printemps avec été

top10esp_ind_saison[, "espece"]**Rubus.ideaicus**:top10esp_ind_saison[, "saison"]printemps 0.010003 *

Hiver avec printemps

top10esp_ind_saison_ref_printemps[, "espece"]**Rubus.ideaicus**:top10esp_ind_saison_ref_printemps[, "saison"]hiver 0.039136 **
top10esp_ind_saison_ref_printemps[, "espece"]**Corylus.avellana**:top10esp_ind_saison_ref_printemps[, "saison"]hiver 1.56e-05 ***
top10esp_ind_saison_ref_printemps[, "espece"]**Crassulaceae.sp1**:top10esp_ind_saison_ref_printemps[, "saison"]hiver 8.54e-05 ***
top10esp_ind_saison_ref_printemps[, "espece"]**Helianthemum.nummularium**:top10esp_ind_saison_ref_p
rintemps[, "saison"]hiver 0.000895 ***
top10esp_ind_saison_ref_printemps[, "espece"]**Pinus.sylvestris**:top10esp_ind_saison_ref_printemps[, "saison"]hiver 6.66e-05 ***
top10esp_ind_saison_ref_printemps[, "espece"]**Poaceae.sp1**:top10esp_ind_saison_ref_printemps[, "saison"]hiver 0.000895 ***
top10esp_ind_saison_ref_printemps[, "espece"]**Poaceae.sp2**:top10esp_ind_saison_ref_printemps[, "saison"]hiver 0.000101 ***
top10esp_ind_saison_ref_printemps[, "espece"]**Rosaceae.sp1**:top10esp_ind_saison_ref_printemps[, "saison"]hiver 0.004413 ***
top10esp_ind_saison_ref_printemps[, "espece"]**Sempervivum.tectorum**:top10esp_ind_saison_ref_printem
ps[, "saison"]hiver 0.000283 ***
top10esp_ind_saison_ref_printemps[, "espece"]**Teucrium.scorodonia**:top10esp_ind_saison_ref_printemps[, "saison"]hiver 0.003019 ***