



Impacts des activités anthropiques sur la biodiversité : une approche spatiale et temporelle par analyse de l'ADN environnemental

Johan Pansu

► To cite this version:

Johan Pansu. Impacts des activités anthropiques sur la biodiversité : une approche spatiale et temporelle par analyse de l'ADN environnemental. Biodiversité. Université de Grenoble, 2014. Français. NNT : 2014GREN058 . tel-01555686

HAL Id: tel-01555686

<https://theses.hal.science/tel-01555686>

Submitted on 4 Jul 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Biodiversité-Ecologie-Environnement**

Arrêté ministériel : 7 août 2006

Présentée par

Johan PANSU

Thèse dirigée par **Pierre TABERLET** et **Philippe CHOLER**

préparée au sein du **Laboratoire d'Ecologie Alpine**
dans l'**École Doctorale Chimie et Sciences du Vivant**

Impacts des activités anthropiques sur la biodiversité : une approche spatiale et temporelle par analyse de l'ADN environnemental

Thèse soutenue publiquement le **12 décembre 2014**,
devant le jury composé de :

M. Rachid CHEDDADI

Directeur de Recherche CNRS, Montpellier, France (Rapporteur)

M. Ludovic ORLANDO

Professeur, Université de Copenhague, Danemark (Rapporteur)

M. Serge AUBERT

Professeur, Université de Grenoble, France (Examinateur, Président)

Mme. Heidy SCHIMANN

Chargée de Recherche INRA, Kourou, France (Examinatrice)

M. Pierre TABERLET

Directeur de Recherche CNRS, Grenoble, France (Directeur de thèse)

M. Philippe CHOLER

Chargé de Recherche CNRS, Grenoble, France (Co-directeur de thèse)

M. Etienne DAMBRINE

Professeur, Université de Savoie, France (Membre invité)



*Moins tu auras de nécessités,
Plus tu auras de liberté.*

Trouvé dans une papillote, décembre 2012

Remerciements

Et voilà ! Après trois années à avoir imaginé ce moment, m'y voici ! C'est donc avec grand plaisir que j'écris ces remerciements ! Tout d'abord (il ne faut pas se le cacher) parce qu'à ce moment-là de la thèse le plus dur est passé, mais surtout pour témoigner ma gratitude à toutes les personnes que j'ai pu rencontrer et qui m'ont permis de vivre, et d'apprécier, cette véritable aventure humaine dont voici l'aboutissement. Ces trois années ont été très riches en découvertes, rencontres et voyages. Peut-être aurais-je dû établir une liste au préalable afin de n'oublier personne mais bon... Au vu du bazar qui s'accumule sur mon bureau depuis quelques mois, je ne suis pas certain que ce fût une meilleure idée ! Donc mes excuses par avance à ceux que je pourrais omettre de citer!

En premier lieu je voudrais remercier toutes les personnes au sein du LECA qui m'ont accueilli et permis de mener à bien ce projet, à commencer par mes directeurs de thèse : Pierre et Philippe. Merci à vous deux pour votre confiance, pour les conditions de thèse que vous m'avez fournies et pour m'avoir fait profiter de vos larges, et très complémentaires, cultures scientifiques. Pierre, merci pour tes conseils, ton optimisme à toute épreuve, pour l'autonomie que tu m'as laissée et pour ton soutien malgré mon côté parfois un peu tête brûlée. Philippe, merci pour ta rigueur scientifique (la seule rigueur qui soit digne et indispensable), ton ouverture d'esprit et pour m'avoir poussé à faire toujours mieux.

Je tiens également à remercier François pour m'avoir fait découvrir le monde de la recherche lors de mon stage de M2 puis, pour son tutorat en enseignement par la suite. Tes talents pédagogiques, ta disponibilité et tes conseils pleins de malice (et de vanne) ont été un véritable exemple.

Un merci tout particulier à Fred B et Lucie qui ont été bien plus que des collègues et compagnons de bureau : un soutien psychologique sans faille et surtout des amis (malgré quelques divergences de goûts musicaux, l'intéressé -et l'orthographe est correct- se reconnaîtra). Je vous suis extrêmement reconnaissant pour m'avoir toujours poussé et motivé (et tout ça sans jamais se prendre au sérieux). Cela a pu coûter quelques bières mais à vrai dire on n'avait pas trop besoin d'excuses pour ça !

Un grand merci également à Ludo (pour ton aide et ton humour trash en n'importe quelle circonstance), Christian, alias Martinez (pour ta bonne humeur et disponibilité inégalées) et Delphine. Merci à vous de m'avoir guidé dans les méandres de la bio mol et pour ces séances de TP mémorables !

Je remercie aussi grandement tous ceux avec qui j'ai eu le plaisir de travailler étroitement au cours de ces trois dernières années, notamment Charline (pour cette excellente et très riche collaboration!), Eric (pour ton aide et pour les projets à venir que j'espère couronnés de succès !), Francesco et Aurélie (pour votre sympathie et vos conseils toujours avisés), Céline (à la fois camarade de thèse et d'« éclopage »), Gillou (pour ton humanité), Alice (qui m'a guidé à mes grands débuts), Marta et Carla (pour votre joie de vivre et votre gentillesse). Et cela toujours dans le sérieux et la bonne humeur !

Je voudrais aussi remercier toutes les autres personnes avec qui j'ai collaboré, qui m'ont permis de m'ouvrir à d'autres disciplines et de découvrir de nouveaux environnements. Merci ainsi à Jérôme C, Amaia, Heidy et les autres membres de la mission en Guyane pour votre accueil, votre bonne humeur et pour toutes les discussions fructueuses que nous avons pu avoir (au labo, en forêt ou dans le carbet-cuisine). Merci à Fabien, Jérôme P et les membres du labo EDYTEM pour m'avoir fait découvrir les joies des études paléoenvironnementales (en particulier cette fameuse démonstration de plongeon involontaire dans le lac du Lauzanier) ainsi que pour les discussions fort intéressantes et les conseils que vous m'avez prodigués. Merci également à Lauric, Seb et Jean-Jacques pour cette excellente collaboration aussi bien sur le plan scientifique qu'humain : tartiflette/vin blanc après le terrain et les combos petit-déj/réunion sont autant de moyens de joindre l'utile à l'agréable et ce fût un plaisir de partager cela avec vous ! Merci aussi à Irène de m'avoir fait confiance pour cette mission en Patagonie ainsi qu'à Alex et Frida pour leur accueil et leur sympathie.

Merci également à toutes les personnes que j'ai rencontrées et avec qui j'ai passé de super moments au labo ou en dehors (souvent autour d'une bière). Merci à Mathieu (pour sa bonne humeur communicative et ses « supers projets ») et Flo A. (pour toutes les discussions d'ordre philosophique... ou pas) ainsi qu'à tous les adeptes du « coin café », haut lieu de la vie du labo et théâtre de nombreux craquages inter-équipes : Stéphane R. (dont le planning d'enseignement m'a boosté à finir ma thèse), Muriel, Sophie(s), Christophe, Tristan, Fred(s), J-C, Nico (et son genou vaillamment sacrifié à la science dans

un champ de Valloire). Et tous les autres avec qui j'ai pu échanger: Loïc, Flo(s), Laure(s), Marti, Wilfried, Julien, Cindy, Florence, Kim, Nancy, Rolland, Serge, Roberto Stéphane B. et Sophie... J'ai également une pensée pour mes petits camarades du bureau 313, les anciens comme les nouveaux (Thibault, Charles, Monica) qui ont enduré les quelques pétages de plomb (toujours dans la bonne humeur !) que j'ai pu effectuer au niveau de ma personne ces dernières semaines.

Je tiens à remercier aussi grandement les membres du jury d'avoir accepté d'évaluer ses travaux et de faire le (parfois très long) déplacement jusqu'à Grenoble.

Evidemment je remercie tous mes potes pour leur soutien et tous ces excellents moments partagés en leur compagnie! A commencé par mes camarades de master et thésards grenoblois: Renaud (ceci marque la fin de nos 15 ans d'études communes... snif), Jérémy (détends-toi, ça va bien se passer) et Hicham (le premier diplômé!). Je pense également aux autres partis faire leur thèse dans d'autres contrées: Camille, Nico, Cyril, Clément et Pierre. Merci également à toute une flopée d'ami(e)s de fac et d'ailleurs : Perrine (qui rêve d'écrire des remerciements donc Perrine « ces remerciements c'est aussi les tiens »), Seb(s), Eléonore, Marie, J-L, Fab, Nico, Elsa(s), Quentin, Thierno, Will, Laurine, Lucas, Mattia, Mathieu, Loïc, Simon, Chloé, Florent, Arthur, Maëlle... Et tous les autres !

Et pour finir, j'adresse mes plus profonds remerciements à ma famille qui est toujours là pour moi, quoi qu'il arrive. Merci à mes grands-parents: Daniel, Marie-Thé, Dédé, Chantal. J'ai une pensée toute particulière pour Nicole, à qui j'avais juré, il y a 20 ans de cela, que je ne ferai jamais de thèse (comme le dit l'adage : il n'y a que les c... qui ne changent pas d'avis). Merci à ma petite sœur Morgane, partie voir du pays à l'autre bout du monde (dommage que tu ne sois pas là mais, à la réflexion je crois que je te souhaite de vivre ça !). Merci à mes parents, Pascal et Fabienne, pour... Ben pour tout en fait ! Les relectures à l'arrache sont bien anecdotiques à côté de tout ce que vous m'avez apporté et je suis très fier de tout ce que vous m'avez transmis. Et enfin merci à Julie qui partage ma vie de semi-nomade depuis quelques temps déjà. Merci pour ton soutien indéfectible et ta patience, qui ont été inestimables, en particulier durant ces dernières semaines !

Table des matières

AVANT-PROPOS.....	3
 CHAPITRE I. INTRODUCTION.....	5
1. ACTIVITES HUMAINES ET BIODIVERSITE A L'ECHELLE DU GLOBE.....	5
a. <i>Une anthropisation généralisée.....</i>	5
b. <i>Une perte globale de biodiversité.....</i>	6
2. LES PERTURBATIONS ANTHROPIQUES A L'ECHELLE LOCALE.....	7
a. <i>Définitions des perturbations</i>	8
b. <i>De l'importance des perturbations en écologie</i>	10
c. <i>Le cas des activités humaines.....</i>	12
d. <i>Facteurs déterminant la réponse des communautés face aux perturbations.....</i>	14
3. DES METHODES D'INVESTIGATION LIMITEES	16
a. <i>Une couverture taxonomique limitée.....</i>	16
b. <i>Nécessité de nouvelles approches pour étudier la biodiversité.....</i>	17
 CHAPITRE II. ANALYSE DE L'ADN ENVIRONNEMENTAL PAR METABARCODING.....	21
1. L'ADN ENVIRONNEMENTAL : UN NOUVEL OUTIL POUR L'EVALUATION A HAUT-DEBIT DE LA BIODIVERSITE	21
a. <i>Barcoding classique</i>	21
b. <i>Metabarcoding environnemental</i>	22
c. <i>De multiples applications en écologie</i>	24
2. PROTOCOLE TYPE ET PARTICULARITES DU METABARCODING ENVIRONNEMENTAL.....	25
a. <i>Echantillonnage et extraction d'ADN.....</i>	25
b. <i>Amplification et séquençage haut-débit</i>	27
c. <i>Filtrage des données et assignation taxonomique.....</i>	29
3. DEMARCHE SCIENTIFIQUE ET STRUCTURE DE LA THESE	31
 CHAPITRE III. EVALUATION DE L'IMPACT DES ACTIVITES ANTHROPIQUES SUR LA BIODIVERSITE DES SOLS PAR METABARCODING ENVIRONNEMENTAL.....	33
1. PROBLEMATIQUE GENERALE.....	33
2. ETUDE DE LA DIVERSITE FONGIQUE DES SOLS DANS LES ILES SUB-ANTARCTIQUES EN LIEN AVEC L'INTRODUCTION D'HERBIVORES EXOTIQUES	35
a. <i>Mise en contexte de l'étude</i>	35
b. <i>Article A</i>	37
c. <i>Matériel et méthodes complémentaires</i>	48
d. <i>Résultats complémentaires.....</i>	55
3. DETERMINATION DES PATRONS DE DISTRIBUTION DES LOMBRICIENS A L'ECHELLE DU PAYSAGE A PARTIR DE L'ADN DU SOL.....	71
a. <i>Mise en contexte de l'étude</i>	71
b. <i>Article B.....</i>	73
c. <i>Matériel et méthodes complémentaires</i>	83
d. <i>Résultats complémentaires.....</i>	90
4. DISCUSSION CHAPITRE III	97
a. <i>Effets des perturbations anthropiques sur la biodiversité des sols.....</i>	97
b. <i>Des effets médiés par les plantes</i>	98
c. <i>Résilience des communautés</i>	99

CHAPITRE IV. UNE APPROCHE PALEOECOLOGIQUE BASEE SUR LE METABARCODING ENVIRONNEMENTAL POUR L'EVALUATION A LONG TERME DE L'IMPACT DES ACTIVITES HUMAINES..... 105

1. PROBLEMATIQUE GENERALE	105
<i>a. De l'importance des données paléocéologiques pour l'étude des perturbations anthropiques</i>	105
<i>b. Détection des perturbations passées</i>	106
<i>c. Reconstruction de la paléobiodiversité</i>	107
<i>d. L'outil ADN en paléocéologie</i>	108
<i>e. Mise en contexte des études</i>	109
2. HISTOIRE DU PASTORALISME ET MISE EN PLACE DES PAYSAGES SOUS L'EFFET DE L'HOMME REVELEES A PARTIR DE L'ADN SEDIMENTAIRE LACUSTRE	113
<i>a. Article C</i>	113
<i>b. Données complémentaires</i>	133
3. IMPACTS A LONG TERME DES ACTIVITES HUMAINES SUR LES COMMUNAUTES VEGETALES : UNE APPROCHE ECOLOGIQUE BASEE SUR L'ADN SEDIMENTAIRE LACUSTRE	149
<i>a. Article D</i>	149
<i>b. Données complémentaires</i>	174
4. DISCUSSION CHAPITRE IV	189
<i>a. Effets à long terme des activités anthropiques en milieu alpin</i>	189
<i>b. Le metabarcoding : un outil complémentaire pour l'étude des paléoenvironnements...</i>	190
<i>c. ... mais un outil encore développement</i>	192

CHAPITRE V. CONSIDERATIONS METHODOLOGIQUES POUR UNE MEILLEURE EVALUATION DE L'IMPACT DES ACTIVITES ANTHROPIQUES SUR LA BIODIVERSITE PAR METABARCODING ENVIRONNEMENTAL 195

1. DISCUSSION AUTOUR DE LA METHODE	195
<i>a. L'échantillonnage : une étape cruciale et adaptable à de nombreuses investigations</i>	195
<i>b. Rémanence de l'ADN et évaluation à court terme des impacts</i>	197
<i>c. Patrons de présence/absence et estimation de richesse</i>	198
<i>d. Une approche quantitative ?</i>	200
2. EN PERSPECTIVES : RESOLUTION TAXONOMIQUE, BASES DE REFERENCES ET APPROCHE « SANS PCR». 203	

CONCLUSION ET PERSPECTIVES..... 205

REFERENCES..... 209

ANNEXE 235

RESUME..... 257

ABSTRACT 258

Avant-propos

Cette thèse s'inscrit dans le cadre de l'écologie des perturbations. Elle vise à mieux comprendre l'impact des activités humaines sur les communautés biologiques et la résilience de ces dernières. Une large gamme de réponses est généralement observée entre groupes taxonomiques, or, les données accumulées jusqu'à aujourd'hui dérivent d'un nombre limité de taxons (Gardner *et al.* 2009). Comme rapporté, un peu durement, par Irwin et ses collaborateurs (2010), nous en sommes toujours à agréger des cas d'études particuliers. Ce manuscrit ne dérive pas de ce postulat mais a pour objectif d'appliquer une nouvelle méthodologie d'évaluation de la biodiversité afin d'améliorer la compréhension des mécanismes impliqués dans la réponse des communautés.

L'utilisation de l'ADN environnemental couplée aux technologies de séquençage à haut-débit est une approche en plein essor pour caractériser les communautés biologiques présentes et passées (Taberlet *et al.* 2012a, Bohmann *et al.* 2014). Les données obtenues permettent d'accéder à l'ensemble de la diversité du vivant (Bik *et al.* 2012), en particulier les fractions généralement « oubliées », faute d'outils adaptés. Cette approche offre aujourd'hui des perspectives inédites pour étudier l'impact des activités anthropiques sur la biodiversité à travers différentes échelles spatiales et temporelles.

Les travaux présentés ici ont été en partie réalisés dans le cadre du projet ANR METABAR qui regroupe également l'Université de Toulouse II et l'Université des Antilles et de la Guyane, dont l'objectif est de développer des protocoles novateurs pour standardiser et automatiser l'acquisition à haut débit de données sur la biodiversité à partir de l'ADN extracellulaire du sol. Ces travaux se divisent en deux parties distinctes.

Dans la première, je me suis intéressé aux changements induits par les activités humaines récentes (de l'ordre de quelques décennies) sur plusieurs communautés du sol, ainsi qu'aux facteurs influençant leurs réponses, dans trois milieux contrastés que sont les Iles Sub-Antarctiques, les Alpes françaises et la forêt tropicale guyanaise. La seconde étude a été réalisée en étroite collaboration avec l'équipe « Ecosystèmes Montagnards » de Irstea de Grenoble tandis que la dernière, encore en cours, est partie intégrante du projet METABAR cité ci-dessus.

La seconde partie offre un changement de perspective temporelle puisqu'elle concerne l'impact à long terme des activités anthropiques sur les milieux de montagne, et, notamment sur les trajectoires des communautés végétales au cours de l'Holocène.

La préservation de l'ADN dans les archives sédimentaires permet de reconstruire à la fois l'histoire du pastoralisme mais également ses effets sur les communautés locales. Ces études ont été rendues possibles grâce au partenariat entre le LECA et le laboratoire EDYTEM de Chambéry dans le cadre du Dispositif de Partenariat en Ecologie et Environnement (DIPEE).

Les travaux présentés ici se concentrent sur la composante taxonomique de la biodiversité. Il faut cependant noter que toutes les autres dimensions de la biodiversité (*e.g.* génétique, phylogénétique, fonctionnelle...) impliquées dans le fonctionnement global des écosystèmes et leurs réponses face aux changements (Naeem *et al.* 2012) sont également affectées par les activités humaines. Nous discuterons au final de l'opportunité de cette approche pour la compréhension des facteurs déterminant les trajectoires de recouvrement et l'évaluation des impacts induits par les perturbations anthropiques sur les communautés biologiques et les écosystèmes dans leur ensemble.

Chapitre I. Introduction

1. Activités humaines et biodiversité à l'échelle du globe

a. Une anthropisation généralisée

L'empreinte humaine sur les écosystèmes est croissante et généralisée (Vitousek *et al.* 1997). Quantifier son ampleur est délicat et dépend principalement de la définition qui en est donnée. L'anthropisation des biomes concernerait néanmoins aujourd'hui plus de 75% de la surface non glacée de la planète (Chown 2010). L'Homme est ainsi devenu, au même titre que le climat et les phénomènes géologiques, un des facteurs majeurs affectant la biosphère et ses processus (Ellis & Ramankutty 2008). En géologie, le concept d'Anthropocène définit une nouvelle ère au cours de laquelle les activités humaines sont devenues l'un des principaux moteurs des processus géologiques à grande échelle (*e.g.* dynamiques atmosphériques, érosion...) (Crutzen 2002). Le début de cette période est sujet à débat (*e.g.* Zalasiewicz *et al.* 2011; Steffen *et al.* 2011; Ruddiman 2013) et dépend principalement de l'interprétation qui en est faite et des phénomènes étudiés. Ce concept a ainsi été, par la suite, étendu à la sphère biotique (Ellis 2011) et il est de plus en plus employé en écologie (*e.g.* Smol & Douglas 2007; Schimel *et al.* 2013; Mielke *et al.* 2014 ; Kueffer & Kaiser-Bunbury 2014). Il caractérise les effets directs, mais également indirects, des populations humaines et de leurs activités sur les patrons écologiques et le fonctionnement des écosystèmes (Ellis 2011). Ces activités, de natures très différentes, ont des impacts aussi bien locaux (urbanisation et transformation des terres, surexploitation des ressources, espèces invasives, déforestation...) que globaux (changements climatiques, altération des cycles biogéochimiques, transfert de polluants à grande échelle...) (Fig. 1) ; de telle manière qu'aucun écosystème terrestre n'est libre d'influence humaine (Vitousek *et al.* 1997; Wright 2005). Si cette pression anthropique sur les systèmes biologiques n'est pas nouvelle (Willis *et al.* 2004, Heckenberger *et al.* 2003), elle s'est considérablement accrue depuis deux siècles suite au développement industriel, à la mondialisation et l'accroissement de la population humaine (Gardner *et al.* 2010; Baird & Hajibabaei 2012; Lenzen *et al.* 2012).

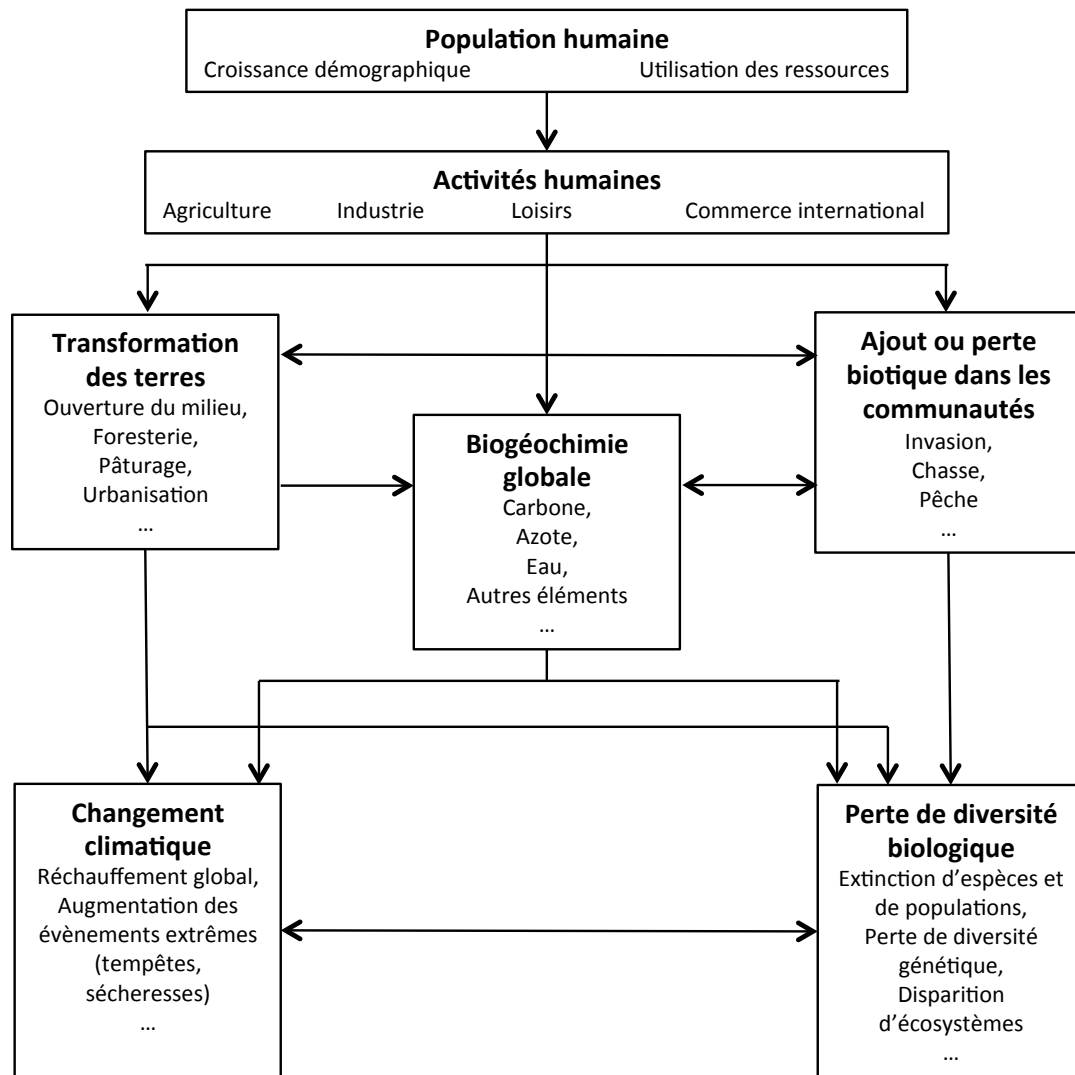


Figure 1. Schéma conceptuel illustrant les effets directs et indirects des activités humaines sur les écosystèmes (d'après Vitousek et al. 1997)

b. Une perte globale de biodiversité

Un des changements d'origines anthropiques irréversibles concerne la perte d'espèces (Chapin *et al.* 2000). L'importance du déclin actuel de biodiversité, souvent qualifié de 6^{ème} crise d'extinction majeure, est estimée 100 à 1000 fois supérieur aux grandes crises d'extinctions passées (*e.g.* Millenium Ecosystem Assessment 2005 ; Lenzen *et al.* 2012). De fait, même les zones reculées, où les impacts directs sont relativement limités, ont connu des changements dans leur biodiversité (Chown 2010), le réchauffement et la modification des régimes de précipitations, modifiant les aires de répartition des espèces (Thuiller 2007). La conversion des terres pour l'expansion des aires urbaines, de l'agriculture ou encore du pastoralisme a profondément modifié la

matrice paysagère et mené à la perte d'habitats de nombreuses espèces (Chapin *et al.* 2000; Myers & Knoll 2001 ; Brooks *et al.* 2002). La forêt tropicale a ainsi régressé de 50% en un siècle et a perdu 1,4% de sa surface entre 2000 et 2005 (Wright 2005; Asner *et al.* 2005). La majeure partie de la surface restante est exploitée (prélèvement sélectif, chasse...) et les zones non perturbées sont le plus souvent fragmentées (Wright 2005; Morris 2010). L'exploitation des ressources biotiques (chasse, pêche...) a, *a priori*, un impact plus limité puisqu'elle n'affecte directement que quelques taxons particuliers. Cependant, elle engendre des perturbations indirectes sur les autres espèces et peut créer des déséquilibres dans l'écosystème en modifiant les relations inter-spécifiques (Harrison *et al.* 2013) ou en entraînant la perte d'un groupe fonctionnel particulier, l'exemple le plus connu étant celui des grands prédateurs (Duffy 2003).

Les flux de personnes et de marchandises permettent aujourd'hui à certaines espèces de franchir les barrières biogéographiques et d'être introduites dans de nouvelles aires (Lenzen *et al.* 2012). Leur présence peut potentiellement perturber les écosystèmes et leurs dynamiques (ultra-dominance, extinction d'espèces, introduction de maladies et parasites, altération de l'environnement abiotique...) et fait aujourd'hui de ce phénomène la seconde cause de diminution de la biodiversité derrière la perte d'habitats (Mack *et al.* 2000). Toutes ces activités ont également un impact indirect *via* les changements qu'elles induisent dans le régime des perturbations naturelles (Dale *et al.* 2001 ; Barlow & Peres 2004). L'érosion des sols va, par exemple, entraîner une augmentation des glissements de terrain (Mack & d'Antonio 1998) alors que l'exploitation des forêts tend à accroître leur sensibilité aux feux (Nepstad *et al.* 1999). De même, les changements climatiques sont supposés être la cause d'une augmentation de la fréquence et de l'intensité des perturbations naturelles (tempêtes, feux, phénomènes El Niño...) (Overpeck *et al.* 1990 ; Vendermeer *et al.* 2000).

2. Les perturbations anthropiques à l'échelle locale

A une échelle plus locale, l'importance des perturbations anthropiques sur les écosystèmes et leur fonctionnement est également largement reconnue (Mack *et al.* 2011 ; McLauchlan *et al.* 2014). Les modifications qu'elles engendrent dans la structure biotique et abiotique des écosystèmes entraînent une altération du fonctionnement de ces derniers. Néanmoins, les mécanismes régissant la réponse des différentes

communautés biologiques à l'échelle locale sont encore souvent relativement mal déterminés (Wardle *et al.* 2004).

a. Définitions des perturbations

Le terme de « perturbation » en écologie regroupe une grande variété de phénomènes biotiques et abiotiques, aussi bien naturels (chutes d'arbres en forêt, feux, tempêtes, sécheresses, herbivorie, prédation...) qu'anthropiques (prélèvements, coupes à blanc, pâturages, feux, fragmentation de l'habitat...) (Sousa 1984 ; Rykiel 1985). Cependant, en donner une définition claire n'est pas chose aisée, celle-ci dépendant grandement du point de vue des auteurs et des études concernées (Rykiel 1985 ; White & Jentsch 2001; Cordonnier 2004). Une définition communément admise considère une perturbation comme un événement discret dans le temps qui va perturber l'écosystème, la communauté ou la structure des populations (par perte de biomasse notamment) et modifier la disponibilité des ressources et du substrat, ou encore l'environnement physique. Cela aura pour conséquence de libérer des ressources pour les autres organismes et de créer une opportunité pour l'installation de nouveaux individus (Sousa 1984 ; White & Pickett 1985). Les perturbations opèrent à différentes échelles spatiales et temporelles et vont entraîner des modifications dans la diversité et la composition des communautés et, par conséquent, dans la structure et le fonctionnement des écosystèmes (cf. encadré 1 ; Sousa 1984 ; Foster *et al.* 1998). Elles sont ainsi des phénomènes fondamentaux pour comprendre la structure et la dynamique des écosystèmes (Rykiel 1985; White & Jensch 2001).

En plus des différences dans leur nature, les perturbations se caractérisent par leur fréquence, leur étendue, leur intensité et leur durée (Shea *et al.* 2004). Ainsi, certaines activités humaines telles que les feux ou les coupes à blanc ont des caractéristiques semblables aux perturbations naturelles dans le sens où elles sont localisées et instantanées (White & Jentsch 2001). En revanche, une grande partie d'entre elles se distinguent par leur durée et leur capacité à affecter de grandes surfaces comme l'utilisation des terres. Il est souvent d'usage de distinguer, en fonction de la durée, la perturbation du stress : la perturbation étant considérée comme un événement soudain et localisé (*e.g.* chute d'arbres) alors que le stress a un effet permanent et éventuellement plus diffus (Dornelas 2010 ; Mackey & Currie 2001)

Encadré 1 : Activités humaines, biodiversité et fonctionnement de l'écosystème

Parce qu'elles engendrent des modifications à la fois dans l'environnement abiotique et les communautés biologiques, les activités humaines influencent les patrons spatiaux des processus écosystémiques (Turner 2010). La richesse et les caractéristiques fonctionnelles des espèces sont des paramètres primordiaux du fonctionnement des écosystèmes (Hector *et al.* 1999 ; Hooper *et al.* 2005 ; Naeem *et al.* 2012). Ils sont notamment impliqués dans la rétention des nutriments (Tilman *et al.* 2004), la productivité (Tilman *et al.* 1997, Tilman *et al.* 2001, Spehn *et al.* 2005), la résistance aux invasions (Naeem *et al.* 2000 ; Levine 2000 ; Kennedy *et al.* 2002), la stabilité des écosystèmes (*e.g.* Stachowicz *et al.* 2002 ; Loreau *et al.* 2003 ; Ives & Carpenter 2007), la résistance aux maladies (Knops *et al.* 1999 ; Keesing *et al.* 2006) ou encore les cycles biogéochimiques (flux d'énergie et cycles des nutriments). De nombreuses études ont montré le lien existant entre la diversité et le fonctionnement des écosystèmes, l'appauvrissement de la diversité entraînant une diminution de la redondance fonctionnelle pouvant mener à une perte de fonctionnalité (*e.g.* Naeem 1994 ; Tilman *et al.* 1997 ; Tilman *et al.* 2001 ; Tilman *et al.* 2004, Hector *et al.* 1999 ; Spehn *et al.* 2005). Récemment, Wagg *et al.* (2014) ont montré la corrélation existant entre la biodiversité des sols et de ses communautés et la multifonctionnalité de l'écosystème. La diminution expérimentale de biodiversité et les changements de composition associés ont notamment été corrélés à un déclin de multiples fonctions et notamment la diversité des plantes, le recyclage des nutriments et leur rétention. L'amplitude des effets dépend en grande partie de l'identité des espèces perdues (Cardinale *et al.* 2006). Une plus grande biodiversité stabilise également l'écosystème et ses propriétés. Les espèces impliquées dans un processus répondant différemment aux perturbations permettent de stabiliser le processus dans son ensemble. Ainsi, les changements induits par les perturbations anthropiques dans les communautés sont à même d'altérer le fonctionnement de l'écosystème dans son ensemble, et par là même les biens et services qu'ils peuvent rendre.

Pour certaines pratiques comme le pâturage, il est clair que des différences dans la durée (d'une journée à plusieurs décennies) vont induire des effets différents et, par conséquent, des réponses différentes. Ces activités humaines peuvent ainsi être vues comme des perturbations anthropiques persistantes (Foster *et al.* 1998 ; MacDougall *et al.* 2013). On adopte alors la définition plus synthétique et plus intégrative donnée par Shea *et al.* (2004) qui définit la perturbation comme un évènement modifiant les opportunités de niches disponibles dans le système. Tout évènement altérant la disponibilité des ressources, sans nécessairement provoquer de pertes directes de biomasse et quelques soit sa durée, peut être considéré, dans cette acceptation, comme

une perturbation. C'est cette dernière définition qui a été retenue dans le cadre de ces travaux.

b. De l'importance des perturbations en écologie

Les premiers travaux sur les perturbations ont profondément changé la notion d'« équilibre » en écologie (Turner *et al.* 1993 ; Turner 2010). En effet, l'acceptation de l'état de non-équilibre des écosystèmes signifie la reconnaissance du fait que les systèmes écologiques sont généralement dans des états de recouvrement depuis la dernière perturbation (Wu & Loucks 1995 ; Turner *et al.* 2003). Historiquement, le cœur des travaux reliant les perturbations à la biodiversité est lié à l'hypothèse des perturbations intermédiaires (« *Intermediate Disturbance Hypothesis* » ou IDH en anglais ; Connell 1978). Celle-ci suggère que la diversité est maximisée dans les sites perturbés à des échelles spatiales et temporelles intermédiaires, les perturbations naturelles permettant alors de limiter la monopolisation des ressources par les espèces les plus compétitrices (cf. encadré 2). Bien que cette hypothèse ait été validée par plusieurs études, par ailleurs sur un nombre restreint de groupes taxonomiques (cf. Shea *et al.* 2004 pour une revue complète), la relation observée *in situ* entre perturbation et diversité n'est pas toujours constante ni courbée dans le sens de l>IDH (*e.g.* Mackey & Currie 2001) suggérant que d'autres paramètres, parfois plus importants, entrent en jeu. Pour exemple, les niches vacantes vont être occupées soit par les espèces encore présentes soit par des espèces environnantes (Connell 1978). La limitation de la dispersion (Hubbell 1979 ; Hubbell *et al.* 1999) restreint cependant le pool régional de taxons candidats. L'habitat environnant peut alors soit améliorer (s'il est favorable) soit exacerber (s'il est hostile) l'impact de la perturbation sur la diversité locale (Gibson *et al.* 2011).

A plus grande échelle, les perturbations sont une des causes primaires de l'hétérogénéité spatiale (Platt 1975 ; Ettema & Wardle 2002). En effet, la répartition spatiale de la biodiversité est contrôlée à la fois par des facteurs environnementaux (*e.g.* climat, caractéristiques physico-chimiques des sols...), par les relations intra et interspécifiques (*e.g.* reproduction/mortalité, prédation, compétition, mutualisme...) mais également par les perturbations (Ettema & Wardle 2002).

Encadré 2 : Hypothèse des perturbations intermédiaires

L'hypothèse des perturbations intermédiaires est, à l'origine, un des mécanismes proposés par Connell (1978) pour expliquer le maintien de la richesse spécifique dans les milieux hyper-diversifiés tels que la forêt tropicale ou encore les écosystèmes coralliens. Les perturbations d'une certaine fréquence, intensité, étendue et durée empêchent la monopolisation des ressources par une ou quelques espèces. Celles-ci ne sont pas complètement destructives, de telle manière que d'autres espèces peuvent exploiter les ressources nouvellement disponibles. Elles maintiennent ainsi les communautés dans un état de non-équilibre permanent qui empêche l'établissement du processus d'exclusion compétitive (Gause 1934) qui éliminerait ultimement les espèces les moins compétitrices ou les moins adaptées. D'après cette hypothèse, la coexistence et/ou les pics de biodiversité sont ainsi maximisés à des niveaux de perturbations intermédiaires. Le compromis entre compétition et colonisation est un des principaux mécanismes impliqués (Shea *et al.* 2004).

Un exemple simple est fourni par Connell (1978) : après une perturbation (*e.g.* tempête), certains arbres d'une forêt sont tués et une trouée dans la canopée apparaît. Dans un premier temps, les graines de quelques espèces vont être capables de coloniser cette zone ; par conséquent la diversité sera faible directement après la perturbation. Si la perturbation est trop fréquente, la communauté consistera uniquement en ces quelques espèces capables de produire des propagules en atteignant rapidement la maturité. Cependant, si le temps est suffisamment long entre les perturbations, les espèces, avec un pouvoir de dispersion moindre ou avec une croissance plus grande, vont pouvoir s'installer, ce qui augmentera la diversité. En revanche, si la fréquence des perturbations décline trop, la diversité diminuera car les meilleurs compétiteurs monopoliseront les ressources. De même, une perturbation trop intense (ou trop longue) va éliminer toutes les espèces présentes et la recolonisation ne se fera qu'à partir d'organismes extérieurs, alors qu'une perturbation modérée va également permettre la recolonisation par les survivants. A l'inverse, si la perturbation est trop faible (ou trop courte), les bons compétiteurs vont continuer à dominer la communauté. Enfin, si l'étendue est trop faible, l'espace perturbé sera limité et potentiellement non-envahi. Cependant, si la perturbation est trop large, seules quelques espèces seront capables de disperser et d'entrer en compétition sur de grandes surfaces (Shea *et al.* 2004).

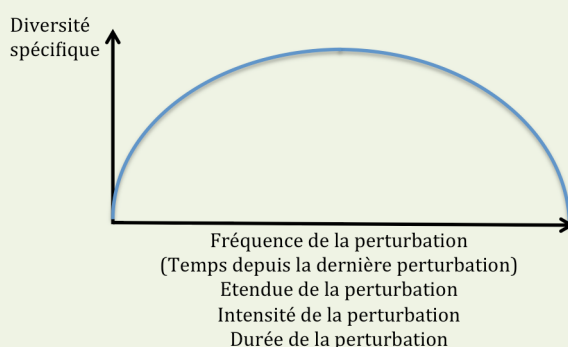


Figure 2. Relation suggérée par l'hypothèse des perturbations intermédiaires (Connell 1978) entre la diversité spécifique et les caractéristiques des perturbations

Par les changements, à différentes échelles spatiales et temporelles, qu'elles vont induire sur les communautés, les ressources et l'environnement abiotique, les perturbations vont contribuer à la création d'une mosaïque d'habitats et à l'hétérogénéité spatiale. En forêt, les perturbations asynchrones créent ainsi des patchs à différents stades de succession, présentant des compositions différentes, ce qui permet de maintenir de plus hauts niveaux de diversité (*e.g.* Molino & Sabatier 2001).

c. Le cas des activités humaines

Les effets des perturbations anthropiques sur la biodiversité sont nombreux et complexes. Certaines pratiques traditionnelles humaines (*e.g.* brûlis, pâturage, prélèvement) favorisent l'hétérogénéité spatiale en permettant le maintien d'une matrice paysagère diversifiée. En milieu de montagne, le pâturage permet par exemple le maintien de milieu ouvert (White & Jentsch 2001 ; Young *et al.* 2005). D'autres activités humaines, à l'instar de certaines catastrophes naturelles (*e.g.* ouragan, tempête), ont la particularité d'affecter de vastes surfaces. Elles se différencient cependant de ces dernières par leur fréquence et/ou leur durée. Ces perturbations anthropiques vont alors imposer des contraintes homogènes, parfois sur de grandes surfaces, qui vont tendre à favoriser un sous-échantillon de la diversité régionale (Foster *et al.* 1998). La diversité beta, qui correspond aux variations de compositions des communautés entre différents sites (cf. encadré 3), s'en trouve alors réduite. Plusieurs études ont montré un phénomène d'homogénéisation biotique à la suite des perturbations humaines à grande échelle (*e.g.* Foster *et al.* 1998 ; Tittensor *et al.* 2007), la conversion des terres étant un exemple typique. Une étude récente en forêt tropicale a ainsi mis en évidence une homogénéisation des communautés bactériennes du sol suite à la conversion de la forêt pour l'agriculture (Rodrigues *et al.* 2013). Un autre patron typiquement observé est le déclin des espèces endémiques, avec des aires de répartition restreintes, ou spécialistes d'un habitat particulier (Lewis 2009). Comme nous l'avons vu auparavant, les espèces qui vont bénéficier des niches nouvellement disponibles proviennent d'un pool restreint et sont souvent des espèces généralistes (Bongers *et al.* 2009 ; Irwin *et al.* 2010). Les perturbations anthropiques favorisent également l'introduction et le maintien d'espèces exotiques dans le milieu (Brown & Gurevitch 2004 ; Irwin *et al.* 2010).

Encadré 3: Mesures de diversité taxonomique

La diversité taxonomique peut être envisagée de différentes manières et de nombreuses mesures existent pour la caractériser. En premier lieu, il convient de distinguer deux composantes de la diversité que sont la richesse et l'équitabilité (ou « *evenness* » en anglais) (Whittaker 1965 ; Magurran 2004). La richesse correspond au nombre de taxons observés (e.g. nombre d'espèces) alors que l'équitabilité se rapporte à l'abondance relative des espèces dans le milieu, une espèce abondante ou représentée par un seul individu n'apportant pas la même contribution. Tandis que le premier nous renseigne sur l'étendue de la répartition des ressources entre espèces, le second fournit des informations sur la manière dont les espèces interagissent pour se partager ces ressources (Whittaker 1972). En second lieu, il est important de noter que les mesures de diversité étant comparatives et non absolues, elles se réfèrent à des unités spatialement et temporellement organisées. Plusieurs niveaux d'organisation sont ainsi couramment reconnus : la diversité alpha (α), la diversité beta (β) et la diversité gamma (γ) (Whittaker 1972 ; Koleff *et al.* 2003 ; Magurran 2004 ; Zinger *et al.* 2012).

La diversité alpha (ou diversité locale) se réfère à la diversité au sein d'un site ou d'un échantillon. Elle est classiquement mesurée par le nombre de taxons observés (e.g. nombre d'espèces, indice de Chao) mais d'autres indices se basent partiellement (e.g. indice de Shannon, indice de Simpson) ou complètement (e.g. indice de Pielou, indice de Berger-Parker) sur l'abondance relative des taxons (Magurran 2004).

La diversité beta (également appelée « *turnover* ») est une mesure de la différence de composition des communautés entre deux ou plusieurs sites ou échantillons (Zinger *et al.* 2012 ; Koleff *et al.* 2003 ; Myers *et al.* 2013). Elle peut se mesurer soit en utilisant uniquement les données présence/absence (e.g. indices de similarité de Jaccard ou de Sorensen) soit en prenant également en compte l'abondance relative des taxons (e.g. indices de dissimilarité de Bray-Curtis) (Ramette 2007 ; Anderson *et al.* 2011 ; Zinger *et al.* 2012). Elle est souvent utilisée pour caractériser les réponses, d'un site à l'autre, des communautés face aux influences anthropiques, aux changements climatiques ou encore aux gradients environnementaux (Myers *et al.* 2013).

La diversité gamma (ou diversité régionale) est similaire à la diversité alpha mais concerne une zone plus large regroupant l'ensemble des unités étudiées. Elle englobe le pool d'espèces potentiellement présent dans les sites étudiés (Whittaker 1972).

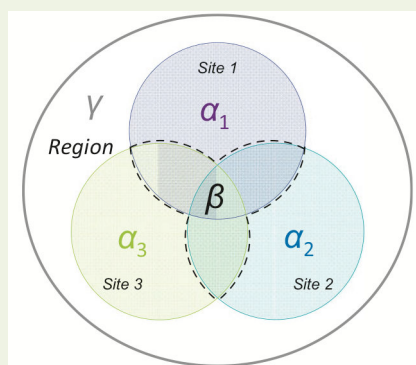


Figure 3. Schéma représentatif des diversités alpha (cercles colorés), beta (pointillés) et gamma (cercle gris) (issu de Zinger *et al.* 2012)

Dans une certaine mesure, celles-ci peuvent contribuer à augmenter la richesse au niveau local mais vont également amplifier le phénomène d'homogénéisation biotique (Winter *et al.* 2009).

Dans l'ensemble, une modification des niveaux de diversité locale et régionale est fréquemment observée mais ces observations ne sont pas uni-directionnelles et dépendent grandement du type de perturbations et des taxons concernés (Irwin *et al.* 2010). L'inconstance de la relation entre diversité et perturbation (Mackey & Currie 2001) en est une preuve flagrante. En forêt tropicale, Lawton *et al.* (1998) ont montré une diminution globale de la diversité de plusieurs groupes taxonomiques avec l'intensification de la perturbation. Cependant certains d'entre eux se sont révélés insensibles aux perturbations extrêmes alors que d'autres montraient des pics de diversité dans les milieux non perturbés ou moyennement perturbés. De même, une faible congruence inter-taxons a été relevée dans la réponse des communautés aux changements d'utilisation des terres en forêt tropicale, particulièrement lorsqu'on étudie l'abondance et la richesse en espèces (Barlow *et al.* 2007). En revanche, les corrélations apparaissent plus fortes si l'on observe la composition ou la structure des communautés, suggérant qu'elles connaissent toutes des modifications qui les différencient des communautés originelles.

d. Facteurs déterminant la réponse des communautés face aux perturbations

Les perturbations anthropiques ne sont pas égales entre elles et leurs impacts varient dans une large gamme d'échelles spatiales et temporelles (Gibson *et al.* 2011). Aussi, les réponses des communautés aux perturbations sont tout aussi complexes (Turner *et al.* 2003) et varient grandement entre groupes taxonomiques. Celles-ci dépendent bien entendu des caractéristiques de la perturbation (intensité, fréquence, étendue et durée) mais également des facteurs abiotiques (*e.g.* conditions climatiques, paramètres du sol) et biotiques liés à la nature des espèces présentes (*e.g.* compétition, facilitation) contrôlant la composition et la dynamique des communautés (Everham & Brokaw 1996 ; Shea *et al.* 2004).

L'importance relative des facteurs abiotiques et biotiques régissant l'assemblage des communautés varie grandement selon les groupes taxonomiques étudiés et les zones géographiques (Drake 1991 ; Green *et al.* 2004). Une première évidence nous

vient de la biogéographie : alors que les patrons de diversité de plantes et d'animaux à l'échelle mondiale sont corrélés principalement à la latitude et à la température, les communautés bactériennes des sols sont contrôlées principalement par des variables édaphiques (Fierer & Jackson 2006). Les micro-organismes ne répondent pas de la même manière que les métazoaires ou les plantes aux gradients macro-écologiques. De même, à l'échelle locale, il n'existe pas d'évidence sur le fait que les facteurs régissant la composition des communautés soient identiques entre groupes taxonomiques ; aussi les réponses aux perturbations peuvent différer d'un groupe à l'autre (Decaëns 2010). Wardle *et al.* (2004) rapportent ainsi une grande variabilité de réponses entre les communautés du sol après l'apport de matière organique, certaines étant principalement dirigées par une régulation « top-down » (*e.g.* par les consommateurs) alors que d'autres ont une régulation de type « bottom-up » (*e.g.* dépendant de la qualité et de la quantité des ressources). Certains groupes taxonomiques vont montrer des réponses différentes aux perturbations suivant les éco-régions où ils se trouvent (Irwin *et al.* 2010) car les règles d'assemblages peuvent également varier d'un biome à l'autre (Gibson *et al.* 2011). En forêt tropicale, l'agrégation spatiale intraspécifique des arbres reflète par exemple une limitation de la dispersion alors qu'en forêt tempérée, où il y a moins d'espèces rares, l'agrégation des individus d'une même espèce répond plutôt à des gradients environnementaux (Myers *et al.* 2013).

Il existe également une part stochastique dans les réponses des communautés aux perturbations. Ces dernières induisent des changements dans la composition des communautés mais toutes les populations ne sont pas affectées de la même manière. Suite à la perturbation, les niches vacantes vont être occupées par les espèces encore en place dans le milieu mais également par celles alentour qui vont être capables de l'envahir. La limitation de la dispersion et du recrutement (Hubbell 1979, Hubbell *et al.* 1999) restreint le pool d'espèces à même d'occuper les niches. La mosaïque paysagère environnante joue alors un rôle très important dans la dynamique de recolonisation et peut entraîner de grandes variations entre sites. Les caractéristiques des espèces en présence et les relations inter-spécifiques dans lesquelles elles sont impliquées vont partiellement diriger les processus de succession (Chazdon *et al.* 2007). Ces changements ont en effet un impact indirect sur la structure des réseaux d'interactions entre espèces (Morris 2010). La perte d'une espèce induisant également une perte de ses interactions, ces relations indirectes sont particulièrement difficiles à prévoir

(Montoya *et al.* 2006). Ainsi, la disparition d'une espèce peut en entraîner d'autres ou, au contraire, être bénéfiques à certains taxons. Enfin, plusieurs études ont également montré l'importance de l'ordre d'arrivée des espèces dans le milieu pour la détermination des règles d'assemblages possibles (Robinson & Dickerson 1987 ; Robinson & Edgemon 1988 ; Drake 1991).

3. Des méthodes d'investigation limitées

a. Une couverture taxonomique limitée

Les réponses des différentes communautés aux perturbations induites par les activités humaines sont difficilement généralisables. La connaissance des facteurs influençant les réponses des communautés, et plus particulièrement leur résilience (cf. encadré 4) face aux perturbations anthropiques, apparaît comme un élément primordial pour le futur. Or, caractériser les changements de diversité induits par l'Homme à partir d'un nombre restreint de taxons est très discutable (Lindenmayer & Likens 2011).

On est encore loin de disposer d'un inventaire complet de la biodiversité (Magurran & Dornelas 2010). Le déficit taxonomique (*e.g.* le ratio entre le nombre de taxons estimés et le nombre de taxons décrits) est peu élevé pour les plantes terrestres et les vertébrés mais il est particulièrement accru chez la méso- (<2 mm) et la microbiote (<0.2mm) (Blaxter 2003). Ces biais sont principalement liés à des soucis d'ordre méthodologique qui limitent l'accès à une grande partie de la diversité biologique. Celle des sols, notamment, est souvent considérée comme une « boîte noire » en écologie. Cela induit *in fine* un fort biais taxonomique dans les études en conservation (Clark & May 2002 ; Lawler *et al.* 2006). Les plantes vasculaires et les vertébrés sont les groupes les plus largement étudiés alors même qu'ils ne représentent qu'une faible fraction de la diversité biologique (Fig. 4). Par conséquent, les modèles de biodiversité dérivent d'un échantillon restreint de groupes taxonomiques (mammifères, oiseaux, plantes vasculaires et quelques insectes charismatiques comme les papillons) (Gardner *et al.* 2009 ; Gardner *et al.* 2010).

Encadré 4 : Réponses des communautés et notion de « résilience » en écologie

Face à une perturbation, la gamme des réponses possibles d'un système va de la résistance à des changements irréversibles en passant par retour plus ou moins complet à l'état originale (Rydgren et al. 2004). La résistance et la résilience sont deux composantes primordiales de la réponse des systèmes face aux perturbations. La « résistance » est une mesure du degré de modification d'une variable (i.e. dans notre cas la composition de la communauté) à la suite d'une perturbation (Fig. 5a ; McCann et al. 2000 ; Allison & Martiny 2008). Elle peut alors être quantifiée *via* le taux de changement dans la composition en espèces (Rydgren et al. 2004).

Concernant la notion de « résilience », deux visions existent en écologie. La définition la plus classique, par ailleurs utilisée dans d'autres domaines, définit la résilience comme la capacité d'un système à retourner à un état d'équilibre (Pimm 1984 ; Gunderson 2000 ; Rydgren et al. 2004) et se mesure, dans le cas des communautés, par la vitesse de retour à l'état initial. Cette définition implique l'existence d'un seul état stable autour duquel le système varie (Fig. 5b ; Gunderson 2000) et est appelé « résilience au sens de l'ingénierie » par certains auteurs (Holling 1973 ; Holling 1996 ; Peterson et al. 1998). La seconde définition de la résilience se base sur l'idée que plusieurs états stables peuvent exister. Le domaine de stabilité regroupe l'ensemble des valeurs des variables du système à partir desquelles celui-ci peut revenir à un équilibre particulier. La résilience de la communauté se mesure alors à la magnitude de perturbations qui peut être absorbée par le système avant que celui-ci ne bascule d'un domaine de stabilité à un autre (Fig. 5c ; Holling 1996). Cette définition est également appelée « résilience écologique » (Holling 1973 ; Holling 1996 ; Peterson et al. 1998). Ces deux visions sont complémentaires et leur utilisation dépend, en premier lieu, de l'échelle d'étude des systèmes (Peterson *et al.* 1998). La résilience écologique est plus particulièrement adaptée à l'étude des macro-systèmes tandis que la résilience au sens de l'ingénieur est utilisée dans des systèmes plus "simples" comme les communautés (Cordonnier 2004). C'est donc cette dernière définition qui sera ici retenue.

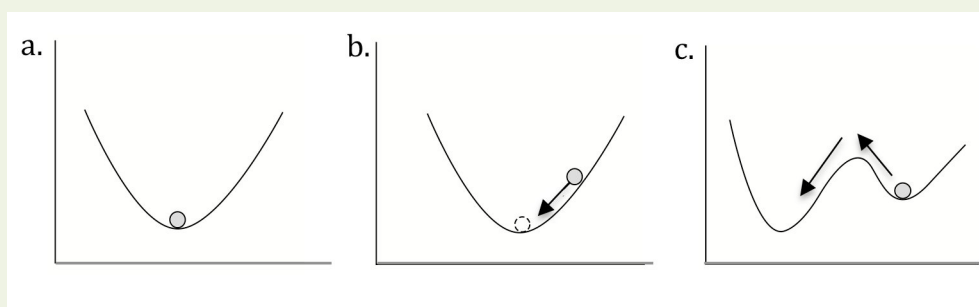


Figure 5. Schéma simplifié de la métaphore d'une balle et d'un paysage souvent utilisée pour caractériser les différences entre les deux types de résilience (e.g. Holling 1973 ; Peterson et al. 1998 ; Gunderson 2000 ; Nyström et al. 2000 ; Beisner et al. 2003). La balle représente le système (dans notre cas une communauté), une vallée représente son domaine de stabilité et les flèches représentent les perturbations. Dans le cas a, le système ne bouge pas, il est résistant. Dans le cas b (résilience « au sens de l'ingénieur »), un équilibre existe lorsque la balle se trouve au fond de la vallée mais les perturbations vont l'amener vers une position transitoire dans la vallée et la résilience se mesure alors par le temps de retour au fond de la vallée. La résilience écologique (cas c) suggère, elle, l'existence de plus d'une vallée, et la résilience est alors définie comme la largeur de la vallée (i.e. du domaine de stabilité) soit la magnitude de perturbations que le système peut supporter avant de basculer dans un autre domaine. Figure adaptée de Nyström et al. (2000)

b. Nécessité de nouvelles approches pour étudier la biodiversité

Les méthodes classiques de suivi de biodiversité sont restreintes à un faible nombre de groupes taxonomiques (Fig. 4) et présentent une faible concordance entre elles et entre écosystèmes (Baird & Hajibabaei 2012). Elles sont, de plus, souvent réalisées sur un nombre limité de petites surfaces, avec peu de réplicats. S'ensuit généralement une longue période d'énumération et d'assignation taxonomique des spécimens récoltés (ce qui peut durer plusieurs années). Les données obtenues sont parfois de qualité limitée et, dans la plupart des cas, invérifiables (Baird & Hajibabaei 2012). L'acquisition de données sur l'ensemble de la biodiversité à partir des approches taxonomiques classiques est inenvisageable de par les coûts financiers, le temps et l'expertise taxonomique requis (Chariton 2012 ; Baird & Hajibabaei 2012).

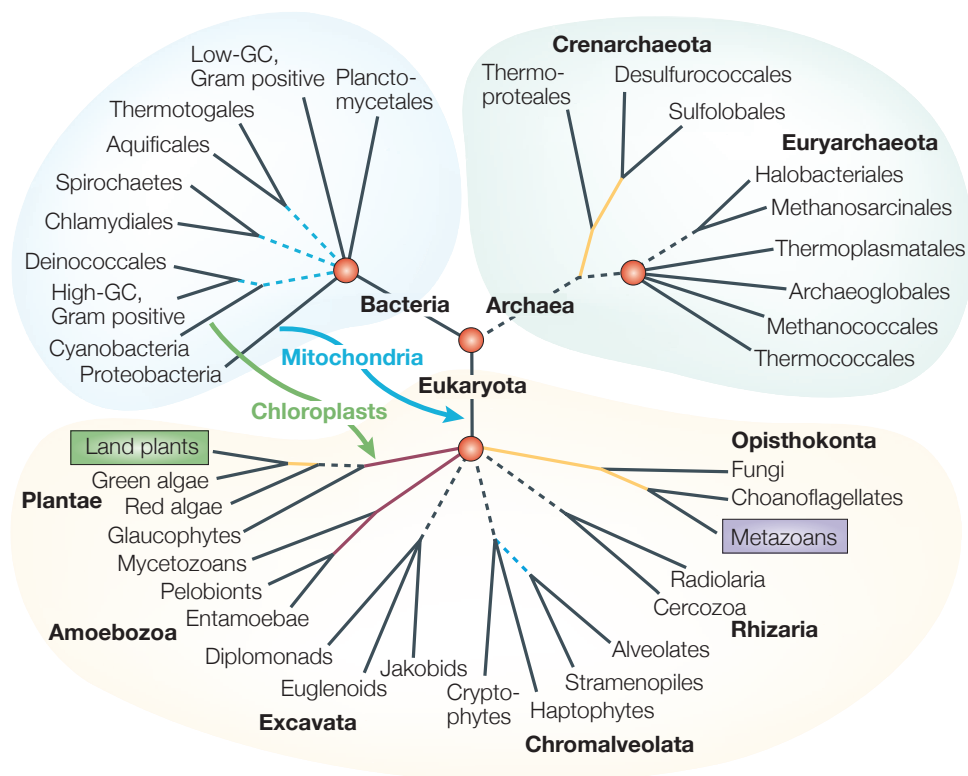


Figure 4. Arbre phylogénétique du vivant (Delsuc et al. 2005). En encadré, les deux groupes taxonomiques les plus étudiés, les Vertébrés représentant eux-mêmes une très faible fraction des Métazoaires.

L'évaluation de l'impact des activités anthropiques sur la biodiversité dans son ensemble requiert des mesures de biodiversité reproductibles sur différents groupes taxonomiques et à des échelles de temps et d'espaces pertinentes (Chariton 2012). Il y a ainsi une nécessité de développer des méthodes de collections de données permettant

de couvrir l'ensemble de la biodiversité, sur de larges échelles, dans une petite fenêtre de temps, avec la meilleure qualité possible et reposant sur une bonne taxonomie (Chiarucci *et al.* 2011). A défaut de pouvoir étudier l'ensemble de la biodiversité pour chaque étude de cas, cela pourrait permettre de caractériser de nouveaux taxons « indicateurs » de changement et de la « bonne santé » des écosystèmes.

Chapitre II. Analyse de l'ADN environnemental par *metabarcoding*

1. L'ADN environnemental : un nouvel outil pour l'évaluation à haut-débit de la biodiversité

L'utilisation d'approches moléculaires pour l'évaluation de la biodiversité offre aujourd'hui de nouvelles perspectives en écologie des perturbations (Poisot *et al.* 2013). L'approche proposée dans ce document de thèse, basée sur l'analyse de l'ADN environnemental et appelée *metabarcoding* environnemental, dérive en grande partie du *barcoding* ADN classique.

a. Barcoding classique

L'utilisation de courtes séquences d'ADN est, depuis une dizaine d'années, largement utilisée pour différencier des spécimens d'animaux, de plantes, de champignons et autres procaryotes (Ji *et al.* 2013). Au sens strict, le *barcoding* (« code-barre » en français) désigne l'utilisation d'une région standardisée de l'ADN pour l'identification rapide et précise d'espèces à partir de spécimens récoltés (Hebert *et al.* 2003a) et a, de ce fait, largement été utilisé en taxonomie (*e.g.* Hebert *et al.* 2004 ; Rougerie *et al.* 2009 ; Puillandre *et al.* 2012). Ces marqueurs génétiques requièrent, entre autres, certaines caractéristiques (Hebert *et al.* 2003a ; Riaz *et al.* 2012): (i) pas ou peu de variation intra-spécifique, (ii) avoir un pouvoir discriminant important, soit une variation inter-spécifique relativement élevée et être, si possible, phylogénétiquement informatif, (iii) être encadré de régions flanquantes conservées pour l'ensemble du groupe taxonomique ciblé, qui vont servir de sites de fixations pour les amorces permettant une amplification des séquences par PCR (Polymerase Chain Reaction, Mullis *et al.* 1986). De nombreux code-barres ont ainsi été proposés à travers le consortium international CBoL (« Consortium for Barcode of Life ») : COI (Cytochrome c Oxydase I, 658 paires de bases (pb)) pour les animaux (Hebert *et al.* 2003a ; Hebert *et al.* 2003b), ITS1 (Internal Transcribed Spacer, 450-700 pb) pour les champignons (Bellemain *et al.* 2010 ; Schoch *et al.* 2012) ou encore la combinaison des marqueurs *rbcL* (ribulose 1,5-bisphosphate carboxylase) et *matK* (maturase K) pour les plantes (respectivement 553 pb et 879 pb, CBoL Plant Working Group 2009). De larges campagnes d'archivages de séquences ADN issus de ressources vivantes ou de

spécimens de musées ont également été lancées de par le monde, notamment dans le cadre de la base de données BOLD (« *Barcode of Life Database* », Ratnasingham & Hebert 2007). La définition du *barcoding* peut cependant être étendue, dans un sens plus large, à l'identification de n'importe quel niveau taxonomique (*e.g.* espèce, genre, famille) par l'utilisation de fragments d'ADN quels qu'il soient (Valentini *et al.* 2009a).

b. Metabarcoding environnemental

Les études en écologie, et *a fortiori* celles portant sur les effets des activités anthropiques, nécessitent l'établissement d'une liste des taxons présents dans le milieu. Si l'avènement du *barcoding* a permis de simplifier et de standardiser l'aspect taxonomique, il n'a pas contribué à réduire l'effort d'échantillonnage (Coissac *et al.* 2012). L'évolution des technologies de séquences à haut-débit (communément appelées HTS pour « High-Throughput sequencing » ou encore NGS pour « Next-Generation Sequencing ») a récemment permis le développement d'approches rapides et innovantes pour l'évaluation de la biodiversité basées sur le principe du *barcoding* (*e.g.* Valentini *et al.* 2009a ; Chariton *et al.* 2010 ; Taberlet *et al.* 2012a ; Baird & Hajibabaei 2012). Les séquenceurs capillaires de type Sanger, utilisés à l'origine en *barcoding* classique, ne permettent de séquencer qu'un seul gène provenant d'un unique individu par échantillon, avec un débit typique de 96 séquences par expérimentation (Kircher & Kelso 2010). A l'inverse, les séquenceurs haut-débit (*e.g.* technologies Illumina, Ion Torrent ou PacBio) peuvent fournir plusieurs millions de séquences par expérimentation (*e.g.* jusqu'à 300 millions par cellule de séquençage pour la technologie Illumina HiSeq) et présentent la particularité de séparer individuellement chaque molécule d'ADN présente dans un échantillon afin de les séquencer en parallèle, permettant le séquençage massif et simultané d'échantillons d'ADN complexes, *e.g.* contenant un mélange de plusieurs gènes provenant d'individus et d'espèces différentes (Shokralla *et al.* 2012 ; Ji *et al.* 2013). Le *metabarcoding* repose ainsi sur ces deux technologies que sont la taxonomie ADN et le séquençage à haut-débit, afin de permettre l'identification simultanée de plusieurs groupes taxonomiques à partir d'un échantillon complexe d'ADN (Taberlet *et al.* 2012b ; Shokralla *et al.* 2012 ; Ji *et al.* 2013).

L'échantillon d'ADN initial peut être issu de deux sources distinctes. Soit il peut provenir d'individus (ou de fragments d'individus) broyés et mélangés ('*bulk sample*',

Hajibabaei *et al.* 2011 ; Yu *et al.* 2012) à partir duquel l'ADN est extrait. A l'instar du *barcoding* classique, cette approche est invasive dans le sens où elle nécessite la capture préalable des individus. Soit, considérant que tous les organismes vivants contaminent leur environnement avec leur ADN, il est possible d'extraire ce dernier d'échantillons environnementaux tels que l'eau (Ficetola *et al.* 2008 ; Dejean *et al.* 2011 ; Thomsen *et al.* 2012a ; Thomsen *et al.* 2012b), les fèces (Deagle *et al.* 2010 ; Bohmann *et al.* 2011 ; Pompanon *et al.* 2012 ; Shehzad *et al.* 2012a ; Shehzad *et al.* 2012b ; Soininen *et al.* 2013), les sols (Andersen *et al.* 2012 ; Bienert *et al.* 2012 ; Taberlet *et al.* 2012b ; Yoccoz *et al.* 2012 ; Baldwin *et al.* 2013), le permafrost (Sønstebo *et al.* 2010 ; Jørgensen *et al.* 2012a ; Willerslev *et al.* 2014), la glace (Jørgensen *et al.* 2012b) ou encore les sédiments aussi bien récents (Fonseca *et al.* 2010 ; Chariton *et al.* 2010 ; Chariton *et al.* 2014) qu'anciens (Lejzerowicz *et al.* 2013 ; Parducci *et al.* 2013 ; Pedersen *et al.* 2013 ; Haouchar *et al.* 2014). Le terme d'« ADN environnemental » désigne ainsi l'ADN pouvant être extrait d'échantillons environnementaux sans isolation préalable des organismes ciblés. On parlera alors de *metabarcoding* environnemental (« *eDNA metabarcoding* », Fig. 6 , Taberlet *et al.* 2012a).

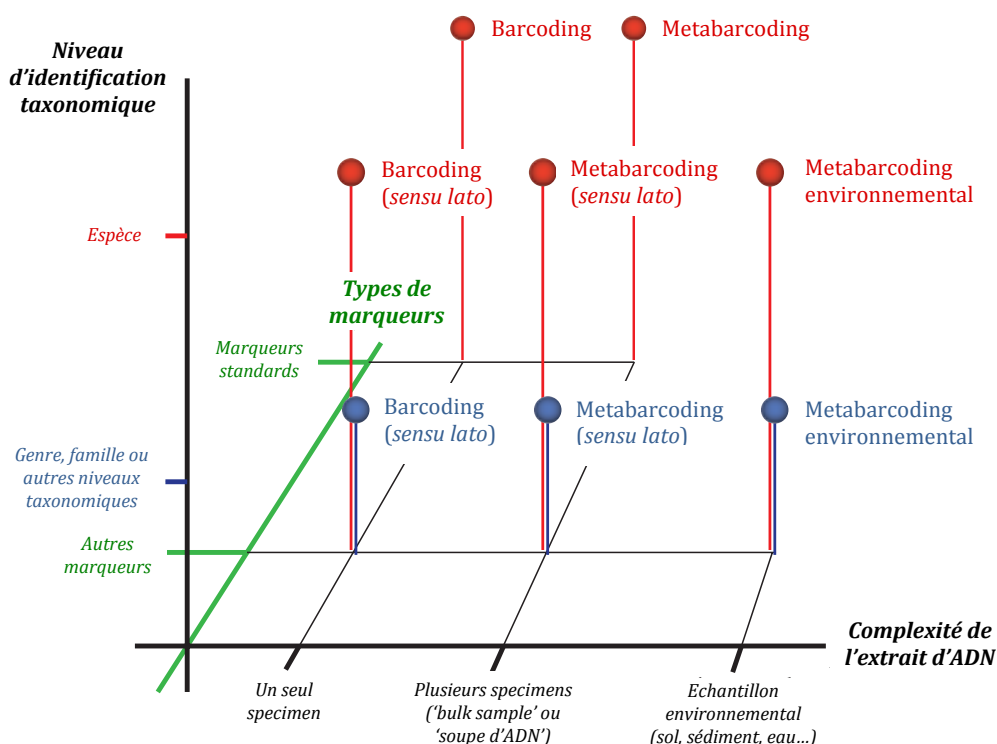


Figure 6. Terminologie suggérée par Taberlet *et al.* (2012a) pour caractériser l'identification taxonomique basée sur l'ADN en fonction du type de marqueur, du niveau d'identification et de la complexité de l'échantillon d'ADN.

A partir de l'ADN extrait, plusieurs marqueurs (permettant de couvrir plusieurs groupes taxonomiques) peuvent être amplifiés par PCR puis séquencés à haut-débit (voir détails ci-dessous). Les données produites peuvent alors servir de proxy pour la biodiversité présente dans les échantillons collectés (Coissac *et al.* 2012). Chaque échantillon de sol, d'eau ou encore de sédiments intégrant de l'information sur l'ensemble de la diversité biologique, il est alors possible d'obtenir des informations sur un grand nombre de groupes taxonomiques en une seule expérimentation à partir d'un échantillonnage unique. Cette approche constitue ainsi une levée du verrou lié à l'effort d'échantillonnage et permet de procéder à un changement d'échelle dans l'acquisition des données de biodiversité (Taberlet *et al.* 2012a). Elle est, de plus, aisément standardisable.

c. De multiples applications en écologie

Le *metabarcoding* environnemental a de multiples applications en écologie et a notamment été utilisé, parfois sous d'autres appellations (*e.g.* « ecogenomics », Chariton *et al.* 2010 ; « ecometagenetics », Porazinska *et al.* 2010, « environmental *barcoding* », Hajibabaei *et al.* 2011) pour décrire la biodiversité dans une large gamme d'environnements (Valentini *et al.* 2009a ; Epp *et al.* 2012 ; Bohmann *et al.* 2014). Il a ainsi été largement utilisé pour la détermination de régime alimentaire (*e.g.* Valentini *et al.* 2009b ; Deagle *et al.* 2010 ; Bohmann *et al.* 2011 ; Pompanon *et al.* 2012 ; Shehzad *et al.* 2012a ; Shehzad *et al.* 2012b ; Soininen *et al.* 2013 ; De Barba *et al.* 2014). D'autres études ont montré son potentiel en biosurveillance pour évaluer la qualité de certains milieux (Baird & Hajibabaei 2012 ; Kermarrec *et al.* 2014) ou encore en écologie de la restauration (Ji *et al.* 2013). Cette approche a, par ailleurs, montré son potentiel en paléoécologie pour la reconstruction des communautés passées aussi bien végétales (*e.g.* Sønstebo *et al.* 2010 ; Jørgensen *et al.* 2012a, Jørgensen *et al.* 2012b ; Parducci *et al.* 2013), animales (Boessenkool *et al.* 2012 ; Willerslev *et al.* 2014) que fongiques (Bellemain *et al.* 2013).

Des méthodologies similaires ont déjà été utilisées pour étudier l'impact de certaines perturbations (pollutions, sécheresses, crues...) sur la diversité des eucaryotes dans les sols ou les sédiments (*e.g.* Chariton *et al.* 2010, Baldwin *et al.* 2013). Elles ont notamment montré la résilience des communautés du sol aux pressions induites par les

changements climatiques sur la disponibilité en eau (Baldwin *et al.* 2013). L'objectif est ici d'étendre ces études aux activités anthropiques et d'évaluer leurs impacts sur les communautés, notamment du sol, et la résilience de celles-ci à travers différentes échelles spatiales et temporelles. D'une part, l'ADN des sols actuels permet d'évaluer l'impact des perturbations relativement récentes (de l'ordre de quelques décennies) en comparant des sites géographiquement proches présentant des histoires de perturbations contrastées. D'autre part, l'ADN retrouvé dans les archives sédimentaires ouvre des perspectives intéressantes pour étudier l'effet à long terme des activités anthropiques, sur plusieurs centaines ou milliers d'années.

2. Protocole type et particularités du *metabarcoding* environnemental

L'approche de *metabarcoding* environnemental dérive pleinement des techniques de microbiologie environnementale développées depuis deux décennies pour caractériser la diversité microbienne (*e.g.* Ogram *et al.* 1987 ; Tringe & Rubin 2005 ; voir Zinger *et al.* 2012 pour une revue) et vise à les appliquer à l'ensemble de la biodiversité (Ji *et al.* 2013). Elle peut être divisée en six étapes (Fig. 7) : (i) échantillonnage sur le terrain (*e.g.* sols, sédiments, eau, air...), (ii) extraction de l'ADN issu des échantillons prélevés, (iii) amplification par PCR de courtes régions d'ADN à l'aide d'amorces universelles pour un groupe taxonomique donné, (iv) séquençage à haut-débit des produits PCR, (v) filtrage des données, (vi) assignation taxonomique par comparaison des séquences filtrées avec une base de données de référence.

a. Échantillonnage et extraction d'ADN

L'échantillonnage se doit d'être aussi représentatif que possible de la zone d'étude et tenir compte de l'hétérogénéité spatiale afin d'intégrer au mieux la diversité locale. Les échantillons issus d'une même zone peuvent être traités indépendamment pour permettre d'étudier la répartition spatiale de la biodiversité du milieu ou être mélangés afin d'obtenir une vue d'ensemble de la diversité du milieu et de minimiser les coûts. L'extraction d'ADN est réalisée, de préférence, le plus rapidement possible après le prélèvement afin de limiter d'éventuels changements dans les communautés microbiennes lors du stockage.

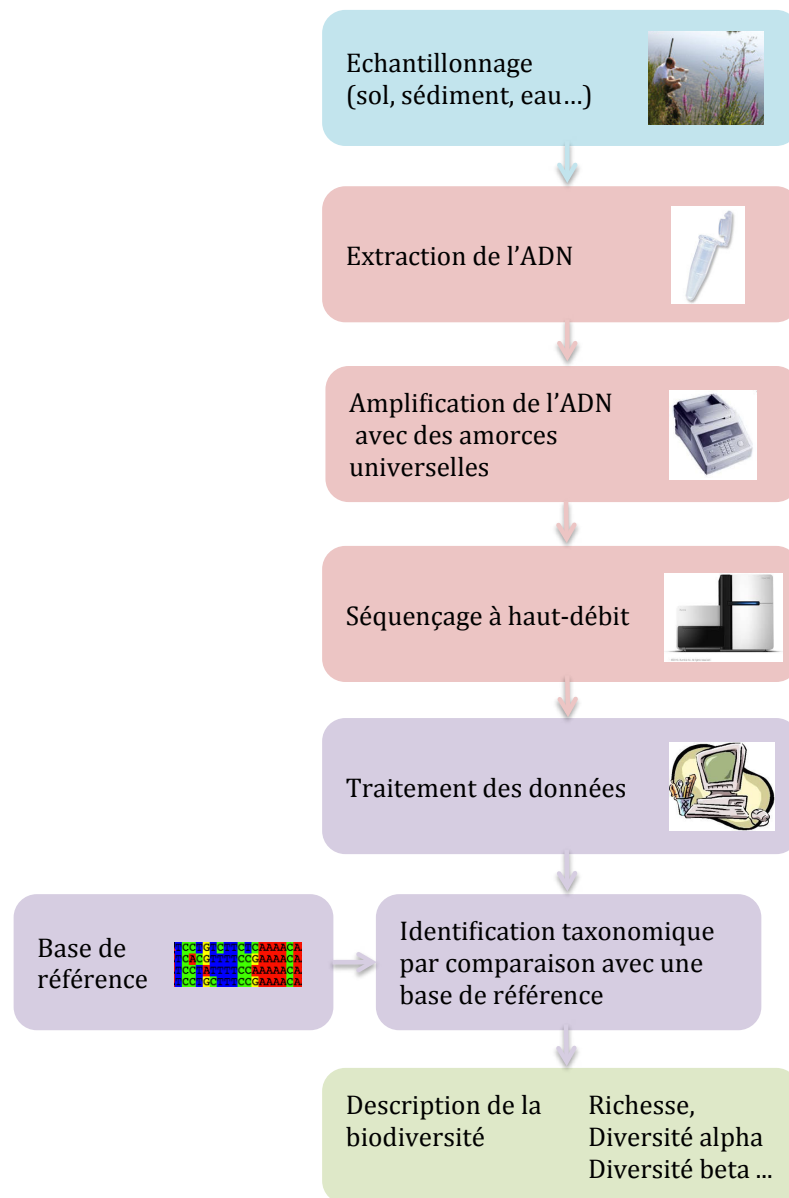


Figure 7. Protocole type en metabarcoding environnemental (d'après Valentini et al. 2009a)

L'ADN contenu dans les échantillons environnementaux contient aussi bien de l'ADN intracellulaire (issu de cellules vivantes ou d'organismes multi-cellulaires) que de l'ADN extracellulaire (Levy-Booth *et al.* 2007; Pietramellara *et al.* 2009). Ce dernier peut avoir de multiples origines : desquamation, décomposition, lyse cellulaire, racines, fèces, urine ont été proposées comme autant de sources potentielles (Lydolph *et al.* 2005 ; Willerslev *et al.* 2014). Dans les sols ou les sédiments, une partie de l'ADN ainsi libéré est adsorbé sur les particules telles que les argiles, la silice et la matière organique où il est

alors protégé des enzymes de dégradation (Pietramellara *et al.* 2009 ; Taberlet *et al.* 2012b). L'utilisation de l'ADN total peut entraîner un biais vers les séquences de macrorestes (Pedersen *et al.* 2013). Aussi, l'utilisation de l'ADN extracellulaire permet-elle de mieux intégrer la diversité locale sur une période plus longue en évitant les changements rapides dans la composition des communautés de micro-organismes liés aux paramètres saisonniers et climatiques (Zinger *et al.* 2009). Enfin, l'extraction de l'ADN extracellulaire peut être réalisée sur de plus gros volumes d'échantillons environnementaux initiaux ; une meilleure représentativité des organismes macroscopiques est alors attendue (Taberlet *et al.* 2012b).

L'utilisation de l'ADN environnemental implique de travailler avec de petites quantités d'ADN dégradé, et ce, même si les conditions de préservation sont particulièrement bonnes (Coolen & Overmann 2007 ; Yoccoz *et al.* 2012). Afin d'éviter de passer à côté d'espèces en réalité présentes (*e.g.* faux négatifs), de multiples extractions et amplifications d'un même échantillon sont souvent réalisées (cf. Ficetola *et al.* accepté en annexe).

b. Amplification et séquençage haut-débit

La dégradation de l'ADN environnemental limite la taille des fragments retrouvés dans les échantillons et, de fait, celle des marqueurs utilisables. En effet, plus les séquences ciblées seront longues, plus faible sera la probabilité de retrouver des fragments entiers correspondant à un taxon référencé dans la base de données (Yoccoz *et al.* 2012). Aussi, l'utilisation de marqueurs longs, tels que ceux proposés par le CBoL, peut conduire à manquer un nombre important d'espèces. L'utilisation de marqueurs courts (<150 pb, Valentini *et al.* 2009a ; Meusnier *et al.* 2008 ; Epp *et al.* 2012), souvent non standardisés (au sens du CBoL) et présentant certaines caractéristiques propres, est alors nécessaire (cf. encadré 5 ; Deagle *et al.* 2014). Un des pendants négatifs à l'utilisation de tels marqueurs, outre le fait que la non-standardisation limite les bases de références, est une résolution taxonomique plus faible (Taberlet *et al.* 2007). Il y a ainsi un compromis à réaliser entre la taille des *metabarcodes* et leur pouvoir discriminant (Chariton 2012). Néanmoins, pour des groupes restreints, par exemple au niveau de la famille ou de l'ordre, certains marqueurs courts se révèlent très résolutifs

(*e.g.* Lumbricidae, Cyperaceae ou Enchytraeidae, Bienert *et al.* 2009 ; De Barba *et al.* 2014 ; Epp *et al.* 2012).

Les séquenceurs nouvelles générations (NGS), bien que différents dans leurs technologies et leurs caractéristiques, ont tous la particularité de paralléliser le processus de séquençage, ce qui permet de produire plusieurs millions de séquences par expérimentation (Kircher & Kelso 2010). Afin de permettre le mélange et le séquençage en parallèle de produits PCR, et ultimement la réassignation des séquences obtenues à leur échantillon initial, chaque échantillon d'ADN est amplifié indépendamment à l'aide

Encadré 5 : Caractéristiques des marqueurs utilisés en metabarcoding environnemental

De par les contraintes associées à l'utilisation de l'ADN issu d'échantillons environnementaux et du séquençage à haut-débit, les marqueurs utilisés en *metabarcoding* environnemental requièrent des caractéristiques différentes des *barcodes* standards. Comme ces derniers, les régions séquencées doivent être quasiment identiques entre individus d'une même espèce, et différentes entre espèces. Elles doivent contenir suffisamment d'information phylogénétique afin de pouvoir assigner les séquences inconnues (*i.e.* non encore référencées dans les bases de données) à un groupe taxonomique (*e.g.* genre, famille...). Les régions flanquantes, servant de sites de fixation des amorces PCR, doivent être extrêmement conservées entre les différents taxons afin de favoriser l'amplification d'un maximum de taxons au sein du groupe taxonomique visé et de limiter les biais d'amplification lors de la PCR. Cette caractéristique est particulièrement importante dans le cas des échantillons environnementaux contenant un mélange d'ADN de plusieurs taxons. De plus, les marqueurs doivent être suffisamment courts pour permettre l'amplification de l'ADN dégradé, les régions supérieures à 150 pb étant généralement difficiles à amplifier à partir d'ADN dégradé (Taberlet *et al.* 2007 ; Valentini *et al.* 2009a). Enfin, ils doivent être préférentiellement situés sur des régions présentes en plusieurs copies par cellule, à l'instar des génomes d'organelles (*i.e.* mitochondrie et chloroplaste), ce qui maximise leur détection dans les échantillons environnementaux.

De tels marqueurs sont évidemment « idéaux » et n'ont pas encore été déterminés, si tant est qu'ils existent (Nielsen *et al.* 2006 ; Valentini *et al.* 2009a) mais certaines de leurs caractéristiques rendent quelques régions parfaitement exploitables en *metabarcoding*. Riaz *et al.* (2012) ont ainsi récemment développé un programme visant à identifier et évaluer de nouveaux *barcodes* pertinents pour les études de *metabarcoding* (ainsi que les amorces PCR associées) à partir de bases de données génétiques existantes (*e.g.* EMBL ou genbank). Ce programme est basé sur deux indices de qualité permettant d'évaluer la couverture taxonomique et le pouvoir discriminant du marqueur en fonction de certaines contraintes telles que la longueur du marqueur ou les taxons ciblés.

d'amorces « taguées » (ajout de 8 bases en position 5' et 3' des amorces *forward* et *reverse* respectivement), spécifiques à chaque échantillon (Binladen *et al.* 2007 ; Valentini *et al.* 2009b). Ainsi, tous les produits PCR peuvent être directement séquencés en une seule expérimentation; la profondeur de séquençage permettant d'avoir une vue détaillée des produits PCR complexes tels ceux issus d'échantillons environnementaux (MacLean *et al.* 2009). Par comparaison, le séquençage capillaire de type Sanger requiert au préalable une fastidieuse étape de clonage des différents fragments d'ADN amplifiés (Shehzad 2011), ce qui limite l'exploration des produits PCR aux taxons les plus fréquents (Zinger *et al.* 2012). En effet, le nombre de fragments clonés puis séquencés étant limités, ceux-ci sont alors issus d'un important sous-échantillonnage parmi les amplicons et correspondent donc principalement aux séquences les plus représentées.

c. Filtrage des données et assignation taxonomique

Le volume de données produit par les NGS ainsi que les particularités du protocole de *metabarcoding* nécessitent un traitement bioinformatique relativement lourd et le développement d'outils appropriés (voir Coissac *et al.* 2012 pour une revue plus spécifique des défis bioinformatiques associés au *metabarcoding*). Parmi ces étapes, on retrouve notamment l'assemblage des séquences dans les cas où le séquençage des fragments se fait dans les deux directions à partir des deux extrémités des fragments (*e.g.* séquençage « *paired-end* »), l'assignation des séquences à leur échantillon d'origine, le traitement des erreurs ou encore l'assignation taxonomique.

Des erreurs de réplication, substitution de bases et insertion/délétion peuvent apparaître lors de l'amplification de la PCR et/ou du séquençage. De même, des chimères de séquences peuvent se former générant un nombre potentiellement important de séquences artefactuelles (*e.g.* Quince *et al.* 2011 ; Reeder & Knight 2010 ; Berry *et al.* 2011 ; Fonseca *et al.* 2012). Ceci peut avoir des conséquences sur l'assignation taxonomique et/ou la formation de MOTUs (« Molecular Operational Taxonomic Unit », Blaxter *et al.* 2005) et entraîner au final une mauvaise interprétation des patrons de diversité (Coissac *et al.* 2012). Plusieurs programmes ont ainsi été développés afin de détecter puis de traiter les erreurs potentielles (*e.g.* *Obiclean*, <http://metabarcoding.org/obitools>; *Denoiser*, Caporaso *et al.* 2010 ; *AmpliconNoise* et *Perseus*, Quince *et al.* 2011 ; *ChimeraSlayer*, Haas *et al.* 2011, APDP, Morgan *et al.* 2013).

En outre, des contaminations sporadiques peuvent également augmenter le nombre de faux positifs. Afin de limiter ces derniers, potentiellement non détectés par les programmes dédiés, certains auteurs considèrent une séquence présente dans un échantillon si celle-ci est confirmée par au moins deux réplicats de PCR indépendants (Willerslev *et al.* 2014 ; cf. Ficetola *et al.* accepté en annexe).

Les marqueurs utilisés en *metabarcoding* sont souvent non standardisés, ce qui limite l'utilisation des bases de données « classiques » en *barcoding* (BOLD). Une des solutions afin de procéder à l'assignation taxonomique des séquences est alors de créer des bases de références spécifiques à chaque étude (*e.g.* Valentini *et al.* 2009b ; Bienert *et al.* 2012). Une seconde approche consiste à générer des bases de références spécifiques à chaque *metabarcode* à partir des bases de données génétiques existantes (*e.g.* Genbank, EMBL, UNITE) par PCR *in silico* (Ficetola *et al.* 2010). Les marqueurs utilisés sont censés contenir suffisamment d'informations taxonomiques pour permettre l'assignation d'une séquence à un certain niveau taxonomique (*e.g.* genre, famille, ordre...) même si celle-ci est absente de la base de références. Une troisième approche, quelque peu différente, consiste à délimiter des taxons sans avoir nécessairement recours à l'identification taxonomique (Coissac *et al.* 2012). Les séquences peuvent alors être regroupées en MOTUs (Molecular Operational Taxonomic Unit) en fonction de leur similarité. Le seuil est généralement défini par le chercheur en fonction des organismes étudiés, des caractéristiques des marqueurs utilisés et des objectifs de l'étude. Il faut cependant noter que certains *metabarcodes* peuvent être très variables dans leur taille rendant l'alignement impossible (*e.g.* boucle P6 de l'intron *trnL* chloroplastique, Taberlet *et al.* 2007). Par conséquent, les méthodes basées sur des alignements multiples de séquences, et éventuellement des arbres phylogénétiques robustes, ne sont pas applicables.

S'il s'agit d'une approche non-idéale, elle peut néanmoins faciliter l'accès à tout un pan de la biodiversité jusqu'alors inaccessible par d'autres méthodes et parfois même non décrit (Blaxter *et al.* 2005). A l'instar de ce qui est pratiqué depuis une décennie en microbiologie environnementale, les MOTUs pourraient être utilisés en écologie des perturbations comme une unité taxonomique à part entière pour décrire les patrons de diversité (*e.g.* Fonseca *et al.* 2010 ; Chariton *et al.* 2010 ; Yu *et al.* 2012 ; Baldwin *et al.* 2013 ; Ji *et al.* 2013) et, par exemple, pour comparer des niveaux de richesse entre différentes localités/traitements (diversité alpha) ou encore les différences de

composition en MOTUs entre localités/traitements (diversité beta) (cf. encadré 3 ; Coissac *et al.* 2012).

3. Démarche scientifique et structure de la thèse

Evaluer la résistance et la résilience des écosystèmes face aux perturbations induites par l'Homme se fait par l'évaluation des changements après perturbations, ceux-ci pouvant concerner la structure taxonomique des communautés ou encore les changements fonctionnels (Calderon-Aguilera *et al.* 2012). Deux types d'approches peuvent être utilisées afin d'étudier l'impact des perturbations sur la biodiversité. La première consiste à effectuer une étude synchronique en étudiant des sites géographiquement proches avec des histoires de perturbations contrastées (Purschke *et al.* 2013). Les patrons spatiaux sont souvent utilisés comme substitut aux patrons temporels, bien qu'ils ne soient pas nécessairement équivalents. Il s'agit d'une estimation indirecte où l'on émet l'hypothèse sous-jacente que les mêmes mécanismes écologiques à l'œuvre dans les différents sites sont similaires (Walker 2010). La seconde approche pour comprendre la dynamique de la biodiversité consiste à en examiner les patrons à différentes périodes de temps (Magurran & Dornelas 2010). Dans cette approche diachronique, on identifie, dans un premier temps, les éléments perturbateurs du système pour observer leurs effets sur la dynamique de la biodiversité, en un ou plusieurs sites donnés (Chazdon *et al.* 2007). Notre hypothèse de travail est que le *metabarcoding environnemental* peut s'avérer être un outil particulièrement adapté à ces deux approches afin de caractériser les changements induits par les activités humaines sur la structure taxonomique des communautés et leur résilience.

Le présent document se divise en trois parties. La première consiste en la réalisation d'études spatialisées visant à déterminer l'impact des activités humaines relativement récentes (introduction d'espèces exotiques et modes d'utilisation des terres) sur la diversité actuelle des communautés de sols, et la réponse de ces dernières, dans des milieux contrastés. L'ADN issu d'échantillons de sols est ici utilisé afin d'étudier ce compartiment difficile d'accès qu'est la diversité des sols.

La seconde partie est une approche temporelle dont le but est d'étudier l'évolution des communautés végétales et des paysages au cours des 10000 dernières années, sous l'influence de différents paramètres parmi lesquels les activités pastorales.

L'ADN ancien conservé dans les archives sédimentaires lacustres est utilisé, dans un premier temps, pour différencier les activités humaines d'autres facteurs abiotiques influençant les trajectoires de végétation et les paysages, puis pour étudier la réponse à long terme de ces derniers face aux perturbations.

Plusieurs éléments de ces deux premières parties sont issus de documents publiés ou soumis et sont présentés ici sous la forme d'article scientifique, tels qu'ils l'ont été lors de leur soumission. Ceux-ci pouvant être lus indépendamment les uns des autres, certaines redondances peuvent apparaître. Chacun de ces articles est, au préalable, replacé dans le contexte de cette thèse et les éléments de discussion présentés à la fin de chaque partie visent à faire une synthèse globale de ces travaux.

Enfin, dans la troisième partie, nous discuterons de la pertinence de l'approche de *metabarcoding* environnemental pour étudier l'impact des activités anthropiques sur la biodiversité et les mécanismes impliqués dans la réponse des communautés.

Chapitre III. Evaluation de l'impact des activités anthropiques sur la biodiversité des sols par *metabarcoding* environnemental

1. Problématique générale

Le rôle primordial des communautés biologiques des sols dans le fonctionnement des écosystèmes (*e.g.* recyclage des nutriments, séquestration du carbone, provision de nourriture...) est unanimement reconnu (*e.g.* Ettema & Wardle 2002 ; McNeill & Winiwarter 2004 ; Wall 2005 ; Wall *et al.* 2010). Cependant, en comparaison avec la biodiversité de surface, la vaste diversité des sols est encore peu connue, à l'exception de certains groupes « ingénieurs » tels que les vers de terre ou les termites (Lavelle *et al.* 1997 ; De Deyn & Van der Putten 2005 ; Decaëns *et al.* 2006 ; Decaëns 2010). Or, celle-ci inclut aussi bien les invertébrés (*e.g.* termites, vers de terre, fourmis, nematodes, isopodes...) que les vertébrés, les racines de plantes, les champignons, les bactéries, les protozoaires ou encore les virus (Wall 2005).

De nombreuses menaces liées aux activités humaines pèsent sur les sols, telles que les invasions biologiques, les changements climatiques qui entraînent des modifications de l'aire de répartition des espèces ou encore l'utilisation des terres (Wolters *et al.* 2000 ; Jeffery & Gardi 2010). Des pratiques agricoles inadaptées peuvent entraîner une dégradation mécanique des sols, mais aussi leur salinisation ou encore un apport intensif de pesticides, de fertilisants et de polluants. De même, la compaction et l'étanchéisation des sols, notamment sous l'effet de l'urbanisation, réduisent l'habitat de certains taxons et limitent leur accès aux ressources telles que l'eau, l'oxygène et la matière organique (Turbé *et al.* 2010).

Les conditions environnementales (*e.g.* pH, température, ressources), les interactions biotiques (*e.g.* compétition, facilitation) et les caractéristiques des organismes (*e.g.* capacité de dispersion, traits fonctionnels) sont autant de paramètres influençant, réciproquement, la distribution géographique de la diversité des sols (Wardle 2006 ; Zinger 2009). Ces facteurs agissent à différentes échelles et les perturbations jouent un rôle prépondérant dans la distribution spatiale des espèces en affectant les caractéristiques biotiques et abiotiques de l'habitat. Cela va, *in fine*, jouer sur la répartition des processus écosystémiques, cette dernière étant intimement liée à la structuration spatiale horizontale de la biodiversité des sols (Ettema & Wardle 2002).

Le manque d'expertise taxonomique, de méthodes d'échantillonnage standardisées et de protocoles expérimentaux adaptés ont ainsi été pointés du doigt pour expliquer les lacunes actuelles dans la compréhension des patrons de diversité du sol (André *et al.* 2002 ; Decaëns *et al.* 2006 ; Decaëns 2010). Le développement de méthodes pertinentes, standardisées et facilement mesurables pour le suivi de la biodiversité des sols est aujourd'hui encouragé par l'Union européenne (Turbé *et al.* 2010). La *barcoding* ADN est utilisé depuis maintenant une décennie comme un outil exploratoire pour caractériser taxonomiquement les groupes peu ou pas connus (Blaxter *et al.* 2005 ; Puillandre *et al.* 2012). Le *metabarcoding* environnemental permet aujourd'hui d'obtenir des informations d'ordre écologique sur la composition de ces communautés et les paramètres qui en sont à l'origine.

Dans cette première partie, l'approche de *metabarcoding* environnemental a été utilisée pour étudier l'impact des activités anthropiques relativement récentes (introduction d'espèces exotiques, utilisation des terres et traitements sylvicoles) sur plusieurs communautés du sol, et la résilience de ces dernières, dans des milieux contrastés. L'ADN du sol a été largement utilisé depuis deux décennies en écologie microbienne pour l'étude *in situ* des communautés fongiques et bactériennes (*e.g.* Ogram *et al.* 1987; Zinger *et al.* 2009 ; Zinger *et al.* 2011 ; Roy *et al.* 2013 ; Blaailid *et al.* 2012 ; Lindahl *et al.* 2013 ; Blaailid *et al.* 2014) et, dans une moindre mesure, pour caractériser les communautés végétales (Hiiesalu *et al.* 2012 ; Taberlet *et al.* 2012 ; Yoccoz *et al.* 2012). La première étude porte, ainsi, sur l'utilisation d'une approche de *metabarcoding* environnemental pour étudier simultanément les trajectoires des communautés végétales et fongiques, suite à l'introduction de mammifères herbivores dans un environnement qui en est naturellement dépourvu : les îles Sub-Antarctiques. Une des opportunités offertes par cette approche est également de pouvoir suivre les communautés animales du sol. Si la faisabilité technique a récemment été démontrée (Bienert *et al.* 2012), son utilisation en recherche et par les gestionnaires nécessite encore des développements. La seconde étude concerne l'application à l'échelle du paysage d'une telle approche pour évaluer l'influence de l'utilisation des terres sur la distribution des communautés lombriciennes en milieu de montagne. Enfin, nous évoquerons une troisième étude, encore en cours, visant à caractériser l'impact des traitements forestiers sur les communautés eucaryotes des sols tropicaux (cf. encadré 6).

2. Etude de la diversité fongique des sols dans les îles Sub-Antarctiques en lien avec l'introduction d'herbivores exotiques

a. Mise en contexte de l'étude

Les invasions biologiques, favorisées par les activités humaines (Lenzen *et al.* 2012), sont aujourd'hui reconnues comme la seconde cause de perte de biodiversité mondiale (Vitousek *et al.* 1997) et entraînent des changements majeurs dans la structure des communautés et le fonctionnement des écosystèmes (Mack *et al.* 2000 ; Courchamp *et al.* 2003). En dépit de leur éloignement, les îles Sub-Antarctiques, au même titre que les autres systèmes insulaires, sont particulièrement menacées par ce phénomène. Le peu de richesse spécifique, et la faible redondance fonctionnelle associée, rendent ces milieux particulièrement sensibles aux changements (Frenot *et al.* 2005). Plusieurs vagues d'introductions d'espèces exotiques, volontaires (*e.g.* pour l'agriculture) ou non, ont eu lieu dans les îles Sub-Antarctiques depuis leur découverte en 1772, principalement associées au trafic maritime et aux déplacements humains (Lebouvier *et al.* 2002 ; Lebouvier & Frenot 2007). Soixante-huit plantes allochtones ont ainsi été recensées sur les îles Kerguelen. Sept, dont le pissenlit (*Taraxacum sp.*) sont aujourd'hui considérées comme invasives au vu de leur large distribution à travers l'archipel (Frenot *et al.* 2001 ; Frenot *et al.* 2005). De même, ces îles comptent aujourd'hui sept espèces de mammifères herbivores alors qu'elles en sont naturellement dépourvues, et parmi elles, le lapin (*Oryctolagus cuniculus*). Ce dernier a été introduit sur l'île principale en 1874 (Kidder 1876), et plus tard sur neuf autres îlots alentours, où sa présence a notamment engendré un déclin rapide de la végétation native et favorisé l'érosion des sols (Courchamp *et al.* 2003 ; Chapuis *et al.* 2004).

Une des conséquences majeures est la tendance à la dominance de l'espèce native *Acaena magellanica* au détriment d'autres espèces, parmi lesquelles *Azorella selago* et *Pringlea antiscorbuta*, plus sensibles aux perturbations biotiques (Chapuis *et al.* 1994 ; Chapuis *et al.* 2004). De plus, l'augmentation des températures (+1.3°C entre les années 1960 et le milieu des années 2000) et les sécheresses (diminution des précipitations de 100 à 250mm par an) tendent à promouvoir l'expansion des plantes non natives (Frenot *et al.* 2005 ; Robin *et al.* 2011). Le surpâturage a également conduit à l'exposition de grandes surfaces de sol au gel, à la pluie et au vent, entraînant une dénudation et une érosion accrues des sols (Courchamp *et al.* 2003). Un programme expérimental

d'éradication du lapin a été mis en place sur deux petites îles de l'Archipel en 1992 et 1994 (l'île Verte et l'île Guillou) afin d'étudier la résilience des communautés végétales et des sols. Ce programme a conduit à l'éradication complète du lapin dans les deux ans suivant le début du traitement (Chapuis *et al.* 2001 ; Chapuis *et al.* 2004). Par ailleurs, si l'impact et la résilience des communautés végétales Sub-Antarctiques face aux perturbations induites par le lapin sont bien connus (Chapuis *et al.* 2004, Frenot *et al.* 2005, Robin *et al.* 2011), les communautés des sols, et particulièrement les communautés fongiques, ont été, par comparaison, très peu étudiées (Dennis *et al.* 2012). Les paramètres abiotiques des sols (*e.g.* pH, C/N...) et la communauté végétale sont parmi les déterminants majeurs des communautés microbiennes du sol (De Deyn & Van der Putten 2005; Yergeau *et al.* 2007 ; Zinger *et al.* 2011; Dennis *et al.* 2012). Aussi, l'altération de ces paramètres, liée aux activités du lapin, amène-t-elle à s'interroger sur l'impact indirect de ce dernier sur les communautés fongiques.

Les approches moléculaires ont été largement utilisées depuis deux décennies en écologie microbienne (*e.g.* Ogram *et al.* 1987, voir Zinger *et al.* 2012 pour une revue) et sont, comme nous l'avons vu, à la base de l'approche de *metabarcoding* environnemental. Historiquement, elles ont notamment permis d'éviter un certain nombre de biais liés aux méthodes traditionnelles basées sur la culture et l'identification subséquente. Seul un faible pourcentage de souches étant cultivables en milieu solide ou liquide (the « *great count plate anomaly* », Staley & Konopka 1985), l'utilisation de méthodes ADN en microbiologie a ainsi permis l'accès à un panel beaucoup plus large de la biodiversité microbienne. Parmi ces approches, celles dites d'« *empreintes moléculaires* » fournissent des informations pertinentes sur les patrons de biodiversité mais ne sont pas utilisables pour l'identification taxonomique et l'évaluation de la richesse (Zinger 2009 ; Bent *et al.* 2007). A l'inverse, le couplage du clonage et du séquençage Sanger a permis l'étude de la composition des communautés ainsi que des patrons d'alpha et beta diversité mais reste limité aux taxons dominants (Zinger *et al.* 2012). Le *metabarcoding* environnemental permet aujourd'hui de décrire la diversité microbienne à un niveau de détail encore supérieur (MacLean *et al.* 2009 ; Zinger *et al.* 2012).

b. Article A

Article A

Long-lasting modification of soil fungal diversity associated with the introduction of alien herbivores to remote sub-Antarctic islands

Article soumis à *Biology Letters*

Pansu Johan^{1,2}, Winkworth Richard³, Hennion Françoise⁴, Gielly Ludovic^{1,2}, Taberlet Pierre^{1,2} & Choler Philippe^{1,2,5}

¹ Univ. Grenoble Alpes, LECA, F-38000 Grenoble, France

² CNRS, LECA, F-38000 Grenoble, France

³ Institute of Fundamental Sciences, Massey University, New Zealand

⁴ UMR Ecobio, Université Rennes 1, CNRS, F-35042 Rennes, France

⁵ Univ. Grenoble Alpes, SAJF, F-38000 Grenoble, France

Contribution personnelle : analyses moléculaires, filtrage des séquences et assignation taxonomique, analyse des données, rédaction du manuscrit

Abstract

During the late 19th century, Europeans introduced rabbits to many of the sub-Antarctic islands, environments that prior to this had been devoid of mammalian herbivores. The impacts of rabbits on indigenous ecosystems are well studied, notably they cause dramatic changes in plant communities and promote soil erosion. However, the responses of fungal communities to such biotic disturbances remain unexplored. We used *metabarcoding* of soil extracellular DNA to assess the diversity of plant and fungal communities at sites on the Kerguelen Islands with contrasting histories of disturbance by rabbits. Our results indicate that fungal richness is naturally low on Kerguelen but that rabbit-induced modification of plant communities and soil properties has impacted on both the composition and diversity of indigenous soil fungal communities. Moreover, we found no indication of recovery in an area where rabbits were eradicated 20 years ago. These results illustrate the long-lasting and cascading effects of introduced herbivores on fungal biodiversity as well as the low resiliency of sub-Antarctic ecosystems.

Introduction

The sub-Antarctic islands are particularly threatened by climate change and human-induced disturbance, especially biological invasions [1]. Following their discovery in 1772, the Kerguelen Islands were home to both whaling and farming communities. As a result, numerous alien plant and animal species became established, among them several deliberately introduced herbivorous mammals. Alien species have caused major changes to the structure and functioning of the indigenous ecosystems [2]. The rabbit, *Oryctolagus cuniculus*, was introduced in 1874 and now occurs at high densities both on the main island, Grande Terre, as well as a number of the smaller ones [3]. Grazing and burrowing by rabbits have had remarkable impacts on the native vegetation, resulting in decreases in abundance of some native species (*e.g.* *Azorella selago* and *Pringlea antiscorbutica*) and the over-dominance of others (*e.g.* *Acaena magellanica*) [3]. Moreover, rabbit-induced increases in erosion and changes to soil characteristics favour invasive plant species [3,4].

While the impacts of alien herbivores on the vegetation are well described [3], their impact on soil communities – in particular soil fungi – are poorly known. Fungal communities play a central role in terrestrial ecosystem functioning (*e.g.* as plant

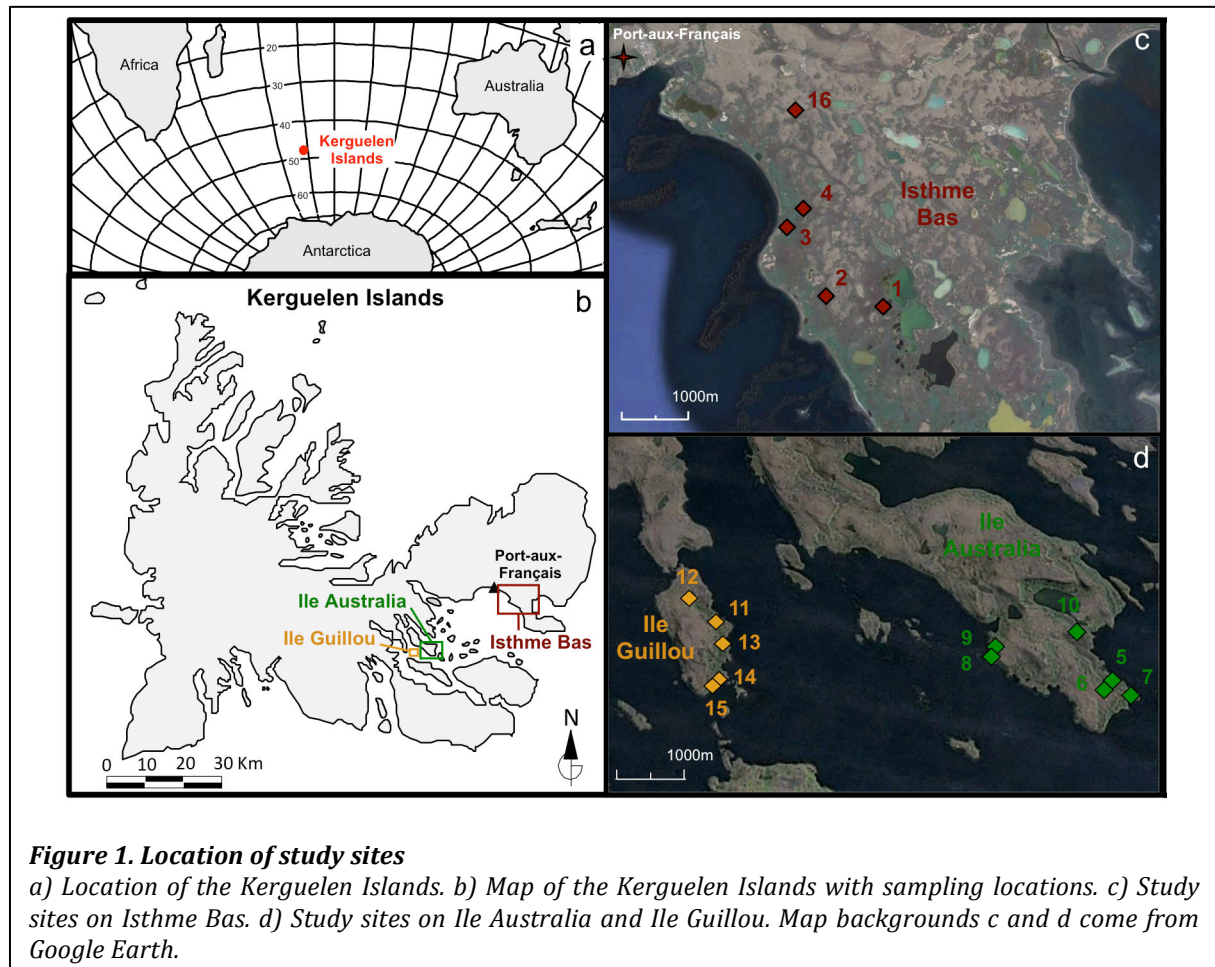
mycorrhizal symbionts, decomposers in nutrient cycling). It has been assumed that sub-Antarctic islands host limited fungal diversity [5] and the relative simplicity of these communities may make them particularly sensitive to disturbance [6]. There is now some evidence linking the distribution of fungal diversity to abiotic conditions and the nature of the aboveground plant community [7,8]. However, the response of fungal communities to biotic disturbance-induced changes to plant communities is poorly explored.

Here, we used soil *metabarcoding* to evaluate the composition of both fungal and plant communities on the Kerguelen Islands [9,10]. We examined soils from locations with contrasting histories of disturbance by rabbits. Specifically, sites that are currently highly disturbed, those from which rabbits were removed 20 years ago and those never impacted by rabbits. We addressed three main questions: (i) How diverse are communities of soil fungi on the sub-Antarctic islands? (ii) How important are biotic and abiotic factors for determining the composition of these communities? (iii) What is the effect of rabbit-induced disturbance on these communities?

Materials and methods

Study sites on the Kerguelen Islands were located on the the Isthme Bas (Grande Terre) where rabbits occur at high density, Ile Guillou where a successful eradication program was undertaken in 1994 [3] and Ile Australia where rabbits have never been reported (Fig. 1; Supplementary Table S1). Within each of the three areas we established 5 or 6 10m×10m plots within which we sampled soils and conducted vegetation surveys.

Soil extracellular DNA was extracted using a previously described method [9]. We used standard methods to PCR amplify established metabarcodes for vascular plants [11] and fungi [12]. Multiple PCR replicates and controls were included. High-throughput sequencing was performed on Illumina HiSeq and MiSeq platforms. The resulting sequences were filtered and clustered using an established workflow. Taxonomic assignments were determined by comparison to large references databases (*e.g.* EMBL) and an exhaustive database for the vascular flora of Kerguelen Islands. DNA sequences are available through Dryad.



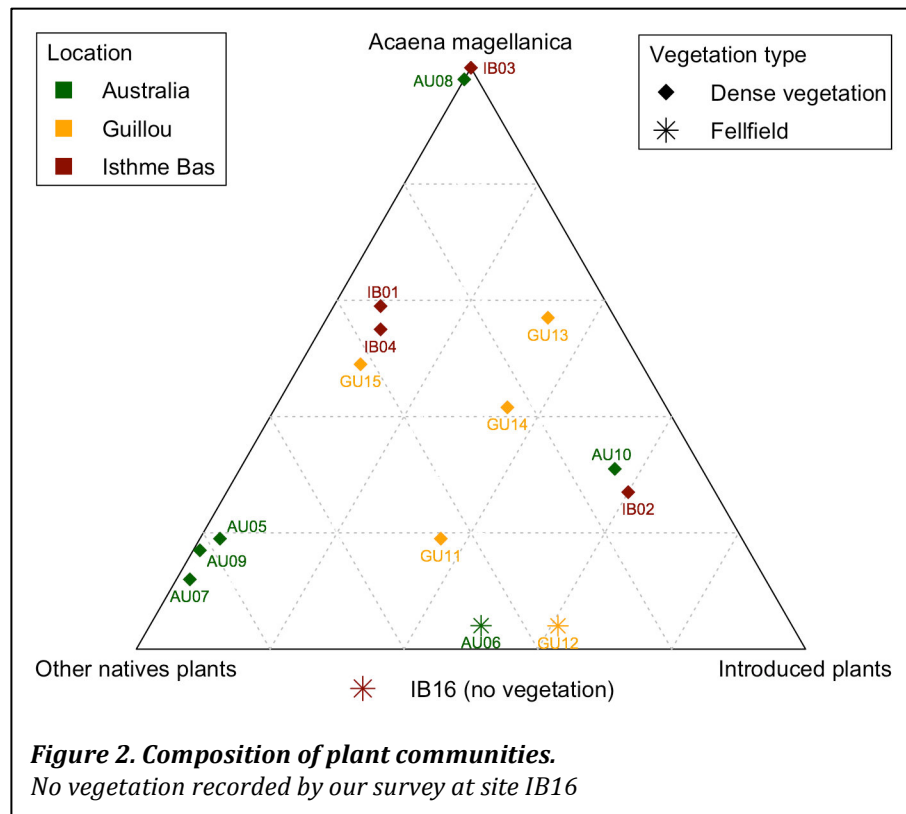
We used non-metric multidimensional scaling to ordinate fungal and plant communities. Vector fitting was used to examine the effects of soil conditions (e.g., soil organic matter content, soil acidity), and in the case of fungal communities biotic parameters (e.g., plant assemblages and their cover), on community composition (Fig. 3a). Procrustes analyses were used to compare ordinations based on different data sets. Finally, we compared the diversity of fungal communities on Kerguelen to those from seven high-altitude plots in the French Alps using a *t*-test. All statistical analyses were completed in R [13, 14].

Full methodological details are given in the supplementary materials.

Results

A robust linear correlation ($P < 0.001$; Supplementary Fig. 3) between the relative abundances of plants obtained by *metabarcoding* and botanical surveys, and the high similarity of community ordinations based on these data (Supplementary Fig. 4 & 5),

indicate that soil DNA is a good proxy of current plant cover on Kerguelen. Vector fitting indicated that the composition of plant communities was correlated with organic matter content ($P < 0.05$; Supplementary Fig. 5). Based on these results we recognised three broad vegetation types within plots, (i) species rich, native-dominated, (ii) species poor, *Acaena*-dominated and (iii) alien-dominated usually *Taraxacum* sp. (Fig. 2; Supplementary Fig. 2 & 4).



Soil *metabarcoding* identified 853 fungal MOTUs on the Kerguelen Islands with per plot diversity significantly lower on Kerguelen than at sites in the Alps (t -test, $P < 0.001$) (Table 1). Soil fungal communities also differed between Kerguelen sites (Supplementary Fig. 6); average MOTUs diversity was markedly higher on Ile Australia than on the Isthme Bas and Ile Guillou (Table 1, Fig. 3b). These communities also differed in composition. Significant correlations ($P < 0.01$) with several vegetation and soil parameters imply two axes of discrimination. These differentiated plots based on the density of plant cover and diversity of the associated plant community (Fig. 3a). Ordinations of fungal and plant communities were more similar to one another than would be expected by chance ($P = 0.001$; Supplementary Fig. 5). This result suggests that

the composition and richness of fungal and plant communities is co-structured across plots. Additional details are provided in the supplementary results.

	Locations	No. of plots	Per plot soil fungal MOTUs diversity	
			Mean	S.D.
Kerguelen Islands	Ile Australia	6	103.83	30.59
	Ile Guillou	5	78.30	10.91
	Isthme Bas	5	69.48	38.17
French Alps	Vercors Massif	3	180.67	19.7
	Arves Massif	4	156.25	37.12

Table 1. *Per plot soil fungal richness in the Kerguelen Islands and French Alps*

Discussion

Our analyses suggest that soil fungal diversity on the Kerguelen Islands is naturally lower than that of grasslands in the French Alps. Lower diversity on the island may be explained by the direct effect of geographic isolation (*e.g.* island vs. continent), although long-distance dispersal of microorganisms has been documented [*e.g.* 15], or climatic differences (*e.g.* limited vs. large seasonal differences). Yet given significant co-structuring of fungal and plant diversity has been reported at the landscape-scale [6,7] the impacts of isolation and climate on fungal diversity may be mediated via plant diversity. That is, because they have limited the pool of plant species on Kerguelen.

On the Kerguelen Islands soil fungal communities appear to reflect the disturbance history of the collection site. Specifically, fungal communities associated with species-rich native plant assemblages on Ile Australia (AU05-AU07, AU09) are distinct, both in terms of MOTUs richness and composition, from those on the Isthme Bas (Figs. 2a & b; Supplementary Fig. 6). We suggest that these differences are the result of contrasting rabbit disturbance histories, a long history of disturbance on Isthme Bas and none on Ile Australia.

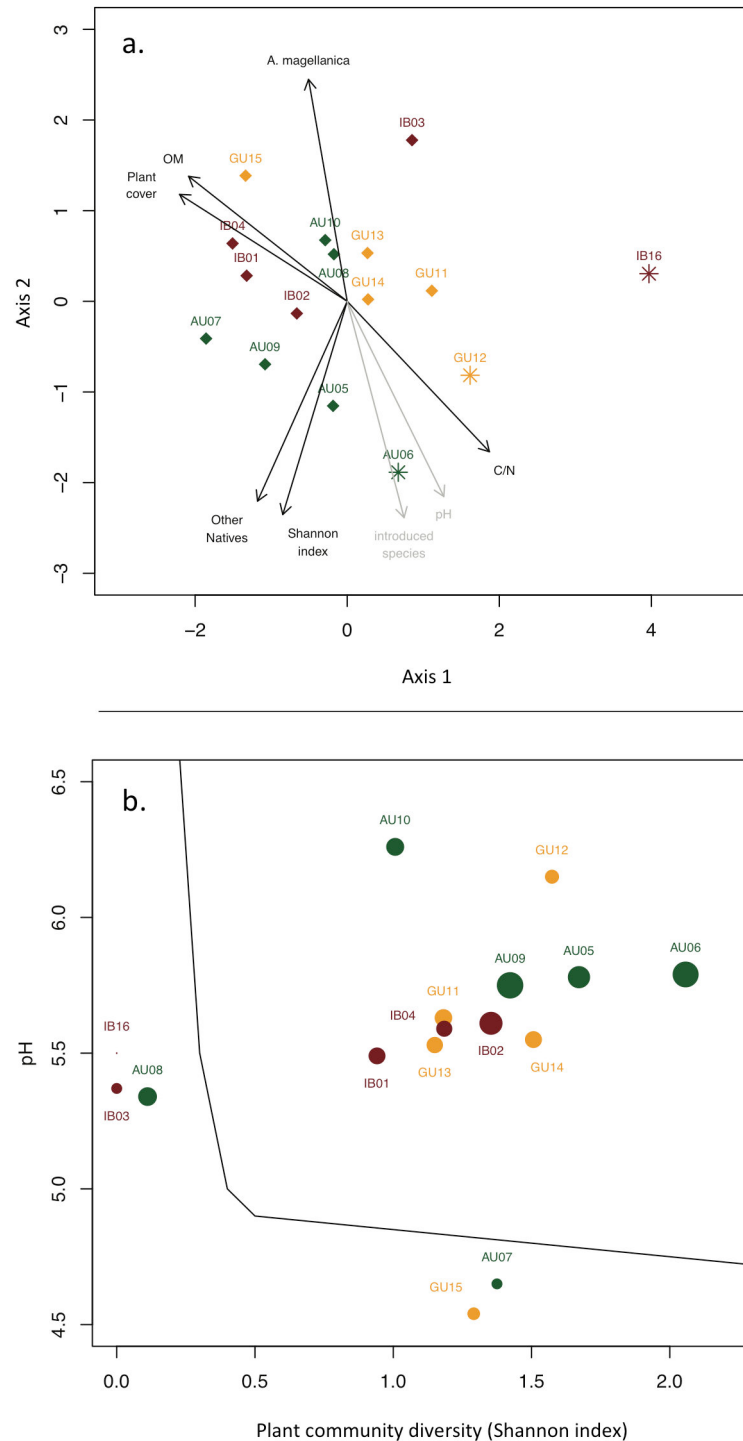


Figure 3. Contribution of local biotic and abiotic parameters in determining soil fungal communities

a) Ordination of soil fungal communities with vector fitting of biotic and abiotic variables. Non-metric multi-dimensional scaling stress = 0.134. Gray and black arrows represent non-significant and significant (at $P < 0.01$) variables respectively. *A. magellanica*, per plot frequency of *A. magellanica*; Other Natives, per plot frequency of other native plants; Introduced Species, per plot frequency of introduced plant; OM, soil organic matter content; C/N, carbon/nitrogen ratio; pH, soil acidity. b) Per plot fungal MOTUs richness relative to soil pH and plant community diversity (Shannon index). Circle size proportional to the number of fungal MOTUs.

However, since on Ile Australia soil fungal communities at naturally disturbed sites (e.g., AU10) are similar to those from Isthme Bas (Fig. 2a & b ; Supplementary Fig. 6) it appears that the community response to disturbance is similar whether the underlying cause is biotic or abiotic. Moreover, to us the co-structuring of soil fungal and plant communities on Kerguelen (Supplementary Fig. 5) implies that the plant community mediates this response. Specifically, we hypothesise that simplification of plant communities as a result of disturbance (e.g., rabbit grazing or erosion) leads to a reduced number of soil microhabitats and in turn to lower soil fungal diversity (Fig. 2b). Two additional observations are consistent with this suggestion. First, that cushion plants are common at undisturbed sites on Ile Australia and this growth form has been shown to modify local abiotic conditions thereby associated soil communities [16]. Second, that the site with the highest fungal diversity on the Isthme Bas (IB02) supports a mixed *Acaena-Taraxacum* plant community rather than a mono-specific stand of *Acaena* (Supplementary Figs. 2 & 6).

Rabbits were eradicated from Ile Guillou 20 years ago [3] but our analyses indicate that soil fungal communities there are yet to recover (Fig 3a; Supplementary Fig. 6). Indeed, both plant and fungal communities on Ile Guillou are more similar to the rabbit-disturbed communities on Isthme Bas than they are to undisturbed ones on Ile Australia (Fig. 2 & 3a; Supplementary Fig. 4, 5 & 6). We suggest that the state of fungal communities on this island reflects the slow recovery of the native vegetation following rabbit eradication. That the vegetation on Ile Guillou has not yet begun to show signs of recovery has been linked to slow growth rates in dominant species (e.g. *Azorella*) as well as soil and climatic changes that favour introduced species (e.g. *Taraxacum* sp.) [2,3].

Soil characteristics have previously been shown to play an important role in structuring microbial communities [7,8]. Here, we identify differences in fungal communities linked to both naturally occurring and rabbit-induced differences in soil parameters. Sites with waterlogged, highly acidic soils are associated with lower fungal diversity on both Ile Australia (e.g. AU07) and Ile Guillou (e.g. GU15). This pattern is therefore independent of rabbit-induced disturbance history (Fig. 3b). Moreover, that soil pH is not correlated with plant community diversity (Supplementary Fig. 5) implies soil acidity has a direct impact on fungal richness. The lowest soil fungal richness was found at a site lacking vegetation (e.g. IB16); in this case we suspect a rabbit-induced cause. Specifically, that at this site rabbit burrowing has promoted substantial erosion

and the associated loss of soil organic matter (Supplementary Table 1, Fig. 3a) has led to lower fungal diversity.

This study illustrates the long-lasting and cascading effects of introduced herbivores on fungal biodiversity as well as the overall sensitivity of remote sub-Antarctic island ecosystems to disturbance by invasive species. The limited diversity of soil fungi detected on Kerguelen suggests fungal communities on these islands lack functional redundancy. If so then they and the functions they support are likely to be particularly sensitive to human-induced changes. The continued growth of DNA reference databases, particularly for soil microorganisms, will help to improve biomonitoring in the sub-Antarctic. Being able to more fully characterise these communities will provide new perspectives on the resistance and resilience of these threatened ecosystems.

Acknowledgments

We thank the 136 IPEV program for funding the field campaign. DNA analyses were performed in the framework of the XX program. We specifically thanks Marine Pouvreau, Marion Lombard, Julie Vingère and Françoise Cardou for their help during the field campaign, Nigel Gilles Yoccoz for providing us with the local plant metabarcode database and Cindy Arnoldi for soil analyses.

References

1. Chown SL *et al.* 2012 Challenges to the future conservation of the Antarctic. *Science* **337**, 158-159.
2. Frenot Y, Chown SL, Whinam J, Selkirk PM, Convey P, Skotnicki M, Bergstrom DM. 2005 Biological invasions in the Antarctic: extent, impacts and implications. *Biol. Rev.* **80**, 45-72.
3. Chapuis JL, Frenot Y, Lebouvier M. 2004 Recovery of native plant communities after eradication of rabbits from the subantarctic Kerguelen Islands, and influence of climate change. *Biol. Conserv.* **117**, 167-179.
4. Courchamp F, Chapuis JL, Pascal M. 2003 Mammal invaders on islands: impact, control and control impact. *Biol. Rev.* **78**, 347-383.

5. Steiman R, Frenot Y, Sage L, Seigle-Murandi F, Guiraud P. 1995 Contribution to the knowledge of the mycoflora of Kerguelen Islands. *Cryptogamie Mycologie* **16**, 277-291.
6. Yergeau E, Bokhorst S, Kang S, Zhou JZ, Greer CW, Aerts R, Kowalchuk GA. 2012 Shifts in soil microorganisms in response to warming are consistent across a range of Antarctic environments. *ISME J.* **6**, 692-702.
7. Zinger L, Lejon DPH, Baptist F, Bouasria A, Aubert S, Geremia RA, Choler P. 2011 Contrasting diversity patterns of crenarchaeal, bacterial and fungal soil communities in an alpine landscape. *Plos One* **6**, e19950.
8. Dennis PG, Rushton SP, Newsham KK, Lauducina VA, Ord VJ, Daniell TJ, O'Donnell, AG, Hopkins DW. 2012 Soil fungal community composition does not alter along a latitudinal gradient through the maritime and sub-Antarctic. *Fungal Ecol.* **5**, 403-408.
9. Taberlet P *et al.* 2012 Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for *metabarcoding* studies. *Mol. Ecol.* **21**, 1816-1820.
10. Yoccoz NG *et al.* 2012 DNA from soil mirrors plant taxonomic and growth form diversity. *Mol. Ecol.* **21**, 3647-3655.
11. Taberlet P *et al.* 2007 Power and limitations of the chloroplast trnL (UAA) intron for plant DNA *barcoding*. *Nucleic Acids Res.* **35**, e14.
12. Epp LS *et al.* 2012 New environmental metabarcodes for analysing soil DNA: potential for studying past and present ecosystems. *Mol. Ecol.* **21**, 1821-1833.
13. Oksanen J *et al.* 2013 vegan: Community Ecology Package. R package version 2.0-7. <http://CRAN.R-project.org/package=vegan>
14. R Development Core Team, 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
15. Bridge PD, Spooner BM. 2012 Non-lichenized Antarctic fungi: transient visitors or members of a cryptic ecosystem? *Fungal Ecol.* **5**, 4381-4394.
16. Roy J, Albert CH, Ibanez S, Saccone P, Zinger L, Choler P, Clément JC, Lavergne S, Geremia RA. 2013 Microbes on the cliff: alpine cushion plants structure bacterial and fungal communities. *Front. Microbiol.* **4**, 64.

*c. Matériel et méthodes complémentaires***Supplementary Materials and Methods***Study sites*

The Kerguelen Islands lie in the Southern Indian Ocean approximately 49°S and 69°E. The islands have a cold, oceanic climate with a mean annual temperature of 4.6°C and averages for the coldest (*e.g.* July) and warmest (*e.g.* February) months of 2.3°C and 7.8°C, respectively (Chapuis *et al.* 2004). Mean annual rainfall is ~760 mm (Meteo-France, 1951-2005 records; Lebouvier & Frenot 2007). However, over the last 50 years annual temperatures and rainfall have increased by ~1.3°C and decreased by ~100-250 mm, respectively (Chapuis *et al.* 2004). Sampling on Kerguelen was conducted at 16 sites during the December 2011 (Fig. 1; Supplementary Table S1). The sites were characterised by one of three rabbit (*Oryctolagus cuniculus*) disturbance histories; specifically, that rabbits are currently present, had been present historically or had never been present. At each site we established a single 100m² (10x10m) plot.

Five sites were sampled on the Isthme Bas (Grande Terre, near Port-aux-Français). As is common on the main island, rabbits occur at high densities in this area and plant communities are often dominated by *Acaena magellanica* at the expense of other native species (Chapuis 1995). At three sites, soils were free-draining; plant communities were a near mono-specific stands of *A. magellanica* (IB03), dominated by the invasive *Taraxacum officinale* (IB02) or plants were absent (*e.g.* bare soil) (IB16). At the two other sites soils were waterlogged with the associated plant communities dominated by *A. magellanica* and *Juncus scheuchzerioides* (IB01 & IB04) (Fig. 1 & 2; Supplementary Table S1 & Fig. S2).

The remaining sites were on rabbit-free islands in the Golfe du Morbihan. We sampled five sites on Ile Guillou (1.45 km²). An eradication program was carried out on this island in 1994 and there has been no sign of rabbits since 1996 (Chapuis *et al.* 2004). At the four sites where soils were free-draining, introduced *Taraxacum* sp. (*e.g.* *T. erythrospermum* and *T. officinale*) were highly abundant with *Festuca contracta* or *A. magellanica* also present (*e.g.* GU11-GU14). The fifth site had waterlogged soils and supported a mixed *A. magellanica* and *J. scheuchzerioides* community (GU15). Rabbits were not introduced to nor have otherwise arrived on Ile Australia (37 km²). On this island we sampled six sites. Four of these (*e.g.* AU05, AU06, AU07, AU09) represent

indigenous plant communities, including one disturbed by natural erosion (AU06); these communities are dominated by *Azorella selago*, *Festuca contracta* and *Blechnum pennamarina*. The remaining two sites were an almost mono-specific *A. magellanica* heathland (AU08) and a second naturally disturbed site dominated by the introduced species, *T. officinale* (AU10) (Fig. 1 & 2; Supplementary Table S1 & Fig. S2).

Soil sampling

A soil sampling procedure should aim to obtain samples that represent as fully as possible local biodiversity (Taberlet *et al.* 2012). Here, we do this by combining 100 soil cores taken on a regular grid of 1 x 1m within each plot; each core consisted of a 20–50 g sample taken from the uppermost 10 cm of soil. To ensure our sampling was representative two sets of cores were collected per plot. Soil corers were cleaned and sterilized between plots to prevent cross contamination.

Extraction of extracellular DNA

The adsorption of extracellular DNA (*e.g.* that released by the lysis of prokaryotic and eukaryotic cells) on to soil colloids and sand particles inhibits their degradation by nucleases (Pietramellara *et al.* 2009). In many environments, but particularly in cold climate areas, such DNA can persist over time and thereby provide a basis for investigating the communities from which it was derived (Taberlet *et al.* 2012; Yoccoz *et al.* 2012). Analysis of extracellular DNA in the soil provides a longer-term view of local diversity than would be obtained using total DNA. Using extracellular DNA reduces the potential impacts of short-term changes that could have marked effects on microbial communities (*e.g.* seasonal) (Zinger *et al.* 2009) and bias towards DNA sequences of living organisms (*e.g.* from rootlets, seeds, insect larvae and fungi) (Taberlet *et al.* 2012).

To limit the potential impacts of storage on microorganism community composition extractions of extracellular DNA were carried out as soon as possible after soil samples were collected. One or two DNA extractions were performed per soil sample (*e.g.* 2 or 4 extractions per plot). This strategy aims to maximise species detection as well as allowing us to differentiate between sequences that result from the actual presence of taxa and those that are likely to be due to sporadic contamination (Giguët-Covex *et al.* 2014). It also allows the effects of variation in sample processing and the overall reproducibility of the results to be monitored.

Soil samples were weighed and transferred to wide-neck barrels. An equal weight of saturated phosphate buffer (Na_2HPO_4 ; 0.12 M; pH 8) was added to each barrel and the mixture vigorously agitated for 15 minutes to release particle-bound DNA (Taberlet *et al.* 2012). A 50ml aliquot of the resulting mixture was filtered using a 0.2 μm filter. Extracellular DNA was extracted from the filtrate using the NucleoSpin® Soil kits (Macherey-Nagel, Düren, Germany) and a modified version of the manufacturer's instructions. We followed Taberlet *et al.* (2012) in omitting the lysis step. In addition, we stored columns on silica gel after the first SW2 buffer washing to allow transport at room temperature from Kerguelen. In addition to soil samples, at least one extraction control (using unused saturated phosphate buffer) was also performed on each extraction day. Just prior to PCR amplification, columns were again washed with SW2 buffer and the DNA eluted in 100 μl of the elution buffer.

DNA amplification and high-throughput sequencing

Markers used for soil *metabarcoding* must be short enough to allow consistent PCR amplification of degraded DNA from environmental samples (Valentini *et al.* 2008). However, there is a trade-off between marker length and taxonomic resolution; shorter markers provide more limited taxonomic resolution (Taberlet *et al.* 2007; Chariton 2012). An additional consideration is that priming sites flanking the marker should be conserved within and specific to the taxonomic group of interest. Meeting these criteria limits the chances of amplification biases from a mixed pool of DNA and amplification of non-target taxa (Valentini *et al.* 2008; Epp *et al.* 2012). Finally, to maximize the chances of detection it is preferable to use markers present in multiple copies per cell (Bellemain *et al.* 2013).

We used the P6 loop region of the chloroplast *trnL* (UAA) intron for *metabarcoding* of plant DNA in our soil samples. This variable region of the chloroplast genome was amplified using the universal primers “*g*” (5'-GGGCAATCCTGAGCCAA-3') and “*h*” (5'-CCATTGAGTCTCTGCACCTATC-3') (Taberlet *et al.* 2007). Amplification products typically range in length between ~10 and ~143 bp. For fungal DNA we used a primer pair proposed by Epp *et al.* (2012). These amplify a portion of Internal Transcribed Spacer 1 (ITS1) in the ribosomal RNA repeat that has been suggested as a standard marker for fungi. Specifically, we use primers “*ITS5*” (5'-GGAAGTAAAAGTCGTAACAAGG-3'; White *et al.* 1990) and “*5.8S_fungi*” (5'-

CAAGAGATCCGTTGTTGAAAGTT-3'; Epp *et al.* 2012). The length of the amplified fragment ranges from ~120 to ~270 bp. To allow multiple PCR products to be sequenced within a single high-throughput sequencing run, 8 nucleotide (nt) tags (differing by at least 5 nt) were added to the 5' end of each primer (Binladen *et al.* 2007; Valentini *et al.* 2008). Two PCR amplifications were carried out per extraction (*e.g.* 4 or 8 PCRs per plot); extraction controls were also amplified twice. Additionally 10 PCR controls were performed. Controls were placed randomly on PCR plates.

PCR amplifications were carried in reaction volumes of 30 μ L containing 2 μ L of DNA extract. Amplification mixtures contained 1.2U of AmpliTaq Gold® DNA Polymerase (Applied Biosystems), 15 mM Tris-HCL, 50 mM KCl, 2 mM MgCl₂, 0.2 mM of each dNTP, 0.2 μ M of each primer and 4.8 μ g of bovine serum albumin (BSA, Roche Diagnostic). Thermocycling conditions were an initial 10min at 95°C for enzyme activation, followed by 45 cycles of 30s at 95°C, 30s at 50°C or 55°C (for plant and fungi amplifications, respectively) and 1min at 72°C with a final elongation step of 7min at 72°C. PCR products were then purified and mixed together. High-throughput sequencing of plant and fungal amplicons was performed with Illumina HiSeq 2500 (2 x 100 bp paired-end reads) and Illumina MiSeq (2 x 250 bp paired-end reads) platforms, respectively.

Processing the high-throughput sequencing data

The OBITOOLS software (<http://metabarcoding.org/obitools>) was used to process high-throughput sequence data. As we performed paired-end sequencing direct and reverse reads from the same DNA fragment were first aligned and merged using the gap-free end algorithm implemented in the *illumina-paired-end* program. Sequences with low alignment quality scores (*e.g.* <40, that is pairs of reads that could not be aligned perfectly for at least 10 nt) were discarded. The resulting sequences were then processed using the following steps. (i) Sequences were filtered according to length using the *obigrep* program; plant sequences <7 bp and fungal sequences <100 bp were discarded. (ii) Sequences were sorted by primers and tags and thereby assigned to individual soil samples using the *ngsfilter* program. Only those with perfect tags and primers containing a maximum of three differences were retained for further analyses (Bienert *et al.* 2012). (iii) We used the *obiuniq* program to merge identical sequences while retaining information about the samples from which they were obtained. Strict singletons (*e.g.* sequences represented only once in the whole dataset) were discarded.

When analysing data from high-throughput sequencing it is usual to cluster highly similar sequences (Zinger *et al.* 2011). However, this is not appropriate when using short DNA markers since single nucleotide differences can differentiate species. Instead, we used the *obiclean* program (Bellemain *et al.* 2013; Giguët-Covex *et al.* 2014) to detect sequences that potentially contained amplification and/or sequencing errors. For each PCR the program creates networks that connect sequences that differ by a single nucleotide substitution or indel. Within this framework, two sequences were considered linked if they differ in frequency by more than 50%; the less frequent is assumed to derive from the other. Each sequence in the network was then assigned to one of three categories, “head” (*e.g.* the most common sequence of a network), “singleton” (*e.g.* a sequence without a variant) or “internal” (*e.g.* sequences that do not fall within either of the other categories and are therefore likely to be errors). Sequences whose status was more commonly “internal” than “head” or “singleton” in the global dataset were discarded (Giguët-Covex *et al.* 2014).

High-throughput sequencing data can be biased by contamination and/or stochasticity in the amplification and sequencing processes. It is therefore necessary to identify and correct for such biases. We considered sequences whose frequencies were highest in control samples to be contaminants and these were discarded. Next a graph-partitioning approach based on Bray-Curtis distances and implemented in the *igraph* package of R (Csardi & Nepusz 2006) was used to filter out outlying PCR replicates as well as those that clustered with control samples. Additionally, since we expect replicate samples from a same plot to cluster together, PCR replicates that did not cluster with others from the same plot were also discarded.

We clustered fungal sequences into molecular operational taxonomic units (MOTUs) using a sequence identity criterion of 97% (Blaalid *et al.* 2013). We used the *sumatra* program to align fungal sequences (Mercier *et al.* submitted, <http://metabarcoding.org/sumatra>) and MCL (van Dongen 2000) to cluster them. All unique plant sequences were considered to be MOTUs. MOTUs were assumed to be present in a PCR replicate if their read count was >5 (Lindhal *et al.* 2013).

Taxonomic assignments

Sequences were assigned to taxa using the *ecoTag* program; this identifies sequences based on matching to a reference database. For fungal MOTUs, taxonomic

assignments were performed using the most frequent sequence of each MOTU and reference databases generated by *in silico* PCR from both EMBL and UNITE using the *ecoPCR* program (Ficetola *et al.* 2010). MOTUs were discarded if the identity score of the best database match was <70%.

Plant sequences were compared to a global plant database generated by *in silico* PCR from EMBL and to an exhaustive database of vascular plants found on the Kerguelen Islands (we thank Nigel Gilles Yoccoz for making this database available). Both databases were used to assign MOTUs to taxa. MOTUs for which the best matches had <95% identity in both databases were excluded from further analyses; in all other cases matches to the local database were used except if a match to the global database had a higher identity score.

Statistical analysis

Prior to statistical analysis we removed fungal and plant MOTUs if they were present in fewer than half the replicates for a given plot. This procedure removes MOTUs derived from DNAs that were at very low frequency within a plot and that therefore potentially represent contaminants or errors. The fungal MOTU matrix was further rarefied to 5000 reads per PCR replicate and the plant MOTUs data transformed to frequencies.

Metabarcoding-based fungal and plant communities were ordinated using non-metric multidimensional scaling (NMDS) of Bray-Curtis distances (Brocard *et al.* 2011). The homogeneity of results from extraction and PCR replicates for a given plot was visually controlled; centroids were calculated for fungal and plant communities at each site. Relationships between centroids and soil parameters (Supplementary Table 1) were assessed using vector fitting. For fungal communities a number of biotic parameters were also examined; specifically, overall plant cover, plant community diversity (Shannon Index) as well as per plot frequencies of *A. magellanica*, other native species and introduced species. The significance of correlations was assessed using 10000 random permutations. Finally, a Procrustes analysis was used to compare the ordinations of fungal and plant communities.

Statistical analyses were performed in R (R Development Core Team 2013) using *vegan* (Oksanen *et al.* 2013) and *ROBITools*, a package dedicated to the analysis of *metabarcoding* data (<https://www.grenoble.prabi.fr/svn/LECASofts/ROBITools/>).

Additional data sets and analyses

For comparison we collected and analysed two additional datasets.

First, we conducted vegetation surveys within each of the 16 Kerguelen plots; these data were used to examine the relationship between the actual plant community and that indicated by *metabarcoding*. Prior to collecting soils at each site we estimated plant species abundances using a contact sampling approach. Two 10-meter linear transects were established within the plot and sampled at 10 cm intervals; for a maximum total of 200 contacts per plot. Statistical analyses of these data were similar to those used on *metabarcoding* data. Plant communities based on botanical surveys were ordinated using NMDS and we examined relationships with soil parameters using vector fitting. A Procrustes analysis was used to compare the ordinations of plant communities based on *metabarcoding* and botanical surveys. We also used linear models to assess the relationship between the relative abundance of taxa identified by these two approaches.

Second we compared fungal diversity from Kerguelen with that of continental areas exposed to cold climates, specifically high-altitude meadows (1500-2500m a.s.l.) in the French Alps. We examined soil samples from 100m² plots on the Vercors Massif (Isère, France; three plots) and the Arves Massif (Savoie, France; four plots). Each soil sample was composed of 30-40 individual soil cores. Protocols for the extraction of extracellular DNA, amplification and sequencing of the fungal metabarcode as well as processing of high-throughput sequencing data were largely identical to those described above for Kerguelen. The one difference being that at most two PCR replicates were performed per sample. We compared the diversity of fungal MOTUs on Kerguelen with that of the French Alps using a *t*-test. For this analysis, the fungal MOTU matrix was rarefied to 5000 reads per PCR replicate but not filtered to remove MOTUs that did not appear in multiple replicates.

d. Résultats complémentaires

Supplementary results

Supplementary results

Table 1: Characteristics of the study sites

Figure 1: Photographs illustrating study sites and soil samples

Figure 2: Per plot plant species abundances based on botanical surveys

Figure 3: Relationship between the relative abundance of plant species as determined by soil *metabarcoding* analyses and botanical surveys on the Kerguelen Islands

Figure 4: Comparison of Kerguelen Island plant communities determined by *metabarcoding* analysis and botanical surveys.

Figure 5: Comparison of botanical survey-based, plant DNA-based and fungal DNA-based ordinations with vector fitting of soil parameters.

Figure 6: Box and whisker plots describing the numbers of rarefied fungal MOTUs per plot

Supplementary Results

Plant communities

High-throughput sequencing of plant DNA amplifications resulted in 1325038 complete reads, filtering reduced this to 997252 reads with 25 unique sequences (Supplementary Table 2). These were assigned to species-level taxa if they shared 100% identity with a reference accession or if identified at a higher taxonomic level and on our survey indicated this taxon was represented by a single species in the corresponding plot. Using this approach we identified 11 of the 18 vascular plant taxa that were recorded in vegetation surveys. The fact that we recovered lower overall diversity using *metabarcoding* than with surveys likely reflects the limited taxonomic resolution of the marker we used. Our marker was unable to differentiate species within *Agrostis*, *Taraxacum* and *Poa*. Sequences that could be assigned to either *Taraxacum* or *Agrostis* were grouped together into the categories “*Taraxacum* sp.” and “*Agrostis* sp.”, respectively. Sequences that could be assigned to Poaceae family but not identified further were placed into the category “other Poaceae”. This difference may also be partly explained by conservative data processing. Our metabarcode identified *Pringlea antiscorbutica* and *Deschampsia antactica* in two and one plots, respectively. However, these data were excluded from further analysis because these species were also identified in corresponding extraction controls and therefore considered stochastic contaminants. That said, our survey data indicates *P. antiscorbutica* and *D. antarctica* occurred in these, and only these, plots suggesting that the *metabarcoding* may have reflected presence rather than contamination.

Statistical analyses comparing data from botanical surveys and *metabarcoding* of extracellular DNA in the soil suggest that the latter provides a good proxy of current plant biodiversity. Specifically, we found a strong correlation between the relative abundances of plant species inferred from *metabarcoding* and survey data ($r^2=0.83$, $P<0.001$; Supplementary Fig. 3). Although some species were missed by *metabarcoding* (“false negatives”) comparison of species lists indicates that this approach did not indicate the presence of plant species not recorded during our surveys (“false positives”).

Yoccoz *et al.* (2012) have suggested that soil DNA content is largely determined by biomass turnover. Since on Kerguelen there is limited plant diversity and the communities are relatively simple (*e.g.* a single layer of vegetation) we suggest that

there is limited potential for large differences between the current aboveground plant community and that which was responsible for extracellular DNA in the soil. That said, we note that *metabarcoding* tends to under-represent some species (*e.g.* *Cerastium fontanum* and *Montia fontana*) while systematically over-representing others (*e.g.* *Azorella selago*). We suspect this reflects growth form differences. Specifically, under-represented taxa are small, weedy species while the over-represented species may form large cushions; in some cases soil cores may have been taken directly through a cushion.

That *metabarcoding* and vegetation surveys provide similar results is also supported by Procrustes analysis of the ordinated communities (Procrustes sum of squares=0.2291, correlation in a symmetric Procrustes rotation=0.978, $P=0.001$; Supplementary Fig. 5). In both cases vector fitting to ordinations of plant communities suggests that species diversity is significantly correlated with soil organic matter content ($P=0.01$ and $P=0.04$ for metabarcode- and survey-based ordinations, respectively) but not to either soil acidity ($P=0.12$ and $P=0.57$) or soil carbon/nitrogen ratio ($P=0.31$ and $P=0.35$).

We distinguished three main plant community types (Fig. 2; Supplementary Fig. 2 & 4), (i) species-rich native communities with high evenness but tending to be dominated by large *A. selago* cushions, *B. penna-marina* and *F. contracta*; (ii) dense *Acaena*-dominated communities that are otherwise species-poor; and (iii) low-cover communities on eroded soils dominated by invasive plant species, especially *Taraxacum* sp. Plant community diversity was highest in plots at undisturbed sites on Ile Australia (*e.g.* AU05, AU06, AU07, AU09). Naturally disturbed sites on this island (*e.g.* AU10) or rabbit-disturbed sites on the Isthme Bas (*e.g.* IB01-IB04, IB16) had lower plant diversity (Supplementary Fig. 2). At rabbit-disturbed sites grazing and burrowing activities have led to the replacement of species diverse native communities, often by dense mono-specific stands of *A. magellanica* or by mixed communities of *A. magellanica* and *Taraxacum* sp. (Chapuis *et al.* 2004). At plots on Ile Guillou where the vegetation is recovering from rabbit disturbance (*e.g.* GU11-GU15) plant communities remain more similar to those on the Isthme Bas than undisturbed communities on Ile Guillou.

Fungal metabarcoding results

We obtained 8602068 complete reads from high-throughput sequencing of the fungal marker from Kerguelen samples and 735295 from samples collected in the

French Alps. Filtered datasets contained 853 and 771 MOTUs respectively. The mean number of rarefied MOTUs per plot in Kerguelen Islands ranges from 12.5 at the no vegetation site (*e.g.* IB16) to over 130 at sites unaffected by rabbits (*e.g.* AU06, AU09) (Table 1; Supplementary Fig. 6). Per plot richness is higher in the French Alps where the number of rarefied MOTUs was 111-201 (Table 1; Supplementary Fig. 6). However, we note that relatively few MOTUs showed strong identity with sequences in the reference database.

Vector fitting to the ordination of soil fungal communities suggests significant correlations between diversity and both soil and plant community parameters (Fig 2a). In terms of soil parameters axes of discrimination were related to organic matter content ($P<0.002$) and soil carbon/nitrogen ratio ($P<0.002$) but not to soil acidity ($P=0.14$). For plant community parameters axes of discrimination were related to plant cover, plant community diversity (Shannon index), per plot frequency of *Acaena* and per plot frequency of other native species (all $P<0.006$); no significant relationship was recovered with the frequency of introduced species ($P=0.5$). From these results we infer two axes of discrimination. The first relates to organic matter content, soil carbon/nitrogen ratio and degree of plant cover at a plot. This axis differentiated communities at plots with eroded or unstable soils and reduced plant cover from those on more stable sites with dense vegetation. The second axis related to plant community diversity at a plot. Specifically, to the Shannon index as well as per plot frequencies of *Acaena* and of other native species. This axis differentiated species poor *Acaena*-dominated plots from more hosting more diverse native communities (Fig. 3a).

Procrustes analysis of the ordinated fungal and plant communities indicate these are more similar to one another than expected by chance whether considering metabarcode- or survey-based estimates of plant communities (Procrustes sum of squares=0.4927-0.6196, correlation in a symmetric Procrustes rotation=0.6167-0.7122, $P=0.001$ -0.004; Supplementary Fig. 5). It is therefore not surprising that in terms of MOTUs diversity soil fungal communities exhibit the same general trends as plant communities. That is, sites on Ile Australia that have not been exposed to rabbits support higher numbers of MOTUs than sites currently disturbed by (*e.g.* Isthme Bas) or recovering from disturbance by rabbits (*e.g.* Ile Guillou) (Supplementary Fig. 6). Mean numbers of rarefied MOTUs per plot on Ile Australia was above 90 for most sites whereas vegetated sites on Ile Guillou and the Isthme Bas generally had ~60-90 MOTUs

per plot (Table 1; Supplementary Fig. 6). Notice that the two plots on Ile Australia where MOTUs diversity is most similar to that of sites on Ile Guillou and the Isthme Bas (*e.g.* AU08 and AU10) are also most similar to these areas in terms of the richness and evenness of their plant communities. Conversely, the plant community at the plot on the Isthme Bas with the highest MOTUs diversity (*e.g.* IB02) is richer and slightly more even than others in this area (Table 1; Fig. 2b; Supplementary Fig. 2; Supplementary Fig. 6). There is an exception to this overall pattern. One plot on each of Ile Australia (*e.g.* AU07) and Ile Guillou (*e.g.* GU15) has lower soil fungal diversity, in each cases ~55 MOTUs. In both cases soils at these sites are also more acidic (Fig. 2b; Supplementary Table 1).

Supplementary references

- Bellemain E *et al.* 2013 Fungal palaeodiversity revealed using high-throughput *metabarcoding* of ancient DNA from arctic permafrost. *Environ. Microbiol.* **15**, 1176–1189.
- Bienert F, De Danieli S, Miquel C, Coissac E, Poillot C, Brun JJ, Taberlet P. 2012 Tracking earthworm communities from soil DNA. *Mol. Ecol.* **21**, 2017–2030.
- Binladen J, Gilbert MTP, Bollback JP, Panitz F, Bendixen C, Nielsen R, Willerslev E. 2007 The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *Plos One* **2**, e197.
- Blaalid R, Davey ML, Kauserud H, Carlsen T, Halvorsen R, Høiland K, Eidesen PB. 2014 Arctic root-associated fungal community composition reflects environmental
- Borcard D, Gillet F, Legendre P. 2011 *Numerical Ecology with R*. Springer, New York.
- Chapuis JL, Barnaud G. 1995 Rehabilitation of islands of Kerguelen archipelago by eradication of the rabbit (*Oryctolagus cuniculus*): Method applied to Ile Verte. *Revue D Ecologie-La Terre Et La Vie* **50**, 377–390.
- Chariton A. 2012 Short and informative DNA products to indirectly measure vascular plant biodiversity. *Mol. Ecol.* **21**, 3637–3739.
- Csárdi G, Nepusz T. 2006 The igraph software package for complex network research. *InterJournal Complex Systems*, **1695**, 5.
- Ficetola GF, Coissac E, Zundel S, Riaz T, Shehzad W, Bessiere J, Taberlet P, Pompanon F. 2010 An *in silico* approach for the evaluation of DNA barcodes. *BMC Genomics* **11**, 434.
- Lebouvier M, Frenot Y. 2007 Conservation and management in the French Sub-Antarctic islands and surrounding seas. *Pap. Proc. R. Soc. Tasmania.* **14**, 23–28.

- Giguët-Covex C *et al.* 2014 Long livestock farming history and human landscape shaping revealed by lake sediment DNA. *Nat. Commun.* **5**, 3211.
- Lindahl BD *et al.* 2013 Fungal community analysis by high-throughput sequencing of amplified markers - a user's guide. *New Phytol.* **199**, 288-99.
- Pietramellara G, Ascher J, Borgogni F, Ceccherini MT, Guerri G, Nannipieri P. 2009 Extracellular DNA in soil and sediment: fate and ecological relevance. *Biol. Fertil. Soils* **45**, 219-235.
- Valentini A *et al.* 2009 New perspectives in diet analysis based on DNA *barcoding* and parallel pyrosequencing: the *trnL* approach. *Mol. Ecol. Resour.* **9**, 51-60.
- van Dongen S. 2000 Graph clustering by flow simulation. University of Utrecht
- White T, Bruns T, Lee S, Taylor J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: *PCR-Protocols A Guide to Methods and Applications* (eds Innis MA, Gelfand DH, Sninski JJ and White TJ), pp. 315-322. Academic Press, San Diego.
- Zinger L, Shahnava B, Baptist F, Geremia RA, Choler P. 2009 Microbial diversity in alpine tundra soils correlates with snow cover dynamics. *ISME J.* **3**, 850-859.

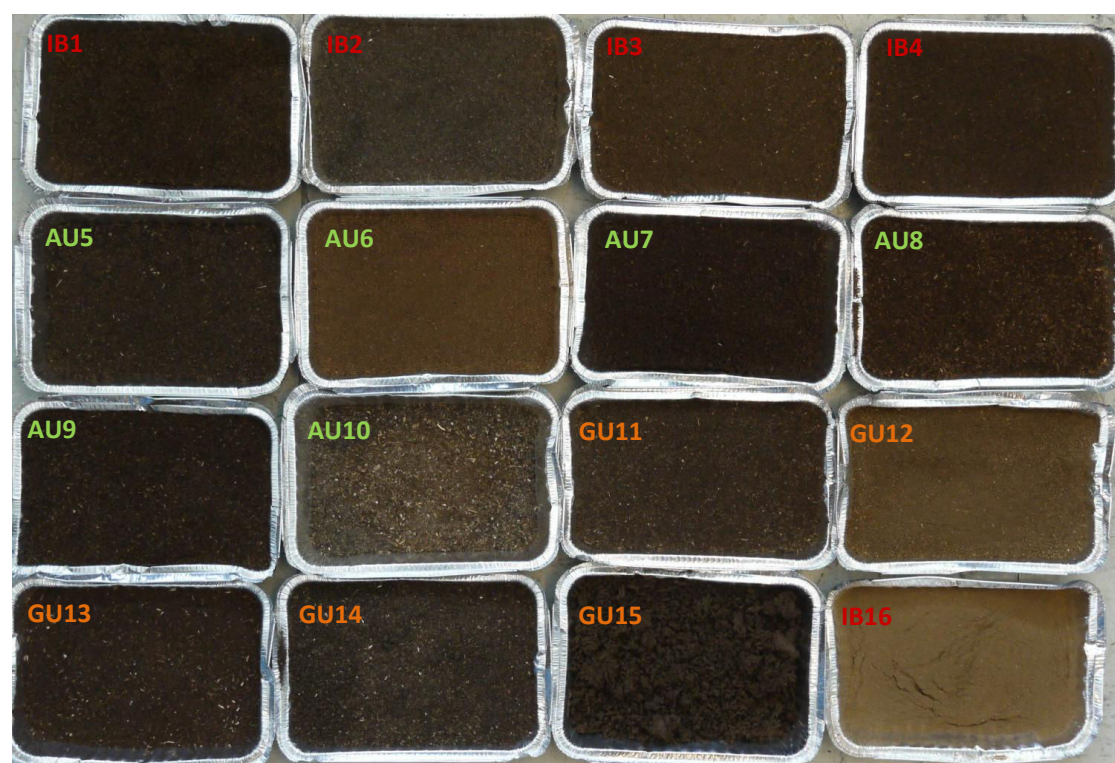
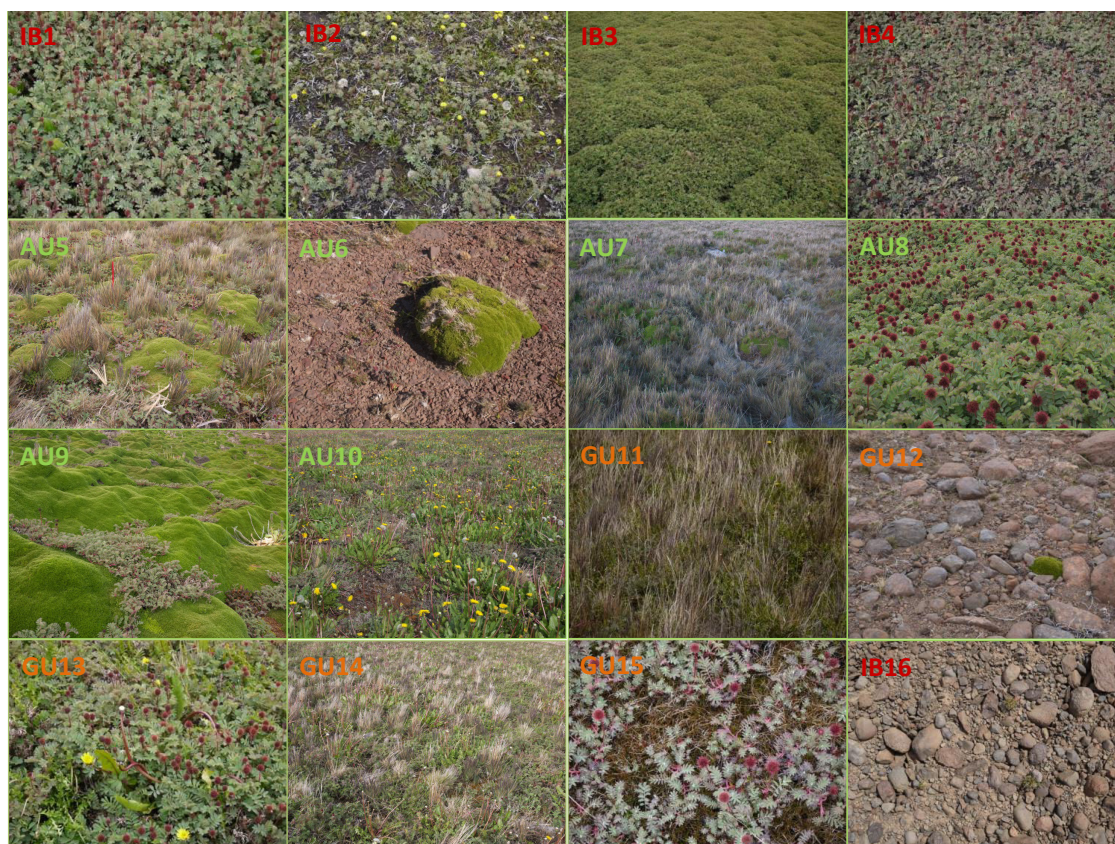
Supplementary Table 1: Characteristics of study sites on the Kerguelen Islands

pH: soil acidity, C/N: Carbon/nitrogen ratio, OM: Organic matter

Location	Rabbit population	Plot	Soil Conditions			Number of PCR replicates
			pH	C/N	OM	
Isthme Bas	High density	IB01	5.49	11.28	53.88	8
		IB02	5.61	11.53	22.84	8
		IB03	5.37	13.39	14.48	8
		IB04	5.59	10.76	54.13	4
		IB16	5.50	14.79	7.89	4
Ile Australia	Never present	AU05	5.78	14.01	25.36	8
		AU06	5.79	14.19	8.51	8
		AU07	4.65	12.23	26.84	4
		AU08	5.34	14.11	36.11	4
		AU09	5.75	12.08	27.39	8
		AU10	6.26	11.21	22.42	4
Ile Guillou	Absent since 1996	GU11	5.63	12.46	20	8
		GU12	6.15	13.23	4.43	4
		GU13	5.53	12.50	24.42	8
		GU14	5.55	11.80	20.62	4
		GU15	4.54	9.56	38.84	4

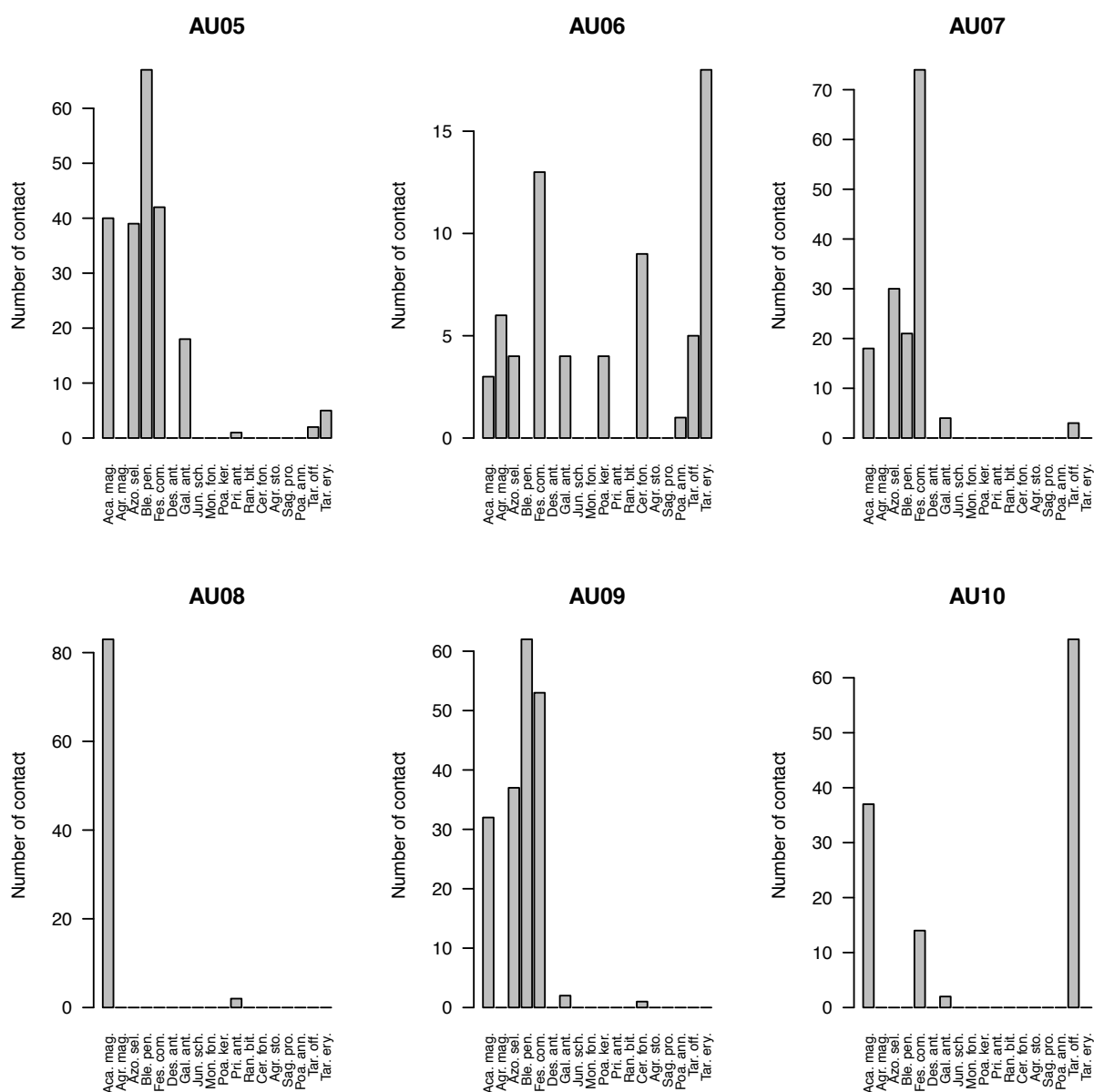
Supplementary Figure 1: Photographs illustrating study sites and soils sampled

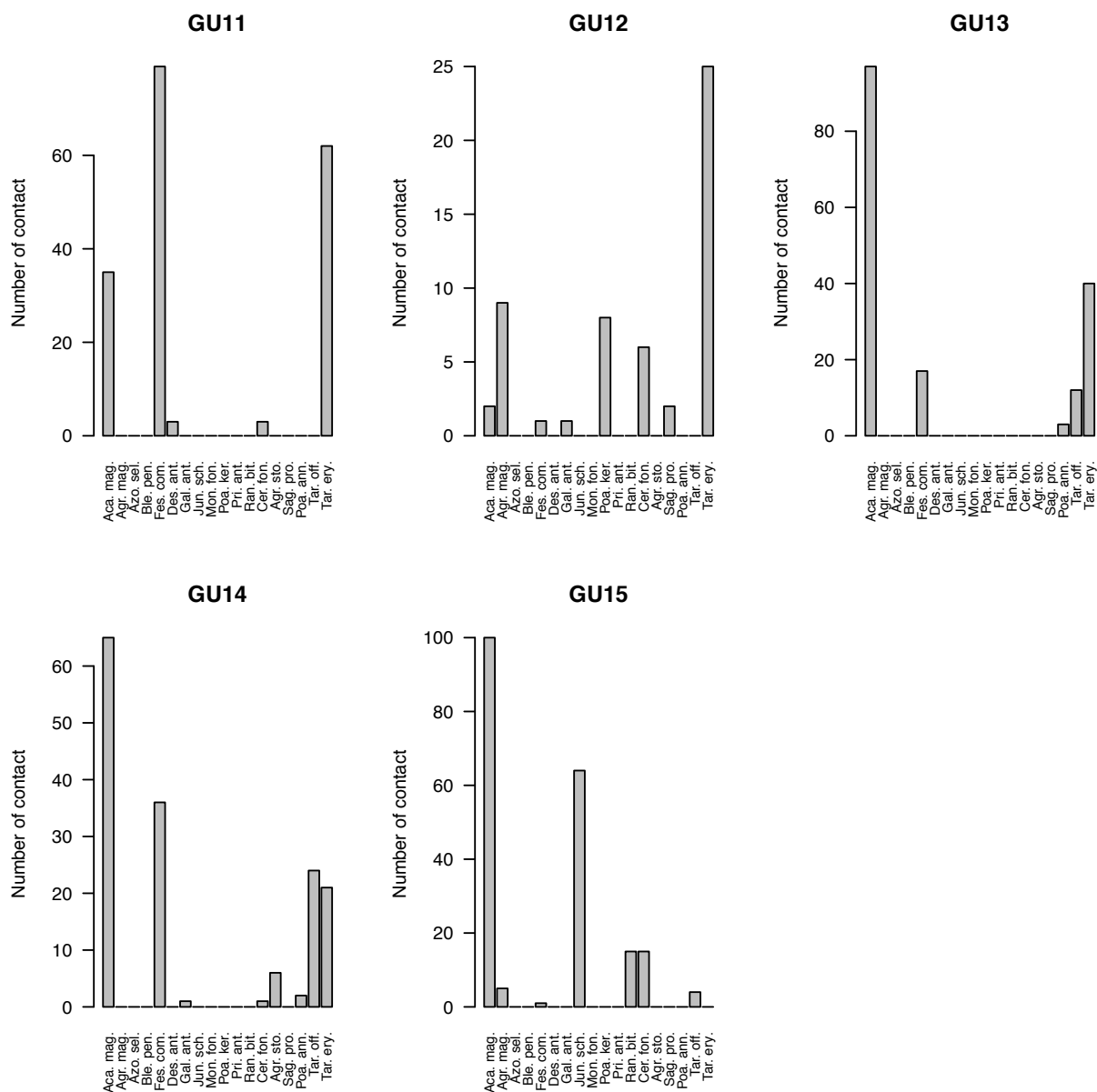
- a. Study sites
- b. Soils sampled (2 mm-sieved topsoils (0-10 cm))

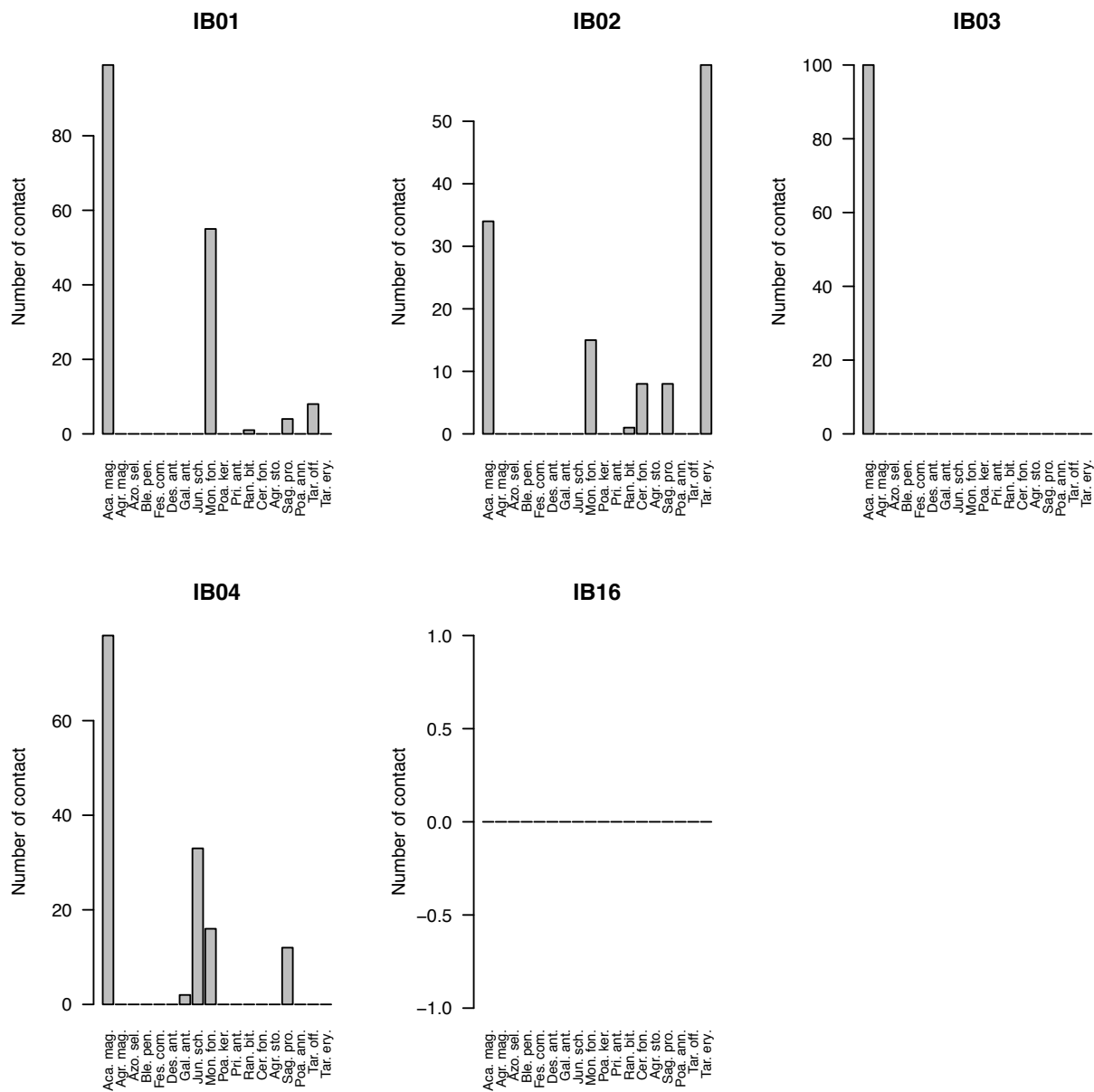


Supplementary Figure 2: Per plot plant species abundances based on botanical surveys

Taxa abbreviations are: Aca. mag., *Acaena magellanica*; Agr. mag., *Agrostis magellanica*; Azo. sel., *Azorella selago*; Ble. pen., *Blechnum penna-marina*; Fes. con., *Festuca contracta*; Des. ant., *Deschampsia antarctica*; Gal. ant., *Galium antarcticum*; Jun. sch., *Juncus scheuchzerioides*; Mon. fon., *Montia fontana*; Poa. ker., *Poa kerguelensis*; Pri. ant., *Pringlea antiscorbutica*; Ran. bit., *Ranunculus biternatus*; Cer. fon., *Cerastium fontanum*; Agr. sto., *Agrostis stolonifera*; Sag. pro., *Sagina procumbens*; Poa. ann., *Poa annua*; Tar. off., *Taraxacum officinale*; Tar. ery., *Taraxacum erythrospermum*.

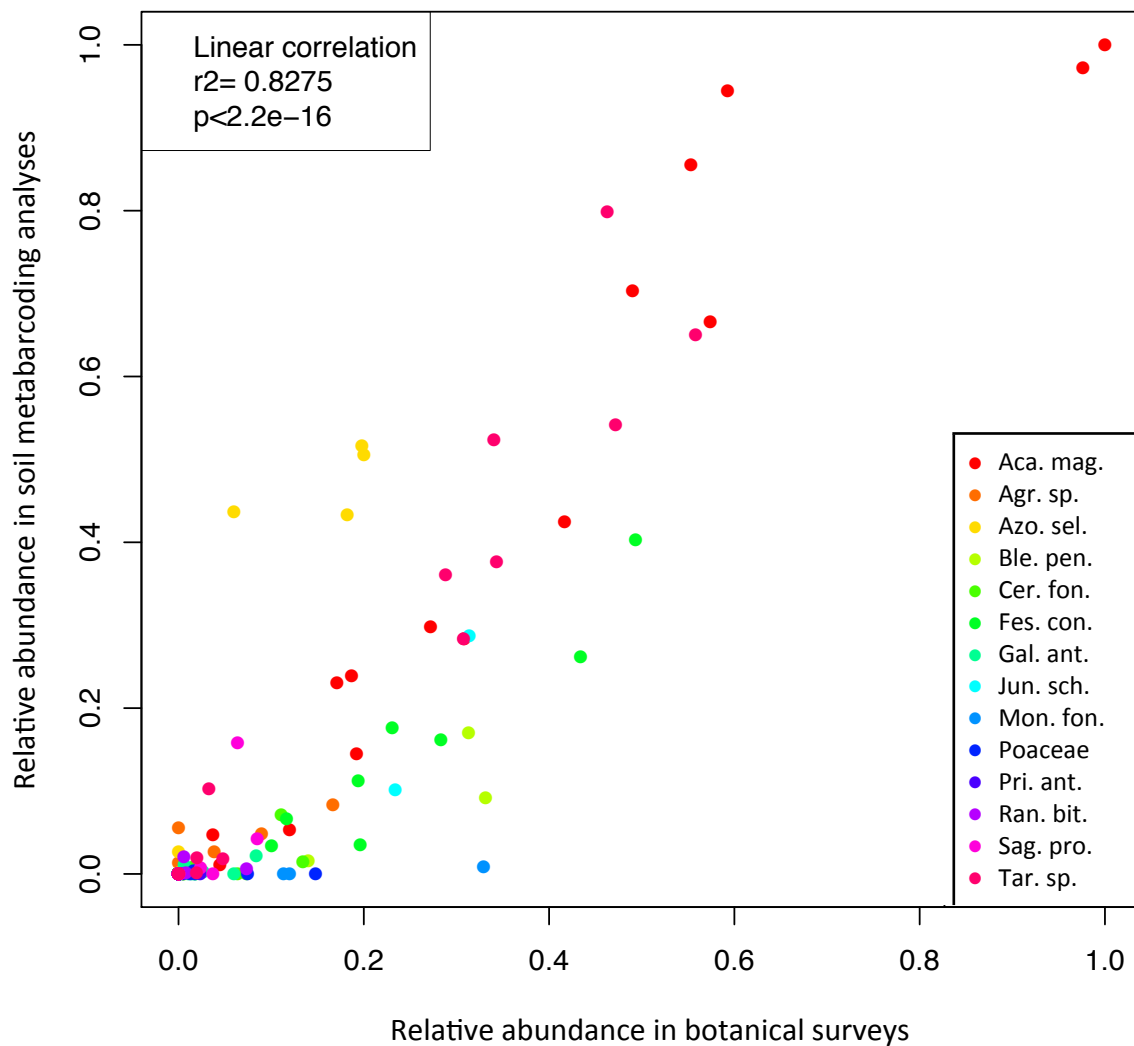






Supplementary Figure 3: Relationship between the relative abundance of plant species as determined by soil *metabarcoding* and botanical surveys on the Kerguelen Islands

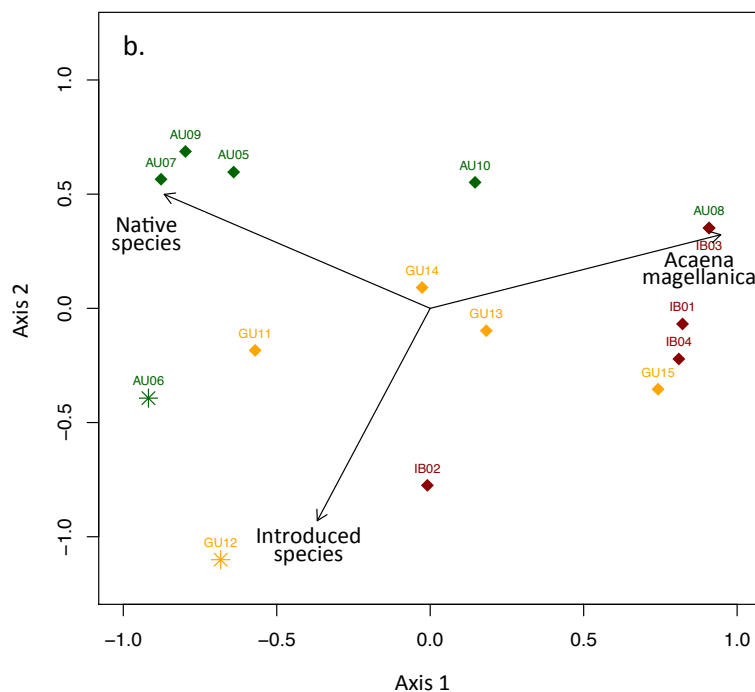
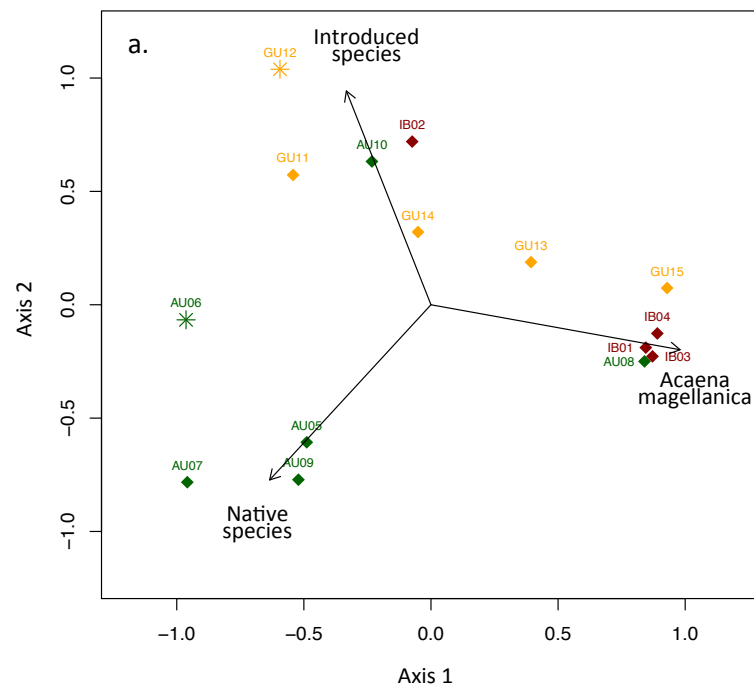
Taxa abbreviations are: *Aca. mag.*, *Acaena magellanica*; *Agr. sp.*, *Agrostis* sp.; *Azo. sel.*, *Azorella selago*; *Ble. pen.*, *Blechnum penna-marina*; *Cer. fon.*, *Cerastium fontanum*; *Fes. con.*, *Festuca contracta*; *Gal. ant.*, *Galium antarcticum*; *Jun. sch.*, *Juncus scheuchzerioides*; *Mon. fon.*, *Montia fontana*; *Poaceae*, Poaceae family; *Pri. ant.*, *Pringlea antiscorbutica*; *Ran. bit.*, *Ranunculus biternatus*; *Sag. pro.*, *Sagina procumbens*; *Tar. sp.*, *Taraxacum* sp.



Supplementary Figure 4: Kerguelen Island plant communities as determined by *metabarcoding* analysis and botanical surveys

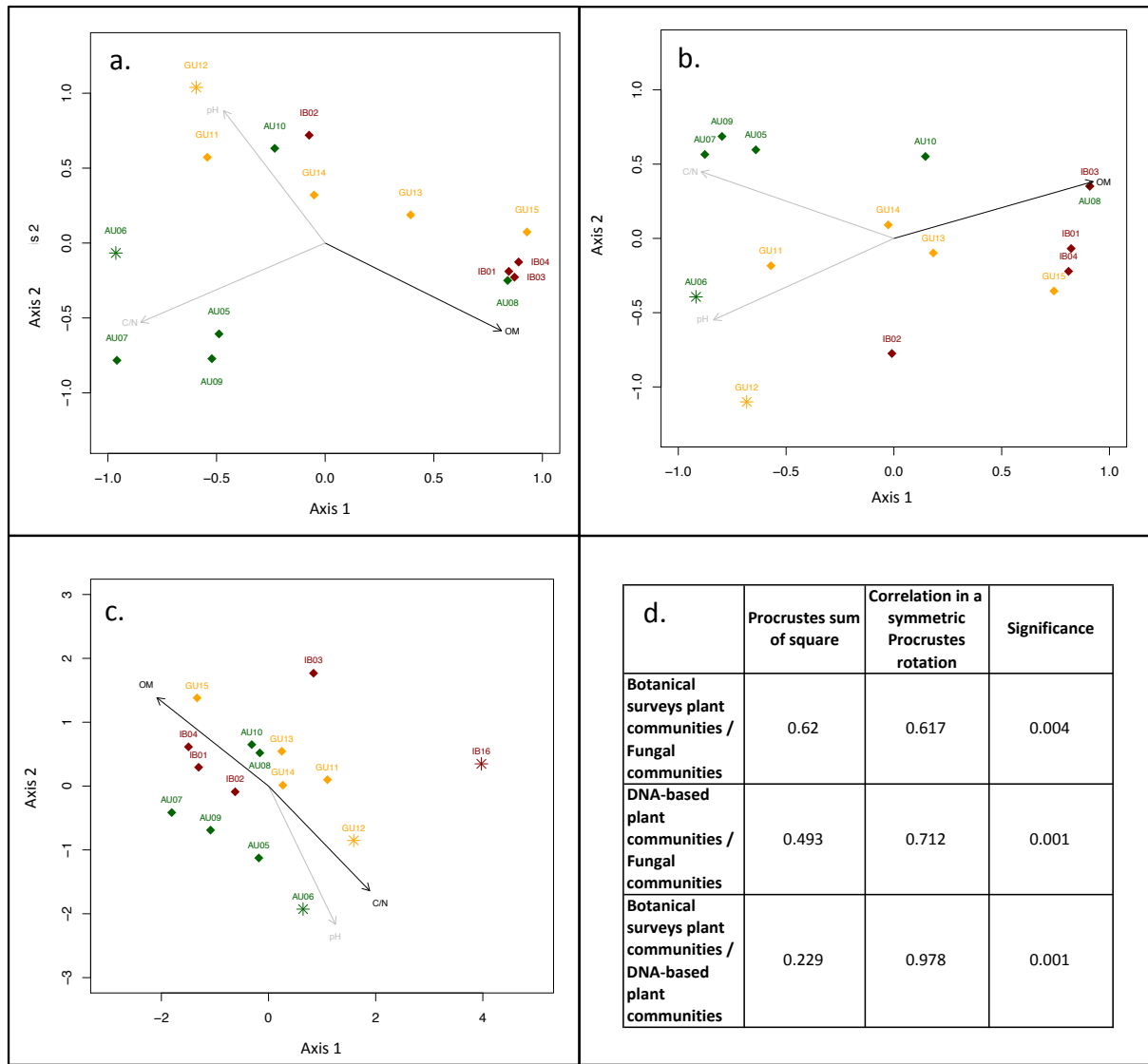
Both approaches resulted in similar ordinations (Procrustes sum of squares = 0.229, correlation in a symmetric Procrustes rotation = 0.978, $P = 0.001$). We particularly distinguish three types of communities: *Acaena*-dominated, dominated by native species and dominated by invasive species.

- Ordination of plant communities based on the frequency of DNA sequences in plant *metabarcoding* analyses (stress = 0.066)
- Ordination of plant communities based on plant species frequency in botanical surveys (stress = 0.081)



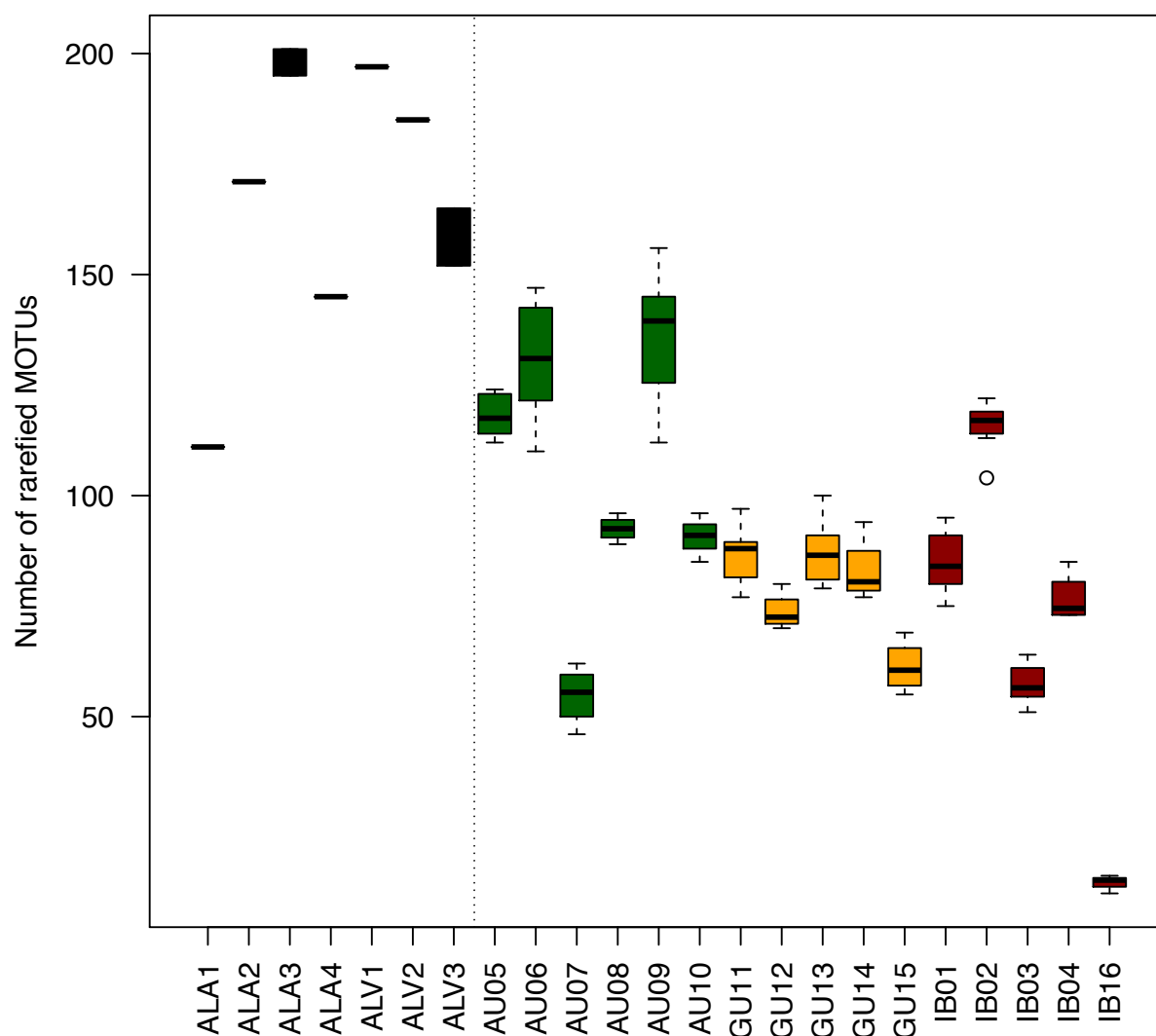
Supplementary Figure 5: Comparison of botanical survey-based, plant DNA-based and fungal DNA-based ordinations with vector fitting of soil parameters. Site IB16 was excluded of these analyses because of the absence of vegetation. Gray arrows and black arrows represent non-significant and significant explanatory variables respectively ($p < 0.05$). Plant communities were not correlated with soil pH.

- a. Ordination of plant communities based on the frequency of DNA sequences in plant *metabarcoding* analyses (stress = 0.066)
- b. Ordination of plant communities based on plant species frequency in botanical surveys (stress = 0.081).
- c. Ordination of fungal communities based on the frequency of DNA sequences in fungal *metabarcoding* analyses (stress = 0.136).
- d. Results from procrustes analyses. Ordinations of fungal and botanical survey-based or soil DNA-based plant communities were more similar than would be expected by chance.



Supplementary Figure 6: Box and whisker plots describing the numbers of rarefied fungal MOTUs per plot

ALA: Arves massif (French Alps); ALV: Vercors massif (French Alps); AU: Ile Australia (Kerguelen Islands); GU: Ile Guillou (Kerguelen Islands); IB: Isthme Bas (Kerguelen Islands)



3. Détermination des patrons de distribution des lombriciens à l'échelle du paysage à partir de l'ADN du sol

a. Mise en contexte de l'étude

La disponibilité des ressources, les perturbations, l'hétérogénéité de l'habitat et les interactions biotiques sont autant de facteurs déterminant la distribution de la biodiversité des sols (Ettema & Wardle 2002 ; Wardle 2006). La composition des communautés dépend ainsi directement des caractéristiques édaphiques (Bardgett *et al.* 2005 ; Fierer & Jackson 2006 ; Zinger *et al.* 2011). Or, les activités humaines ont la particularité de modifier en profondeur la disponibilité des nutriments et les paramètres physico-chimiques des sols (Curry 2004 ; Stauffer *et al.* 2014). L'utilisation des terres est ainsi considérée comme un des facteurs majeurs influençant les patrons de biodiversité des milieux actuels (Dambrine *et al.* 2007).

En milieu alpin, la disparition de certaines pratiques traditionnelles au cours du siècle dernier a induit de profonds changements dans les paysages, l'exemple typique étant la fermeture des milieux ouverts consécutivement à la diminution des troupeaux (Brooks *et al.* 2002 ; Young *et al.* 2005). Dans le massif du Vercors (Isère, France), la majorité des forêts a ainsi été exploitée au cours du XVIIIème siècle pour la production de bois de chauffage, ce qui a favorisé le développement d'hêtraies en taillis. Depuis le début du XXème siècle, ces forêts anciennes ont été progressivement transformées, parfois par des plantations de conifères. Par ailleurs, l'abandon des prairies pâturées et des terres cultivées a mené à la recolonisation naturelle de ces milieux (Redon *et al.* 2014).

L'étude, à l'échelle du paysage, de l'impact de ces changements sur la biodiversité des sols est aujourd'hui limitée par des contraintes techniques et les biais liés aux méthodes d'échantillonnage. Cette étude a pour objectif d'évaluer l'application d'une approche de *metabarcoding* environnemental pour caractériser les communautés lombriciennes (Lumbricidae) et les facteurs régissant leurs répartitions à l'échelle du paysage. Cette famille fait partie des organismes dits « ingénieurs » des sols et joue un rôle crucial dans leur aération, l'infiltration de l'eau et le recyclage de la matière organique (Lavelle *et al.* 1997). Ils ont ainsi été largement étudiés, leur taxonomie et leur écologie étant, comparativement aux autres taxons du sol, bien connues (Decaëns *et al.* 2013). De ce fait, ils constituent un groupe candidat idéal pour l'évaluation de nouvelles approches de caractérisation des communautés animales du sol.

Les communautés lombriciennes de trois milieux à forts enjeux dans le massif du Vercors (hêtraies, sapinières anciennes, plantations de pins et prairies) ont été déterminées à la fois par *metabarcoding* environnemental et par une approche classique. L'objectif est ici (i) de valider une approche basée sur l'ADN du sol pour la caractérisation à grande échelle de ces communautés et (ii) de déterminer les facteurs influençant les patrons de distribution observés en lien avec l'utilisation actuelle des terres par l'Homme.

b. Article B

Article B

Landscape-scale distribution patterns of earthworms inferred from soil DNA

Article soumis à *Soil Biology and Biochemistry*

Pansu Johan ^{1,2}, De Danieli Sébastien³, Puissant Jérémy³, Gonzalez Jean-Maxime³, Gielly Ludovic^{1,2}, Cordonnier Thomas³, Zinger Lucie⁴, Brun Jean-Jacques³, Choler Philippe^{1,2,5}, Taberlet Pierre^{1,2} & Cécillon Lauric³

¹ Univ. Grenoble Alpes, LECA, F-38000 Grenoble, France

² CNRS, LECA, F-38000 Grenoble, France

³ Irstea EMGR, Centre de Grenoble, 2 rue de la Papeterie-BP 76 Saint-Martin-d'Hères, F-38402, France

⁴ Laboratoire Evolution & Diversité Biologique, UMR CNRS-UPS 5174, Toulouse, France

⁵ Univ. Grenoble Alpes, SAJF, F-38000 Grenoble, France

Contribution personnelle : participation à la mise en place de l'étude, rédaction du projet de financement, échantillonnage du sol, analyses moléculaires, filtrage des séquences et assignation taxonomique, analyse des données, rédaction du manuscrit.

Abstract

Assessing land-use effect on the diversity of soil biota has long been hampered by difficulties in collecting and identifying soil organisms over large areas. Recently, environmental DNA-based approaches coupled with next-generation sequencing were developed to study soil biodiversity. Here, we optimized a protocol based on soil DNA, to examine the effects of land-use on earthworm communities in a mountain landscape. This approach allowed an efficient detection of earthworm diversity and highlighted a significant land-use effect on the distribution patterns of earthworms that was not revealed by a classical survey. Our results show that the soil DNA-based earthworm survey at the landscape scale improves over previous approaches, and opens a way towards large-scale assessment of soil biodiversity and its drivers.

Short communication

Earthworm communities play an important role in terrestrial ecosystems as ecosystem engineers (Lavelle *et al.* 1997): they regulate the aeration, water infiltration, and nutrient cycling in soils. Their distribution is mainly influenced by soil properties and vegetation type (Curry 2004; Salomé *et al.* 2011). At the landscape scale, human land-use induces a strong spatial heterogeneity of earthworm communities by altering their habitat biogeochemistry and food supply (Grossi *et al.* 1995; Grossi & Brun 1997; Curry 2004; Stauffer *et al.* 2014). As the heterogeneity of earthworm communities translates to some extent into the spatial patterns of ecosystem functions (Ettema & Wardle 2002; Blouin *et al.* 2013; Hedde *et al.* 2013), earthworms are increasingly used as bioindicators of soil quality (Römbke *et al.* 2005; Pérès *et al.* 2011).

Assessment of land-use effect on the diversity of soil biota is difficult to implement over large areas due to technical constraints, bias linked with current methods of extraction, and the general lack of taxonomic skills. Current International Standard for earthworm sampling is based on handsorting and/or chemical expellant, (NF EN ISO 23611-1 2011). This method is time consuming (Bartlett *et al.* 2006), and its efficiency depends on soil parameters, season, species characteristics and life stages (Coja *et al.* 2008; Lawrence & Bowers 2002). Moreover, taxonomic assignment is often difficult, especially for juveniles, and does not account for cryptic species. During the past decade, DNA *barcoding* was successfully used for earthworm identification (Rougerie *et al.* 2009; James *et al.* 2010; Decaëns *et al.* 2013). This approach provides a

more accurate estimation of taxonomic richness by accounting for both juveniles and cryptic diversity (King *et al.* 2008; Richard *et al.* 2010; Klarica *et al.* 2012). Despite these progresses in the taxonomic identification of specimens, problems inherent to earthworm sampling methodologies limit our understanding of earthworm diversity.

Recently, Bienert *et al.* (2012) demonstrated the potentiality of a more integrative *metabarcoding* approach based on soil extracellular DNA to identify earthworm species. Persistence of DNA in soils allows overcoming constraints linked to earthworm sampling methodologies, while benefiting from the taxonomic precision of DNA *barcoding*. The authors identified soil sampling as the critical step towards an exhaustive inventory of earthworm diversity using environmental *metabarcoding* (Bienert *et al.* 2012). Here, we optimized this non-invasive soil DNA approach for surveying earthworm communities across a mountain range, to assess its potential for a landscape-scale application. Results were compared to those obtained by a classical earthworm survey and confronted to soil parameters potentially influencing earthworm distribution patterns.

The study was conducted in “Quatre Montagnes” area, a 255km² landscape in the Vercors mountain range (Northern French Alps). Eighteen homogeneous 1ha plots were investigated, reflecting 3 contrasting historical land-use trajectories (see Supplementary material): young spruce plantation (n=6), old beech coppice (n=6) and pasture (n=6) (Table 1 and Supplementary Table S1). In each plot, basic characteristics of topsoils (0-5 cm) were determined (Tables 1 & S1). A classical earthworm survey was performed in June 2012 and October 2013 in three 0,5m² sub-plots in a 20x20m square, adapted from Zaborski (2003; Supplementary data). For the assessment of earthworm diversity based on soil DNA, a composite soil sample, containing litter and organo-mineral soil (ca. 20 cm depth), was collected using the same sampling scheme (30 soil cores from three 0,5m² sub-plots). As this low-intensity sampling may not represent the actual composition of earthworm communities in 1ha plots (*e.g.* Rossi *et al.* 1997), a high-intensity soil DNA sampling scheme (duplicate composite samples from 100 soil cores distributed on two regular grids of 100x100m) was also performed on a subset of twelve 1ha plots (Supplementary data).

	Land cover		
	Old beech coppice	Young spruce plantation	Pasture
Elevation (m a.s.l.)	1397 (100)	1363 (109)	1573 (56)
Organic C (g.kg ⁻¹)	88.4 (22.8) ^a	28.9 (10.2) ^b	62.7 (24.2) ^a
Total N (g.kg ⁻¹)	5.33 (1.0) ^a	1.69 (0.5) ^b	5.82 (2.22) ^a
CN ratio	16.7 (3.9) ^a	16.8 (2.5) ^a	10.8 (0.4) ^b
Available P (g.kg ⁻¹)	0.04 (0.02)	0.05 (0.02)	0.05 (0.02)
pH (H₂O)	6.3 (1.0) ^a	4.7 (0.4) ^b	6.1 (1.2) ^a
Na (g.kg ⁻¹)	0.010 (0.002) ^a	0.004 (0.002) ^b	0.008 (0.006) ^{ab}
Ca (g.kg ⁻¹)	6.2 (2.5) ^a	1.0 (0.4) ^b	5.4 (4.0) ^a
Humus index	3.8 (0.4) ^a	2.8 (0.8) ^b	1.3 (0.5) ^c

Table 1. Topsoil (0-5 cm) parameters of the three land-uses (n=6).
Values are means (standard deviations). Letters indicate significant differences ($P < 0.05$) between land-uses (one-way ANOVA tests).

For both soil sampling intensities, soil extracellular DNA was extracted following Taberlet *et al.* (2012). Four extractions per plot were performed, each of them being amplified twice using the *ewDE* metabarcode (F: 5'-ATTCGGTTGGGGCGACC-3', R: 5'-CTGTTATCCCTAAGGTAGCTT), a short DNA marker targeting a 70 base pairs (bp) region on the mitochondrial 16S gene (Bienert *et al.* 2012). This marker was chosen because classical barcodes (*e.g.* Cytochrome Oxydase I, ca. 650 bp; Hebert *et al.* 2004) are too long to be retrieved entirely in soil extracellular DNA, this latter being subjected to degradation/fragmentation processes (Yoccoz *et al.* 2012, Bienert *et al.* 2012). Amplicons were sequenced on Illumina HiSeq2000 platform (2 x 100 bp pair-end). Data were subsequently filtered according to the protocol described in the Supplementary material.

Despite the conservative data filtering procedure, a total of 16 earthworm molecular operational taxonomic units (MOTUs) was detected across the 18 plots with the low-intensity sampling and 31 for the high-intensity sampling scheme across 12 plots; Table 2). Taxonomic assignment of MOTUs to a reference DNA database with >95% identity allowed discriminating 7 different species (9 for the high-intensity sampling; Table 2). Eight MOTUs could not be identified to the species level due to relatively incomplete DNA databases (19 for the high-intensity sampling), while the few remaining MOTUs belonged to the same morphospecies. Overall, most species that were visually identified were retrieved using soil DNA (Tables S2, S3 & S4), especially those

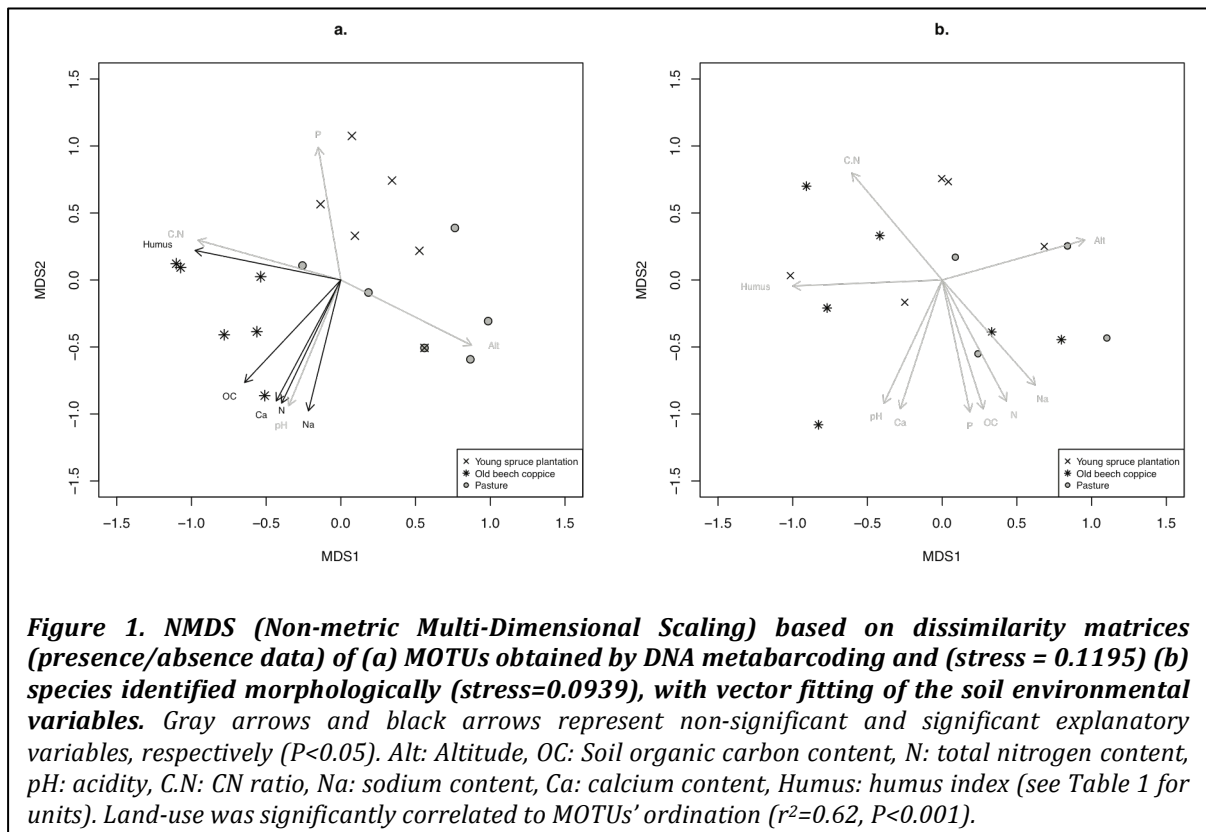
that were well represented in reference databases and with the high-intensity sampling. This sampling scheme allowed detecting on average 4 additional species per plot. As suggested by Bienert and colleagues (2012), extracting DNA from the O and A horizons simultaneously allowed the detection of epigeic species, while still detecting endogeic and anecic species (Tables S2 & S3). Interestingly, we found similar earthworm communities in duplicate composite samples collected on 1ha plots (analysis of similarities $r=0.966$ and $r=0.823$, $p<0.001$ for presence/absence and relative abundance data respectively). This illustrates both the detection power and the replicability of the soil DNA-based approach for surveying earthworm communities on large plots, and hence, its relevance for land management applications.

	n	MOTUs	MOTUs assigned to species level (similarity threshold > 0.95)	Morphological assignment
Low-intensity sampling				
Overall	18	16	7	13
Old beech coppice	6	8	5	7
Young spruce plantation	6	10	6	8
Pasture	6	11	6	11
High-intensity sampling				
Overall	12	31	9	-
Old beech coppice	4	17	7	-
Young spruce plantation	3	16	8	-
Pasture	5	22	7	-

Table 2. Number of MOTUs and earthworm species identified by the soil DNA approach and the classical earthworm survey.
Values are reported as overall number or discriminated per land-use. For the soil DNA approach, values are reported for the low- and high-intensity sampling schemes.

Distribution patterns of earthworm communities were ordinated using a non-metric multidimensional scaling. First, MOTUs presence/absence patterns clearly correlated with land-use ($P<0.001$ for low- and high-intensity sampling, Fig.1a & S1), unlike those from the classical earthworm survey (Fig. 1b). Similar results were obtained using relative abundance of MOTUs or taxonomic assignment to a reference DNA database with identity >95% (Fig. S2). Using soil DNA, plots located in old beech coppice were well differentiated from the others. Their soils displayed a significantly

higher humus index (Ponge & Chevalier 2006) than the other land-use types and a high organic carbon (SOC) and cation contents (Table 1). Humus index ($P<0.001$) and SOC ($P<0.002$), total nitrogen ($P<0.03$), sodium and calcium contents ($P<0.05$) were significantly correlated with the ordination of communities (Fig. 1a), underlining the influence of soil parameters on earthworm distribution (*e.g.* Curry 2004).



This study highlights the potential of environmental DNA *metabarcoding* to assess distribution pattern of earthworm communities and their drivers at the landscape scale. The field protocol is relatively easy to implement and provides a consistent fingerprint of earthworm communities. Collected DNA samples are also suitable to retrieve data from a large range of soil taxonomic groups using appropriate metabarcodes. The current limitation of the approach is that it relies on barcode reference databases that are often incomplete. Further improvements of databases are required to extend these kinds of analysis to other taxonomic groups. Nonetheless, we provided evidence that MOTUs-based analysis is a promising complementary way for studying earthworm biodiversity patterns, and more broadly those of soil biota.

Acknowledgments

This research was conducted on the long term research site "Zone Atelier Alpes", a member of the ILTER-Europe network. It was jointly funded by "Zone Atelier Alpes" through metaLOVER project and by the French National Research Agency through MetaBar (ANR-11-BSV7-020-01) and FORGECO (ANR-09-STRA-02-01) research programs. Authors specifically thank E. Mermin, P. Tardif, N. Daumergue and T. Goïtré (Irstea) for their help on the field, and G. Loucougaray (Irstea) for providing basic soil information on three pasture plots.

References

- Bartlett, M.D., Harris, J.A., James, I.T., Ritz, K., 2006. Inefficiency of mustard extraction technique for assessing size and structure of earthworm communities in UK pasture. *Soil Biology & Biochemistry* 38, 2990-2992.
- Bienert, F., *et al.* 2012. Tracking earthworm communities from soil DNA. *Molecular Ecology* 21, 2017-2030.
- Blouin, M., *et al.* 2013. A review of earthworm impact on soil function and ecosystem services. *European Journal of soil science* 64, 161-182.
- Coja, T., Zehetner, K., Bruckner, A., Watzinger, A., Meyer, E., 2008. Efficacy and side effects of five sampling methods for soil earthworms, Annelida, Lumbricidae. *Ecotoxicology and Environmental Safety* 71, 552-565.
- Curry, J.P., 2004. Factors affecting the abundance of earthworms in soils. In: Edwards, C.A., Ed., *Earthworm Ecology*. CRC Press, Boca Raton, pp. 91-113.
- Decaëns, T., Porco, D., Rougerie, R., Brown, G.G., James, S.W., 2013. Potential of DNA *barcoding* for earthworm research in taxonomy and ecology. *Applied Soil Ecology* 65, 35-42.
- Ettema, C.H., Wardle, D.A., 2002. Spatial soil ecology. *Trends in Ecology & Evolution* 17, 177-183.
- Grossi, J.-L., Chenavier L., Delcros, P., Brun J.-J., 1995. Effects of landscape structure on vegetation and some animal groups after agriculture abandonment. *Landscape and Urban Planning* 31, 291-301.
- Grossi, J.-L., Brun, J.-J., 1997. Effect of climate and plant succession on lumbricid populations in the French Alps, *Soil Biology & Biochemistry* 29, 329-333
- Hebert, P.D.N., Ratnasingham, S., de Waard, J.R., 2004. *Barcoding* animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270 (Suppl 1), 96-99.

- Hedde, M., Bureau, F., Delporte, P., Cécillon, L., Decaëns T., 2013. The effects of earthworm species on soil behaviour depend on land use. *Soil Biology & Biochemistry* 65, 264-273.
- King, R.A., Tibble, A.L., Symondson, W.O.C., 2008. Opening a can of worms: unprecedented sympatric cryptic diversity within British lumbricid earthworms. *Molecular Ecology* 17, 4684-4698.
- Klarica, J., Kloss-Brandstätter, A., Traugott, M., Juen, A., 2012. Comparing four mitochondrial genes in earthworms - Implications for identification, phylogenetics, and discovery of cryptic species. *Soil Biology & Biochemistry* 45, 23-30.
- Lawrence, A.P., Bowers, M.A., 2002. A test of the 'hot' mustard extraction method of sampling earthworms. *Soil Biology & Biochemistry* 34, 549-552.
- Lavelle, P., *et al.* 1997. Soil function in a changing world: the role of invertebrate ecosystem engineers. *European Journal of Soil Biology* 33, 159-193.
- NF EN ISO 23611-1, 2011. Qualité du sol - Prélèvement des invertébrés du sol - Partie1: tri manuel et extraction au formol des vers de terre. AFNOR.
- Pérès, G., *et al.* 2011. Earthworm indicators as tools for soil monitoring, characterization and risk assessment. An example from the national Bioindicator programme (France). *Pedobiologia* 54, 77-87.
- Ponge, J.F., Chevalier, R., 2006. Humus index as an indicator of forest stand and soil properties. *Forest Ecology and Management* 233, 165-175.
- Richard, B., Decaëns, T., Rougerie, R., James, S.W., Porco, D., Hebert, P.D.N., 2010. Re-integrating earthworm juveniles into soil biodiversity studies: species identification through DNA *barcoding*. *Molecular Ecology Resources* 10, 606-614.
- Römbke, J., Jansch, S., Didden, W., 2005. The use of earthworms in ecological soil classification and assessment concepts. *Ecotoxicology and Environmental Safety* 62, 249-265.
- Rossi, J.P., Lavelle, P., Albrecht, A., 1997. Relationships between spatial pattern of the endogeic earthworm *Polypheretima elongata* and soil heterogeneity. *Soil Biology & Biochemistry* 29, 485-488.
- Rougerie, R., *et al.* 2009. DNA barcodes for soil animal taxonomy. *Pesquisa Agropecuaria Brasileira* 44, 789-802.
- Stauffer M., Leyval C., Brun J.-J., Leportier P., Berthelin J. 2014. Effect of willow short rotation coppice on soil properties after three years of growth as compared to forest, grassland and arable land uses. *Plant Soil* 377, 423-438.
- Salomé, C., Guenat, C., Bullinger-Weber, G., Gobat, J.M., Le Bayon, R.C., 2011. Earthworm communities in alluvial forests: Influence of altitude, vegetation stages and soil parameters. *Pedobiologia* 54, 89-98.

- Taberlet, P., *et al.* 2012. Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for *metabarcoding* studies. *Molecular Ecology* 21, 1816-1820.
- Yoccoz, N.G., *et al.* 2012. DNA from soil mirrors plant taxonomic and growth form diversity. *Molecular Ecology* 21, 3647-3655.

c. Matériel et méthodes complémentaires

Supplementary material and methods

Study site

The elevation of the 18 study plots ranged between 1200 and 1600m a.s.l., with a mean annual temperature over the “Quatre Montagnes” landscape of +6.4°C (Autrans, 1050m), and a mean annual precipitation of 1492 mm (Autrans, 1050m). Young spruce (*Picea abies*) plantations are ca. 50 to 60 years old. They were previously managed as temporary meadows (land temporarily cultivated with herbaceous forest crops for mowing, FAO 2014) or cropland, whereas old beech (*Fagus sylvatica*) coppice are ancient forests (at least > 200 years-old). Some firs (*Abies alba*) can be found in old beech coppice. Since the beginning of the XXth century, old beech coppice gradually declined, partly replaced by conifer plantations (Redon *et al.* 2014). Pastures are grazed by cattle during summer. Their surface is currently decreasing in this mountain area due to abandonment and subsequent recolonization (Redon *et al.* 2014). Soils were Endoleptic to Epileptic Cambisols (Calcaric), (Hypereutric), or (Dystric) depending on soil pH (IUSS Working Group WRB 2007).

Soil basic characteristics and earthworm sampling

In each plot, basic characteristics of topsoils (0-5 cm) were determined on composite (5 soil cores), sieved (< 2mm) samples (Tables 1 and S1). The measured properties included organic C and total N contents, CN ratio, available P (Olsen), pH (H₂O), Ca and Na cation contents, and an assessment of the humus form through the humus index. All reference methods used for measuring basic soil properties are listed in Table S1.

For the classical earthworm survey, we used the extraction method following Zaborski (2003) by combining allyl isothiocyanate (AITC) and handsorting in three 0,5m² sub-plots in a 20x20m square situated in the center of the plot. We irrigated 0,5 m quadrats with two applications of 10 L of AITC solution (10g/L) at 10 minutes intervals. Earthworms emerging at the soil surface were collected during 10 min after each application. Individuals were stored in alcohol and later morphologically identified according to Bouché (1972).

Soil sampling for metabarcoding analyses

For the soil DNA approach, soil cores containing both litter and organo-mineral soil (O and A soil horizons, comprising most of the soil profile in these shallow Cambisols), were sampled in October 2011 (high-intensity sampling) and October 2013 (low-intensity sampling) and pooled together. Coring samplers were sterilized between each plot to avoid cross-contamination. Soil samples were stored in refrigerated areas and extracellular DNA was extracted in the week following sampling.

DNA extraction

The use of extracellular DNA adsorbed by soil particles was recently proposed to describe soil biodiversity (*e.g.* Bienert *et al.* 2012; Taberlet *et al.* 2012). Total DNA is susceptible to be biased towards sequences from living cells, thus a better representativity of the local diversity is expected with extracellular DNA. Besides, this DNA could be extracted from large amount of soil (several kilograms) allowing a wider coverage of studied area in a minimum of time (Taberlet *et al.* 2012).

Four extracellular DNA extractions were performed per soil sample. Then, two PCR per extraction were carried out leading to a total of eight PCR replicates per site. This strategy allows maximizing the detection of species and the detection of possible sporadic contamination (Ficetola *et al.* accepted), but also to monitor variability introduced by the method and the replicability of the results.

To extract extracellular DNA, we weighted each soil samples and the same weight of saturated phosphate buffer (Na_2HPO_4 ; 0.12 M; pH 8) was added and mixed thoroughly for 15 minutes to release DNA bound to particles (Taberlet *et al.* 2012). A small fraction of the mixture was centrifuged (10 minutes at 10000rpm) and supernatant was used as starting material in the following extraction steps, carried out using a commercial kit (NucleoSpin® Soil; Macherey-Nagel, Düren, Germany), skipping the lysis step and following manufacturer's instructions (Taberlet *et al.* 2012). Four extraction controls were performed.

DNA amplification and high-throughput sequencing

Markers used for environmental DNA *metabarcoding* analysis must be short enough to allow amplification of degraded DNA from environmental samples (<150bp) (Valentini *et al.* 2008). Thus, there is a need to find a trade-off between marker length

and taxonomic resolution (Taberlet *et al.* 2007; Chariton 2012). Here, we used the *ewDE* primers pair proposed by Bienert and colleagues (2012) (*ewD*: 5'-ATTCGGTTGGGGCGACC-3' and *ewE*: 5'-CTGTTATCCCTAAGGTAGCTT-3') that targets a short variable region (about 70bp) in the mitochondrial 16S gene. To assign sequence reads to the relevant sample after high-throughput sequencing, 8bp tags with at least 5 differences between them were added to the 5' end of each primer (Binladen *et al.* 2007; Valentini *et al.* 2009).

DNA amplifications were carried out in a final volume of 30 μ L containing 2 μ L of DNA extract. The amplification mixture contained 1.2 U of AmpliTaq Gold® DNA Polymerase (Applied Biosystems), 15 mM Tris-HCL, 50 mM KCl, 2 mM $MgCl_2$, 0.2 mM of each dNTP, 0.2 μ M of each primer and 4.8 μ g of bovine serum albumin (BSA, Roche Diagnostic). After 10min at 95°C for enzyme activation, PCR mixtures underwent 45 cycles: 30s at 95°C (denaturation), 30s at 55°C (hybridization) and 1min at 72°C (elongation). A final elongation step, 7min at 72°C, was finally performed. PCR products were purified and mixed together. High-throughput sequencing was performed on an Illumina HiSeq 2000 platform (2 x 100 bp paired-end reads). Two PCR per extraction were carried out and four PCR controls were randomly placed on PCR plates.

For the high-intensity sampling scheme, DNA extraction and amplification were similar to the protocol described above except for some points: (i) from the two soil samples per site, we performed two extractions per sample and two PCR per extraction (*e.g.* 8 PCR replicates per site), (ii) elongation steps were skipped in PCR cycles and no controls were performed. Data filtering after NGS was identical to those described below.

High-throughput sequencing data filtering

Data were filtered using the OBITOOLS software (<http://metabarcoding.org/obitools>). Direct and reverse reads corresponding to the same DNA fragment were first aligned and merged thanks to a gap-free end algorithm implemented in the *IlluminaPairEnd* program. Sequences with a low alignment quality score were removed (<40 *e.g.* pair of reads that can not align perfectly on at least 10 bp at each read end). Those shorter than 55 bp and/or containing ambiguous nucleotides were then discarded using the *obigrep* program. Afterward, sequences were assigned to the relevant PCR by the *ngsfilter* program thanks to tagged primers. Only sequences

containing perfect tags and primers with a maximum of 3 errors were considered (Bienert *et al.* 2012). Then, we used the *obiuniq* program to merge strictly identical sequences within each PCR sample. Strict singletons (sequences occurring only once in the dataset) were removed. To reduce computation time, sequences with a count <100 reads in the whole dataset were removed for the low-intensity sampling. This stringent filtering did not alter the subsequent results (data not shown). The *obiclean* program allows detecting sequences corresponding to putative amplification and sequencing errors (Bellemain *et al.* 2013). For each PCR product, the program creates networks that connect 2 sequences differing by a single difference (indel or substitution). Here, two sequences were linked the occurrence of the less frequent sequence is lower than 50% of its related most frequent sequence. A status is then assigned to each sequence as follows: 'head' (most common sequence of a network), 'singleton' (sequence without a variant) or 'internal' (sequences that are neither 'head' nor 'singleton' and that most likely correspond to errors). Only sequences that were more often 'head' or 'singleton' than 'internal' in the global dataset were kept for the following steps (Giguët-Covex *et al.* 2014).

Taxonomic assignment was performed with *ecoTag* program that finds highly similar sequence(s) in a suitable database. Here, sequences were compared to an earthworm database generated by *in silico* PCR from EMBL by *ecoPCR* program (Ficetola *et al.* 2010). Only sequences assigned to an earthworm taxa with an identity score >80% were kept for subsequent analysis. This step allows discarding Enchytraeidae sequences also amplified by the used primers pair.

Afterward, potentials contaminants, *e.g.* sequences having their highest proportion in negative control samples, were also removed. A graph-partitioning approach (*igraph* package in R, Csardi & Nepusz 2006) based on Bray-Curtis distance was used to detect, and subsequently filter out, outlying PCR and those clustering with control samples. Additionally, we expected that replicates from a same plot be clustered together; consequently PCR products not clustered with those from the same sample were also discarded.

It is often difficult to disentangle rare MOTU's from potential errors at low abundance. Here, we considered a sequence as genuine in a given plot if it was found in half of the replicates, or more, with a count representing at least 0.5% of the total number of lumbrician reads retrieved in the PCR product (Ficetola *et al.* accepted). This

threshold was chosen empirically as it represents a breakpoint in the number of MOTUs detected. This conservative data filtering procedure can make missing some rare taxa but allows to discard potential errors.

Finally, sequences with 4 differences (corresponding to ca. 95% similarity) or less were clustered together into molecular operational taxonomic unit (MOTU) using *sumatra* program for alignment (<http://metabarcoding.org/sumatra>) and infomap (Edler & Rosvall 2013) for clustering.

Statistical analysis

The effect of land-use type on basic soil characteristics was assessed by analyses of variance (ANOVA) followed by Tukey's honestly significant difference (HSD) post-hoc test. Each null hypothesis was rejected at $p < 0.05$. Normality and homoscedasticity of data and residuals were respectively assessed using Shapiro-Wilk and Levene tests. Olsen P, pH, Ca and Na contents were log transformed and humus index was square root transformed to fulfill ANOVA's assumptions.

Here, we used both presence/absence data and relative abundance data to assess earthworm distribution pattern among the three land-uses. Distance matrices of earthworm communities, inferred by *metabarcoding* and earthworm sampling, were calculated with Bray-Curtis distance and ordinated using a non-metric multidimensional scaling (NMDS). We then used a vector fitting method to relate the ordination to land-uses and soil parameters (Borcard *et al.* 2011). All statistical analyses were performed using R software version 2.15 (R Development Core Team 2012), using the *vegan* package (Oksanen *et al.* 2013).

Supplementary references

- Andersen, K., *et al.* 2011. Meta-barcoding of 'dirt' DNA from soil reflects vertebrate biodiversity. *Molecular Ecology* 21, 1966–1979.
- Bellemain, E., *et al.* 2013. Fungal palaeodiversity revealed using high-throughput *metabarcoding* of ancient DNA from arctic permafrost. *Environmental Microbiology* 15, 1176–1189.
- Binladen, J., *et al.* 2007. The Use of Coded PCR Primers Enables High-Throughput Sequencing of Multiple Homolog Amplification Products by 454 Parallel Sequencing. *Plos One* 2, e197.

- Borcard, D., Gillet, F. & Legendre, P., 2011. Numerical Ecology with R. Springer, New York.
- Chariton, A., 2012. Short and informative DNA products to indirectly measure vascular plant biodiversity. *Molecular Ecology* 21, 3637-3639.
- Csárdi, G., Nepusz T., 2006. The igraph software package for complex network research. *InterJournal, Complex Systems*, 1695.
- Edler D., Rosvall, M., 2013. The MapEquation software package, available online at <http://mapequation.org>
- FAO 2014. Land-use and irrigation – codes and definitions. FAO statistics division. Food and Agriculture Organization of the United Nations. www.fao.org/fileadmin/templates/ess/ess_test_folder/Definitions/LandUse_list.xls
- Ficetola, G. F., Coissac, E., Zundel, S., Riaz, T., Shehzad, W., Bessière, J., Taberlet, P., Pompanon, F., 2010. An In silico approach for the evaluation of DNA barcodes. *BMC Genomics* 11, 434.
- Ficetola, G.F., *et al.* Replication levels, false presences, and the estimation of presence/absence from eDNA *metabarcoding* data. *Molecular Ecology Resources*, accepted
- Giguët-Covex, C., *et al.* 2014. Long livestock farming history and human landscape shaping revealed by lake sediment DNA. *Nature communications* 5, 3211.
- NF ISO 10390, 2005. Soil quality, determination of pH. AFNOR.
- NF ISO 10694, 1995. Qualité du sol, dosage du carbone organique et du carbone total après combustion sèche (analyse élémentaire). AFNOR.
- NF ISO 11263, 1995. Qualité du sol - Dosage du phosphore - Dosage spectrométrique du phosphore soluble dans une solution d'hydrogénocarbonate de sodium. AFNOR.
- NF ISO 13878, 1998. Soil quality, determination of total nitrogen content by dry combustion ("elemental analysis"). AFNOR.
- NF X 31-108, 2002. Qualité des sols, détermination des cations Ca⁺⁺, Mg⁺⁺, K⁺, Na⁺ extractibles par l'acétate d'ammonium, méthode par agitation. AFNOR.
- Oksanen, J., Blanchet, F.J., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Wagner, H., 2013. vegan: Community Ecology Package. R package version 2.0-7. <http://CRAN.R-project.org/package=vegan>
- R Development Core Team, 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

- Redon, M., Luque, S., Gosselin, F., Cordonnier T., 2014. Is generalisation of uneven-aged management in mountain forests the key to improve biodiversity conservation within forest landscape mosaics? *Annals of Forest Science*, 1-10.
- Valentini, A., Pompanon F., & Taberlet P., 2009. DNA *barcoding* for ecologists. *Trends in Ecology & Evolution* 24, 110-117.
- Zaborski, E.R., 2003. Allyl isothiocyanate: an alternative chemical expellant for sampling earthworms. *Applied Soil Ecology* 22, 87-95.

c. Résultats complémentaires

Supplementary data

Table S1: Site description and basic soil characteristics

Table S2: Earthworm identified by classical survey

Table S3: Earthworm identified by *metabarcoding* with low-intensity sampling

Table S4: Earthworm identified by *metabarcoding* with high-intensity sampling

Figure S1: Ordination of earthworm communities based on presence/absence of MOTUs obtained with high-intensity sampling

Figure S2: Ordination of earthworm communities based on relative abundance of MOTUs with the low-intensity sampling

Figure S3: Ordination of earthworm communities based on species identified by *metabarcoding* with the low-intensity sampling

Supplementary Table S1: Site description and basic soil characteristics

Land-use	Site	Sampling intensity for soil DNA extraction	Latitude (°N)	Longitude (°E)	Altitude (m a.s.l.)	Organic carbon (g.kg ⁻¹)	Total nitrogen (g.kg ⁻¹)	C/N	P (Olsen; g.kg ⁻¹)	pH (H ₂ O)	Na (g.kg ⁻¹)	Ca (g.kg ⁻¹)	Humus index
Reference method	-	-	-	-	-	NF ISO 10694 (1995)	NF ISO 13878 (1998)	-	NF ISO 11263 (1995)	NF ISO 10390 (2005)	NF X 31-108 (2002)	NF X 31-108 (2002)	Ponge and Chevalier (2006)
Old beech coppice	B10	Low	45.207265	5.612166	1420	51.6	3.83	13.5	0.026	7.55	0.00902	7.23	3
	B11	Low	45.197934	5.610156	1275	109	4.57	23.9	0.025	4.92	0.0121	2.88	4
	B12	Low & high	45.120958	5.614527	1436	79.4	5.27	15.1	0.027	5.9	0.00812	5.16	4
	B13	Low & high	45.163805	5.594072	1350	83.4	5.57	15	0.079	6.76	0.00803	7.82	4
	B14	Low & high	45.160339	5.592288	1340	115	6.28	18.4	0.052	7	0.0131	9.69	4
	B26	Low & high	45.122103	5.621081	1563	92.2	6.45	14.3	0.037	5.68	0.00791	4.31	4
Young spruce plantation	S11	Low	45.113028	5.502597	1310	32.7	1.82	18	0.030	5.1	0.00543	1.18	3
	S12	Low & high	45.199803	5.549023	1300	36	2.18	16.5	0.028	4.93	0.00501	1.47	2
	S13	Low & high	45.045872	5.479051	1400	11.4	0.901	12.7	0.041	4.24	0.00131	0.241	3
	S14	Low & high	45.053775	5.478231	1305	24.8	1.57	15.8	0.051	5.15	0.00198	1.14	3
	S18	Low	45.190295	5.586754	1570	40.3	2.3	17.5	0.084	4.76	0.00453	1.19	2
	S08	Low	45.079552	5.507217	1290	28	1.39	20.1	0.049	4.29	0.00283	1.04	4
Pasture	P17	Low & high	45.098134	5.600683	1470	93.6	8.76	10.7	0.084	6.52	0.00359	7.37	2
	P19	Low & high	45.189606	5.586697	1550	57.1	5.16	11.1	0.043	7.16	0.00598	5.67	2
	P24	Low & high	45.198722	5.593213	1616	34	3.28	10.4	0.050	4.86	0.00343	0.893	1
	P25	Low & high	45.19243	5.595775	1586	52.2	5.13	10.2	0.031	5.38	0.004	2.66	1
	P26	Low & high	45.192182	5.593234	1610	48.3	4.27	11.3	0.020	5.18	0.0116	3.42	1
	P27	Low	45.189677	5.59339	1605	91	8.31	11	0.040	7.71	0.0196	12.1	1

Supplementary Table S2: Earthworm identified by classical survey

Genus	Species	Ecology	B10	B11	B12	B13	B14	B26	P17	P19	P24	P25	P26	P27	S08	S11	S12	S13	S14	S18
<i>Allolobophora</i>	<i>cupulifera</i>	endogeic	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
<i>Aporrectodea</i>	<i>caliginosa</i>	endogeic	2	0	0	4	0	0	0	0	0	0	0	12	0	0	0	1	3	0
<i>Aporrectodea</i>	<i>icterica</i>	endogeic	0	0	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0
<i>Aporrectodea</i>	<i>longa</i>	anecic	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0
<i>Aporrectodea</i>	<i>nocturna</i>	anecic	0	0	3	1	0	9	0	0	0	0	0	5	2	2	2	1	9	0
<i>Aporrectodea</i>	<i>rosea</i>	endogeic	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0
<i>Aporrectodea</i>	<i>sp</i>	-	25	20	2	20	3	37	0	0	1	1	0	48	5	2	8	10	78	2
<i>Lombricus</i>	<i>castaneus</i>	epigeic	0	0	8	0	0	0	0	0	0	2	1	1	1	0	3	0	0	3
<i>Lombricus</i>	<i>friendi</i>	epigeic	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
<i>Lombricus</i>	<i>meliboeus</i>	epigeic	0	0	0	0	0	0	3	0	0	0	0	0	1	0	6	1	0	2
<i>Lombricus</i>	<i>rubellus</i>	epigeic	0	0	2	1	9	0	7	0	0	0	0	0	0	1	0	5	7	0
<i>Lombricus</i>	<i>sp</i>	-	24	14	50	16	29	16	10	0	7	4	2	4	0	2	1	19	37	17
<i>Lombricus</i>	<i>terrestris</i>	anecic	0	4	0	0	8	0	4	0	5	2	2	11	0	0	0	4	0	2
<i>Octolasion</i>	<i>cyaneum</i>	endogeic	0	5	12	0	1	0	0	0	0	0	0	3	0	0	0	0	0	0
<i>Octolasion</i>	<i>lacteum</i>	endogeic	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>Octolasion</i>	<i>sp</i>	-	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table S3: Earthworm identified by *metabarcoding* with low-intensity samplingColumn heading :

B10: Mean number of reads detected in site B10 ; B11: Mean number of reads detected in site B11...

Best identity: best identity by comparison to the reference database

Rank: taxonomic rank of the best assignment

Scientific name: scientific name of the best assignment

			Sample																	
Best identity	Rank	Scientific name	B10	B11	B12	B13	B14	B26	P17	P19	P24	P25	P26	P27	S08	S11	S12	S13	S14	S18
0.92	species	<i>Lumbricus meliboeus</i>	31699	9104	89813	30854	27994	39354	18475	0	0	56595	0	0	0	0	0	0	0	0
0.986	species	<i>Aporrectodea longa</i>	1077	0	0	0	0	0	55870	0	11071	44885	0	0	53977	38751	0	49614	16996	16723
1	species	<i>Aporrectodea caliginosa</i>	0	0	0	0	0	0	36460	2579	53864	57269	41928	4994	0	46015	8131	0	42314	22168
0.892	species	<i>Allolobophora chlorotica</i>	0	0	0	0	0	0	10558	61594	0	0	0	0	0	0	0	0	0	10994
1	species	<i>Aporrectodea rosea</i>	1726	16767	0	4679	0	0	27920	0	33743	38785	102936	0	0	131383	0	0	27961	0
0.986	species	<i>Octolasion cyaneum</i>	48094	107287	1261	0	2158	7750	20569	0	0	0	0	0	0	91279	0	0	308	0
0.986	species	<i>Lumbricus castaneus</i>	0	8173	18863	0	0	0	0	0	0	22823	0	0	3189	7999	0	4455	0	764
1	species	<i>Aporrectodea icterica</i>	0	0	0	0	0	0	0	0	0	0	72399	26751	22175	14388	38461	0	0	0
1	species	<i>Aporrectodea sp. h1</i>	0	0	0	0	0	0	57776	0	0	0	0	37321	0	0	46602	0	0	0
0.857	species	<i>Lumbricus rubellus</i>	0	38798	711	64453	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	species	<i>Aporrectodea rosea</i>	0	0	0	0	0	0	23733	0	15391	15772	0	0	0	0	0	0	0	0
0.986	genus	<i>Octolasion</i>	0	0	10267	33939	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	species	<i>Lumbricus terrestris</i>	27420	0	0	0	6895	0	0	0	0	0	0	0	0	0	0	0	0	0
0.868	genus	<i>Aporrectodea</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4697	18112	0
0.877	species	<i>Allolobophora chlorotica</i>	0	0	0	0	0	0	0	0	0	0	14531	0	0	0	0	0	0	0
0.859	subfamily	<i>Lumbricinae</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	539	0	6875

Table S4: Earthworm identified by *metabarcoding* with high-intensity samplingColumn heading :

B10A: Mean number of reads detected in site B10 from soil sample A ;

B10B: Mean number of reads detected in site B10 from soil sample B...

Best identity: best identity by comparison to the reference database

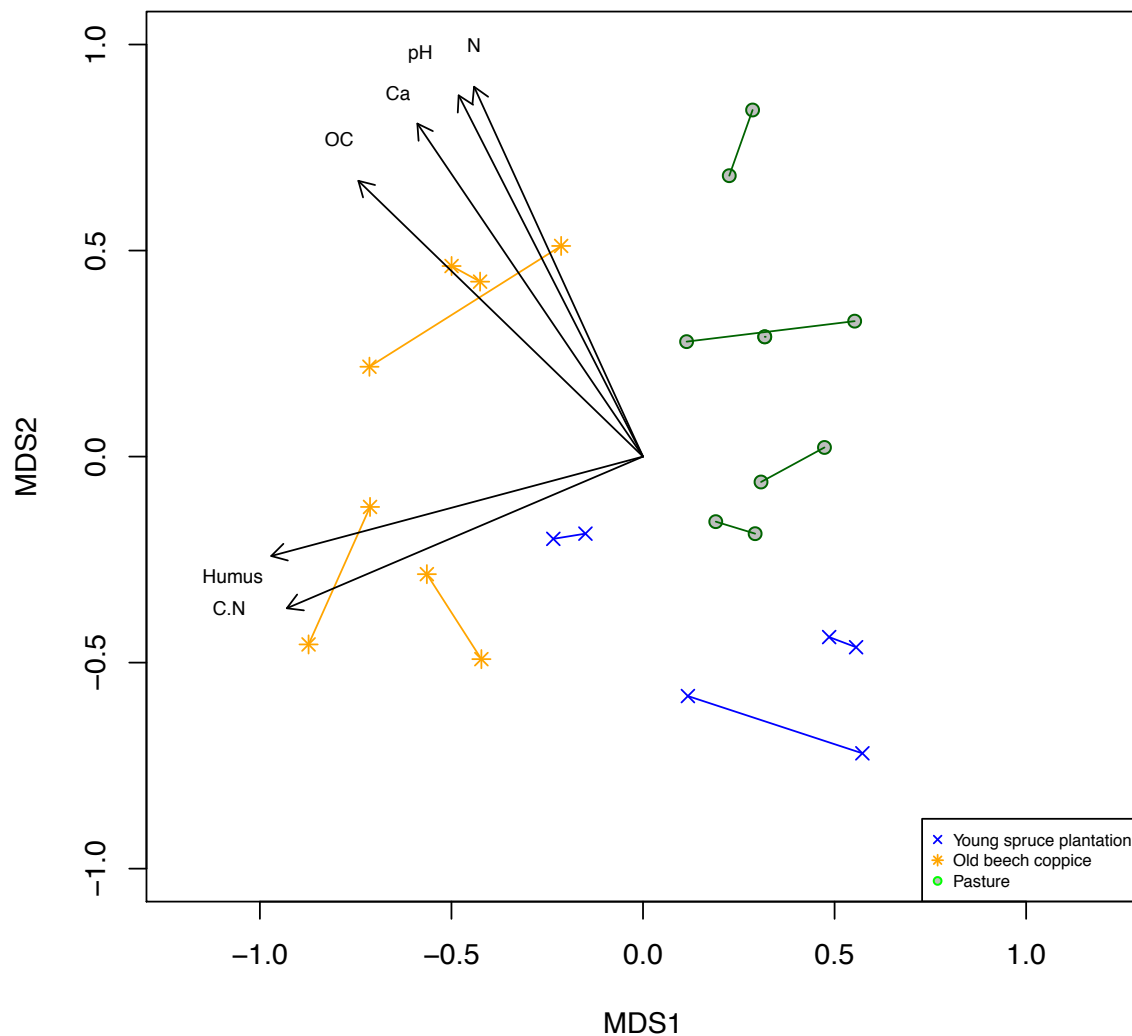
Rank: taxonomic rank of the best assignment

Scientific name: scientific name of the best assignment

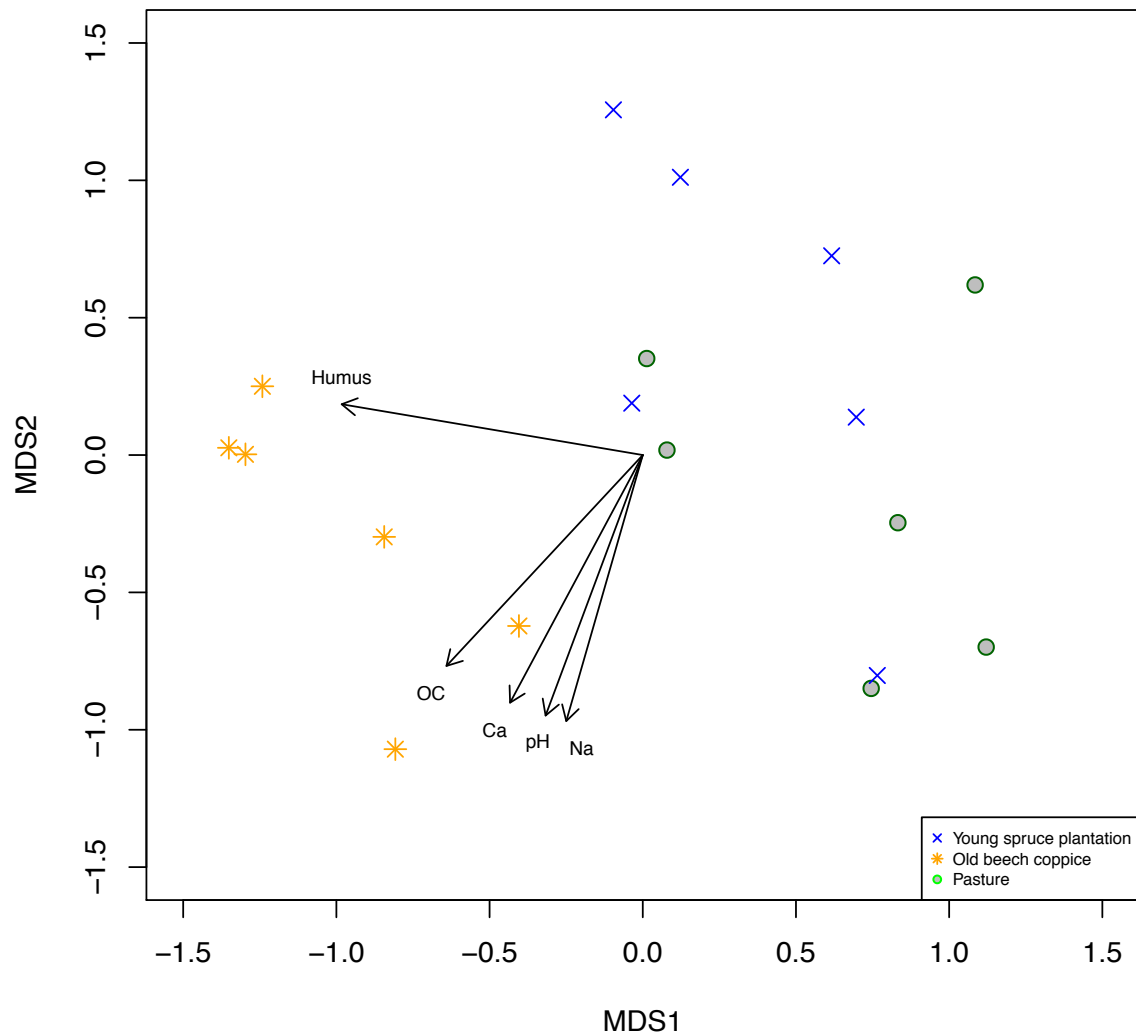
			Sample																									
Best identity	Rank	Scientific name	B12A	B12B	B13A	B13B	B14A	B14B	B26A	B26B	P17A	P17B	P19A	P19B	P24A	P24B	P25A	P25B	P26A	P26B	S12A	S12B	S13A	S13B	S14A	S14B		
0.92	species	<i>Lumbricus meliboeus</i>	7456	22612	0	1251	836	5224	6955	2045	0	0	0	0	186	0	0	0	0	0	571	298	0	0	5	15		
0.986	species	<i>Octolasion cyaneum</i>	940	1619	8286	17831	8297	8077	1816	9407	3021	544	2405	704	2461	5943	2375	2342	1622	1466	674	3050	64	168	4	0		
1	species	<i>Aporrectodea caliginosa</i>	0	0	91	0	4440	3952	0	0	1013	6505	2157	1791	421	1607	3179	553	2408	5684	3089	1085	5655	1257	5	48		
1	species	<i>Aporrectodea rosea</i>	0	588	133	234	1749	1665	1675	778	1132	1075	2076	2021	607	29	7953	1950	2900	145	2546	2374	0	0	0	0		
1	species	<i>Aporrectodea longa</i>	0	0	0	0	0	0	0	0	692	168	2685	3082	1060	3421	0	951	1205	1534	0	0	2608	3397	373	76		
0.857	species	<i>Lumbricus rubellus</i>	2393	0	1145	424	0	0	405	3301	0	0	0	0	0	0	0	0	363	2467	2363	1010	0	0	0	0		
0.986	species	<i>Lumbricus castaneus</i>	4224	1409	0	0	0	0	0	217	0	0	2457	542	0	0	616	0	942	296	633	574	57	143	67	0		
0.878	subfamily	<i>Lumbricinae</i>	0	0	0	0	0	0	0	0	0	0	97	1739	890	698	3126	2094	2492	4024	225	0	0	0	0	0		
0.859	subfamily	<i>Lumbricinae</i>	0	0	0	0	0	0	225	752	0	0	0	0	3571	76	0	0	886	2645	100	2566	392	547	236	993		
0.986	species	<i>Lumbricus terrestris</i>	0	0	222	325	682	1269	0	0	426	698	122	450	0	0	3728	0	0	0	0	0	0	0	0	0		
1	species	<i>Dendrodrilus rubidus</i>	0	0	0	0	2402	0	1449	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	68	283		
0.877	species	<i>Allolobophora chlorotica</i>	0	0	0	0	0	0	0	0	0	0	602	797	424	141	2455	21351	903	192	0	0	0	0	0	24		
0.932	species	<i>Aporrectodea longa</i>	0	0	0	0	0	0	0	0	4532	3009	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
0.921	species	<i>Aporrectodea caliginosa</i>	0	0	0	0	0	0	0	0	0	84	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
1	species	<i>Aporrectodea sp. h1</i>	0	0	0	0	0	0	0	0	2223	4028	634	1194	0	0	691	0	0	0	0	0	0	0	0	0		
1	species	<i>Aporrectodea rosea</i>	0	0	0	0	0	0	255	374	0	0	245	376	0	0	229	0	90	62	0	0	0	0	0	0		
0.96	species	<i>Lumbricus castaneus</i>	0	0	0	0	91	3429	0	0	713	208	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
1	species	<i>Lumbricus rubellus</i>	0	0	47	1495	0	0	0	0	0	0	0	0	0	0	0	0	0	0	102	81	0	0	119	0		
1	species	<i>Aporrectodea icterica</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	140	1988	0	0	139	2767	0	0	0	0		
0.868	genus	<i>Aporrectodea</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	747	1824	306	226		
0.947	species	<i>Aporrectodea caliginosa</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	305	0	0	0	0	0	0		
0.947	species	<i>Aporrectodea caliginosa</i>	0	0	0	0	0	0	0	0	0	142	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
0.986	genus	<i>Octolasion</i>	274	221	0	0	612	258	1744	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
0.872	subfamily	<i>Lumbricinae</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	159	73	0	0	0	0	0		
0.895	species	<i>Aporrectodea caliginosa</i>	0	0	0	0	0	0	0	0	0	0	0	0	25	0	0	0	0	170	0	0	0	0	0	0		
1	species	<i>Lumbricus rubellus</i>	2212	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
0.947	species	<i>Aporrectodea sp. h1</i>	0	0	0	0	0	0	0	0	83	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
0.857	species	<i>Octolasion cyaneum</i>	0	0	0	0	94	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
0.872	subfamily	<i>Lumbricinae</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	55	0	0	0	0		
0.855	species	<i>Octolasion cyaneum</i>	0	0	0	0	0	0	0	82	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
0.855	species	<i>Octolasion cyaneum</i>	0	0	0	0	0	0	0	109	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		

Supplementary Figure S1: Ordination of earthworm communities based on presence/absence of MOTUs obtained with high-intensity sampling (stress=0.1458).

Bars connect the two soil samples from the same plot. Vector fitting was computed on centroids of each plot. Only significant explanatory variables were represented ($P < 0.05$). A significant correlation was found between land-use and ordination of earthworm communities ($r^2 = 0.63$, $P < 0.001$).

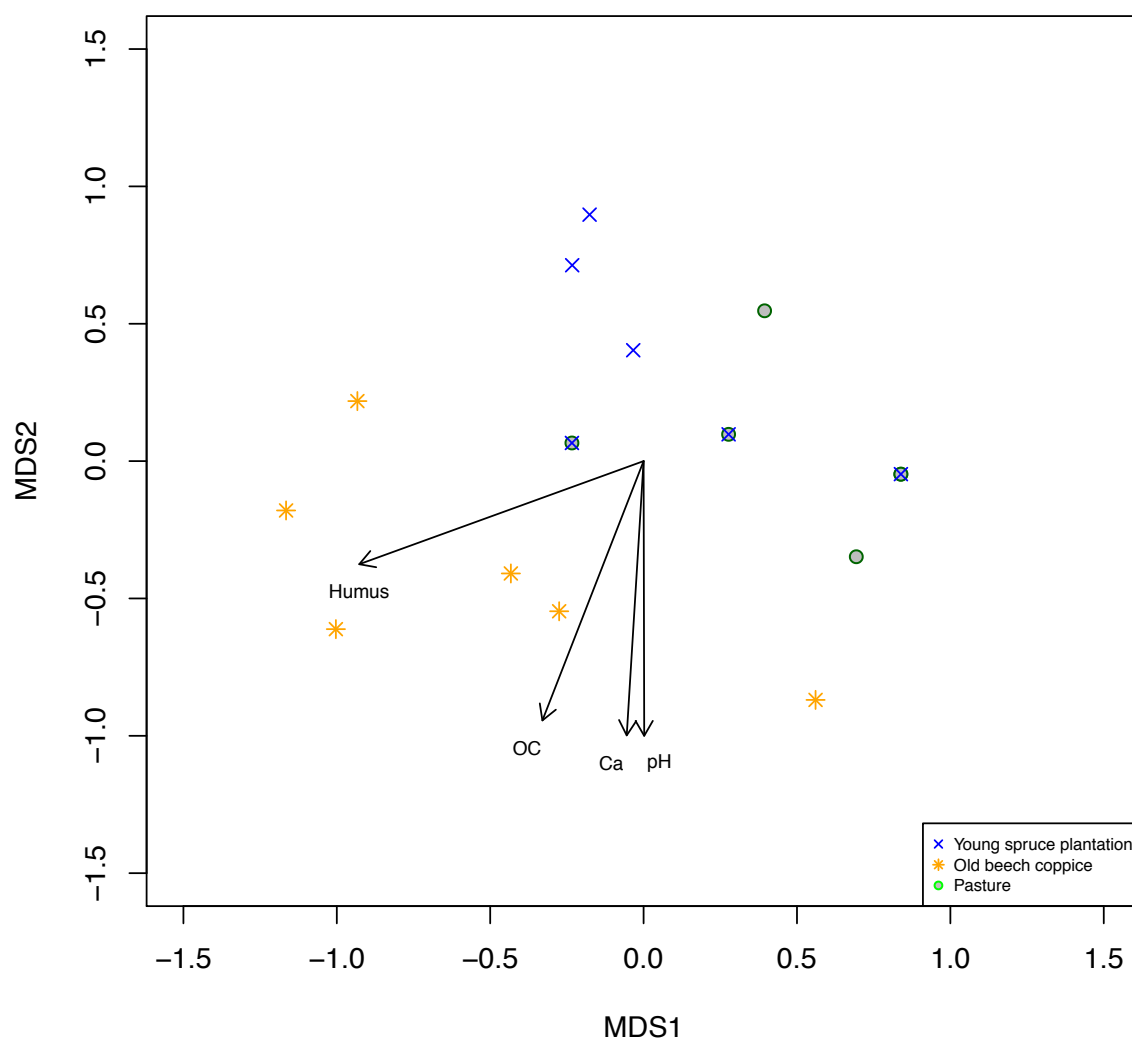


Supplementary Figure S2: Ordination of earthworm communities based on relative abundance of MOTUs with the low-intensity sampling (stress=0.1129). Only significant explanatory variables were represented ($P < 0.05$). A significant correlation was found between land-use and ordination of earthworm communities ($r^2 = 0.57$, $P < 0.001$).



Supplementary Figure S3: Ordination of earthworm communities based on species identified by *metabarcoding* with the low-intensity sampling (presence/absence data; stress=0.0940). Only significant explanatory variables were represented ($P < 0.05$). A significant correlation was found between land-use and ordination of earthworm communities ($r^2 = 0.46$, $P < 0.001$).

A significant part of MOTUs could not be assigned to the species level leading to a reduced estimation of species diversity. This could explain the similarity between some spruce plantations and pasture plots.



4. Discussion chapitre III

Quels sont les facteurs biotiques et abiotiques influençant la structure des communautés du sol et comment sont-ils affectés par les activités humaines ? Ces questions sont primordiales pour déterminer l'impact des activités anthropiques sur les communautés. La réponse est particulièrement compliquée à évaluer car les sols sont spatialement plus hétérogènes dans leurs propriétés biotiques et abiotiques, et tous ces facteurs interagissent avec la perturbation (De Deyn & Van der Putten 2005). L'hétérogénéité perçue par les organismes du sol diffère de celle perçue par les macro-organismes de surface du fait des différences de dimension d'habitats. Les réponses des communautés de surface aux perturbations ne reflètent ainsi pas nécessairement celles des sols.

L'approche de *metabarcoding* basée sur l'ADN du sol, utilisée ici, a permis d'obtenir les profils de distribution, et les patrons de diversité de plusieurs groupes taxonomiques, et d'en étudier les principales relations à l'environnement biotique et abiotique. Si cette approche est largement répandue en microbiologie (*e.g.* Lentendu *et al.* 2011 ; Zinger *et al.* 2011; Lindahl *et al.* 2013 ; Botnen *et al.* 2014), elle n'est encore que très peu utilisée pour caractériser les communautés du sol autres que fongiques ou bactériennes. Malgré les possibilités offertes pour l'évaluation sur de grandes surfaces de la biodiversité des sols, et de l'impact des activités humaines, son utilisation n'est pas triviale. La mise en application de cette approche, et ses limites, pour étudier l'impact sur les communautés et les mécanismes impliqués dans leur réponse sont, par ailleurs, discutées dans le chapitre V.

a. Effets des perturbations anthropiques sur la biodiversité des sols

Les activités humaines, en l'occurrence ici l'introduction d'herbivores et l'utilisation des terres, ont pour particularité commune d'avoir profondément modifié les sols en affectant à la fois leurs propriétés physico-chimiques et leur diversité. La diversité des organismes des sols est, par ailleurs, connue pour être affectée directement par la réduction de la matière organique et indirectement par la réduction de la diversité des plantes (Turbé *et al.* 2010). Dans l'article A, l'introduction du lapin dans les îles Kerguelen à la fin du XIXème siècle a engendré une modification et une diminution du couvert végétal ainsi qu'une altération physique des sols (Bardgett & Wardle 2003) *via*

l'érosion et la création de galeries, ce qui a mené notamment à une réduction des apports en matière organique. Il en a résulté une altération des communautés fongiques à la fois en terme de composition mais également de diversité. De même, dans l'article B, nous avons constaté une modification des paramètres édaphiques, parmi lesquels le carbone organique et les nutriments, sous l'effet de la gestion des terres par l'Homme. Dans les deux cas, les modifications de l'environnement abiotique et du couvert végétal ont entraîné des changements dans les communautés du sol, en l'occurrence, ici, les champignons et les lombriciens.

b. Des effets médiés par les plantes

Les études présentées ici (études A et B, encadré 6) illustrent l'influence des changements de végétation, associés aux activités humaines, sur la diversité souterraine. Les communautés du sol et celles de surface interagissent et s'influencent mutuellement (Wardle *et al.* 2004). Les plantes produisent le carbone organique nécessaire aux décomposeurs ainsi que des ressources pour les organismes associés aux racines (herbivores, pathogènes, symbiotes...). En retour, les décomposeurs recyclent la matière organique morte et régulent indirectement la croissance des plantes et la composition des communautés en contrôlant la disponibilité des nutriments. Quant aux organismes de la rhizosphère et leurs consommateurs, ils influencent directement les plantes et contrôlent les flux d'énergie et de nutriments entre plantes et décomposeurs (Wardle *et al.* 2004).

A l'échelle de nos sites d'études, la composition et la diversité des communautés végétales, et les interactions avec le compartiment aérien, vont ainsi influencer la biodiversité des sols. L'altération à long terme de la structure des communautés végétales induit des changements dans les apports de la litière en modifiant sa qualité, sa composition et sa décomposabilité (Bardgett & Wardle 2003). Les activités humaines vont alors indirectement modifier les apports de matière organique mais également la concentration des autres nutriments et des métabolites secondaires dans la rhizosphère (Wardle *et al.* 2004). Il est important de noter que ces apports varient au cours de l'année en fonction du stade de développement de la végétation et que cela peut résulter en de grandes variations intra-annuelles dans les communautés du sol, notamment microbiennes (Jaeger *et al.* 1999 ; Bardgett *et al.* 2005 ; Björk *et al.* 2008 ; Zinger *et al.* 2009 ; Puissant *et al.* soumis). La dynamique saisonnière des communautés du sol

n'étant pas prise en compte dans ces analyses, l'ampleur de l'impact anthropique sur les communautés édaphiques peut ainsi être potentiellement mésestimée.

A une échelle plus fine, la biodiversité des sols est fortement influencée par les conditions abiotiques locales et les interactions biotiques, celles-ci agissant comme un filtre pour les espèces du pool régional (De Deyn & Van der Putten 2005). La composition végétale impacte directement la structure physique des sols et la disponibilité des micro-habitats ; les plantes en coussin sont par exemple connues pour modifier localement les conditions abiotiques (Roy *et al.* 2013). Les relations entre les taxons du sol n'ont pas été directement étudiées ici mais constituent également un élément primordial dans la composition des communautés souterraines. Les plus gros organismes vont, par exemple, influencer sur la diversité des organismes de plus petites tailles en promouvant la dispersion et en créant des modifications d'habitat (Wardle 2006). Le *metabarcoding* environnemental, permettant d'étudier simultanément la diversité de plusieurs groupes taxonomiques en un point donné (Bik *et al.* 2012), offre, sous réserve de quelques adaptations (cf. partie V.1.a), l'opportunité d'étudier plus finement les réseaux d'interactions souterraines.

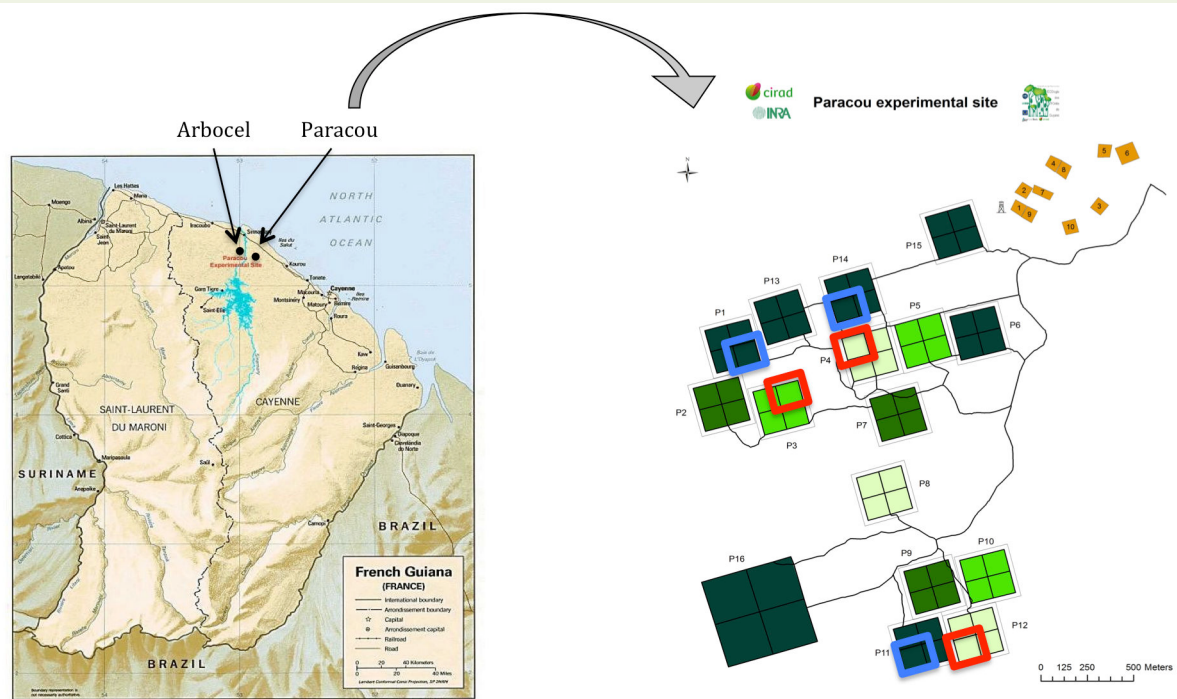
c. Résilience des communautés

Les activités humaines, qu'ils s'agissent ici de l'utilisation des terres ou de l'introduction d'herbivores, affectent les communautés biologiques en agissant comme une succession de perturbations. L'absence de retour un état pré-anthropique, aussi bien pour les communautés végétales que fongiques, sur l'île Guillou, après l'éradication du lapin, illustre l'impact durable que peuvent avoir ces activités, même après leur arrêt, sur la biodiversité. Dans la même idée, l'étude préliminaire menée en Guyane (cf. encadré 6) montre que les perturbations anthropiques, même instantanées, si elles sont suffisamment intenses, peuvent engendrer un héritage qui peut perdurer pendant des décennies, et qui ne concerne pas uniquement les communautés ciblées directement par l'Homme.

Encadré 6. Une étude de cas en forêt tropicale

Les deux études présentées précédemment se sont concentrées sur des écosystèmes froids et tempérés. Une étude, encore en cours, réalisée au cours de cette thèse, vise à étendre l'application du *metabarcoding* environnemental à des milieux plus complexes tel que celui de la forêt tropicale. Celui-ci s'est révélé très prometteur pour caractériser la biodiversité végétale en forêt guyanaise, en dépit de la rapide dégradation de l'ADN dans ces sols (Yoccoz *et al.* 2012).

Cette étude s'est déroulée sur les sites expérimentaux (CIRAD) de Paracou (5°29'N, 52°92'W) et Arbocel (5°30'N, 53°W) en Guyane française et a notamment pour but de caractériser l'impact des traitements forestiers passés (environ 30 ans) sur la diversité actuelle des communautés eucaryotes des sols tropicaux. Neuf parcelles d'un hectare ont été échantillonnées (Fig. 8). Parmi elles, trois servant de témoins se situent dans des zones non perturbées, trois sont des parcelles ayant subi un prélèvement sélectif en 1984 (prélèvement de 20 à 45 arbres de plus de 40cm de diamètre) et trois sont des forêts secondaires régénérées suite à une coupe à blanc en 1976.



Le protocole d'échantillonnage dit de « forte intensité » décrit dans l'article B a été utilisé. Ainsi, deux échantillons composites de sol issus de 100 prélèvements répartis uniformément sur la zone (*i.e.* un prélèvement tous les 10m) ont été récoltés. Deux extractions par échantillon de sol et deux amplifications par PCR ont été réalisées soit un total de 8 réplicats de PCR par site. Les amorces universelles utilisées (F : 5'-TCACAGACCTGTTATTGC-3' ; R : 5'-TYTGTCTGSTTRATTSCG-3') ciblent une courte région de l'ARN ribosomique 18S permettant de caractériser l'ensemble de la diversité eucaryote. De tels marqueurs sont peu résolutifs mais offrent un compromis intéressant au vu de leur large couverture taxonomique pour obtenir des profils de distribution de la micro-biote eucaryote des sols (Chariton *et al.* 2010 ; Baldwin *et al.* 2013). Les résultats préliminaires (Fig. 9, résultat non publié) illustrent l'héritage des coupes à blanc sur les assemblages eucaryotes des sols. Si les niveaux de diversité n'apparaissent pas différents ($P>0.05$), les communautés eucaryotes des sols de forêts secondaires sont clairement différenciées de celles des deux autres traitements. Les différences de composition végétale, mises en évidence conjointement par les relevés botaniques sur les arbres de diamètre supérieur à 10cm et par *metabarcoding* (caractérisées notamment par une diminution des lianes et des espèces de sous-bois et la forte présence d'espèces héliophiles et/ou pionnières en forêt secondaire), expliquent en partie ce patron.

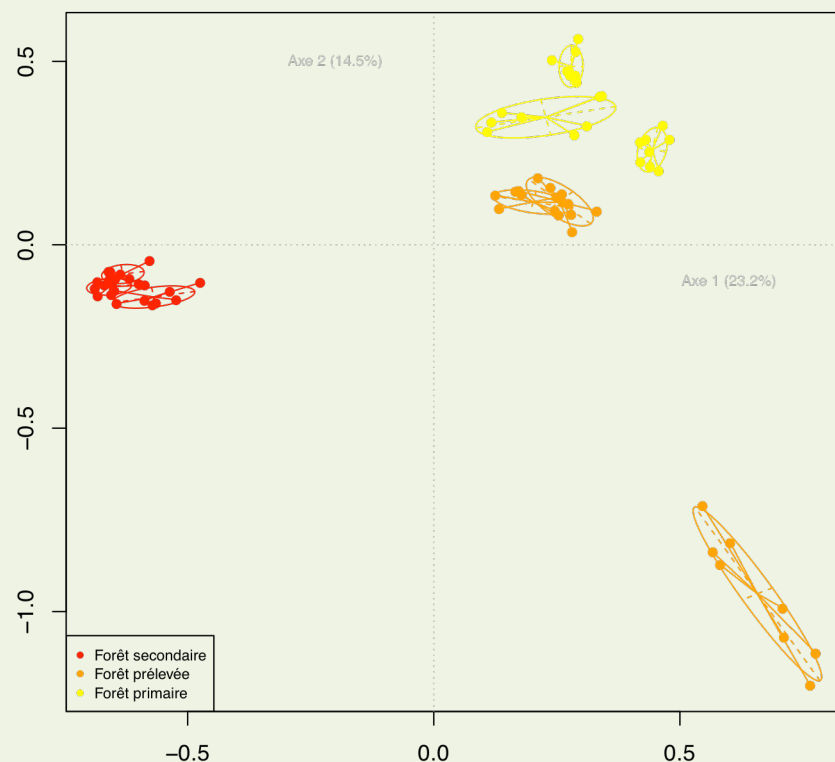


Figure 9. Ordination des communautés eucaryotes des sols tropicaux. Analyse inter-classes réalisée à partir des données raréfiées en présence/absence. Chaque point représente un réplicat de PCR.

Les changements dans les communautés, induits par les perturbations humaines, peuvent se maintenir sur de longues périodes (Fig. 10 ; Turner & Dale 1998 ; Foster *et al.* 2003 ; Calderon-Aguilera *et al.* 2012). Si les niveaux de richesse locale peuvent se rétablir rapidement (cf. encadré 6; Chazdon *et al.* 2003 ; Chazdon *et al.* 2007), des différences de composition peuvent perdurer, et ce pour plusieurs raisons. Tout d'abord, comme nous l'avons vu dans le cas de l'île Guillou, la synergie entre plusieurs événements perturbateurs peut être un des facteurs limitant le retour du système aux conditions pré-perturbations (Paine *et al.* 1998). Dans ce cas précis, l'interaction avec un autre type de perturbations anthropiques, à savoir les changements climatiques, explique en partie ce phénomène.

Les opportunités de niche engendrées par les perturbations peuvent permettre l'établissement de taxons, notamment des espèces invasives, auparavant exclus du milieu par exemple par compétition, et qui vont pouvoir se maintenir par la suite (Brown & Gurevitch 2004 ; Irwin *et al.* 2010). C'est, par exemple, le cas du pissenlit (*Taraxacum sp.*) aux Iles Kerguelen mais de telles interactions existent aussi dans les communautés du sol, en particulier chez les champignons où des phénomènes d'exclusion compétitive ont été mis en évidence (Kennedy & Bruns 2005). Le pool d'espèces capables d'occuper le milieu après la perturbation est restreint aux taxons capables de survivre à la perturbation et à ceux provenant du milieu environnant qui vont alors pouvoir envahir le système. La limitation de la dispersion et du recrutement va alors agir de telle manière que la composition des communautés, immédiatement après la perturbation, va être différente de celle observée auparavant. Les caractéristiques des espèces, notamment leur potentiel de compétition, ainsi que leur ordre d'arrivée dans le milieu vont également jouer un rôle primordial pouvant mener à des changements durables dans la composition des communautés (Robinson & Dickerson 1987 ; Drake 1991). Kennedy *et al.* (2009) ont ainsi montré l'importance des effets de priorité sur la dynamique de colonisation des racines par les champignons ectomycorhiziens.

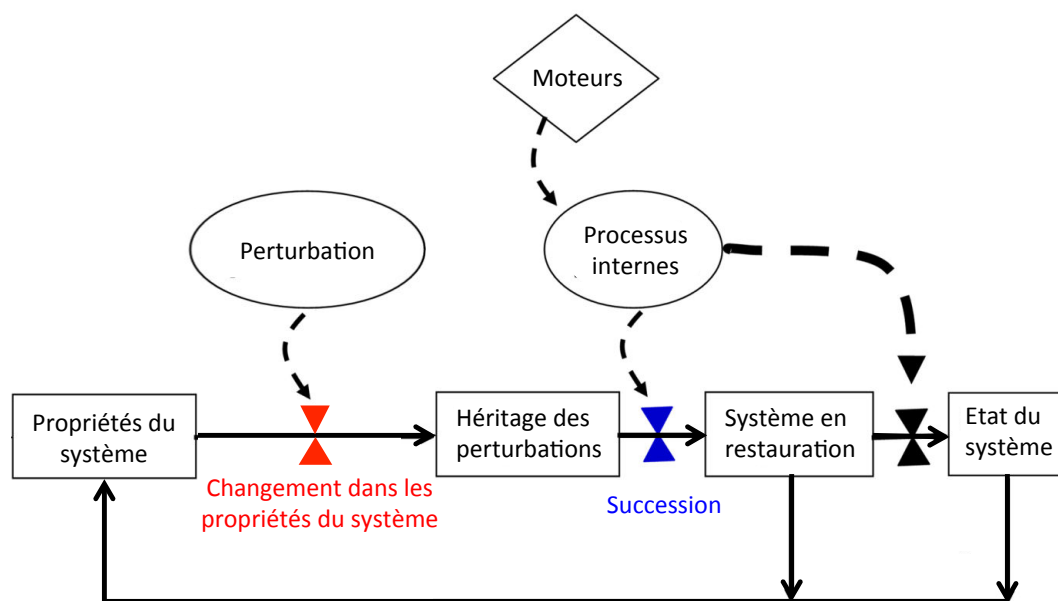


Figure 10. Schéma simplifié de l'effet des perturbations anthropiques (adapté de Peters et al. 2011). Les perturbations induisent des changements dans les propriétés biotiques et abiotiques du système. Cet héritage va alors influencer les processus de succession et engendrer des modifications durables dans la composition des communautés.

L'héritage des perturbations est encore plus important lorsque celles-ci touchent les conditions abiotiques du système et notamment les sols (e.g. Compton & Boone ; 2000 ; Jussy *et al.* 2002 ; Foster *et al.* 2003). Par exemple, les feux ont un effet persistant sur les taux de nitrification dans les forêts de pins américaines (Deluca *et al.* 2006). Les résultats des études A et B montrent un effet des activités humaines (utilisation des terres et introduction d'herbivores exotiques) sur les quantités de matière organique des sols, associé à des changements de végétation. La végétation passée peut ainsi influencer les communautés du sol à travers l'effet qu'elle exerce localement sur certaines caractéristiques des sols. Les plantes apparaissent ainsi à la fois comme des indicateurs des perturbations passées mais également comme des moteurs de la diversité des sols. Dans certains cas, ces effets peuvent perdurer pendant des siècles (Seastedt & Adams 2001). Plusieurs études ont ainsi mis en évidence l'héritage de l'agriculture gallo-romaine sur la biodiversité des forêts actuelles (Dupouey *et al.* 2002 ; Dambrine *et al.* 2007). Cet impact très ancien a modifié irréversiblement la structure des sols et la disponibilité des nutriments et il est à la base de différences dans les communautés de plantes mais également de celles du sol comme les champignons ectomycorhiziens (Diedhiou *et al.* 2009). Les communautés de plantes qui se sont développées à la suite de l'afforestation des parcelles agricoles abandonnées diffèrent

ainsi des forêts anciennes à cause de propriétés édaphiques différentes mais également des différences de potentiels de colonisation des espèces dans ces nouvelles forêts (Compton & Boone 2000 ; Dupouey *et al.* 2002). L'augmentation des plantes rudérales est, par exemple, favorisée par les fortes teneurs en nutriments (Dupouey *et al.* 2002) alors que les espèces strictement liées aux forêts anciennes sont limitées par leur faible capacité de dispersion (Hubbell *et al.* 1999 ; Dupouey *et al.* 2002 ; Dambrine *et al.* 2007).

Les perturbations anthropiques peuvent ainsi impacter durablement la biodiversité des sols en modifiant les paramètres édaphiques et la végétation de surface. Dans certains cas, un phénomène d'hystérèse peut même être observé, faisant basculer le système dans un état alternatif stable (cf. encadré 4 ; Nystrom *et al.* 2000 ; Beisner *et al.* 2003 ; Scheffer *et al.* 2001 ; Scheffer *et al.* 2003 ; MacDougall 2013). Ces effets s'apprécient sur des périodes de temps très longues qui dépassent généralement le cadre des études écologiques classiques. Les études paléo-écologiques offrent alors une perspective à long terme intéressante pour étudier la résilience des écosystèmes face aux activités humaines.

Chapitre IV. Une approche paléoécologique basée sur le *metabarcoding* environnemental pour l'évaluation à long terme de l'impact des activités humaines

1. Problématique générale

a. De l'importance des données paléoécologiques pour l'étude des perturbations anthropiques

L'influence des activités humaines sur les communautés biologiques n'est pas un phénomène récent et certaines pratiques parfois très anciennes (*e.g.* pâturage, chasse, agriculture) ont des conséquences sur la distribution et la composition des assemblages actuels (Dupouey *et al.* 2002 ; Heckenberger *et al.* 2003 ; Willis *et al.* 2004 ; Heckenberger *et al.* 2007 ; Heckenberger *et al.* 2008 ; Barlow *et al.* 2012). Les données sur le long terme sont ainsi particulièrement importantes pour comprendre l'impact des perturbations anthropiques sur la biodiversité. L'évaluation de l'importance écologique d'une perturbation passe par l'évaluation de sa capacité de résilience. Or, comme nous l'avons vu précédemment (cf. partie III.4.d.), l'effet des activités humaines sur la biodiversité peut perdurer pendant des décennies voire des siècles (Dupouey *et al.* 2002 ; Dambrine *et al.* 2007). Les études paléoécologiques sont à même de fournir des perspectives concernant les processus écologiques influençant les communautés biologiques, et les réponses de ces dernières, sur de longues périodes de temps (Willard & Cronin 2007 ; Dearing *et al.* 2010). Elles donnent ainsi accès à des informations de référence cruciales concernant la structure de l'écosystème avant la néolithisation et les premiers impacts humains à grande échelle (Turner *et al.* 2003 ; Willis & Birks 2006). En outre, elles permettent de quantifier les régimes de perturbations passés afin de contextualiser les perturbations actuellement subies par le système (Marlon *et al.* 2012). Cela permet, *in fine*, de discriminer les variations « naturelles » du système des changements plus profonds induits par une modification des régimes de perturbations. Enfin, la connaissance de l'histoire des communautés permet de mieux comprendre les trajectoires suivies après perturbations, qu'elles soient anthropiques ou naturelles (Drake 1991 ; Foster *et al.* 1998 ; Turner *et al.* 2003 ; Willard & Cronin 2007). Finalement, ces informations peuvent servir à calibrer des modèles d'évolution et au

développement de stratégies de conservation (Willis & Birks 2006 ; van Leeuwen *et al.* 2008; Hu *et al.* 2009).

b. Détection des perturbations passées

L'étude paléo-environnementale des perturbations implique à la fois la détection, à partir d'archives environnementales, desdites perturbations et de leurs effets sur les communautés et/ou l'environnement abiotique. Le développement de techniques permettant à la fois de reconstruire les perturbations et la dynamique des écosystèmes terrestres est ainsi un élément clé en paléoécologie (McLauchlan *et al.* 2014).

Actuellement, la connaissance des sociétés humaines passées et l'étude de leurs impacts sur l'environnement sont principalement fondées sur des données archéologiques, documentaires (*e.g.* récits, cadastres) ou encore sur les informations obtenues à partir d'archives naturelles. Les études paléoécologiques menées en Europe datent les premiers impacts humains à grande échelle au cours du Néolithique, en lien avec le développement de l'agriculture et l'ouverture des milieux (Behre 1988). Aujourd'hui encore, la majeure partie des perturbations anthropiques en milieu alpin est liée aux pratiques agricoles (Mocci *et al.* 2008 ; Mazier *et al.* 2009; Röpke *et al.* 2011). Plusieurs outils ont été développés pour reconstruire les régimes des perturbations naturelles passées tels que les fossiles ou les pollens (*e.g.* Tinner *et al.* 1996 ; Röpke *et al.* 2011 ; David *et al.* 2010a ; David *et al.* 2010b), la dendochronologie (*e.g.* Malik & Witusba 2012), les charbons (*e.g.* Tinner *et al.* 2005 ; Carcaillet *et al.* 2009 ; Blarquez & Carcaillet 2010 ; Rey *et al.* 2013) ou encore la sédimentologie (*e.g.* Giguet-Covex *et al.* 2011 ; Giguet-Covex *et al.* 2012) (cf. Table 1, McLauchlan *et al.* 2014).

Perturbation	Intensité	Fréquence	Mesure de reconstruction
Feux	Faible	Elevée	Dendrochronologie (cicatrices enregistrées dans le bois)
	Elevée	Faible	Dates d'établissement des arbres, charbons sédimentaires
Insectes	Elevée	Faible	Démographie des arbres, fossiles d'insectes dans les sédiments
Vent	Elevée	Faible	Arbres abattus, matériaux clastiques dans les sédiments
Sécheresse	Variable	Variable	Dendrochronologie (cercle de croissance), niveau des lacs, salinité des lacs
Déforestation	Elevée	Faible	Pollen sédimentaire, taux de sédimentation, démographie des arbres
Mouvement de terrain	Variable	Variable	Taux de sédimentation

Table 1. Mesures de reconstruction des principales perturbations naturelles en paléo-environnement (d'après McLauchlan *et al.* 2014)

Les activités anthropiques induisent des changements particuliers dans les communautés végétales (*e.g.* Tinner *et al.* 1996 ; Gobet *et al.* 2003). La reconstruction des pratiques agricoles à partir d'archives sédimentaires passe ainsi par la recherche, *via* les pollens et les macro-restes, de taxons végétaux cultivés (*e.g.* céréales, Rösche *et al.* 1992) mais aussi de taxons « indicateurs » d'activités humaines (*e.g.* plantain et aulne pour le pastoralisme; Wick & Tinner 1997 ; David *et al.* 2010a) ou d'assemblages botaniques particuliers (Mazier *et al.* 2009). L'ouverture des milieux, sous l'effet de l'intensification des perturbations anthropiques, est, elle, caractérisée par la réduction des espèces ligneuses (*e.g.* Wick & Tinner 1997 ; Colombaroli *et al.* 2010) mais aussi par une érosion accrue (Giguët-Covex *et al.* 2011). Plus directement, la présence passée de troupeaux peut-être inférée à partir des spores de champignons coprophiles (Davis & Schafer 2006 ; Cugny *et al.* 2010 ; Etienne *et al.* 2013). De même, l'histoire des feux peut être déterminée par l'accumulation des charbons dans les sédiments (Blarquez & Carcaillet 2010 ; Colombaroli *et al.* 2010 ; Colombaroli *et al.* 2013).

Quoi qu'il en soit, ces indicateurs d'activités humaines sont souvent indirects et peuvent être influencés par d'autres paramètres environnementaux. Un des principaux défis dans ce domaine consiste ainsi à différencier les forçages environnementaux tels que le climat ou les perturbations naturelles de ceux d'origines humaines (Anderson *et al.* 2006).

c. Reconstruction de la paléobiodiversité

La reconstruction des communautés biologiques terrestres passées est traditionnellement inférée à partir des informations contenues dans les archives naturelles tels que les sédiments lacustres ou les tourbières (*e.g.* Tinner *et al.* 1996 ; Birks & Birks 2000 ; de Beaulieu *et al.* 2006 ; Gaillard *et al.* 2008 ; Blarquez & Carcaillet 2010 ; David *et al.* 2010a ; David *et al.* 2010b ; Marquer *et al.* 2014). Les communautés animales sont ainsi principalement reconstruites à partir de fossiles et des macrorestes. Une des principales limitations de cette méthode concerne la sporadicité de ce type de données dans les archives sédimentaires (Birks & Birks 2000 ; Jørgensen *et al.* 2012a ; Willerslev *et al.* 2014). Leur identification est, de plus, largement dépendante de la nature et du degré de dégradation des éléments récoltés (Boessenkool *et al.* 2014).

Concernant la végétation passée, elle peut être inférée à partir de macro-restes mais également à partir des pollens. Ces enregistrements ont une longue histoire et ont fourni des informations cruciales pour retracer l'évolution des communautés végétales sous l'effet conjugué du climat, des perturbations naturelles et des activités humaines (*e.g.* David 2010a ; Miehe *et al.* 2014 ; Marquer *et al.* 2014). Les assemblages polliniques incluent à la fois des grains de pollens des environs mais également de sources plus distantes (Seppä & Bennett 2003). Le signal inféré à partir de ces enregistrements est donc régional. De plus, la capacité de cette approche pour détecter les faibles densités de populations ou les taxons produisant peu de pollens est, dans certains cas, limitée (Birks & Birks 2000 ; Jørgensen *et al.* 2012a ; Parducci *et al.* 2013). Les données issues de macrofossiles végétaux sont alors un excellent complément puisqu'elles offrent une résolution spatiale plus fine (*e.g.* à l'échelle du bassin versant). Néanmoins, les tailles d'échantillons obtenus sont souvent limitées (Jørgensen *et al.* 2012a ; Willerslev *et al.* 2014 ; Bossenkool *et al.* 2014).

d. L'outil ADN en paléoécologie

En complément des marqueurs morphologiques, l'utilisation de l'ADN issu d'archives naturelles s'est développée au cours des 10 dernières années pour l'étude des paléo-environnements terrestres. Ces archives incluent aussi bien le permafrost (*e.g.* Willerslev *et al.* 2003 ; Willerslev *et al.* 2014) que les sédiments non gelés (*e.g.* Hebsgaard *et al.* 2009 ; Thomsen *et al.* 2009), la glace (*e.g.* Willerslev *et al.* 2007 ; Jørgensen *et al.* 2012b) voire même des coprolithes et des déjections de rongeurs (*e.g.* Poinar *et al.* 1998 ; Murray *et al.* 2012 ; Wood *et al.* 2012). L'ADN d'une grande variété de taxons terrestres a ainsi pu être extrait puis séquencé, parmi lesquels des plantes (*e.g.* Willerslev *et al.* 2003 ; Jørgensen *et al.* 2012a ; Parducci *et al.* 2013, Pedersen *et al.* 2013), des champignons (Lydolph *et al.* 2005 ; Bellemain *et al.* 2013), des mammifères (Boessenkool *et al.* 2012 ; Haile *et al.* 2009 ; Willerslev *et al.* 2014) et des micro-organismes tels que les nématodes (Willerslev *et al.* 2014) ou des bactéries fécales (Madeja *et al.* 2010).

Les nouvelles technologies permettent aujourd'hui de séquencer massivement cet ADN ancien issu d'archives naturelles. Plusieurs études ont démontré la pertinence d'approches basées sur l'ADN environnemental et les NGS pour caractériser les paléo-communautés (*e.g.* Sønstebo *et al.* 2010 ; Jørgensen *et al.* 2012a ; Jørgensen *et al.* 2012b ;

Bellemain *et al.* 2013 ; Willerslev *et al.* 2014). Celles-ci ont, dans certains cas, permis la détection de taxons même en l'absence de macrofossiles visibles (*e.g.* Jørgensen 2012a ; Parducci *et al.* 2012 ; Willerslev *et al.* 2014).

Les lacs alpins sont considérés comme des sentinelles des changements environnementaux (Adrian *et al.* 2009) et leurs archives sédimentaires accumulent des informations sur la diversité biologique du lac (*e.g.* cyanobactéries, protistes, invertébrés) (Coolen & Gibson 2009 ; Savichtcheva *et al.* 2011 ; Stoof-Leichenring *et al.* 2012 ; Domaizon *et al.* 2013) mais également sur celle du bassin versant *via* les pollens, les spores, les macrorestes ou encore de l'ADN. L'ADN ancien retrouvé dans les sédiments lacustres a plusieurs origines supposées : les fèces, les morceaux de peaux et les macrorestes sont des sources potentielles pour l'ADN d'origine animale (Lydolph *et al.* 2005), alors que l'ADN d'origine végétal est supposé provenir principalement des tissus somatiques comme les racines fines et les feuilles et, dans une moindre mesure, des pollens (Jørgensen *et al.* 2012a). En effet, l'ADN contenu dans les grains de pollen est présent en faible quantité et le taux de succès de son amplification direct est très faible, notamment lorsqu'il est extrait à partir d'échantillon environnementaux, laissant supposer une faible contribution des pollens à l'ADN végétal contenu dans les sédiments (Parducci *et al.* 2005 ; Jørgensen *et al.* 2012a ; Parducci *et al.* 2013). L'ADN ancien retrouvé dans les sédiments lacustres pourraient ainsi être utilisé comme un marqueur des communautés animales et végétales présentes dans le bassin versant (Jørgensen *et al.* 2012a; Parducci *et al.* 2013; Boessenkool *et al.* 2014).

L'utilisation de l'ADN retrouvé dans les archives sédimentaires permet ainsi potentiellement d'inférer à la fois les perturbations biotiques, qu'elles soit induites par l'Homme (*e.g.* pâturage) ou non (*e.g.* invasion d'insectes), et leurs conséquences locales (*e.g.* à l'échelle du bassin versant) sur de multiples communautés.

e. Mise en contexte des études

Les deux études présentées dans cette partie se basent sur des données collectées dans le lac d'Anterne (Haute-Savoie, France, 2063m a.s.l.). Ce lac, actuellement situé au-dessus de la limite supérieure des arbres, et ses alentours, ont fait l'objet de nombreuses études aussi bien sédimentologiques (Giguet-Covex *et al.* 2011), que géochimiques (Arnaud *et al.* 2002 ; Giguet-Covex *et al.* 2012 ; Giguet-Covex *et al.* 2013), climatiques (Millet *et al.* 2009), archéologiques (Rey 2008 ; Rey 2013) ou encore

palynologiques (David 2010a ; David 2010b). Par conséquent, l'histoire des sols et l'évolution du paysage est relativement bien documentée (Giguet-Covex & Rey 2013). Ce lac constitue ainsi un modèle d'étude sur lequel peuvent être confrontées différentes méthodes de reconstruction paléo-environnementale. Les données récoltées, notamment archéologiques, suggèrent une présence humaine précoce autour de ce lac relativement isolé. L'hypothèse d'une activité pastorale ancienne dans le bassin versant a ainsi été émise pour expliquer la forte augmentation de l'érosion durant la période romaine, une période supposément chaude et sèche et par conséquent peu propice à de tels phénomènes (Giguet-Covex *et al.* 2011). De même, l'étude pollinique d'une petite tourbière située aux alentours du lac montre des changements importants, attribués à l'action anthropique, aux alentours de 4500 cal. BP avec une intensification supposée aux alentours de 1500 cal. BP. Néanmoins, aucune preuve directe n'a pu valider une telle hypothèse. Les deux objectifs principaux des études présentées dans cette partie sont ainsi d'utiliser l'outil moléculaire sur les archives sédimentaires du lac d'Anterne afin (i) de retracer précisément l'histoire des activités pastorales dans ce milieu de haute montagne au cours de l'Holocène et (ii) d'évaluer l'impact de ces perturbations anthropiques sur le paysage et les trajectoires végétales du bassin versant.

Les conditions environnementales, parmi lesquelles la température, l'humidité, l'acidité et l'oxygénation, sont des éléments clés pour la préservation à long terme de l'ADN, ce qui limite le nombre de sites disponibles pour les études basées sur l'ADN sédimentaire (Lindhal *et al.* 1993 ; Corinaldesi *et al.* 2008 ; Haile *et al.* 2009). Les conditions froides et anoxiques, que l'on retrouve dans le lac d'Anterne, sont supposées favoriser la préservation de l'ADN extracellulaire (Coolen & Overmann 2007 ; Coolen & Gibson 2009 ; Anderson-Carpenter *et al.* 2011 ; Corinaldesi *et al.* 2008 ; Corinaldesi *et al.* 2011). La composition sédimentaire du lac d'Anterne (argiles, sables, substances humiques et complexes organo-minéraux) favorise l'adsorption de l'ADN extracellulaire sur la matrice sédimentaire et sa préservation des enzymes de dégradation (Pietramellara *et al.* 2009 ; Coolen & Overmann 2007). Les sédiments du lac étant majoritairement clastiques, la compaction est rapide, ce qui limite l'épaisseur de sédiment à l'interface avec colonne d'eau, soumise aux mélanges par la circulation hydrique. Après la compaction, la percolation verticale des liquides est limitée par la présence de sédiments stratifiés et saturés, assurant une succession temporelle fiable (Parducci *et al.* 2013). Ces caractéristiques font du lac d'Anterne un lieu, *a priori*, adapté

pour l'utilisation de l'ADN sédimentaire à des fins paléoécologiques.

Dans l'article C, une approche de *metabarcoding* environnemental a ainsi été appliquée sur 47 tranches de carottes sédimentaires issues du lac d'Anterne, couvrant les 10000 dernières années, afin de retracer l'histoire des activités pastorales dans le bassin versant, et son impact sur le paysage, en particulier les flux érosifs et le couvert végétal. L'ADN animal constitue un marqueur direct du pastoralisme, qui, par comparaison aux indicateurs d'activités pastorales tels que les molécules fécales ou les spores de champignons coprophiles (Hausmann *et al.* 2002 ; D'anjou *et al.* 2012 ; Etienne *et al.* 2013), permet de discriminer les espèces domestiques. Les données animales et végétales obtenues à partir de l'ADN sédimentaire ont ainsi été combinées aux données sédimentologiques préalablement obtenues afin de caractériser l'effet des activités humaines sur les dynamiques d'érosion au cours de l'Holocène.

L'étude suivante se concentre sur la reconstitution des communautés végétales, à partir de l'ADN sédimentaire, et de leurs trajectoires au cours de l'Holocène. L'objectif est, premièrement, de montrer la pertinence des données végétales inférées par *metabarcoding* environnemental pour les études des trajectoires paléoécologiques. L'utilisation de taxons indicateurs (Iversen 1964) est un outil d'interprétation largement utilisé pour caractériser les changements de végétation passés (Mazier *et al.* 2009), particulièrement dans les reconstructions basées sur l'ADN sédimentaire (*e.g.* Sønstebo *et al.* 2010 ; Jørgensen *et al.* 2012a ; Parducci *et al.* 2013). L'approche comparative, elle, se base sur l'identification des caractéristiques des assemblages actuels, et l'utilisation de comparaisons numériques entre ces assemblages modernes et ceux inférés à partir d'études environnementales, pour discriminer les différents milieux et/ou pratiques humaines (Mazier *et al.* 2009). Dans cette étude, les assemblages de plantes inférés par *metabarcoding* ont ainsi été comparés à ceux issus de plus de 3000 relevés botaniques dans le but d'estimer les préférences d'habitats des espèces. Le second objectif est d'étudier les facteurs susceptibles d'influencer les trajectoires à long terme des communautés végétales, et plus particulièrement le rôle des activités anthropiques. Pour se faire, les données végétales obtenues ont été corrélées avec des indicateurs de développement des sols, de l'évolution du climat et des activités humaines.

**2. Histoire du pastoralisme et mise en place des paysages sous l'effet de l'Homme
révélées à partir de l'ADN sédimentaire lacustre**

a. Article C

Article C

Long livestock farming history and human landscape shaping revealed by lake sediment DNA

Publié dans *Nature Communications*

(DOI: 10.1038/ncomms4211)

Charline Giguet-Covex^{1,2*}, Johan Pansu^{1*}, Fabien Arnaud², Pierre-Jérôme Rey²,
Christophe Griggo², Ludovic Gielly¹, Isabelle Domaizon³, Eric Coissac¹, Fernand David⁴,
Philippe Choler^{1,5}, Jérôme Poulénard² & Pierre Taberlet¹

¹ LECA, CNRS UMR 5553, Université Joseph Fourier, 38041 Grenoble Cedex 9, France

² EDYTEM, Université de Savoie, CNRS Pôle Montagne, 73376 Le Bourget du Lac, France

³ INRA, UMR042 CARRTEL, Thonon les Bains, France

⁴ CEREGE, Université Aix-Marseille, 13545 Aix en Provence Cedex 4, France

⁵ SAJF, UMS 3370 CNRS – Université Joseph Fourier, 38041, Grenoble Cedex 9, France

* Co-premiers auteurs

Contribution personnelle : participation à la mise en place de l'étude et aux analyses moléculaires, filtrage des séquences et assignation taxonomique, analyse des données ADN, co-rédaction du manuscrit

Abstract

The reconstruction of human-driven, Earth-shaping dynamics is important for understanding past human/environment interactions and for helping human societies that currently face global changes. However, it is often challenging to distinguish the effects of the climate from human activities on environmental changes. Here we evaluate an approach based on DNA *metabarcoding* used on lake sediments to provide the first high-resolution reconstruction of plant cover and livestock farming history since the Neolithic period. By comparing these data with a previous reconstruction of erosive event frequency, we show that the most intense erosion period was caused by deforestation and overgrazing by sheep and cowherds during the Late Iron Age and Roman Period. Tracking plants and domestic mammals using lake sediment DNA (lake *sedDNA*) is a new, promising method for tracing past human practices, and it provides a new outlook of the effects of anthropogenic factors on landscape-scale changes.

Introduction

The concept of the Anthropocene¹ recognizes that humankind has become one of the many factors capable of affecting global geological processes. However, the questions of when, where and how human societies became such a geological force remain to be resolved. Landscape shaping, which affects erosion, is most likely the oldest human-caused geological process^{2,3}. In particular, the domestication of animals has long been a determinant of the modification of landscapes⁴. However, because changes in climate also influence landscape transformations, it is often challenging to disentangle anthropogenic effects from those of a climatic origin. Past reconstructions of landscape changes and human activities are required to meet this challenge. Currently, our knowledge of ancient societies and their effects on the environment is mainly based on sparse archaeological data as well as documentary and natural archives⁵⁻¹³. Even recently developed proxies of pasturing activities (*e.g.* those utilizing fecal molecules and coprophilous fungi spores in lake sediments) are indirect and not specific to particular domestic animal species⁵⁻⁷. Thus, information available on the nature of past human activities is scattered or imprecise. It is therefore necessary to address this methodological gap in order to better understand the impact of human activities on local-to-global geological processes. Recent developments in the combined use of high-throughput sequencing¹⁴ and *metabarcoding* approaches enable the assessment of the

biodiversity of current ecosystems using extracellular DNA that is contained in environmental samples^{15,16}. DNA from a variety of taxa, including fecal bacteria, plants and domestic mammals, has been successfully extracted and sequenced directly from natural archives (frozen soils, cave sediments and lake sediments), even in the absence of visible macrofossils¹⁷⁻²⁵. The application of DNA *metabarcoding* to lake sediments has also proven its potential to trace past vegetation cover²². However, this tool has never been used with the objective of detecting the presence of mammals and plants in ancient ecosystems and to reconstruct a detailed picture of human land use since the Neolithic Period.

Here we test whether the application of the *metabarcoding* approach to high-resolution lake sediment archives can provide reliable and novel information regarding human occupation. We use the well-described sediments of the sub-alpine Lake Anterne²⁶, which is located in the North French Alps (2063-2478 m asl) above the present treeline ecotone (Fig. 1). The catchment area of Lake Anterne has steep slopes made of calcshales and black shales, which are easily erodible rocks (see Methods). Thus, Lake Anterne sediments provide a good record of erosional processes (Fig. 1d). We combine results from the *metabarcoding* approach with a previous reconstruction of the frequency of erosive events²⁶ to document the effects of human activities on erosion dynamics. In particular, we investigate the hypothesis of an anthropogenic origin of the strong increase in erosion that occurred during the Roman Period, which is a supposed warm and dry climatic period²⁷, expected to correspond to unfavorable conditions for erosion. Our reconstruction of the livestock farming history shows periods of grazing activity with sheep and/or cowherds that mainly occurred at the end of the Iron Age, the Roman Period, the Medieval Times and the Modern Period. During the Roman Period, deforestation strongly affected *Alnus* sp. and the frequency of erosive events is the highest of the last 10,000 years. Together, these results suggest that the most intense grazing pressure on landscape and that erosion in this sub-alpine area occurred during the Roman occupation.

Methods

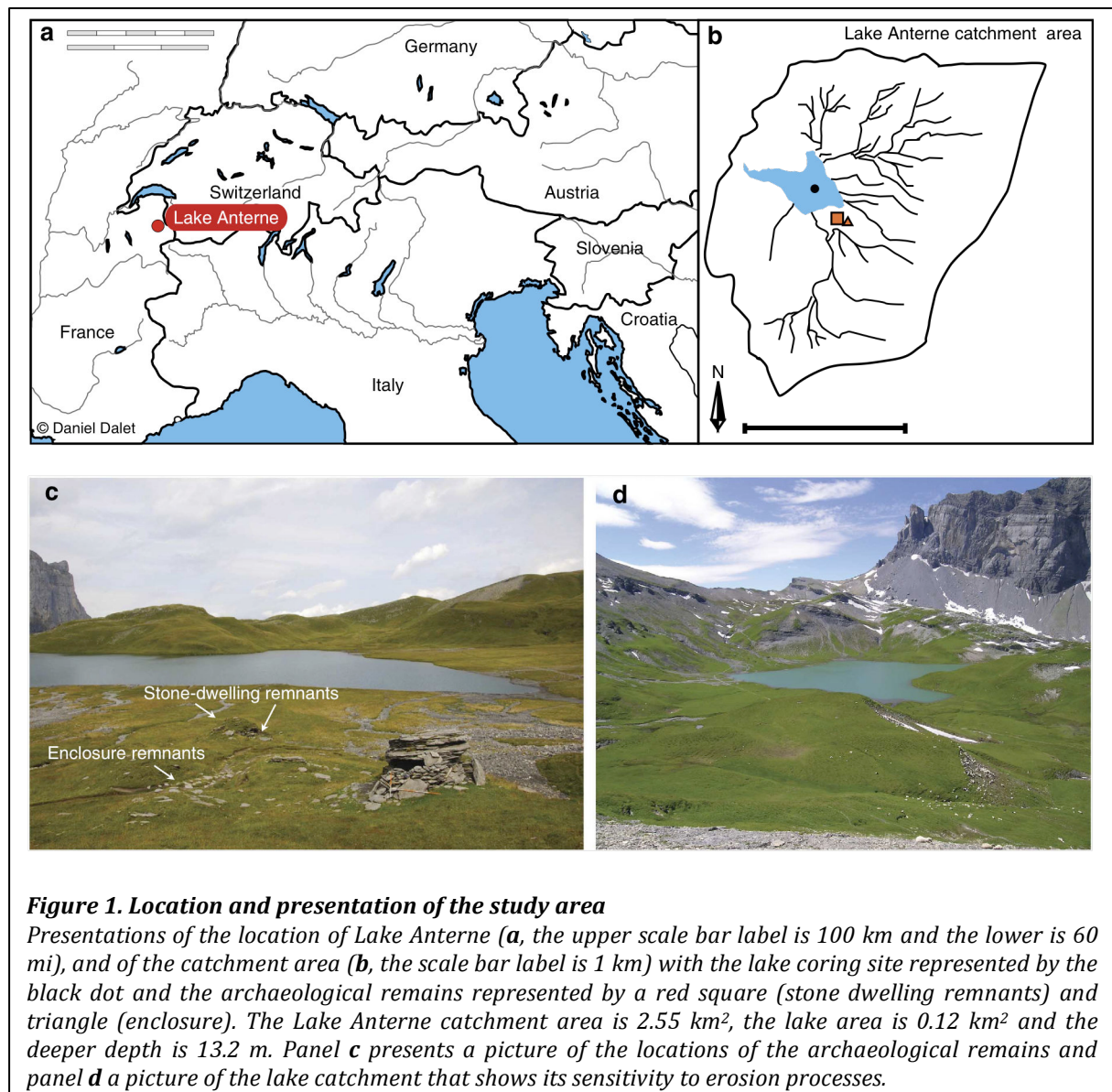
Study site

Lake Anterne is a high-altitude lake (2063 m asl, 0.12 km² and 13.2 m maximum water depth) that is located in the northern French Alps (Fig. 1a & 1b). Lake Anterne has

a small catchment (2.55 km²) with steep slopes that are composed of easily erodible rocks (calcshales, namely Bajocian; and black shales, namely Bathonian-Oxfordian). The following three main soil types were identified in these rocks: leptosols (~50% of the catchment area), cambisols (~25%) and acid soils (podzosols/stagnosols; ~15%)²⁶. These soil types represent different stages of soil development (from less- to more-developed soils). Given the presence of easily erodible material on steep slopes, the lake sedimentation is mainly due to erosion processes (Fig. 1d). This sedimentation consists of clastic laminations that were formed by erosive events (mainly floods triggered by precipitation events)³³. The lake is frozen 6 to 7 months per year, which favors anoxic conditions at the lake bottom. During the summer, anoxic conditions also occur periodically (Giguet-Covex unpublished data) and are most likely due to organic matter decomposition by aerobic bacteria. During the year, the temperature at the lake bottom varies between 4 and 8°C (Giguet-Covex unpublished data). These anoxic conditions and the low mean annual temperature favor DNA preservation³⁰. A 20-m-long composite core (ANT-07) was retrieved in September 2007 from the deepest part of the lake and far from active deltas to obtain a continuous sequence that was as long as possible (Fig. 1b). This core, which was used for DNA analyses, was shrink-wrapped to limit oxidation of the sediments and stored in a refrigerated area (4°C; similar temperature as the lake bottom). A total of 10,160 years of environmental history could be reconstructed from this core²⁶. For the present study, we focused on the last 6,400 years (16.4 m of the sediment core), which represents the period that recorded the main changes in erosion patterns²⁶.

Dating of the sediment core

The age–depth model was based on 18 ¹⁴C measurements on terrestrial macroremains. AMS ¹⁴C analyses were performed by the Poznan Radiocarbon Laboratory and the LMC14 Laboratory (CNRS). All dates were calibrated (2 sigma) using the Radiocarbon Calibration Program REV 6.0.1³⁸ and Intcal09 calibration curve³⁹ (Supplementary Table 3). The age–depth model was then constructed without taking into account instantaneous deposits (thicker than 5 mm)²⁶ and by fitting a cubic spline curve using the MCAgeDepth software⁴⁰ (Supplementary Fig. 2).



Frequency of erosive events

To determine the frequency of erosive events, the sedimentary deposits that resulted from these events were visually identified and documented in terms of stratigraphic location. Visual identification was based on the sedimentological features of these deposits. These features are characterized by the succession of three facies that become finer upward^{36,41} from bottom to top as follows: coarse and well-sorted sediments; a mixture of silts and fine sands; and a layer of fine silts and clays. The thickness of these erosive deposits varies from 0.5 to 27 cm. The frequency was then computed as a running sum over a 100-yr time window, as estimated by our age-depth model that was based on ¹⁴C dating²⁶.

Sediment core sampling and DNA extraction

Metabarcoding was performed on samples from a 16.4-m-long sediment core (ANT-07) that corresponded to the most recent 6,400 years. Sediment core sampling was performed in an isolated laboratory at Université de Savoie, where no DNA research had previously been performed (as suggested in reference 42). Forty-four slices were sampled from the core, and the edges of the slices were removed to avoid contamination from water spreading along the core. Each sample was divided into two subsamples. The weight of the sediment that was obtained and used for extractions was 13.71 +/- 1.15 g. Two DNA extractions were performed per subsample (*e.g.* four extractions per sediment slice), and two PCRs were performed per extraction (*e.g.* eight PCRs per sediment slice). This strategy allowed for the detection of possible sporadic contamination⁴². The presence of a DNA sequence in only one PCR could be due to sporadic contamination, PCR bias or stochasticity due to low DNA concentration. Four extraction controls were also processed. Each of the controls was amplified two times with the primer pairs that were used in this study (see below). These controls allow potential contaminations to be identified in the PCR products.

Extracellular DNA extraction and amplification were performed in two rooms that were specifically dedicated to ancient DNA at the Laboratory of Alpine Ecology (Université de Grenoble, France). The extraction of this DNA from the core subsamples was performed by mixing the sediment with 15 mL of saturated phosphate buffer (0.12 M Na₂HPO₄; pH ≈ 8) for 15 minutes to release the DNA bound to particles, and the mixture was then centrifuged (10 minutes at 10000 g). The resulting supernatant was recovered and used as starting material in the extraction steps that were carried out using a commercial kit (NucleoSpin® Soil; Macherey-Nagel, Düren, Germany) by skipping the lysis step and following the manufacturer's instructions¹⁶. The DNA extract was eluted in 100 µL of SE buffer.

DNA amplification and high-throughput sequencing

We targeted two taxonomic groups as follows: mammals and plants. Mammalian DNA was amplified using a new primer pair (MamP007F, 5'-CGAGAAGACCCTATGGAGCT-3'; and MamP007R, 5'-CCGAGGTCRCCCCAACC-3'), which led to the amplification of a 60- to 84-bp fragment of the mitochondrial 16S gene. These

primers were designed using the *ecoPrimers* program⁴³. Plant DNA was amplified using universal primers (*g*, 5'-GGGCAATCCTGAGCCAA-3'; and *h*, 5'-CCATTGAGTCTCTGCACCTATC-3') that targeted the P6 loop region of the chloroplast *trnL* (UAA) intron⁴⁴. The length of the amplified fragment is highly variable among species and ranges from 10 to 146 bp (mean length of 45 bp). To assign sequence reads to the relevant sample after high-throughput sequencing, 8-bp tags (with at least 5 differences between each) were added to the 5' end of each primer^{15,45}.

All DNA amplifications were carried out in a final volume of 30 µL containing 2 µL of DNA extract. The amplification mixture contained 1 U of AmpliTaq Gold® DNA polymerase (Applied Biosystems), 15 mM Tris-HCl, 50 mM KCl, 2 mM MgCl₂, 0.2 mM of each dNTP, 0.1 µM of each primer and 4.8 µg of bovine serum albumin (BSA, Roche Diagnostic). To restrict the amplification of human DNA when amplifying mammalian DNA⁴⁶, we designed a human-specific blocking oligonucleotide (MamP007_B_Hum1, 5'-GGAGCTTTAATTTATTAATGCAAACAGTACC-C3-3') and added it to the PCR mixture at a concentration of 2 µM. For both primer pairs, the PCR mixture was denatured at 95°C for 10 minutes, followed by 45 cycles of 30 s at 95°C and 30 s at 50°C (no elongation step). The PCR products were then purified and mixed together (equivolume mixes) before sequencing, which was performed using an Illumina HiSeq 2000 platform (2 x 100-bp, paired-end reads). Two different runs were performed as follows: the first run included 9 samples (called 'RTM') to test the method; and the second run included 35 samples (called 'ANT') to improve the resolution of our record. This second run also integrated the control samples.

Sequence analysis and taxa assignment

Sequence data were analyzed using the OBITOOLS software (<http://www.grenoble.prabi.fr/trac/OBITools>). First, the direct and reverse reads that corresponded to the same DNA fragment were aligned and merged using the *SolexaPairEnd* program, with the quality being taken into account when merging the two reads. The *ngsfilter* program was then used to assign sequences to the relevant sample based on the associated primers and tags. Only sequences containing perfect tags and primers with a maximum of 3 errors were considered. We used the *obiuniq* program to merge strictly identical sequences and kept the information about the original samples they originated from. Afterward, the sequences were filtered using the *obigrep* program

according to their length as follows: sequences shorter than 50 bp for the MamP007 and 10 bp for the gh primers were removed; and sequences that had an occurrence lower than 100 over the entire dataset were removed. Because we used short markers, clustering sequences at a high level of similarity, as is usually performed with this type of data⁴⁷, was not suitable because a single difference in the nucleotide sequence often indicated different species. The *obiclean* program⁴⁸ was developed to detect artifactual sequences that contained amplification and/or sequencing errors. For each PCR product, the program creates networks that connect sequences with one indel or substitution difference. Two sequences are linked if the occurrence of the less-frequent sequence is less than 20% of that of the other sequence. A status is then assigned to each sequence as follows: “head” (most common sequence of a network), “singleton” (sequence without a variant) or “internal” (sequences that are neither “head” nor “singleton” and that most likely correspond to amplification or sequencing errors). Only sequences that were more often “head” or “singleton” than “internal” in the global dataset were kept for the following steps. Finally, the sequences were assigned to the relevant taxa using the *ecoTag* program, which finds highly similar sequences in a suitable database (created by *in silico* PCR from Genbank with the *ecoPCR* program⁴⁹). Each sequence was then assigned to unique taxa. If no complete match between a unique sequence in the database and the query sequence was found, the allotted taxa corresponded to the last common node in the NCBI taxonomic tree of all the taxids that annotated the sequences and matched the query sequence⁵⁰. Only sequences corresponding to the *Viridiplantae* phylum when using plant primers and to the *Mammalia* class when using the mammalian primers were saved for the following steps.

Another filtering step was performed to disentangle sequences that were genuinely released by organisms from the lake catchment from potential sporadic contamination. A sequence was considered genuine (with a high confidence level) at a given time if it was found at least 2 times among the 8 PCR replicates and with a minimum of 5 read numbers. This criterion obviously dismissed many putative taxa that were present at low abundance or had heterogeneous distribution¹⁷, which is why sequences that were retrieved in one replicate were represented as ‘sequences potentially present’.

Results

Plant DNA and landscape changes

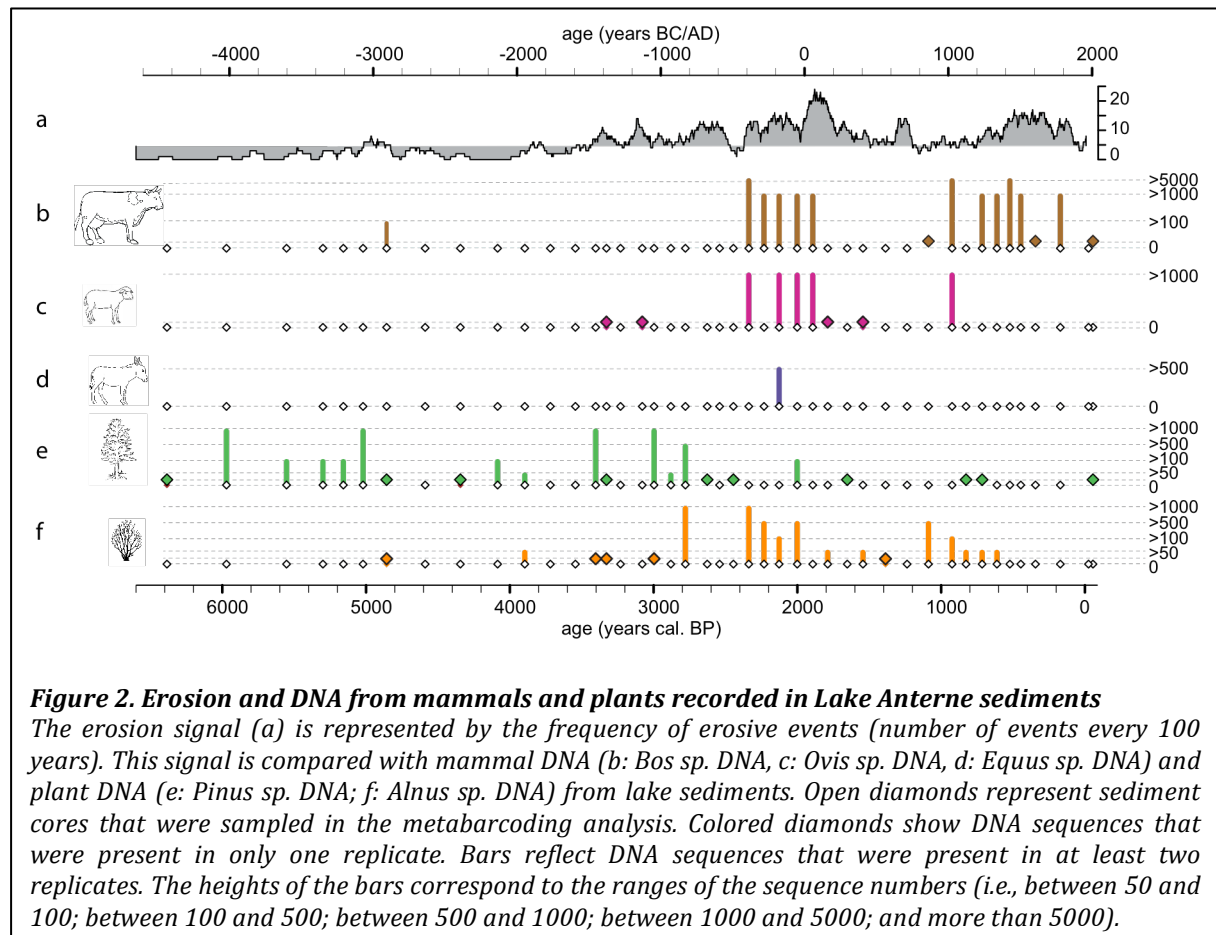
For the first run, which included 9 samples, we obtained 3,151,601 reads that represented 68,137 unique sequences. After the different filtering steps, as described in the methods, the number of sequences was reduced to 2,140,829 (428 unique sequences). For the second run, which included 35 samples, 5,575,081 reads (386,880 unique sequences) were obtained after high-throughput sequencing. This number was further reduced to 4,059,103 after filtering (909 unique sequences). Plant contaminations, mainly from a plant sequence from the Lamiales order (42% of total read numbers after filtering), were detected in the control samples of the second run. Overall, a large dataset of plant taxa is obtained. For this study, we focused on *Pinus* sp. and *Alnus* sp. (Supplementary Table 1 & 2; Fig. 2) to document deforestation dynamics in relation to pasturing practices. Only insignificant contaminations (one read) were observed regarding these plant taxa of interest (Supplementary Table 2). *Pinus* sp. DNA was mainly present before 2800 cal. yr BP. More precisely, three phases of high DNA content were recorded as follows (Fig. 2): between 6000 and 5000 cal. yr BP; at 3400 cal. yr BP; and between 3000 and 2800 cal. yr BP. High occurrences of *Alnus* sp. DNA were recorded after the decline of *Pinus* sp. DNA at 2800 cal. yr BP, between 2350 and 2000 cal. yr BP and at approximately 1000 cal. yr BP.

Mammal DNA and livestock farming history

For the first and second runs, we obtained 967,653 and 1,029,717 reads, respectively, after high-throughput sequencing (81,011 and 81,684 unique sequences, respectively). After filtering, the number of sequences was reduced to 165,373 (19 unique sequences) for the first run and 78,910 (29 unique sequences) for the second run.

For this study, three mammalian genera that are known to be associated with human activities were selected (Supplementary Tables 1 & 2; Fig. 2): *Bos* sp., *Ovis* sp. and *Equus* sp. Only insignificant contaminations (one or two reads) were observed regarding these genera (Supplementary Table 2). Cow DNA was detected in at least two PCRs during the Neolithic Period (4800 cal. yr BP), at the end of the Iron Age/onset of the Roman Period (2400-1900 cal. yr BP) and during the Medieval/Modern times (approximately 1000 AD; between 1200 and 1500 AD; and at 1750 AD). In at least two

PCRs, sheep DNA was found at the end of the Iron Age/onset of the Roman Period and at approximately 1000 AD. *Equus* sp. DNA was also recorded at 2100 cal. yr BP. Over the last 6500 years, the highest cow and sheep DNA contents were observed at the end of the Iron Age, onset of the Roman Period and the Medieval-Modern times (Supplementary Table 2).



Discussion

In this study, we evaluated whether DNA *metabarcoding* applied to lake sediment was a reliable approach for tracing the history of human activities and landscape changes. In practice, the reliability of the lake sediment DNA record can be affected by taphonomical processes, which implicate several factors. First, physicochemical conditions and sediment composition must be favorable to avoid the complete degradation of DNA. In Lake Anterne, sediment compounds, such as clays, sands, humic substances and organo-mineral complexes, that allow DNA adsorbance, together with

cold temperature and periodic anoxic conditions on the lake bottom (see Methods), are likely to have favored DNA preservation²⁸⁻³⁰. Moreover, DNA was present in all samples, even the oldest, suggesting that the DNA was sufficiently preserved to successfully reconstruct the presence of the taxa of interest. Second, DNA mixing through water circulation in lake sediment layers must be a minor process to maintain the DNA stratigraphic position. Lake sediments are permanently saturated with water, which prevents percolation of liquids and, thus, DNA leaching. Before sediment compaction, vertical advection of pore water can occur within the first centimeters at the water/sediment interface. However, in the case of Lake Anterne, the sediments are mainly clastics, which induce fast compaction and limit the thickness where mixing is possible. After compaction, vertical advection becomes minimal^{19,22}. Thus, we assume that the temporal succession of our DNA record was not disturbed.

The reliability of a lake *sedDNA* record also depends on potential bias linked to the analytical techniques (possible contamination; amplification and sequencing errors). However, strict precautions were taken in the laboratory at each step of the experiment to avoid contamination. All of the laboratory steps (sediment core sampling, DNA extraction and DNA amplification) were performed in a room where no DNA research had previously been performed and in rooms that were specifically dedicated to ancient DNA. Additionally, replicates during the sampling, extraction and amplification steps allowed an optimal chance to obtain mammal and plant DNA sequences and enabled us to detect sporadic contamination, which is expected to be present in only one of eight replicates. Appropriate data treatment was also performed to filter out amplification and/or sequencing errors. These measures and the quasi-absence of contamination in the negative controls confirmed the quality of our data. Moreover, the temporal consistency of the plant and mammal DNA analyses (no random distribution of the data) further strengthens the confidence in our lake *sedDNA* data.

Finally, other results from independent archives and proxies provided an independent validation of a portion of our lake *sedDNA* results. Indeed, the pattern of tree cover that was revealed by our lake *sedDNA* is concordant with previous regional palynological studies that were performed near the Anterne catchment (Survilly, 2235 m asl; Ecuellen, 1850 m asl)^{31,32}. Moreover, in support of the reconstruction of livestock farming that was obtained using mammal DNA, one enclosure and two stone dwelling remnants were found near the lake (Fig. 1b & 1c and Supplementary Fig. 1). The

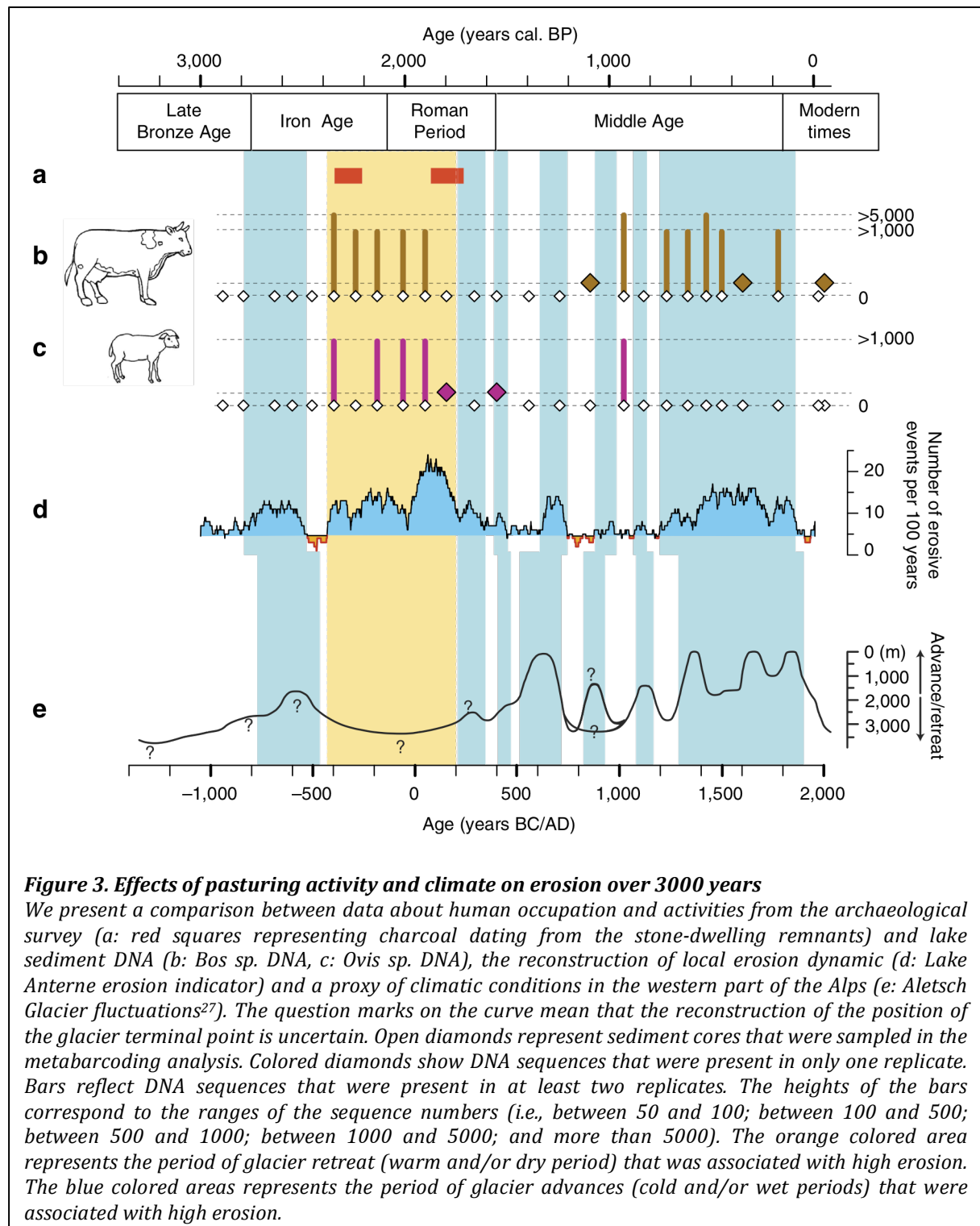
foundation of the dwelling closest to the lake was dated to the Second Iron Age (2,335-2,155 cal. yr BP), but, according to ceramic analysis (unpublished data) and ^{14}C dating on charcoal (1,865-1,705 cal. yr BP), the most intense occupation most likely occurred at the beginning of the Roman Period. One hundred and fifty-two bone remains that were found at the site were determined (unpublished data) as follows: one hundred and forty-three were attributed to sheep (*Ovis aries*) or goat (*Capra hircus*), six were attributed to sheep, and three were attributed to domestic rooster (*Gallus gallus domesticus*). The presence of these bones allowed, at least in part, for the validation of our mammal DNA results. The lack of cattle bones may have occurred because these animals were not eaten on site or because of the limited archaeological research in this zone. The transition at approximately 1200 AD from a mix of sheep and cow herds to a dominance of cowherds has been previously documented through historical archives^{5,33} of the same period (between 1300 and 1500 AD) in the French and Swiss Alps, which again validates our DNA results. However, our study is the first to track such land use changes using a natural archive. This approach enabled a more precise detection of past land-use practices, especially for prehistoric periods and locations where no written traces were left by inhabitants.

We did not record sheep DNA in the lake sediments that corresponded to the present day, although we have evidence of the presence of sheep at a low density in the catchment periodically during the summer. This result highlights the need of a relatively high amount of mammalian DNA that was transferred to the lake to be detected in the sediments (see Supplementary Discussion about mammalian DNA detection) and shows that the absence of DNA in lake sediments should be treated with caution, as this does not necessarily highlight the absence of the targeted species. In contrast, no cattle herd is currently in the catchment, but we record the presence of cow DNA in one of eight PCRs of the recent lake sediments, which may be explained by sporadic contamination or stocking of older cow DNA in soils because it is known that DNA can persist in alpine soils for at least 50 years³⁴. Thus, these results, which are contradictory to expectations, reveal that lake *sed*DNA allows the reconstruction of past livestock history with a high level of confidence when the stocking rate (in animal unit/day/ha) is relatively high and/or the practice induces important transfer of domestic mammal DNA (mainly from feces, skin flakes and chitinous material³⁵) to the lake. Based on these considerations,

the main phases of livestock farming in the sub-alpine Lake Anterne catchment occurred during the end of the Iron Age, the Roman Period and the Medieval-Modern times.

The second objective of our study was to integrate our lake *sed*DNA record in the context of the erosion dynamic history to bring a novel perspective regarding human/environment interactions, in particular during the Roman Period (0-200 AD), which was when the highest frequency of erosive events of the entire Holocene Period was recorded²⁶ (Fig. 3). As shown by the reconstruction of Aletsch glacier fluctuations (Swiss Alps), the Roman Period was characterized by a warm and/or dry climate²⁷ (Fig. 3), *e.g.* conditions that are not favorable to erosion. Based on these facts, we previously hypothesized that human-associated activities were the primary determinants of the increased frequency of erosive events³⁶ during this period at Lake Anterne. Our lake *sed*DNA record (Fig. 2 & 3) provides new evidence to support the “human impact” hypothesis of this erosion paroxysmal phase and provides a more comprehensive picture regarding the nature and intensity of human practices during this period. The presence of cattle and sheep DNA in lake sediments is indicative of intense pasturing activity with mixed herds in the catchment. Grazing has strong effects on soil erosion through direct processes, such as soil compaction, reduction of infiltration and detachment of particles, as well as indirect processes, such as the reduction of vegetation cover and development of bare and unstable soils³⁷. The disappearance of *Alnus* sp. (Fig. 2) may reflect subalpine forest clearance, and this change in the landscape vegetation suggests significant grazing pressure that was most likely associated with a high stocking rate. The direct effects of trampling and grazing by herds, as well as the drastic reduction of the tree cover on soil stability, helps to explain the extensive erosion that was recorded in lake sediments of this period.

‘Natural History’, written by Pliny the Elder in 77-79 AD, describes the manufacturing of one of the most popular cheeses in Rome, which was made from cow milk in the Centronian Alps (corresponds approximately to the North French Alps) (see Supplementary Note 1). Pliny’s statement suggests the development of a pastoral activity that was profitable enough to engage in an important cheese trade with Rome, which was approximately 1000 km away. The high amount (high number of reads in two of eight replicates; supplementary Table 2) of cattle DNA in the sedimentary record is consistent with a boom in this type of economy.



The temporal depth of our natural archive shows that this activity most likely began at the end of the Iron Age (400 BC) (gFi. 3) and, thus, occurred prior to the Roman invasion of the region in 76 BC. Furthermore, the presence of equine DNA at 200 BC (Fig. 2) suggests that intensive farming activities had already been organized by the Gallic people in this high-altitude pasture. Nevertheless, because we recorded an expansion of

Alnus sp. and a lower increase in the frequency of erosive events than during the Roman Period, grazing pressure and, consequently, farming activity were most likely less important (Fig. 2). The Roman intensification of livestock farming, which required more space for grazing, most likely explains the complete disappearance of *Alnus* sp.^{31,32} (Fig. 2). Finally, at the end of the Iron Age, the green alder shrub expansion, which stabilized the soil cover, and the dry climatic conditions^{27,36} supported a major role of sheep and cowherds in catchment erosion.

The *metabarcoding* approach applied on lake sediments reveals valuable and novel data on the occupation dynamics of the European Alps, as it provided the first high-resolution reconstructed history of livestock farming over six millennia at the species level. Nevertheless, independent and complementary data (sedimentological, palynological, archaeological and historical) are certainly important when validating and interpreting lake *sedDNA* analyses. By combining sedimentological data and lake *sedDNA* evidence, we showed that overgrazing by sheep and cowherds associated with deforestation triggered the most important erosion crisis of the Holocene in the Anterne catchment, which occurred during the Roman Period. Finally, we provided evidence for an important role of livestock farming on erosion dynamics, which allowed for the discrimination of the effects of climate and anthropogenic drivers. Lake sediment DNA (*sedDNA*) would be of great interest to the archaeological and anthropological communities. In addition to classical paleoenvironmental approaches, the use of lake *sedDNA* will likely become a new approach for reconstructing Holocene paleoenvironments to better understand the human/climate/environment interactions during the Anthropocene.

Acknowledgments

We thank the French National Research Agency for funding the fieldwork and sedimentological measurements through the Pygmalion research program (ANR BLAN07-2_204489). Archaeological investigations were led thanks to the Pygmalion program and grants from the Haute Savoie Service of Archaeology under the responsibility of Joël Serralongue, who we would also like to thank. The Université de Savoie and Joseph Fourier as well as the CNRS jointly funded C. Giguët-Covex's salary and DNA measurements in the framework of the DiPEE Chambéry-Grenoble (Retromont

program) and the MetaBar research program (ANR 11 BSV7 020 01). We would also like to thank ASTERS, the manager of the Haute Savoie natural reserves, for constant help since the onset of the paleoenvironmental studies on Lake Anterne.

References

1. Crutzen, P. J. Geology of mankind. *Nature* **415**, 23-23 (2002).
2. Oldfield, F. & Dearing, J.A. The Role of Human Activities in Past Environmental Change. In *Paleoclimate, global change and the future* (eds Alverson, K.D., Bradley, R.S., Pedersen, T.F.) (Springer Verlag, 143-162, 2003).
3. Ruddiman, W.F. The Anthropogenic greenhouse era began thousands of years ago. *Climatic Change* **61**, 261-293 (2003).
4. Diamond, J. Evolution, consequences and future of plant and animal domestication. *Nature* **418**, 700-707 (2002).
5. Hausmann, S. *et al.* Interactions of climate and land-use documented in the varved sediments of Seeburgsee in the Swiss Alps. *The Holocene* **12**, 279-289 (2002).
6. D'Anjou, R.M., Bradley, R.S., Balascio, N.L. & Finkelstein, D.B. Climate impact on human settlement and agricultural activities in northern Norway revealed through sediment biogeochemistry. *PNAS* **109**, 20332-20337 (2012).
7. Etienne, D., Wilhelm, B., Sabatier, P., Reyss, J.-L. & Arnaud, F. Influence of sample location and livestock numbers on *Sporormiella* concentrations and accumulation rates in surface sediments of Lake Allos, French Alps. *Journal of Paleolimnology* **49**, 117-127 (2013).
8. Wick, L. & Tinner, W. Vegetation Changes and Timberline Fluctuations in the Central Alps as Indicators of Holocene Climatic Oscillations. *Arctic and Alpine Research* **29**, 445-458 (1997).
9. Schmidt, R., Koinig, K. A., Thompson, R. & Kamenik C. A multi proxy core study of the last 7000 years of climate and alpine land-use impacts on an Austrian mountain lake (Unter Landschitzsee, Niedere Tauern. *Palaeogeography, Palaeoclimatology, Palaeoecology* **187**, 101-120 (2002).
10. Walsh, K., Mocci, F. & Palet-Martinez, J. Nine thousand years of human/landscape dynamics in a high altitude zone in the southern French Alps (Parc National des Ecrins, Hautes-Alpes). *Preistoria Alpina* **42**, 9-22 (2007).
11. Deschler-Erb, S. Subsistence Strategies in the Neolithic and Early Bronze Age Hill Settlement of Thaur-Kiechlberg/Tyrol: The Example of the Animal Bones. *Mining in European History and its Impact on Environment and Human Societies*, 221-225 (2010).
12. Curdy, P. & Chaix, L. Les premiers pasteurs du Valais. *Globe*, 93-116 (2009).

13. Röpke, A., Stobbe, A., Oeggl, K., Kalis, A.J. & Tinner, W. Late-Holocene land-use history and environmental changes at the high altitudes of St Antönien (Switzerland, Northern Alps): Combined evidence from pollen, soil and tree-ring analyses. *The Holocene* **21**, 485-498 (2011).
14. Glenn, T.C. Field guide to next-generation DNA sequencers. *Molecular Ecology Ressources* **11**, 759–769 (2011).
15. Valentini, A., Pompanon, F. & Taberlet, P. DNA *barcoding* for ecologists. *Trends in Ecology and Evolution* **24**, 110-117 (2009).
16. Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C. & Willerslev, E. Towards next-generation biodiversity assessment using DNA *metabarcoding*. *Molecular Ecology* **21**, 2045–2050 (2012).
17. Willerslev, E. *et al.* Diverse Plant and Animal Genetic Records from Holocene and Pleistocene Sediments. *Science* **300**, 791-795 (2003).
18. Madeja, J., Wacnik A., Wypasek, E., Chandran, A. & Stankiewicz, E. Integrated palynological and molecular analyses of late Holocene deposits from Lake Milkowskie (NE Poland): Verification of local human impact on environment. *Quaternary International* **220**, 147-152 (2010).
19. Anderson-Carpenter, L. L., McLachlan, J. S., Jackson, S. T., Kuch, M., Lumibao, C. Y. & Poinar, H. N. Ancient DNA from lake sediments: Bridging the gap between paleoecology and genetics. *BMC Evolutionary Biology* **11**, 1-15 (2011).
20. Jorgensen, T. *et al.* Island in the ice: detecting past vegetation on Greenlandic Nunataks using historical records and sedimentary ancient DNA *Meta-barcoding*. *Molecular Ecology* **21**, 1980-1988 (2011a).
21. Jorgensen, T. *et al.* A comparative study of ancient sedimentary DNA, pollen and macrofossils from permafrost sediments of northern Siberia reveals long-term vegetational stability. *Molecular Ecology* **21**, 1989-2003 (2011b).
22. Parducci, L. *et al.* Glacial Survival of Boreal Trees in Northern Scandinavia. *Science* **335**, 1083-1086 (2012).
23. Hebsgaard, M.B. *et al.* The farm beneath the sand – an archeological case study on ancient ‘dirt’ DNA. *Antiquity* **83**, 430-444 (2009).
24. Haile, J. *et al.* Ancient DNA chronology within sediment deposits: Are paleobiological reconstructions possible and is DNA leaching a factor. *Molecular Biology Evolution* **24**, 982-989 (2007).
25. Haile, J. *et al.* Ancient DNA reveals late survival of mammoth and horse in interior Alaska. *PNAS* **106**, 22352-22357 (2009).
26. Giguet-Covex, C. *et al.* Changes in erosion patterns during the Holocene in a currently treeless subalpine catchment inferred from lake sediment geochemistry (Lake Anterne, 2063 m asl, NW French Alps): the role of climate and human activities. *The Holocene* **21**, 651-665 (2011).

27. Holzhauser, H., Magny, M. & Zumbühl, H.J. Glacier and lake-level variations in west-central Europe over the last 3500 years. *The Holocene* **15**, 789-801 (2005).
28. Crecchio, C. & Stotzky, G. Insecticidal activity and biodegradation of the toxin from *Bacillus thuringiensis* subsp. *kustuki* bound to humic acids from soil. *Soil Biology & Biochemistry* **30**, 463-70 (1998).
29. Pietramellara, G. *et al.* Extracellular DNA in soil and sediment: fate and ecological relevance. *Biology and Fertility of Soils* **45**, 219-235 (2009).
30. Coolen, M.J.L. & Gibbson, J.A.E. Ancient DNA in lake sediments offers a novel window into past aquatic ecosystems and the environmental changes that triggered past species successions. *Pages News* **17**, 104-106 (2009).
31. David, F. Expansion of green alder (*Alnus alnobetula* (Ehrh) K.Koch) in the northern French Alps: a palaeoecological point of view. *Compte Rendu de l'Académie des sciences, Biologies* **333**, 424-428 (2010a).
32. David, F. An example of the consequences of the human activities on the evolution of the subalpine landscapes. *Compte Rendu Palevol* **9**, 229-235 (2010b).
33. Carrier, N. & Mouthon, F. Paysans des Alpes - Les communautés montagnardes au Moyen Âge, 420 pp (*Presses universitaires de Rennes*, 2010)
34. Yoccoz, N.G. *et al.* DNA from soil mirrors plant taxonomic and growth form diversity. *Molecular Ecology* **21**, 3647-3655 (2012).
35. Lydolph, M.C. *et al.* Beringian paleoecology inferred from permafrost-preserved fungal DNA. *Applied and Environmental Microbiology* **71**, 1012-1017 (2005).
36. Giguët-Covex, C. *et al.* Frequency and intensity of high altitude floods over the last 3.5 ka in NW European Alps. *Quaternary Research* **77**, 12-22 (2012).
37. Trimbel, S.W. & Mendel, A.C. The cow as a geomorphic agent - a critical review. *Geomorphology* **13**, 233-253 (1995).
38. Stuiver, M. & Reimer, P.J. Extended ¹⁴C data base and revised CALIB 3.0 ¹⁴C age calibration program. *Radiocarbon* **35**, 215-230 (1993).
39. Reimer, P.J. *et al.* IntCal09 and Marine09 radiocarbon age calibration curves, 0-50,000 years cal BP. *Radiocarbon* **51**, 1111-1150 (2009).
40. Higuera, P.E., Brubaker, L.B., Anderson, P.M., Hu, F.S. & Brown, T.A. Vegetation mediated the impacts of postglacial climate change on fire regimes in the South central Brooks Range, Alaska. *Ecological Monographs* **79**, 201-219 (2009).
41. Arnaud, F. *et al.* Flood and earthquake disturbance of ²¹⁰Pb geochronology (Lake Anterne, North French Alps). *Terra Nova* **14**, 225-232 (2002).
42. Cooper, A. & Poinar, H. N. Ancient DNA: do it right or not at all. *Science* **289**, 1139-1139 (2000).
43. Riaz, T. *et al.* ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Research* **39**, 1-11 (2011).

44. Taberlet P. *et al.* Power and limitations of the chloroplast *trnL* (UAA) intron for plant DNA *barcoding*. *Nucleic Acids Research* **35**, 1-8 (2007).
45. Binladen, J. *et al.* The Use of Coded PCR Primers Enables High-Throughput Sequencing of Multiple Homolog Amplification Products by 454 Parallel Sequencing. *PLoS ONE* **2**, e197 (2007). doi:10.1371/journal.pone.0000197
46. Boessenkool, S. *et al.* Blocking human contaminant DNA during PCR allows amplification of rare mammal species from sedimentary ancient DNA. *Molecular Ecology* **21**, 1806-1815 (2011).
47. Zinger, L., Coissac E., Choler, P. & Geremia, R.A. Assessment of microbial communities by graph partitioning in a study of soil fungi in two alpine meadows. *Applied and Environmental Microbiology* **75**, 5863-5870 (2009).
48. Bellemain, E. *et al.* Fungal palaeodiversity revealed using high-throughput *metabarcoding* of ancient DNA from arctic permafrost », *Environmental microbiology*, **15**, 1176-89 (2013).
49. Ficetola, F. *et al.* An In silico approach for the evaluation of DNA barcodes. *BMC Genomics* **11**, 434 (2010).
50. Shehzad, W. *et al.* Carnivore diet analysis based on next-generation sequencing: application to the leopard cat (*Prionailurus bengalensis*) in Pakistan. *Molecular Ecology* **21**, 1951-65 (2012).

b. Données complémentaires

Supplementary material

Supplementary discussion

Note 1: Extract of Pliny the Elder « natural history »

Figure S1: Map of archaeological sites near the lakeshore

Figure S2: Age-depth model

Table S1: Recap chart of the taxonomic assignment of sequences used in the study

Table S2: Plant and mammalian read numbers in the Lake Anterne sediment samples

Table S3: Core and archaeological site datings

Supplementary discussion

The following several parameters could represent important factors that affect the amount of mammal DNA that is transferred to the lake: i) pastoral charge (stocking rate in animal unit/day/ha), ii) animal behavior and iii) pastoral practices (*e.g.* driving of herds in the areas of maximum production of forage resources by shepherds; concentration of herds for milking; access to rivers and lakes for drinking; and the presence of enclosures that concentrate herds close to the lake). Indeed, parameters that depend on species, human needs, environmental conditions and climate, induce differences in the amounts of animal DNA that are mobilized by water erosion and then accumulate in lake sediments. In particular, pastoral practices and animal behavior control the quantity of DNA that is transferred to the lake because they can trigger “diffuse or point” sources of DNA. Moreover, mammal DNA production is most likely different according to the considered domestic animal species due to different biomass production. For an equivalent stocking rate, it is likely that cattle (milking, concentration of animals near rivers and higher biomass production) and equine (high biomass production) are better recorded than sheep.

Currently, a sheep flock is present from time to time during summer with a scattered distribution in the catchment. Thus, the stocking rate is low. Furthermore, the scattered distribution of sheep induces a diffuse source of DNA from feces, skin flakes and chitinous materials, *e.g.* the main sources of mammalian DNA³¹. Together, this low stocking rate and diffuse source of DNA most likely trigger low sheep DNA transfer toward the lake. At the sediment surface, sheep DNA was not observed. Thus, this result raises the question of the limit of detection of DNA when its quantity is low. Stocking rate, animal behavior and pastoral practices, therefore, represent important factors that affect the recording sensitivity of the sedimentary DNA tool. Consequently, we can deduce neither a stocking rate nor a grazing pressure (depending on the number of animals, their distribution in the catchment, shepherd practice and animal route) from the amount of mammal DNA that is preserved in lake sediments. Our record of domestic mammal sedimentary DNA reflects periods of a relatively high stocking rate and/or particular pastoral practices in the catchment and imply more or less important grazing pressure on plant cover and soil erosion. As an example, the presence of the remains of one non-dated enclosure in the catchment (Supplementary Figure S1) shows that, in the past (most likely the Iron Age and/or Roman period because the dwelling contains

bones that were dated to these periods), domestic mammals were concentrated near the lake at least during the night. This point source of DNA that is easily and quickly transferable to the lake could explain the high quantity of domestic mammalian DNA that is found in lake sediments during this/these periods. However, this does not exclude the possibility of a high stocking rate also contributing to the high DNA content. The presence of sheep DNA originating during the Bronze Age and the end of the Roman period (1800-1500 cal. yr BP) in only 1 in 8 PCRs could be associated with sporadic contamination, low DNA concentration linked to a particular practice or a low stocking rate. However, we favor the second assumption (low DNA concentration) because sheep DNA occurs simultaneously with peaks in erosion, suggesting a response to grazing pressure. Moreover, for samples from the Bronze Age, we record a decrease in *Pinus* sp. DNA, which could have been triggered by deforestation for grazing activity. Exemplary evidence (DNA occurrences in 1 in 8 PCRs) of *Alnus* sp. expansion also suggests grazing activity during this period^{8,31}.

Note 1: Extract of Pliny the Elder « natural history »Livre XI

XCVII. *Laus caseo Romae, ubi omnium gentium bona comminus iudicantur, e prouinciis Nemausensi praecipua, Lesurae Gabalicoque pagis, sed breuis ac musteo tantum commendatio. duobus Alpes generibus pabula sua adprobant: Delmaticae Docleatem mittunt, Ceutronicae Vatusicum.*

« The types of cheese that are most esteemed in Rome, where the various good things of all nations are to be judged by comparison, are those which come from the provinces of Nemausus (Nîmes, in France. Hardouin speaks of goat milk cheeses made in the neighborhood known *as fromages de Baux*) and, more specifically, the villages of Lesura and Gabalis (most likely the modern Losere and Gevaudan); however, its excellence is only very short-lived, and it must be eaten while it is fresh. The pastures of the Alps are known for two sorts of cheese; the Dalmatic Alps send us the Docleatian cheese, and the Centronian Alps the Vatusican (He perhaps refers to the modern *Passy cheese*). »

Livre VIII

LXX . *non degeneres existimandi etiam minus laudato aspectu: plurimum lactis Alpinis, quibus minimum corporis, plurimum laboris capite, non ceruice, iunctis.*

« We must not look upon those types as having degenerated, the appearance of which is not so favorable. Those that are bred in the Alps, although very small of body, give a great quantity of milk and are capable of enduring much labor; they are yoked by the horns, and not by the neck. »

Figure S1: Map of archaeological sites near the lakeshore

The map shows the location of the two dwelling stones and the enclosure that were discovered in the southern part of the Lake Anterne catchment. The locations of the charcoals that were used for the radiocarbon dating and the bone remains are also presented.

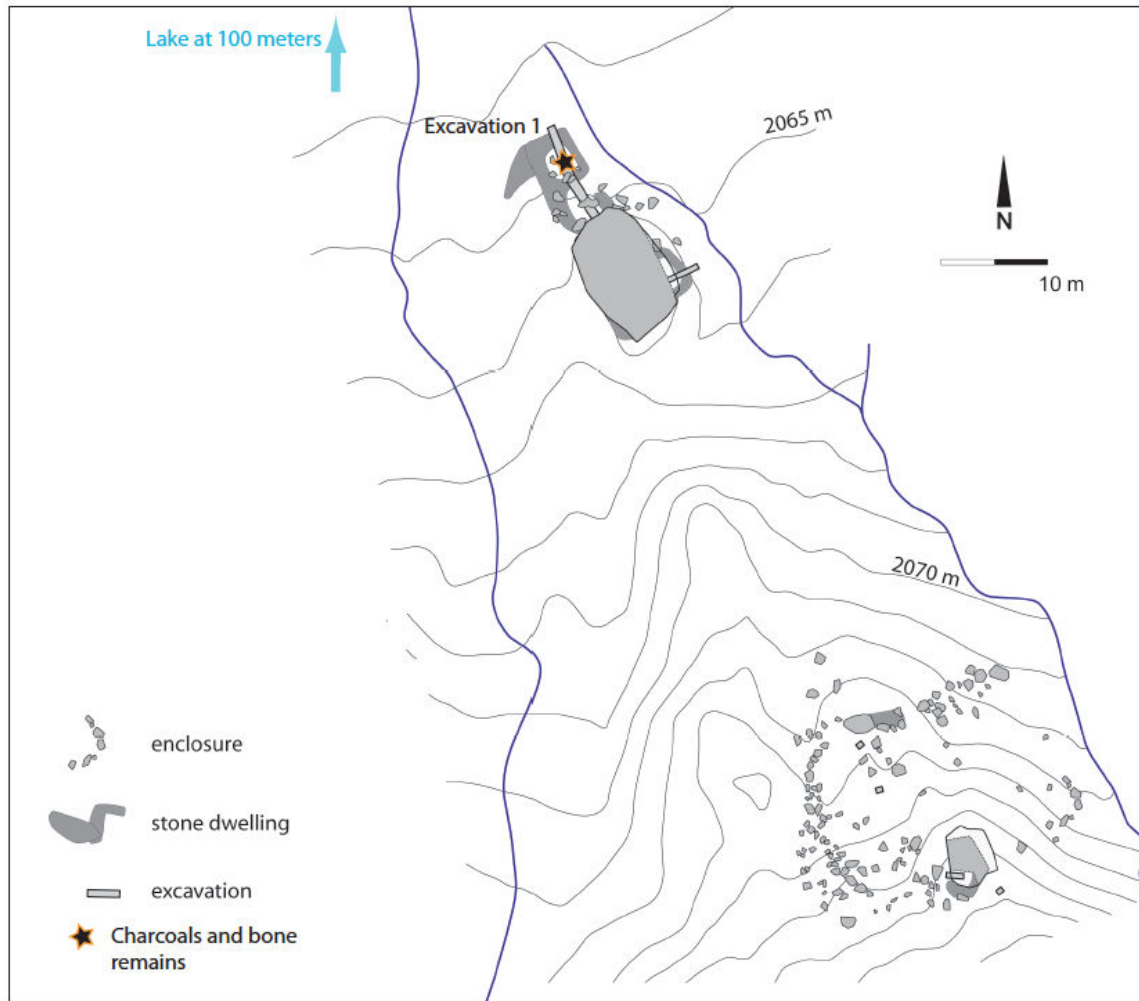


Figure S2: Age-depth model

Presentation of the age-depth model of the lake sediment core over the last 6500 years. This model was built using 20 radiocarbon dates that were taken from two different cores (ANT-01 and ANT-07) and fitted to a cubic spline curve. The different lithological units of the core are also presented on the left side of the figure.

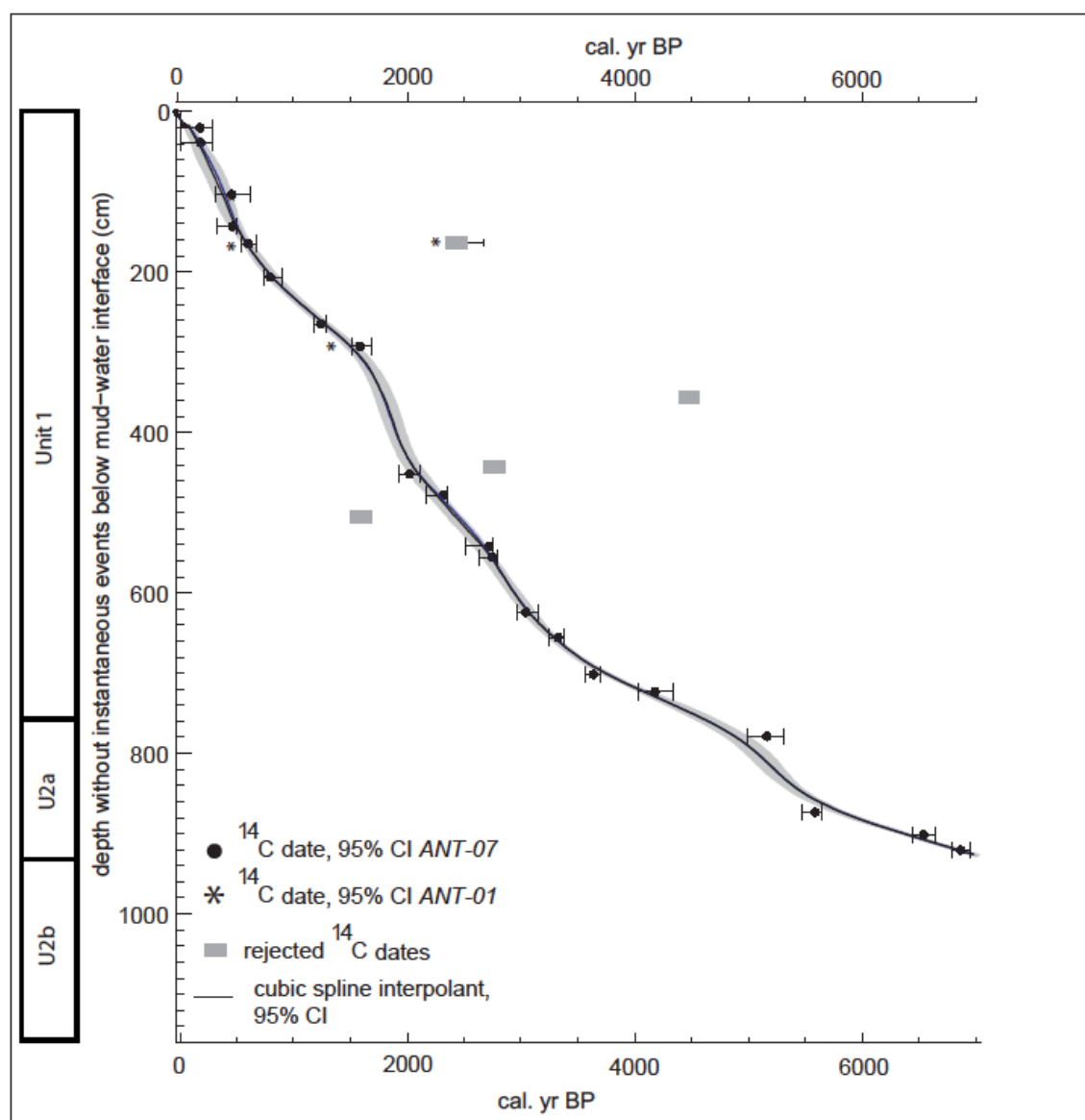


Table S1: Recap chart of the taxonomic assignment of sequences used in the study
List of sequences used for this study and their associated taxa in the database. The most probable taxa according to the site location are also presented.

MOTU number	Sequences	Most similar sequence(s) in public database using BLAST				Putative taxon identification taking into account the Anterne lake location	
		species name(s)	Query coverage (%)	Max identity (%)	Accession number(s)	Scientific name(s)	Common name(s)
1	ttaactaaccaaccaagagagaatag atttaaccattaaggaataacaacaaat ctccatgagtggttagttc	<i>Bos taurus</i> / <i>Bos javanicus</i> / <i>Bos indicus</i>	100%	100%	KC153977.1 / FJ997262.1 / JN817305.1	<i>Bos taurus</i>	Cattle
2	ttaactaagtaactcaaggaaaataaa ttcaaccaccaagggaataacaacact ccttatgagtaacagttc	<i>Ovis aries</i>	100%	100%	HE577850.1	<i>Ovis aries</i>	Sheep
3	ttaattaactgattcacaaaaacaac acacaaacctaacctcagggaacac aaaactttgattgaatcagcaattc	<i>Equus caballus</i> / <i>Equus przewalskii</i>	100%	100%	JN398377.2 / JN398403.1	<i>Equus caballus</i>	Horse
4	atcctgtttccgaaaacaataaaac aaattgaagggttcataaagcgagact aaaaag	<i>Alnus viridis</i> / <i>Alnus sinuata</i> / <i>Alnus sieboldiana</i> / <i>Alnus firma</i>	100%	100%	GQ244548.1 / AY147067.1 / AB168042.1 / AB063524.1	<i>Alnus viridis</i>	Mountain alder
5	atcacgtttccgaaaacaataaaac aaattgaagggttcataaagcgagact aaaaag	<i>Alnus viridis</i> / <i>Alnus sinuata</i> / <i>Alnus sieboldiana</i> / <i>Alnus firma</i>	91%	100%	GQ244548.1 / AY147067.1 / AB168042.1 / AB063524.1	<i>Alnus viridis</i>	Mountain alder
6	atctgtttccgaaaacaataaaac aaattgaagggttcataaagcgagact aaaaag	<i>Alnus viridis</i> / <i>Alnus sinuata</i> / <i>Alnus sieboldiana</i> / <i>Alnus firma</i>	93%	100%	GQ244548.1 / AY147067.1 / AB168042.1 / AB063524.1	<i>Alnus viridis</i>	Mountain alder
7	atccggttcataagacaatgtttctct cctaagataggaagg	<i>Pinus amamiana</i> / <i>Pinus aristata</i> / <i>Pinus arizonica</i> / <i>Pinus brutia</i> / <i>Pinus bungeana</i> / <i>Pinus canariensis</i> / <i>Pinus caribaea</i> / <i>Pinus cembra</i> / <i>Pinus cembroides</i> / <i>Pinus chiapensis</i> / <i>Pinus coulteri</i> / <i>Pinus cubensis</i> / <i>Pinus culminicola</i> / <i>Pinus dalatensis</i> / <i>Pinus densata</i> / <i>Pinus densiflora</i> / <i>Pinus devoniana</i> / <i>Pinus discolor</i> / <i>Pinus douglasiana</i> / <i>Pinus echinata</i> / <i>Pinus edulis</i> / <i>Pinus elliotii</i> / <i>Pinus engelmannii</i> / <i>Pinus fenzeliana</i> / <i>Pinus flexilis</i> / <i>Pinus fragilissima</i> / <i>Pinus gerardiana</i> / <i>Pinus glabra</i> / <i>Pinus greggii</i> / <i>Pinus halepensis</i> / <i>Pinus hartwegii</i> / <i>Pinus heldreichii</i> / <i>Pinus hwangshanensis</i> / <i>Pinus jeffreyi</i> / <i>Pinus johannis</i> / <i>Pinus kesiya</i> / <i>Pinus koraiensis</i> / <i>Pinus krempfii</i> / <i>Pinus kwangtungensis</i> / <i>Pinus lambertiana</i> / <i>Pinus latteri</i> / <i>Pinus lawsonii</i> / <i>Pinus leiophylla</i> / <i>Pinus lumholtzii</i> / <i>Pinus massoniana</i> / <i>Pinus maximartinezii</i> / <i>Pinus monophylla</i> / <i>Pinus montezumae</i> / <i>Pinus morrisonicola</i> / <i>Pinus mugo</i> / <i>Pinus muricata</i> / <i>Pinus nelsonii</i> / <i>Pinus nigra</i> / <i>Pinus occidentalis</i> / <i>Pinus palustris</i> / <i>Pinus patula</i> / <i>Pinus peuce</i> / <i>Pinus pinaster</i> / <i>Pinus pinceana</i> / <i>Pinus pinea</i> / <i>Pinus ponderosa</i> / <i>Pinus pringlei</i> / <i>Pinus pseudostrobus</i> / <i>Pinus pumila</i> / <i>Pinus pungens</i> / <i>Pinus quadrifolia</i> / <i>Pinus radiata</i> / <i>Pinus remota</i> / <i>Pinus rigida</i> / <i>Pinus roxburghii</i> / <i>Pinus sabiniana</i> / <i>Pinus serotina</i> / <i>Pinus strobiformis</i> / <i>Pinus sylvestris</i> / <i>Pinus taeda</i> / <i>Pinus taiwanensis</i> / <i>Pinus thunbergii</i> / <i>Pinus tropicalis</i> / <i>Pinus wallichiana</i> / <i>Pinus yunnanensis</i> / <i>Pinus yunnanensis</i>	100%	100%	JN854226.1 / FJ899567.2 / JN854225.1 / JN854224.1 / JN854223.1 / FJ899572.2 / JN854222.1 / AF327588.1 / JN854220.1 / JN854219.1 / JN854215.1 / JN854214.1 / JN854213.1 / JN854211.1 / JN854209.1 / JX504689.1 / JN854208.1 / JN854207.1 / JN854205.1 / JN854204.1 / JN854203.1 / JN854202.1 / JN854201.1 / JN854212.1 / FJ899576.2 / JN854200.1 / EU998741.4 / JN854199.1 / JN854198.1 / JN854197.1 / JN854206.1 / JN854195.1 / JN854194.1 / JN854193.1 / JN854192.1 / JN854191.1 / JX504690.1 / EU998742.4 / JN854153.1 / EU998743.4 / JN854190.1 / JN854188.1 / JN854187.1 / JN854186.1 / JN854185.1 / JN854184.1 / EU998745.4 / JN854183.1 / JN854182.1 / JN854181.1 / JN854180.1 / EU998746.4 / JN854179.1 / JN854177.1 / JN854176.1 / JN854175.1 / JQ739417.1 / FJ899583.2 / JN854174.1 / JN854173.1 / FJ899555.2 / JN854189.1 / JN854169.1 / JN854168.1 / JN854167.1 / JN854166.1 / JN854165.1 / JN854164.1 / JN854163.1 / JN854162.1 / JN854161.1 / JN854160.1 / JN854159.1 / JX504686.1 / FJ899561.2 / JN854157.1 / JX504692.1 / JN854156.1 / JN854154.1 / JN854152.1 / JN854151.1	<i>Pinus cembra</i> / <i>Pinus mugo</i> / <i>Pinus sylvestris</i>	Swiss Pine / Mountain pine / Scots pine

Table S2: Plant and mammalian read numbers in the Lake Anterne sediment samples

The number of sequences for the different taxa of interest is presented for each replicate. In the replicate names, A and B correspond to sample replicates, 1 and 2 correspond to extraction replicates, and a and b correspond to PCR replicates. Crosses signify that no DNA was amplified in the sample.

sample information							number of DNA sequences						
core section name - depth (cm)	master core depth (cm)	Age cal. BP	Age up cal. BP	Age low cal. BP	Run	Replicate name	Bos	Ovis	Equus	Total mammal	Alnus	Pinus	Total viridiplantae
ANT 07 n°8 - 3.5	3.5	-56	-56	-56	2	ANT01A1a	1	0	0	5	1	227	8933
						ANT01A1b	0	0	0	2	0	0	7024
						ANT01A2a	0	0	0	1	0	0	1680
						ANT01A2b	1	0	0	3	0	0	6123
						ANT01B1a	2	2	0	8	0	0	6649
						ANT01B1b	0	0	0	0	0	0	4761
						ANT01B2a	2	0	0	1393	0	0	3714
						ANT01B2b	7	0	0	10	0	0	6939
ANT 07 B1 - 9	9	-25	-23	-27	1	RM12A1a	0	0	0	1	0	0	31758
						RM12A1b	0	0	0	1	0	0	46654
						RM12A2a	0	0	0	1	0	0	35255
						RM12A2b	0	0	0	992	0	0	28025
						RM12B1a	0	0	0	0	1	0	34104
						RM12B1b	x	x	x	x	0	0	29213
						RM12B2a	0	0	0	2	0	1	40769
						RM12B2b	0	0	0	0	1	0	35186
ANT 07 n°8 - 82.5	87.3	170.99	206	91	2	ANT02A1a	2	0	0	2	1	0	9406
						ANT02A1b	1	0	0	1	1	0	7508
						ANT02A2a	48	0	0	48	0	0	1420
						ANT02A2b	1	0	0	1	0	0	7933
						ANT02B1a	5	0	0	7	0	0	7336
						ANT02B1b	1	0	0	1	1	0	5857
						ANT02B2a	1	0	0	1	0	0	3169
						ANT02B2b	3014	0	0	4010	0	0	8740
ANT 07 C1 - 49.5	180.3	343.43	406	238	1	RM02A1a	0	0	0	0	0	0	30447
						RM02A1b	44	0	0	746	1	0	38228
						RM02A2a	0	0	0	2	0	0	37233
						RM02A2b	0	0	0	1259	1	0	26719
						RM02B1a	0	0	0	2	0	0	41333
						RM02B1b	0	0	0	1	0	0	21947
						RM02B2a	0	0	0	0	0	0	34998
						RM02B2b	2	0	0	1493	0	1	34308
ANT 07 C1 - 105	236.3	445.55	498	353	2	ANT03A1a	10	0	0	11	1	0	12659
						ANT03A1b	2	0	0	3	0	0	13821
						ANT03A2a	1	0	0	1	0	0	1847
						ANT03A2b	12	0	0	12	1	0	14385
						ANT03B1a	2973	0	0	2975	3	1	12479
						ANT03B1b	0	0	0	289	3	0	10111
						ANT03B2a	2	0	0	2	0	0	4592
						ANT03B2b	11	0	0	11	0	0	11829
ANT 07	276.9	521.07	566	444	1	RM03A1a	4	0	0	4	0	0	28237

C1 - 145.6						RM03A1b	4	0	0	4	0	1	35280
						RM03A2a	2	0	0	2	1	0	28274
						RM03A2b	233	0	0	233	0	0	26464
						RM03B1a	2531	0	0	2531	0	1	32914
						RM03B1b	5047	0	0	5047	0	1	29528
						RM03B2a	3586	0	0	3586	0	0	32639
						RM03B2b	2071	0	0	2071	0	0	27969
ANT 07 B2 - 114.9	337	611.24	659	558	2	ANT04A1a	3323	0	0	5226	88	0	20016
						ANT04A1b	0	0	0	0	2	0	17179
						ANT04A2a	0	0	0	0	0	1	2650
						ANT04A2b	142	0	0	142	53	0	15912
						ANT04B1a	2	0	0	2	0	0	14086
						ANT04B1b	x	x	x	x	0	0	10124
						ANT04B2a	1	0	0	2	22	0	5413
ANT 07 B2 - 156.2	378.3	714.71	773	672	2	ANT04B2b	4335	0	0	4337	2	1	12442
						ANT05A1a	1	0	0	1	0	1	10833
						ANT05A1b	1	0	0	1	1	327	10506
						ANT05A2a	0	0	0	0	0	0	2268
						ANT05A2b	0	0	0	1	22	0	10545
						ANT05B1a	3	0	0	3	0	0	6684
						ANT05B1b	1	0	0	1	0	1	5157
ANT 07 C2 - 96.9	423.6	826.32	885	782	2	ANT05B2a	992	0	0	992	11	0	3993
						ANT05B2b	1544	0	0	1544	0	0	9139
						ANT06A1a	1	1	0	2	1	0	15356
						ANT06A1b	0	0	0	0	1	5	18794
						ANT06A2a	0	0	0	0	19	0	2223
						ANT06A2b	0	0	0	0	1	0	9160
						ANT06B1a	0	0	0	0	0	0	9180
ANT 07 C2 - 114.9	441.1	923.12	982	878	2	ANT06B1b	x	x	x	x	26	0	9107
						ANT06B2a	1	0	0	2	0	0	3897
						ANT06B2b	2	0	0	2	2	1	12472
						ANT07A1a	12	2263	0	2276	8	0	31513
						ANT07A1b	0	0	0	1	2	0	26464
						ANT07A2a	1	0	0	1	0	0	3969
						ANT07A2b	20	5	0	25	4	1	20582
ANT 07 C2 - 172.9	490.1	1087.4	1138.5	1040	1	ANT07B1a	11085	6	0	11094	3	0	23853
						ANT07B1b	0	0	0	0	137	0	17138
						ANT07B2a	1	0	0	1	85	1	8390
						ANT07B2b	7	0	0	7	3	2	23067
						RM06A1a	0	0	0	1	419	0	30026
						RM06A1b	1	0	0	2	9	1	33820
						RM06A2a	0	0	0	1	349	0	31519
ANT 07 B3 - 86.9	519.9	1236.9	1282	1193	2	RM06A2b	24	0	0	25	334	0	25887
						RM06B1a	1	0	0	1106	803	2	31198
						RM06B1b	0	0	0	0	404	0	30166
						RM06B2a	0	0	0	0	20	2	35616
						RM06B2b	2	1	0	3	163	0	28734
						ANT08A1a	1	0	0	2	0	4	15937
						ANT08A1b	0	0	0	0	0	1	13305
ANT 07 B3 - 86.9	519.9	1236.9	1282	1193	2	ANT08A2a	0	0	0	0	0	0	2984
						ANT08A2b	0	0	0	0	0	1	12421
						ANT08B1a	2	1	0	3	0	0	11571
						ANT08B1b	0	0	0	0	0	0	7849
						ANT08B2a	0	0	0	0	0	0	2652
						ANT08B2b	2	0	0	3	0	0	9741
ANT 07	561.5	1386.6	1448	1344	2	ANT09A1a	0	0	0	0	0	0	7235

B3 - 128.5						ANT09A1b	0	0	0	0	0	0	0	7218
						ANT09A2a	0	0	0	0	0	66	0	968
						ANT09A2b	0	0	0	0	0	2	1	4921
						ANT09B1a	1	1	0	2	1	1	0	8222
						ANT09B1b	0	0	0	0	0	0	0	6882
						ANT09B2a	x	x	x	x	0	0	0	3012
						ANT09B2b	x	x	x	x	1	0	0	6981
ANT 07 C3 - 81.3	615.5	1543.2	1631	1500	2	ANT10A1a	0	3	0	4	2	0	0	9734
						ANT10A1b	0	2	0	2	1	2	0	10934
						ANT10A2a	0	0	0	0	0	0	0	2874
						ANT10A2b	0	3	0	3	85	0	0	11898
						ANT10B1a	0	2960	0	2963	0	1	0	9689
						ANT10B1b	0	0	0	0	0	0	0	5757
						ANT10B2a	0	0	0	0	22	0	0	4599
ANT 07 C3 - 117.3	651.8	1652.6	1758	1610	2	ANT10B2b	3	1	0	4	48	2	0	10339
						ANT11A1a	0	0	0	0	0	0	0	15242
						ANT11A1b	0	0	0	0	0	0	0	19344
						ANT11A2a	0	0	0	0	0	0	0	2494
						ANT11A2b	0	0	0	0	0	1	0	9544
						ANT11B1a	0	0	0	2	0	0	0	11622
						ANT11B1b	0	0	0	0	0	0	0	6598
ANT 07 C3 - 173.8	708.3	1788.4	1904	1737	2	ANT11B2a	0	0	0	0	0	0	0	5149
						ANT11B2b	2	0	0	3	0	293	0	9856
						ANT12A1a	0	1	0	1	0	3	0	2180
						ANT12A1b	x	x	x	x	0	2	0	992
						ANT12A2a	0	0	0	0	18	0	0	3468
						ANT12A2b	x	x	x	x	8	1	0	2112
						ANT12B1a	x	x	x	x	1	0	0	4942
ANT 07 B4 - 155.5	790.1	1892.5	1999	1827	2	ANT12B1b	0	113	0	113	0	0	0	1199
						ANT12B2a	0	1	0	1	1	0	0	8256
						ANT12B2b	0	1	0	1	0	0	0	4755
						ANT13A1a	2	1309	0	1314	0	1	0	6212
						ANT13A1b	0	0	0	0	0	2	0	3885
						ANT13A2a	0	1	0	1	2	0	0	6143
						ANT13A2b	0	0	0	0	0	0	0	3399
ANT 07 B5 - 14.5	858.1	2000.8	2078	1916	1	ANT13B1a	252	0	0	252	1	0	0	5318
						ANT13B1b	1	0	0	1	0	0	0	1106
						ANT13B2a	5	1285	0	2745	0	0	0	9566
						ANT13B2b	1585	1	0	1589	2	0	0	5097
						RM09A1a	0	156	0	272	627	0	0	6873
						RM09A1b	2	0	0	3	427	0	0	8922
						RM09A2a	1	1743	0	3362	274	205	0	7560
ANT 07 C4 - 180.5	915.1	2126.1	2182	2040	2	RM09A2b	1250	0	0	1251	182	59	0	6456
						RM09B1a	3	965	0	968	200	186	0	8112
						RM09B1b	2528	0	0	2530	233	0	0	6954
						RM09B2a	1	0	0	1	232	0	0	8343
						RM09B2b	1	3384	0	3386	133	0	0	7909
						ANT14A1a	0	0	0	0	2	0	0	6444
						ANT14A1b	0	0	680	680	3	2	0	4327
ANT 07 C4 - 180.5	915.1	2126.1	2182	2040	2	ANT14A2a	0	0	331	331	2	0	0	5152
						ANT14A2b	1	0	0	2	76	0	0	2925
						ANT14B1a	0	0	0	57	85	0	0	6931
						ANT14B1b	2	274	0	389	39	0	0	2053
						ANT14B2a	2038	1119	2	3161	112	0	0	11852
						ANT14B2b	468	2	0	470	105	0	0	5531
ANT 07	949.5	2231.5	2270	2148	2	ANT15A1a	0	1	0	1	122	2	0	6467

B5 - 102.4						ANT15A1b	1	0	0	1	103	2	4377
						ANT15A2a	10	0	0	10	259	0	6956
						ANT15A2b	0	0	0	1	33	0	3138
						ANT15B1a	2640	0	0	2641	363	0	5314
						ANT15B1b	1	0	0	1	28	0	1545
						ANT15B2a	1	1	0	2	941	0	14976
						ANT15B2b	0	0	0	0	401	0	7386
ANT 07 B5 - 132.5	979.9	2337.8	2367	2245	1	RM08A1a	8	1246	0	1255	1992	0	30495
						RM08A1b	5	560	0	565	459	0	43251
						RM08A2a	3	0	0	4	904	0	28408
						RM08A2b	4	0	0	4	923	0	27358
						RM08B1a	3758	4010	0	7770	715	0	33021
						RM08B1b	10530	2	0	12020	284	0	19406
						RM08B2a	3	0	0	686	854	1	31959
						RM08B2b	1943	3555	0	5499	815	0	28376
ANT 07 B5 - 160	1007.1	2444.9	2475	2341	2	ANT16A1a	1	0	0	389	1	2	5594
						ANT16A1b	0	0	0	0	0	2	3537
						ANT16A2a	1	0	0	1	0	0	5657
						ANT16A2b	0	0	0	0	0	0	3692
						ANT16B1a	1	0	0	1	0	0	9684
						ANT16B1b	x	x	x	x	0	0	2505
						ANT16B2a	3	1	0	5	0	1	16168
						ANT16B2b	3	0	0	3	1	159	7157
ANT 07 C5 - 93.8	1039.3	2540.1	2565	2435	2	ANT17A1a	0	0	0	0	1	1	7323
						ANT17A1b	0	0	0	0	0	1	4640
						ANT17A2a	0	0	0	0	0	1	6354
						ANT17A2b	0	0	0	590	0	0	3466
						ANT17B1a	0	0	0	0	1	0	8163
						ANT17B1b	x	x	x	x	0	0	1914
						ANT17B2a	2	1	0	5	0	0	13086
						ANT17B2b	0	0	0	0	1	1	6362
ANT 07 C5 - 124	1069.3	2626.2	2647	2530	2	ANT18A1a	0	0	0	0	0	3	11591
						ANT18A1b	0	0	0	0	0	1	4767
						ANT18A2a	0	0	0	0	0	296	5461
						ANT18A2b	x	x	x	x	1	0	4088
						ANT18B1a	x	x	x	x	2	0	7302
						ANT18B1b	x	x	x	x	0	0	2080
						ANT18B2a	2	0	0	2	1	2	15920
						ANT18B2b	x	x	x	x	0	2	6488
ANT 07 B6 - 72.5	1135.1	2778.6	2804	2702	2	ANT19A1a	0	0	0	0	125	5	14845
						ANT19A1b	0	0	0	0	44	6	9023
						ANT19A2a	0	0	0	0	7	0	12398
						ANT19A2b	0	0	0	0	6	102	7725
						ANT19B1a	0	0	0	1	1186	0	12608
						ANT19B1b	0	0	0	0	18	0	2706
						ANT19B2a	4	2	0	11	1218	638	27445
						ANT19B2b	1	0	0	2	430	1	10467
ANT 07 B6 - 102.5	1165.6	2879.5	2916	2823	2	ANT20A1a	0	0	0	1	1	9	14373
						ANT20A1b	0	0	0	0	0	5	6633
						ANT20A2a	0	0	0	0	0	0	11956
						ANT20A2b	0	0	0	0	0	0	5648
						ANT20B1a	1	0	0	1	3	0	13421
						ANT20B1b	0	1	0	1	0	0	3734
						ANT20B2a	1	0	0	1	3	0	34048
						ANT20B2b	0	1	0	3	0	0	15837
ANT 07	1203.1	2996.5	3054	2945	2	ANT21A1a	0	0	0	0	1	1138	6508

B6 - 140.5						ANT21A1b	0	0	0	0	1	820	3258
						ANT21A2a	0	0	0	214	0	2	3199
						ANT21A2b	x	x	x	x	0	1	2197
						ANT21B1a	0	0	0	0	46	0	4970
						ANT21B1b	0	1	0	1	0	0	1390
						ANT21B2a	1	0	0	1	2	3	10434
						ANT21B2b	1	0	0	1	0	96	5498
ANT 07 C6 - 68.6	1225	3078.7	3142	3023	2	ANT22A1a	0	3	0	3	0	1	11808
						ANT22A1b	0	1	0	1	0	1	6188
						ANT22A2a	0	0	1	1	0	0	7188
						ANT22A2b	0	0	0	1	0	0	3795
						ANT22B1a	0	0	0	0	1	0	6569
						ANT22B1b	0	1	0	1	0	0	2139
						ANT22B2a	1	4	0	5	0	0	19582
						ANT22B2b	0	661	0	661	1	0	8808
ANT 07 C6 - 130	1285.1	3229.7	3283	3166	2	ANT23A1a	0	0	0	0	0	3	8622
						ANT23A1b	0	0	0	0	0	1	6448
						ANT23A2a	0	0	0	0	0	0	7526
						ANT23A2b	x	x	x	x	0	1	5222
						ANT23B1a	0	0	0	0	1	1	10185
						ANT23B1b	0	2	0	2	0	0	2650
						ANT23B2a	0	3	0	3	1	0	21193
						ANT23B2b	0	0	0	1576	1	0	9723
ANT 07 C6 - 147.4	1303.3	3327.4	3383	3264	2	ANT24A1a	0	228	0	228	0	1	3334
						ANT24A1b	0	2	0	2	0	0	2674
						ANT24A2a	0	0	0	0	0	0	2324
						ANT24A2b	0	0	0	0	2	0	5269
						ANT24B1a	0	0	0	0	1	0	6665
						ANT24B1b	0	0	0	0	143	0	4286
						ANT24B2a	0	0	0	0	0	0	2441
						ANT24B2b	x	x	x	x	1	99	5021
ANT 07 C6 - 173	1327.8	3402.5	3465	3339	1	RM04A1b	0	1	0	1	0	1	24179
						RM04A2a	0	0	0	1	2	8	16336
						RM04B1a	0	0	0	0	1	1	21371
						RM04B1b	1	0	0	1	1	1	19632
						RM04A1a	0	0	0	3	1061	1	23204
						RM04A1b	0	0	0	0	4	0	21683
						RM04A2a	0	0	0	1	1	1132	28176
						RM04A2b	0	0	0	2	1	1	17124
ANT 07 B7 - 65.5	1346.4	3543.5	3604	3481	2	ANT25A1a	2	1	0	3	0	0	7203
						ANT25A1b	0	0	0	0	0	0	5934
						ANT25A2a	0	0	0	0	0	0	3422
						ANT25A2b	0	0	0	1	0	0	9215
						ANT25B1a	0	0	0	1	0	0	7164
						ANT25B1b	0	0	0	1	2	0	5896
						ANT25B2a	0	0	0	0	0	0	4298
						ANT25B2b	0	0	0	2	0	0	6087
ANT 07 B7 - 83.5	1363.8	3717.3	3783	3660	2	ANT26A1a	0	2	0	2	1	0	4299
						ANT26A1b	0	0	0	0	0	1	4222
						ANT26A2a	0	0	0	0	0	0	2979
						ANT26A2b	2	0	0	4	1	1	7459
						ANT26B1a	0	2	1	4	0	0	4939
						ANT26B1b	0	1	0	2	0	1	3130
						ANT26B2a	0	0	0	0	1	4	4462
						ANT26B2b	0	1	0	1	1	0	4626

ANT 07 B7 - 109.3	1390.2	3895.3	3970	3840	2	ANT27A1a	0	0	0	0	1	0	5328
						ANT27A1b	0	0	0	1	5	0	5588
						ANT27A2a	0	0	0	0	2	1	3350
						ANT27A2b	2	0	0	2	3	0	10887
						ANT27B1a	0	0	0	0	1	0	12101
						ANT27B1b	0	0	0	0	3	0	8004
						ANT27B2a	0	0	0	0	29	28	3870
						ANT27B2b	0	0	0	0	2	45	5789
ANT 07 B7 - 123.5	1402.9	4084.1	4168	4028	2	ANT28A1a	1	0	0	1	0	0	6318
						ANT28A1b	0	0	0	1	0	0	5909
						ANT28A2a	0	0	0	0	0	127	3899
						ANT28A2b	1	0	0	1	0	0	13660
						ANT28B1a	0	0	0	0	0	0	10164
						ANT28B1b	0	0	0	0	1	135	7527
						ANT28B2a	0	0	0	1	0	42	3749
						ANT28B2b	0	0	0	0	0	1	6260
ANT 07 B7 - 141	1421.9	4343.7	4439	4283	2	ANT29A1a	0	0	0	1	0	0	4911
						ANT29A1b	0	0	0	0	0	1	4407
						ANT29A2a	0	0	0	0	0	0	3583
						ANT29A2b	0	0	0	1	0	1	6113
						ANT29B1a	0	0	0	0	0	0	5474
						ANT29B1b	1	0	0	1	1	1	3028
						ANT29B2a	1	0	0	1	0	252	3549
						ANT29B2b	0	0	0	0	0	1	4602
ANT 07 B7 - 163.5	1442.1	4588.8	4685	4512	2	ANT30A1a	0	1	0	1	0	0	6504
						ANT30A1b	0	0	0	0	1	0	5898
						ANT30A2a	0	0	0	0	0	1	4449
						ANT30A2b	0	0	0	0	0	0	14277
						ANT30B1a	0	0	0	1	0	0	13328
						ANT30B1b	0	0	0	2	1	2	9414
						ANT30B2a	x	x	x	x	0	2	5943
						ANT30B2b	0	0	0	0	0	1	9767
ANT 07 C7 - 116.3	1473.3	4856.3	4966	4765	2	ANT31A1a	2	0	0	2	4	0	14427
						ANT31A1b	2	0	0	8	1	1	12423
						ANT31A2a	0	0	0	1	5	1	8343
						ANT31A2b	1	0	0	3	1	1	26397
						ANT31B1a	7	0	0	5603	2	0	19769
						ANT31B1b	4	0	0	8466	1	0	13951
						ANT31B2a	493	0	0	494	1	264	8231
						ANT31B2b	0	0	0	4	0	0	11740
ANT 07 C7 - 148.3	1505.4	5020.7	5131	4908	1	RM05A1a	x	x	x	x	0	1	35189
						RM05A1b	x	x	x	x	0	0	39710
						RM05A2a	x	x	x	x	0	1	39318
						RM05A2b	x	x	x	x	0	1	30709
						RM05B1a	0	0	0	0	0	2080	41133
						RM05B1b	1	0	0	1	0	2487	30921
						RM05B2a	0	0	0	1	0	1967	46193
						RM05B2b	0	0	0	11535	0	1	36425
ANT 07 B8b - 5	1532.1	5157.4	5252	5034	2	ANT32A1a	0	1	0	5	0	0	11817
						ANT32A1b	0	0	0	2	0	4	11122
						ANT32A2a	0	0	0	0	0	0	9101
						ANT32A2b	0	0	0	8844	0	333	28887
						ANT32B1a	0	0	0	2	0	0	15546
						ANT32B1b	0	0	0	0	0	1	11711
						ANT32B2a	0	0	0	2	0	0	8617
						ANT32B2b	0	0	0	0	0	249	12625

ANT 07 B8b - 33	1559.6	5299.5	5378	5194	2	ANT33A1a	0	0	0	0	0	2	4655
						ANT33A1b	0	0	0	0	1	94	3699
						ANT33A2a	0	0	0	0	0	83	4411
						ANT33A2b	0	0	0	0	0	192	10984
						ANT33B1a	0	0	0	0	0	3	8041
						ANT33B1b	0	0	0	0	0	184	4528
						ANT33B2a	0	0	0	0	1	1	3676
						ANT33B2b	0	0	0	0	1	1	4591
ANT 07 B8b - 65	1592.1	5552.6	5612	5442	2	ANT34A1a	0	0	0	1	0	116	8869
						ANT34A1b	0	1	0	1	0	0	243
						ANT34A2a	0	0	0	0	0	47	4921
						ANT34A2b	0	0	0	0	0	239	17954
						ANT34B1a	0	0	0	0	1	0	13366
						ANT34B1b	0	0	0	0	1	235	9109
						ANT34B2a	0	0	0	195	1	63	6550
						ANT34B2b	0	0	0	0	0	8	9398
ANT 07 C8 - 59.9	1622.2	5970.3	6014	5889	1	RM07A1a	0	0	0	1	0	514	41013
						RM07A1b	0	0	0	8	0	821	49063
						RM07A2a	0	0	0	7244	0	1199	44068
						RM07A2b	0	0	0	4	0	0	32297
						RM07B1a	0	0	0	6	1	1027	41345
						RM07B1b	0	0	0	1	0	427	32169
						RM07B2a	0	0	0	3637	0	965	46168
						RM07B2b	0	0	0	3	0	266	45552
ANT 07 B8b - 116.2	1643.3	6384.9	6438	6320	2	ANT35A1a	0	1	0	2	0	0	7785
						ANT35A1b	0	0	0	0	0	0	336
						ANT35A2a	0	0	0	0	0	0	7317
						ANT35A2b	0	0	0	0	0	0	19574
						ANT35B1a	0	0	0	1	0	1	14425
						ANT35B1b	0	0	0	3	0	0	8183
						ANT35B2a	0	0	0	0	0	40	6535
						ANT35B2b	0	0	0	0	0	1	464
total read numbers							68726	25903	1015	159751	19601	20471	44237 18
Control samples					2	TEMOIN1a	0	0	0	0	1	0	4632
						TEMOIN1b	x	x	x	x	0	0	4809
						TEMOIN2a	0	0	0	0	1	0	841
						TEMOIN2b	0	0	0	0	0	1	2408
						TEMOIN3a	0	1	0	1	0	0	3766
						TEMOIN3b	0	0	0	0	1	0	2374
						TEMOIN4a	0	0	0	0	0	0	1469
						TEMOIN4b	1	1	0	2	0	0	4276

Table S3: Core and archaeological site datings

List of ^{14}C dates that were obtained for the macroremains from the lake sediment core and charcoals from the stone dwelling near the lakeshore. Samples in bold correspond to rejected dates. Four macroremains, marked by stars, originated from a previous core that was obtained in 2001 (ANT 01).

sample name	laboratory	cumulated depth (cm)	depth (cm) (without events)	sample type	lithology	radiocarbon age	age cal. BP 2 σ range
ANT 07 B1	POZ 25944	46	20,2	tree bark	flood deposit	190+/-30	150+/-150
ANT 07 B1b	POZ 30276	80,8	36,2	wood and leafs	flood deposit	170+/-25	145+/-145
ANT 01 01A*	Gif-a 101287	208,5	104,8		flood deposit	430+/-80	470+/-160
ANT 07 B2	SacA 12209	269,6	142,4	leafs	flood deposit	395+/-30	420+/-95
ANT 01 02 A2*	Gif-a 101288	327,5	163,1		flood deposit	2410+/-35	2520+/-180
ANT 01 02 A2*	POZ 794	329	164,6		flood deposit	640+/-60	610+/-70
ANT 07 C2	SacA 12210	415,3	206,1	leafs	flood deposit	885+/-30	820+/-90
ANT 07 B3b	POZ 30277	536	264,5	twig and leaf	flood deposit	1300+/-30	1235+/-60
ANT 01 02 A2*	POZ 720	608,7	292,3		flood deposit	1680+/-35	1610+/-90
ANT 07 C3	SacA 12211	714,7	356,6	tree bark	flood deposit	4005+/-30	5170+/-125
ANT 07 B5b	POZ 30278	883,3	442,4	plant macrofossils	flood deposit	2650+/-35	2790+/-55
ANT 07 C4b	POZ 25946	905,8	451,4	conifer shell	flood deposit	2055+/-35	2025+/-95
ANT 07 B5b	POZ 30279	960,8	477,8	wood	flood deposit	2295+/-35	2260+/-100
ANT 07 B5b	SacA 10729	1009,1	504,3	wood	flood deposit	1690+/-30	1610+/-80
ANT 07 C5b	POZ 30280	1088,5	541,9	charcoals	flood deposit	2565+/-30	2630+/-125
ANT 07 B6b	POZ 25945	1123,8	555,6	wood-Alnus/corylus	flood deposit	2620+/-35	2730+/-110
ANT 07 C6b	SacA 10726	1222,1	624	wood	flood deposit	2910+/-20	3055+/-90
ANT 07 C6b	POZ 30282	1294,3	655,3	leaf and tree bark	flood deposit	3100+/-30	3320+/-70
ANT 07 B7b	SacA 12212	1369,2	701	wood and leafs	flood deposit	3390+/-30	3635+/-70
ANT 07 C7b	POZ 30283	1403,75	723	charcoals	laminations	3795+/-35	4180+/-170
ANT 07 C7b	POZ 30284	1480,75	778,2	charcoals	flood deposit	4530+/-50	5240+/-195
ANT 07 B8b	POZ 30286	1612,45	872,8	small root	flood deposit	4840+/-40	5565+/-90
ANT 07 B8b	SacA 16863	1644,65	901,1	charcoals (conifer shell)	laminations	5745+/-40	6545+/-105
RSLA 1 niv. 2	POZ 33446	Archaeological site		charcoal (Pinus sp.)		1835+/-30	1785+/-80
RSLA 1 niv.5b	SacA 20689	Archaeological site		charcoals (conifer shell)		2230+/-30	2245+/-90

**3. Impacts à long terme des activités humaines sur les communautés végétales :
une approche écologique basée sur l'ADN sédimentaire lacustre**

a. Article D

Article D

**Reconstructing long-term human impacts on plant communities :
an ecological approach based on lake sediment DNA**

soumis à *Molecular Ecology*

Johan Pansu^{1,2*}, Charline Giguët-Covex^{1,2,3,4*}, Gentile Francesco Ficetola^{1,2}, Ludovic Gielly^{1,2}, Frédéric Boyer^{1,2}, Fabien Arnaud^{3,4}, Jérôme Poulenard^{3,4}, Pierre Taberlet^{1,2} & Philippe Choler^{1,2,5}

¹ Univ. Grenoble Alpes, LECA, F-38000 Grenoble, France

² CNRS, LECA, F-38000 Grenoble, France

³ Univ. Savoie, EDYTEM, F-73000 Chambéry, France

⁴ CNRS, EDYTEM, F-73000 Chambéry, France

⁵ Univ. Grenoble Alpes, SAJF, F-38000 Grenoble, France

* Co-premiers auteurs

Contribution personnelle : participation à la mise en place de l'étude et aux analyses moléculaires, filtrage des séquences et assignation taxonomique, analyse des données ADN, rédaction du manuscrit

Abstract

Paleoenvironmental studies are essential to understand biodiversity changes over long timescale and to assess the relative importance of anthropogenic and environmental factors. Sedimentary DNA (*sedDNA*) is an emerging tool in the field of paleoecology and has proved to be a complementary approach to pollen and macroremains for investigating past community changes. *SedDNA*-based reconstructions of ancient environment often rely on indicator taxa or expert knowledge, but quantitative ecological analyses might provide more objective information. Here, we analysed *sedDNA* to investigate plant community trajectories in the catchment of a high-elevation lake in the Alps over the last 6400 years. We combined data on past and present plant species assemblages along with sedimentological and geochemical records to assess the relative impact of human activities, through pastoralism, and abiotic factors (temperature and soil evolution). Over the last 6,400 years, we identified strong variations in plant communities, mostly related to soil evolution and pastoral activities. An abrupt vegetation change corresponding to the establishment of an agro-pastoral landscape was detected during the Late Holocene, about 4,500 years ago, with the replacement of mountain forests and tall-herb communities by heathlands and grazed lands. Our results highlight the importance of anthropogenic activities in mountain area for the long-term evolution of local plant assemblages. *SedDNA* data, associated with other paleoenvironmental proxies and present plant assemblages, appear as a relevant tool to reconstruct past plant cover history. Their integration, in complement to classical tools, offers interesting perspectives for a better understanding of long-term ecosystem dynamics under the influence of Human-induced and environmental drivers.

Introduction

The « Anthropocene » concept was originally defined by geologists to characterize a new period in which human activities became one of the most important factors affecting geological process at the global scale (Crutzen 2002; Steffen *et al.* 2011; Wilkinson 2005; Ruddiman 2012). However, biodiversity and ecological processes are also affected. Thus, this definition was extended to the earth biotic component as a period when ecosystems' functioning is primarily influenced by human activities (Ellis 2011). Understanding how and when human activities began modifying ecosystems, and

how ecosystems responded in these human modified-landscapes is a key challenge (Seddon *et al.* 2014).

Because ecological monitoring rarely extends beyond a few decades, long-term studies are crucial to understand the causes and the dynamics of environmental changes (Seddon *et al.* 2014). Lakes are recognized as sentinels of environmental changes (Adrian *et al.* 2009). Their sediments accumulate information about biological communities living both within the water body (*e.g.* cyanobacteria, diatoms, protists, invertebrates; Lotter 1998; Heiri *et al.* 2003; Millet *et al.* 2009; Coolen & Gibson 2009), as well as in the neighbouring terrestrial environment (*e.g.* Birks & Birks 2000; Gaillard *et al.* 2008; Gauthier *et al.* 2010; Tinner *et al.* 2003). The history of vegetation and human impact in the areas surrounding lakes is usually inferred from pollen and macroremains (*e.g.* Birks & Birks 2000; David *et al.* 2010; Anderson-Carpenter *et al.* 2011; Rey *et al.* 2013). The combination of these proxies is essential to reconstruct trajectories of plant assemblages: pollen provides information representing wide spatial scales, due to the wind transport, while macrofossils mainly represent the local plant assemblages (Birks & Birks 2000; Seppä & Bennett 2003). These latter are more sporadic and their identification level is variable, depending of the nature and the preservation state of plant remains (Jørgensen *et al.* 2012; Boessenkool *et al.* 2014).

A complementary approach based on sediment DNA (*sedDNA*) is now emerging (*e.g.* Willerslev *et al.* 2003; Sønstebo *et al.* 2010; Jørgensen *et al.* 2012). The DNA of a variety of taxa, including plants and mammals, can be directly obtained from recent and past sediments (Chariton *et al.* 2010; Hofreiter *et al.* 2012; Pedersen *et al.* 2013; Willerslev *et al.* 2014) even in absence of visible macrofossils. This *sedDNA* constitutes a potential source of information to assess past lake communities (Stoof-Leichenring *et al.* 2012; Ravasi *et al.* 2012; Domaizon *et al.* 2013), but also species assemblages from the surrounding terrestrial ecosystems (Jørgensen *et al.* 2012; Parducci *et al.* 2013; Pedersen *et al.* 2013). Recent developments of high-throughput sequencing now allow massive sequencing of ancient DNA contained in sedimentary archives.

The DNA *metabarcoding* approach seeks to identify multiple taxa from environmental DNA samples in a single experiment (Taberlet *et al.* 2012a). Recent studies demonstrated the technical feasibility of *sedDNA*-based *metabarcoding* analyses and its reliability for paleoecological studies (*e.g.* Jørgensen *et al.* 2012; Parducci *et al.* 2013; Boessenkool *et al.* 2014; Willerslev *et al.* 2014; Giguët-Covex *et al.* 2014). Several

of them compared DNA-based reconstruction with pollen and/or macro-remains based reconstructions and concluded that *metabarcoding* from lake sediments is a complementary tool to classical approaches for investigating past flora (Jørgensen *et al.* 2012; Parnetti *et al.* 2013; Boessenkool *et al.* 2014).

Paleoecological trajectories of plant communities are often studied using indicator taxa (Mazier *et al.* 2009). Multivariate analyses of ecological communities are a powerful tool that allows an objective understanding of the processes determining environmental variation (Legendre & Legendre 2012), but they have been seldom applied on *sedDNA* data. Here, we proposed to combine these two approaches then relying on *metabarcoding* data from subalpine lake, along with present plant species assemblages. Our purposes are to infer past vegetation dynamics in the catchment over the Holocene and to assess the relative importance of human impacts, natural evolution of communities, and variation in climatic conditions in the observed trajectories. First, we analysed time-series (chronosequences) of plant DNA preserved in lake sediments. Second, we used information on species' current habitat preference and other proxies (grazing activity, development of soils, climate change) to reconstruct past landscape history and the potential drivers of vegetation changes.

Material and methods

Study site and sampling

The study was carried out in Lake Anterne (North French Alps, 2063m a.s.l., above the present day treeline). Catchment vegetation is mostly composed of subalpine grasslands dominated by *Festuca nigrescens*, *Carex sempervirens*, and *Nardus stricta*, with scattered pockets of heathland dominated by *Vaccinium uliginosum* and *Rhododendron ferrugineum* and alpine meadows dominated by *Sesleria coerulea* and *Kobresia myosuroides* (Fig. S1 in Supporting Information). Three main soil types were identified in the catchment: leptosols (~50% of the catchment area), cambisols (~25%) and acid soils (podzols/stagnosols, ~15%) (Giguet-Covex *et al.* 2011). Catchment is also characterized by the presence of steep slopes and highly erodible rocks (shales, calcshists and black shales). Consequently, Anterne lake sediments are mainly of detrital origin, *e.g.* made of particles from rocks and soils of the catchment. They are characterized by the alternations of coarse-grained laminae and silty clay laminae corresponding to flood deposits triggered by precipitation events (Giguet-Covex *et al.*

2012). In detrital sediments a high proportion of pollen grains can be degraded, crumpled, broken or corroded, which makes their identification more difficult (Cushing 1967; Birks 1970; Wilmshurst & McGlone 2005; Doyen *et al.* 2013). In such context, ancient sediment DNA could be an interesting approach to reconstruct past vegetation cover. Lake bottom is submitted to anoxic conditions almost all the year. In winter a six to seven months long ice cover allows organic matter decomposition by aerobic bacteria and prevents water column mixing (no water input in the lake and protection from the wind). Anoxic conditions can also occur in summer when the flood activity is low, preventing also the water column mixing (Giguet-Covex, unpublished data). Lake bottom temperature monitoring shows variations between 4 and 8°C (unpublished data). These cold temperatures, combined with quasi-anoxic conditions are favourable conditions for DNA preservation (Coolen & Overmann 2007; Corinaldesi *et al.* 2008; Coolen & Gibson 2009).

The 20.2 m long core (ANT-07) used for the plant DNA study was retrieved in the deeper part of the lake (13 m) in 2007. Core parts specifically dedicated to DNA analyses were shrink-wrapped to limit oxidation of the sediments, stored in a refrigerated area (4°C; similar temperature as the lake bottom) and opened only for DNA extractions. They cover the last 10165 years according to the age-depth model based on 18 ¹⁴C dates on terrestrial macroremains (Giguet-Covex *et al.* 2011). Soils and erosion histories have been described elsewhere (Giguet-Covex *et al.* 2011; Giguet-Covex *et al.* 2012; Fig. S2). Moreover, a previous *metabarcoding* study established presence of DNA attributed to domestic animals in sediments records corresponding to the late Iron Age/Roman period and the Middle Ages/Early Modern Period (Giguet-Covex *et al.* 2014).

Extracellular DNA extraction

From the sediment core, 47 slices of 1cm thickness have been sampled with a focus on the last 6000 years. Sampling of core slices was carried out in a laboratory in University of Savoie where no DNA research were previously done. To avoid potential contamination, we removed edges of slices that could contain contaminating DNA, in particular due to water circulation along the coring tube. Each sample has been divided in two subsamples, and two extractions per subsample (*e.g.* 4 per sample) were carried out. Two different PCRs per extraction were then performed in order to obtain 8

amplification replicates per core slice (Taberlet *et al.* 2012b). This strategy allows maximizing species detection, as well as revealing potential sporadic contaminations.

*Sed*DNA includes both intracellular and extracellular DNA. Results from total DNA extraction and sequencing can thus be biased towards sequences from macroremains (Pedersen *et al.* 2013). We targeted extracellular DNA. Adsorption of extracellular DNA on sediment particles protects it against nuclease degradation (Pietramellara *et al.* 2009). The depurination kinetics of adsorbed DNA in marine sediments was proved to proceed at much slower rates than those theoretically predicted or estimated for terrestrial environments (Corinaldesi *et al.* 2008; Corinaldesi *et al.* 2011). Extracellular DNA was taken off from about 15 g of sediment by mixing it during 15 min with 15 ml of saturated phosphate buffer (Na_2HPO_4 ; 0.12 M; pH $\approx$ 8). Two ml of the mixture was then centrifuged (10min at 10000 g). Four hundred μl of resulting supernatant was kept as starting material for the following extraction steps, using the NucleoSpin® Soil kit (Macherey-Nagel, Düren, Germany), skipping the cell lysis step and following manufacturer's instructions (Taberlet *et al.* 2012b). The DNA extract was eluted in 100 μl of SE buffer and diluted 10 times before being used as PCR template. Four extraction controls were performed. DNA extraction and amplification were performed in rooms specifically dedicated to ancient DNA studies in University Grenoble Alpes.

DNA amplification and high-throughput sequencing

Plant DNA was amplified with universal primers “g” and “h” (Taberlet *et al.* 2007) targeting a short variable region on chloroplast DNA. To assign sequence reads to the relevant sample, 8 bp tags (with at least 5 differences between them) were added to the 5' end of primers (Binladen *et al.* 2007; Valentini *et al.* 2009). DNA amplifications were carried out in a final volume of 30 μl containing 3 μl of diluted DNA extract. The amplification mixture contained 1.2 U of AmpliTaq Gold® DNA Polymerase (Applied Biosystems), 15 mM Tris-HCl, 50 mM KCl, 2 mM of MgCl_2 , 0.2 mM of each dNTP, 0.2 μM of each primer and 4.8 μg of bovine serum albumin (BSA, Roche Diagnostic). After 10 min at 95°C for polymerase activation, PCR mixture underwent 45 cycles: 30 s at 95 °C, 30 s at 50 °C, and 1 min at 72 °C. A final elongation step (7 min at 72 °C) was performed. Eight PCR controls were carried out, in addition to extraction controls, and randomly distributed on PCR plates. PCR products were then purified and mixed together before

sequencing. This later was performed by 2 x 100 bp pair-end sequencing on Illumina HiSeq 2000 platform.

Sequences filtering and taxa assignment

DNA sequences were filtered using the OBITOOLS software (<http://metabarcoding.org/obitools>). First, the direct and reverse reads corresponding to the same sequence were aligned and merged thanks to a gap-free end algorithm implemented in the *IlluminaPairEnd* program. Sequences with a low alignment quality score were removed (<40 *e.g.* pair of reads that can not align perfectly on at least 10 bp at each read end). Then, the *ngsfilter* program assigned the assembled sequences to the relevant PCR. Only sequences containing both primers (with a maximum of 3 mismatches per primer) and exact tag sequences were selected. Sequences containing ambiguous nucleotides and those shorter than 7 bp were discarded by *obigrep* command. We used the *obiuniq* program to merge together strictly identical sequences, keeping the information about the original samples they come from. Afterwards, singletons (*e.g.* sequences occurring once in the dataset) were filtered out. The *obiclean* program allows detecting sequences corresponding to putative amplification and sequencing errors. Within each PCR product, the program creates networks connecting 2 sequences differing by a single difference (indel or substitution) are linked if the occurrence of the less frequent sequence is lower than 50% of its related most frequent sequence. A status is then assigned to each sequence within each PCR: “head” (most common sequence of a network), “singleton” (sequence without variant) or “internal” (sequences that are neither “head” nor “singleton”, consequently corresponding probably to an amplification or sequencing error). Only sequences that are more often “head” or “singleton” than “internal” across the whole dataset were kept for the following steps. Finally, sequences were assigned to the relevant taxa using *ecoTag* program that finds highly similar sequence(s) in a suitable database. Here, sequences were compared to global plant database generated by *in silico* PCR from EMBL with the *ecoPCR* program (Ficetola *et al.* 2010) and to a database grouping most of arctic and boreal plant taxa (Willerslev *et al.* 2014). By default, sequences were assigned to the global database reduced to alpine taxa. If identity score was higher for the global database as a whole or to the arctic/boreal database, sequences were reassigned to a taxon of these databases.

Potential contaminants, identified by having their highest occurrence in negative control samples, were removed. Then, the Bray-Curtis distance was calculated between all PCR. A graph-partitioning approach (*igraph* package in R, Csardi & Nepusz 2006) was used to detect, and subsequently filter out, outlying PCR and those clustering with control samples. Additionally, for each sample, outliers PCR (*e.g.* PCR that did not cluster with other PCR from the same sample) were also removed.

In the final dataset, we kept only sequences with a perfect match after taxonomic assignment and whose occurrence was higher than 1000 reads over the whole dataset (*e.g.* corresponding to less than 0.02% of the most frequent sequence). As it is difficult to disentangle rare MOTU's from potential errors at low abundance, we considered a sequence as genuine in a PCR product if its count >10 reads. This criterion obviously dismisses many putative taxa present at low abundance but allows also filtering out potential errors.

Attesting the presence of a DNA sequence in sediments at a given timespan requires caution and stringent filtering. Due to the low quantity of initial template DNA and to potential sporadic contaminations, amplification process is relatively stochastic. Detection probability of a taxon present during a given time may be low. On the other hand, false positives can never be excluded, for instance because of occasional contamination or sequence misidentification. To overcome these issues, we performed eight replicates per samples, and a sequence was considered as genuine in a given timespan if it was detected in at least 3 replicates (Fig. S3). A few haplotypes of exotic origin or detected as contaminants in control samples and removed. The taxonomic assignment of these sequences includes *Actinidia*, *Myriophyllum*, *Trifolium*, *Capsicum*, *Acer*, *Tilia*, *Malae* and *Pooideae*. Except the exotic *Actinidia* and *Myriophyllum*, which is an aquatic taxon, these sequences were found only once in control samples. The conservative filtering procedure used here prevented the detection of these false positives. Thus, the sequences used for reconstructing past plant communities come from endogenous DNA, and unlikely from contemporaneous contaminations.

We then used occupancy models (Miller *et al.* 2011) to evaluate detection probability and the probability of false detection of each remaining MOTU (Table S1). This analysis therefore provides a measure of the reliability of the pattern of presence/absences. Simulation analysis showed that this approach could successfully estimate these parameters in environmental DNA studies (Ficetola *et al.* submitted).

Thus, we estimated the probability of non-detecting each MOTU when it is actually present, and the probability of obtaining three false positives in a given sample. Occupancy modelling was performed using the *unmarked* package in R (Fiske & Chandler 2011).

Data analysis

To investigate plant community trajectories over the last 6000 years, we focused on points containing at least 10 taxa identified up to the genus or species level. Redundancy analysis (RDA) based on presence/absence of taxa was performed to identify community changes among periods and the taxa specifically associated to them (Mazier *et al.* 2009; Borcard *et al.* 2011). For the most contrasted periods, we selected a set of indicator taxa (species or genera) for which barcodes were available in the databases in order to investigate their long-term dynamics.

In RDA we also assessed the potential role of temperature, type of soil and human activities on the evolution of plant communities over time. Available temperature variation reconstruction from the lake Anterne, based on chironomid records (Millet *et al.* 2009), cover the last 2000 years only. Consequently, we used the mid-European decadal isotope-climate record from VanGrafenstein *et al.* (1999) extended to the present, as a measure of long-term variation of temperature. Calcium content in sediments is used as a proxy for carbonated rocks dissolution and thus for past soil acidity (Giguët-Covex *et al.* 2011, Fig. S2). The presence of pastoralism (presence of cow and/or sheep herds), assessed from *sedDNA* (Giguët-Covex *et al.* 2014), was considered as a proxy of human impact, as pastoralism is the main human activity in this area. Significance of RDA was assessed through 10,000 permutations.

Ecology of taxa identified by DNA *metabarcoding* was derived from a multivariate analysis of 3077 mountain vegetation relevés (*e.g.* an exhaustive list of species with visual estimate of local species' abundance) from the Northern part of the French Alps. Relevés were extracted from a large floristic database maintained by the Conservatoire Botanique National Alpin. Relevés corresponding to mountain forests, grasslands and meadows above 1500 m a.s.l. were retained. A non-metric multidimensional scaling (NMDS) based on Bray-Curtis distance was performed in order to summarize current pattern of species' co-occurrence (Borcard *et al.* 2011). NMDS was limited to 982 alpine plant species, detected in at least 5 relevés. A non-metric multidimensional scaling

(NMDS) based on Bray-Curtis distance was performed in order to summarize current pattern of species' co-occurrence (Borcard *et al.* 2011). NMDS was limited to 982 alpine plant species, detected in at least 5 relevés. We then used a vector fitting approach to relate the ordination scores of 223 species with habitat preferences (light, temperature, nitrogen, pH and moisture), as given by Ellenberg indicator values. The vector fitting was tested for significance by means of 1000 permutations. Habitat preferences of species identified by *metabarcoding* were estimated by the projection of their NMDS scores along habitat vectors. For taxa only identified at the genus level, we used the mean score of all species belonging to that genus (Fig. S4). Statistical analyses were performed using *vegan* package in R (Oksanen *et al.* 2013).

Results

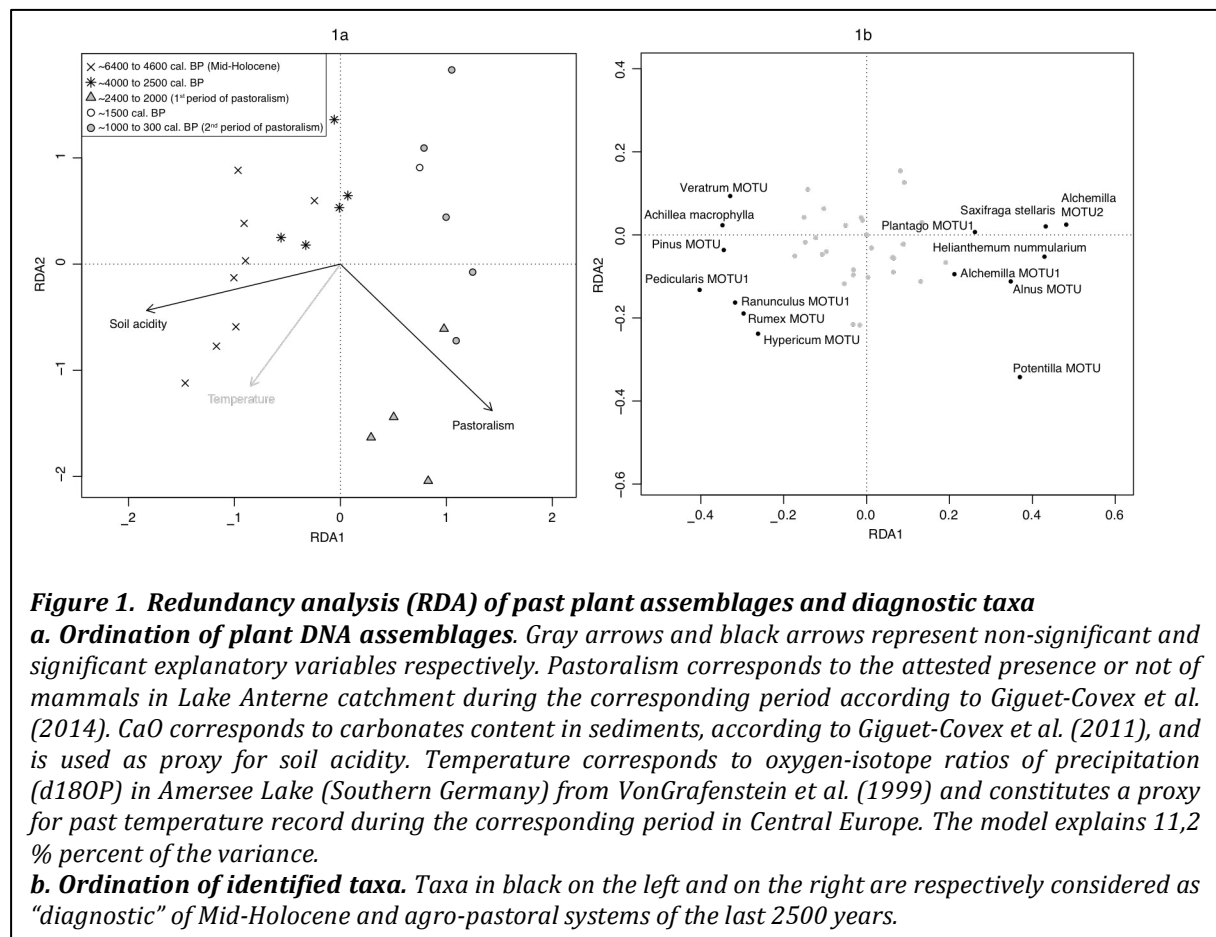
Plant DNA identification

High-throughput sequencing provided 28,019,267 usable reads corresponding to 219,266 unique sequences. After filtering procedure, we obtained 17,759,158 reads and 144 unique sequences. Eighty of them (corresponding to 17,264,015 reads) cross at least once the threshold of three replicates per core slice (Appendix S1, Fig. S3). Among these latters, 30 plant families and 38 genera were identified and 9 sequences were assigned up to the species level. For most of identified taxa, detection probability was high (generally >0.3), while the probability to obtain 3 false positives was always <5%, allowing reliable inference on presence/absence patterns (Table S1).

Plant community changes

Among the 47 slices of sediments, 23 contained at least 10 well-identified taxa and were used to investigate vegetation dynamics over the Holocene. Among them, 8 are between ≈ 6400 to 4600 cal. BP, 5 between ≈ 3900 to 3550 cal. BP, 4 between ≈ 2400 and 2000 cal. BP, 1 around 1500 cal. BP and 5 between 1000 to 300 cal. BP. When two sequences, or more, were assigned to the same genus (*e.g. Alchemilla, Ranunculus, Plantago...*), they were considered separately as their barcodes belonged to different species. Therefore, 49 sequences (corresponding to 39 taxa identified at the genus or the species level) were available for ecological analyses. The redundancy analysis of plant DNA chronosequences (time-series) showed that these groups of dates were well clustered (Fig. 1a). The first RDA axis discriminated the mid-Holocene climatic optimum

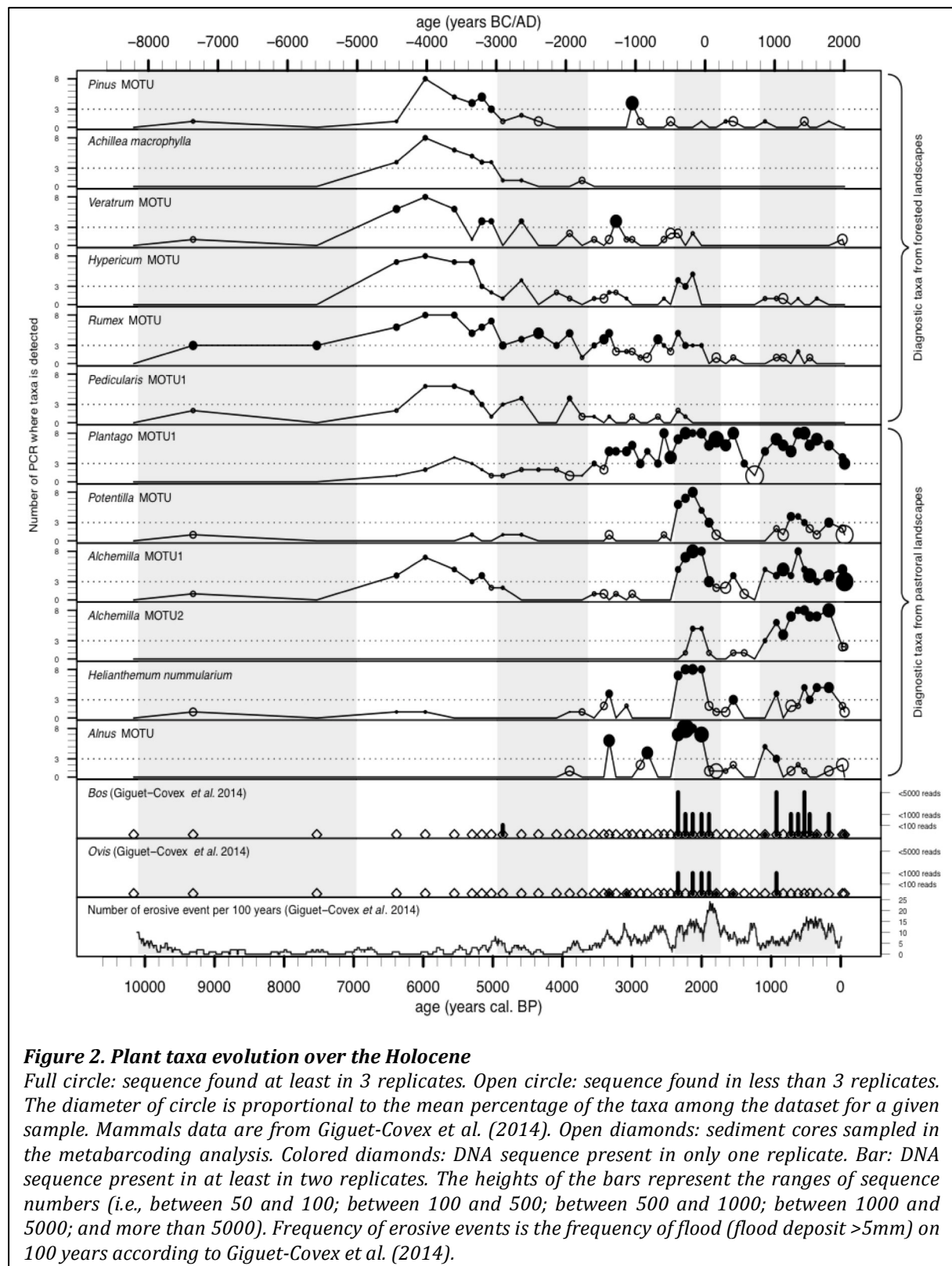
(≈ 6400 to 4600 cal. BP) from the recent period, between the Iron Age and the Modern times (≈ 2400 to 300 cal. BP), where two important phases of pastoral activities and erosion are identified (Fig. 2).



The predictors used in RDA (Fig. 1a & b) explained 11.2% of the variation (permutation $P=0.002$). The pure effect of soil acidity (carbonate content of eroded soils) and pastoralism (proved presence of herds) accounted respectively for 26% and 14% of explained variation, while the pure effect of temperature explained no variation. Furthermore, the joint effect of soil acidity and pastoralism accounted for 42%.

Plant taxa chronosequences (Fig. 1b & Fig. 2) allowed identifying diagnostic taxa of the two main periods. *Achillea macrophylla* and taxa from genera *Pinus*, *Veratrum*, *Hypericum*, *Rumex*, *Pedicularis* and *Ranunculus* were characteristics of the mid-Holocene optimum. Agro-pastoral periods, Iron Age-Roman period (≈ 2400 - 1800 cal. BP) and Middle Ages-Modern Times (≈ 1200 - 150 cal. BP) are characterized by the presence of

Helianthemum nummularium, *Saxifraga stellaris* and taxa from genera *Plantago*, *Alnus*, *Alchemilla* and *Potentilla* (Fig. 1b & Fig. 2).



Ecological characteristics of these periods were inferred by comparison of diagnostic taxa with current species assemblages (Fig. S4). The Holocene Climatic Optimum corresponded to vegetation currently observed in relatively warmer, more shaded and nutrient-rich environment, by comparison to those found for the last 2500 years (Fig. 3). No trends were found for pH and moisture. Other taxa detected during the mid-Holocene included several *Ericaceae*: *Vaccinium* genus (including *V. uliginosum*) and *Rhodoreae* tribe (Fig. S3), typical from developed acid soils (podzols) as *Pinus* trees.

For the recent period, most identified taxa belong to low stature plants from grazed grasslands and meadows, some on acidic soils and others on basic ones. Some species of these genera, such as *Plantago alpina* and *P. lanceolata*, are well-known indicators of pastoralism (Wick *et al.* 2003). *Alnus* sp. was detected at the beginning of the two agro-pastoral periods but disappeared respectively around 2000 and 1000 cal. BP (Fig. 2) and *Vaccinium* sp. was punctually found.

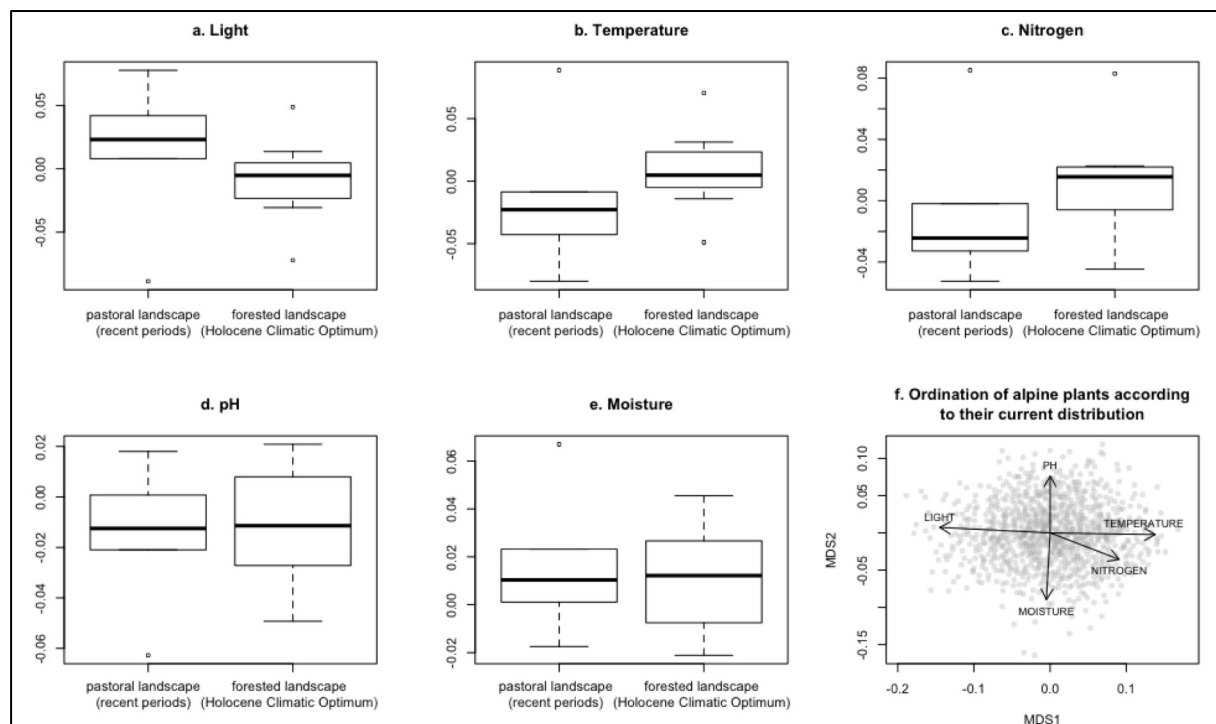


Figure 3. Ecological characteristics of the mid-Holocene and agro-pastoral systems

Boxplot reflecting ecological characteristics of the mid-Holocene period and agro-pastoral systems (last 2500 years), inferred from current distribution of diagnostic taxa: a) Light; b) Temperature; c) Nitrogen; d) pH; e) moisture. f) NMDS ordination of 982 alpine plants according to their current distribution in Northern French Alps and their relation with abiotic parameters (stress value=0.1305256). All fittings are significant ($P < 0.001$).

Discussion

Plant community evolution

Sedimentary DNA is supposed to come from mainly from surrounding environment and to have a local origin (Jørgensen *et al.* 2012; Parducci *et al.* 2013). Here, it provided fine-scale information about past plant biodiversity and community dynamics. Even if some limitations exist for taxonomic resolution, it was possible to identify diagnostic taxa showing, in particular, an abrupt distribution change from the pre-human and the agro-pastoral period (Fig. 1 & 2). The integration between *metabarcoding* data and current information on species' distribution allowed obtaining reliable ecological information from some taxa at a level other than species.

Our results demonstrated an important phase of vegetation shift between the mid-Holocene and the last 2500 years (Fig. 1 & Fig. 2). They are congruent with palynological data obtained from two neighbouring peatbogs (David 2010a ; 2010b). Comparison of indicator taxa with their current distribution shows a good congruence between past species distribution and present ones: diagnostic taxa from these two periods currently display strong ecological differences (Fig. 3). Herbaceous taxa from the mid-Holocene include large-leaved species (*e.g.* *Veratrum sp.*, *Achillea macrophylla*) that are characteristics of tall-herb communities usually found in mountain forests understory or fringes. We hypothesized that subalpine Pine forests and tall-herb communities were the dominant vegetation in the lake catchment before human impacts. This period (≈ 6400 to 4600 cal. BP) corresponds to the hypsithermal, characterized by forest expansion all over Europe (*e.g.* Nicolussi *et al.* 2005; Prentice *et al.* 2011; Marquer *et al.* 2014), which is congruent with our results. The subsequent opening of the landscape around 4600 cal. BP was also observed in Survilly peatbog (2235m), with the regression of *Pinus* trees (David 2010a).

The last 2500 years are marked by two successive waves of pastoralism followed by a decrease of human activity. They were characterized by presence of heliophilous taxa from open environment, typically subalpine meadow: *Helianthemum nummularium*, *Potentilla sp.*, *Alchemilla sp.* *Plantago* genus contains several taxa recognized as pastoralism indicators. A similar plant cover was reconstructed in surrounding peatbogs at this period (David 2010a ; 2010b). DNA analyses proved the presence of cattle and/or sheep herds in the lake Anterne catchment during the late Iron Age and the Roman period, as well as during the Middle Ages and the Modern times (Fig. 1, Giguët-Covex *et*

al. 2014). *Hedysarum hedysaroides*, and *Potentilla* were found specifically during these two periods, some of these species being associated with pastoral activities. These taxa disappear between the two pastoral activities phases (≈ 1800 to 1200 cal. BP) and are absent from the two most recent points, which is congruent with the absence of mammal DNA (Fig. 2) and suggest the agricultural abandonment of the catchment at these times. *Alnus* was also detected at the beginning of pastoral activities but disappeared afterward likely related to grazing intensification of pasturing and shrubs clearance (Wick & Tinner 1997). Such decrease of *Alnus* was also observed in Ecuellas peatbog (1855 m) around 1960 cal. BP (David 2010b). Our results suggest a landscape mosaic corresponding to open environments such as subalpine grasslands and alpine meadows dominated by low stature plants with remnants pockets of heathlands (*Vaccinium* and *Alnus*). Otherwise, the detection of taxa belonging to different communities, currently present on different sides around the lake, suggests that sediment DNA offers a good representation of the catchment as a whole. We hypothesize this human shaped landscape resulted from the combination of logging, burning and grazing (Gobet *et al.* 2003; Blarquez *et al.* 2010; Rey *et al.* 2013).

Role of anthropogenic and abiotic factors

Previous studies highlighted the important role of Human in landscape shaping (Gobet *et al.* 2003; Röpke *et al.* 2011). Here, pastoralism is significantly correlated with plant community changes during the last 6400 years (Fig. 1a). Its importance might be underestimated in our analysis, as data used to infer past farming history only take into account periods where mammal DNA was found with certainty (Giguet-Covex *et al.* 2014). However, other data suggest that the first human impact in the catchment might have occurred 1000 years before the first attested presence of herds (Covex *et al.* 2011; Giguet-Covex *et al.* 2014): regular detection of *Plantago* sp. since 3500 cal. BP, punctual presence of *Alnus* sp. at 3230 and 2779 cal. BP (Fig. 2) but also the increase in the frequency and thickness of deposits of erosive events around 3400 cal. BP that will never return to previous conditions afterwards. This increase is concomitant with those of the Rock-Eval pyrolysis oxygen index that reflects the erosion of deep soil horizons (Fig. S2; Giguet-Covex *et al.* 2011), indicating an important and nearly continuous soil destabilization since this period. Reconstruction of past fire activities based on charcoal remains (*e.g.* Blarquez & Carcaillet 2010; Marlon *et al.* 2012; Rey *et al.* 2013) would

provide additional information on human activities associated with landscape clearance around the lake.

Soil acidity of eroded soils was also strongly related to past plant communities (Fig. 1a) even if no difference was found using ecology of indicator taxa. It highlights the importance of soil properties in plant community changes over long timescale. Even if soil characteristics are here reduced to acidity, others parameters (non-reconstructed from lake sediments) have probably an effect. During the Mid-Holocene, calcium content in terrigenous sediments reaches its minimum (CaO, Fig. S2). The low calcium content is concordant with the presence of subalpine dwarf-shrub heath community, under acid soils (podzols). These soils have experienced first a complete dissolution of CaCO_3 , then a leaching of bases leading to acidification. Then, the strong decrease of *Pinus* (Fig. 2) and *Vaccinium* (Fig. S3) is concomitant with silicium content rises (Fig. S2), which could reflect major erosion of previously formed podzols or the formation and erosion of another type of acid soil called stagnosols (hydromorphic soils). The degradation of acid soils lasts until 3200 cal. BP. During the main periods of livestock farming, a progressive soil degradation with first erosion of soil surface horizons (increase of Hydrogen Index) followed by those of deeper horizons (increase of Oxygen Index) is recorded (Fig. S2; Giguet-Covex *et al.* 2011). After these periods, the increase of calcium content suggests the appearance of new raw soils (carbonated bedrock) probably linked to the previous degradation of developed soils (Fig. S2). These results suggest an indirect impact of human activities on plant communities by modifying soils characteristics.

Climate was a key element of biome structuring and played important role during the deglaciation and the first vegetation processes (Van Grafenstein *et al.* 1999). In this study we focused on temperature, a climatic parameter for which reliable proxies exist over long time. Although past altitudinal variations of some species were correlated with temperature (*e.g.* Holtmeier & Broll 2005; Colombaroli *et al.* 2010) and the climatic dependance of human activities was observed in other alpine areas (Tinner *et al.* 2003; Röpke *et al.* 2011), no significant relationship was found here between mean annual temperature and changes of plant communities during the last 6000 years (Fig. 1). The paleoclimatic reconstruction used here was inferred in Alpine area and has an appropriate resolution, but it referred to a low-altitude lake. Yet, temperature changes are known to have greater amplitude in mountains areas (Beniston 2005). That might partly explain why we did not find any correlation between mean temperatures *per se*

and vegetation changes during the last 6000 years (Fig. 1). The role of other parameters, such as precipitation and humidity, more difficult to infer, but may be relevant. Anyway, this highlights the importance of anthropogenic factors for determining local plant assemblages.

These results illustrate the capacity of human societies to irreversibly modify mountain vegetation and landscapes. In Europe, the first large-scale human impacts on the environment occurred during the Neolithic with sedentarisation, development of agriculture and the associated forest clearance (Behre 1988). The onset of strong human impacts on plant cover and ecosystems at high altitude has been dated to the Bronze Age (Schmidt *et al.* 2002 ; Koinig *et al.* 2003 ; Lotter *et al.* 2006 ; Walsh *et al.* 2007 ; Roepke & Krause 2013). Around lake Anterne, introduction of livestock in the system induced a dramatic shift in plant community. Pastoralism was a severe perturbation to shift the vegetation towards an alternative state (Miehe *et al.* 2014). Some species that have disappeared were not recorded anymore even in periods with supposed absence of livestock (*e.g.* *Eryngium* and *Pinus* genera, *Achillea macrophylla*...). Effects of human activities persist over long timescale because disturbance affects also physical and chemical parameters of the environment (Scheffer *et al.* 2001), preventing a possible return to forest mountain landscape during climatically favourable periods. Ecological and historical trajectories are here closely related; and the legacy of past disturbances, particularly on soils, could play an important role. The relative importance of disturbance history and soils characteristics in plant community trajectories also highlights the limits of species distribution models only based on species climate envelopes ranges (Pearman *et al.* 2008). Our approach helps understanding the coupled dynamics of soils, human activities, climate and vegetation.

Perspectives

SedDNA-based paleoecological reconstructions are still in their infancy (Boessenkool *et al.* 2014). Here, we showed the relevance of these data, combined with other proxies and present plant assemblages, to reconstruct past plant cover history and its potential drivers of changes. Multiproxy approaches are essential for the understanding of the long-term evolution of ecosystems (Birks & Birks 2000; Röpke *et al.* 2011), thus we argue that integration of *sedDNA* in paleoecological studies, along

with pollen and macroremains, will certainly improve the understanding of past human-environment interactions.

The *sed*DNA-based approach has currently some limitations but offers, in addition to a relatively fast and easy protocol, some interesting perspectives. Here, only sequences assigned with a perfect assignment score (100%) were considered, consequently data are highly dependent of the completeness of the reference database. Besides, the taxonomic resolution of the *trnL* P6 loop does not always allow the identification to the species level, barcodes being often shared by several species. However, this fragment represents only a very limited part of the chloroplast DNA, less than one thousand of the whole molecule. The fact that such a small part of chloroplast DNA can be reliably amplified demonstrates that a very large number of chloroplast DNA fragments are present in the DNA extract, and thus in the core slice. The next challenge will be to extract more efficiently the taxonomic information from these DNA fragments contained in the core. One possibility will be to implement a PCR-free approach (Taberlet *et al.* 2012c; Zhou *et al.* 2013), and to sequence directly the DNA extracts by a shotgun approach. An ongoing project consists to sequence the whole chloroplast DNA for about 4500 species of the Alpine flora. Shotgun sequencing combined with the availability of an exhaustive reference database containing whole chloroplast DNA molecules has the potential of allowing a much more precise reconstruction of past plant communities, with identifications to the species level, and opens unprecedented avenues in paleoecology.

Acknowledgements

We thank the French National Research Agency for funding the fieldwork and sedimentological and geochemical measurements through the Pygmalion research program (ANR BLAN07- 2_204489). DNA measurements were performed in the framework of the Retromont program (funded by DiPEE Chambéry-Grenoble and the University Grenoble Alpes Savoie) and the MetaBar research program (ANR 11 BSV7 020 01). Université de Savoie, Université Grenoble-Alpes and CNRS jointly funded C. Giguet-Covex's salary. We would also like to thank ASTERS, the manager of the Haute Savoie natural reserves, for constant help since the onset of the paleoenvironmental studies on Lake Anterne and the National Alpine Botanical Conservatory (CBNA) for

providing botanical relevés.

References

- Adrian R, O'Reilly CM, Zagarese H *et al.* (2009) Lakes as sentinels of climate change. *Limnology and Oceanography*, **54**, 2283-2297.
- Anderson-Carpenter LL, McLachlan JS, Jackson ST *et al.* (2011) Ancient DNA from lake sediments: Bridging the gap between paleoecology and genetics. *BMC Evolutionary Biology*, **11**, 30.
- Behre KE (1988) The rôle of man in European vegetation history. In Huntley & Webb T (Eds.), *Vegetation history*, 633–672. Dordrecht: Springer Netherlands.
- Beniston M (2005) Mountain climates and climatic change: An overview of processes focusing on the European Alps. *Pure Applied and Geophysics*, **162**, 1587-1606.
- Binladen J, Gilbert MTP, Bollback JP *et al.* (2007) The Use of Coded PCR Primers Enables High-Throughput Sequencing of Multiple Homolog Amplification Products by 454 Parallel Sequencing. *Plos One*, **2**, e197.
- Birks HJB (1970) Inwashed pollen spectra at Loch Fada, Isle of Skye. *New Phytologist*, **69**, 807-820.
- Birks HH, Birks HJB (2000) Future uses of pollen analysis must include plant macrofossils. *Journal of Biogeography*, **27**, 31–35.
- Blarquez O, Carcaillet C, Mourier B, Bremond L, Radakovitch O (2010) Trees in the subalpine belt since 11 700 cal. BP: origin, expansion and alteration of the modern forest. *Holocene*, **20**, 139-146.
- Blarquez O, Carcaillet C (2010) Fire, fuel composition and resilience threshold in subalpine ecosystem. *PloS One*, **5**, e12480.
- Boessenkool S, McGlynn G, Epp LS *et al.* (2014) Use of ancient sedimentary DNA as a novel conservation tool for high-altitude tropical biodiversity. *Conservation Biology*, **28**, 446–455.
- Borcard D, Gillet F, Legendre P (2011) *Numerical Ecology with R*. Springer, New York.
- Chariton AA, Court LN, Hartley DM, Colloff MJ, Hardy CM (2010) Ecological assessment of estuarine sediments by pyrosequencing eukaryotic ribosomal DNA. *Frontiers in Ecology and the Environment*, **8**, 233-238.
- Colombaroli D, Henne PD, Kaltenrieder P, Gobet E, Tinner W (2010) Species responses to fire, climate and human impact at tree line in the Alps as evidenced by palaeo-environmental records and a dynamic simulation model. *Journal of Ecology*, **98**, 1346–1357.
- Coolen MJL, Gibson JAE (2009) Ancient DNA in lake sediment records. *Pages News*, **17**, 104-106.

- Coolen MJL, Overmann J (2007) 217 000-year-old DNA sequences of green sulfur bacteria in Mediterranean sapropels and their implications for the reconstruction of the paleoenvironment. *Environmental microbiology*, **9**, 238–249.
- Crutzen PJ (2002) Geology of mankind. *Nature*, **415**, 23-23.
- Csárdi G, Nepusz T (2006) The igraph software package for complex network research. *InterJournal Complex Systems*, **1695**, 5.
- Cushing EJ (1964) Evidence for differential pollen preservation in late quaternary sediments in Minnesota. *Review of Palaeobotany and Palynology*, **4**, 87-101.
- David F (2010a) An example of the consequences of human activities on the evolution of subalpine landscapes. *Comptes Rendus Palevol*, **9**, 229–235.
- David F (2010b) Expansion of green alder (*Alnus alnobetula* [Ehrh] K. Koch) in the northern French Alps: a palaeoecological point of view. *Comptes Rendus Biologies*, **333**, 424–428.
- Domaizon I, Savichtcheva O, Debroas D *et al.* (2013). DNA from lake sediments reveals the long-term dynamics and diversity of *Synechococcus* assemblages. *Biogeosciences Discussions*, **10**, 2515–2564.
- Doyen E, Vanni  re B, Berger JF *et al.* (2013) Land-use changes and environmental dynamics in the upper Rhone valley since Neolithic times inferred from sediments in Lac Moras. *The Holocene*, **24**, 1320-1335.
- Ellis EC (2011) Anthropogenic transformation of the terrestrial biosphere. *Philosophical Transactions of the Royal Society of London Series A*, **369**, 1010-1035.
- Ficetola GF, Coissac E, Zundel S *et al.* (2010) An in silico approach for the evaluation of DNA barcodes. *BMC Genomics*, **11**, 434.
- Ficetola GF, Pansu J, Bonin A *et al.* Replication levels, false presences, and the estimation of presence / absence from eDNA metabarcoding data. *submitted*
- Fiske I, Chandler R (2011) unmarked: An R package for fitting hierarchical models of wildlife occurrence and abundance. *Journal of Statistical Software*, **43**, 1-23.
- Gaillard MJ, Sugita S, Bunting MJ *et al.* (2008) The use of modelling and simulation approach in reconstructing past landscapes from fossil pollen data: a review and results from the POLLANDCAL network. *Vegetation History and Archaeobotany*, **17**, 419–443.
- Gauthier E, Bichet V, Massat C *et al.* (2010) Pollen and non-pollen palynomorph evidence of Medieval farming activities in southwestern Greenland. *Vegetation History and Archaeobotany*, **19**, 427-438.
- Giguet-Covex C, Arnaud F, Enters D *et al.* (2012) Frequency and intensity of high-altitude floods over the last 3.5 ka in northwestern French Alps (Lake Anterne). *Quaternary Research*, **77**, 12-22.
- Giguet-Covex C, Arnaud F, Poulenard J *et al.* (2011) Changes in erosion patterns during the Holocene in a currently treeless subalpine catchment inferred from lake

- sediment geochemistry (Lake Anterne, 2063 m a.s.l., NW French Alps): The role of climate and human activities. *Holocene*, **21**, 651-665.
- Giguët-Covex C, Pansu J, Arnaud F *et al.* (2014) Long livestock farming history and human landscape shaping revealed by lake sediment DNA. *Nature communications*, **5**, 3211.
- Gobet E, Tinner W, Hochuli PA, van Leeuwen JFN, Ammann B (2003) Middle to Late Holocene vegetation history of the Upper Engadine (Swiss Alps): the role of man and fire. *Vegetation History and Archaeobotany*, **12**, 143-163.
- Heiri O, Wick L, van Leeuwen JFN, van der Knaap WO, Lotter AF (2003) Holocene tree immigration and the chironomid fauna of a small Swiss subalpine lake (Hinterburgsee, 1515 m asl). *Paleogeography, Paleoclimatology, Paleoecology*, **189**, 35-53.
- Hofreiter M, Collins M, Stewart JR (2012) Ancient biomolecules in Quaternary palaeoecology. *Quaternary Science Review*, **33**, 1-13.
- Holtmeier FK, Broll G (2005) Sensitivity and response of northern hemisphere altitudinal and polar treelines to environmental change at landscape and local scales. *Global Ecology and Biogeography*, **14**, 395-410.
- Jørgensen T, Haile J, Møller P *et al.* (2012) A comparative study of ancient sedimentary DNA, pollen and macrofossils from permafrost sediments of northern Siberia reveals long-term vegetational stability. *Molecular Ecology*, **21**, 1989-2003.
- Koinig KA, Shotyk W, Lotter AF, Ohlendorf C, Sturm M (2003) 9000 years of geochemical evolution of lithogenic major and trace elements in the sediment of an alpine lake—the role of climate, vegetation, and land-use history. *Journal of Paleolimnology*, **30**, 307-320.
- Legendre P, Legendre L (2012) *Numerical Ecology*, 3rd Edition edn. Elsevier, Amsterdam.
- Lotter AF (1998) The recent eutrophication of Baldeggersee (Switzerland) as assessed by fossil diatom assemblages. *Holocene*, **84**, 395-405.
- Lotter AF, Heiri O, Hofmann W *et al.* (2006) Holocene timber-line dynamics at Bachalpsee, a lake at 2265 m asl in the northern Swiss Alps. *Vegetation history and archaeobotany*, **15**, 295-307.
- Marlon JR, Bartlein PJ, Gavin DG *et al.* (2012) Long-term perspective on wildfires in the western USA. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, E535-E543.
- Marquer L, Gaillard M-J, Sugita S *et al.* (2014) Holocene changes in vegetation composition in northern Europe: why quantitative pollen-based vegetation reconstructions matter. *Quaternary Science Review*, **90**, 199-216.
- Mazier F, Galop D, Gaillard MJ *et al.* (2009) Multidisciplinary approach to reconstructing local pastoral activities: an example from the Pyrenean Mountains (Pays Basque). *The Holocene*, **19**, 171-188.

- Miehe G, Miehe S, Böhner J *et al.* (2014) How old is the human footprint in the world's largest alpine ecosystem? A review of multiproxy records from the Tibetan Plateau from the ecologists' viewpoint. *Quaternary Science Review*, **86**, 90-209.
- Miller DA, Nichols JD, McClintock BT *et al.* (2011) Improving occupancy estimation when two types of observational error occur: non-detection and species misidentification. *Ecology*, **92**, 1422-1428.
- Millet L, Arnaud F, Heiri O *et al.* (2009) Late-Holocene summer temperature reconstruction from chironomid assemblages of Lake Anterne, northern French Alps. *The Holocene*, **19**, 317-328.
- Nicolussi K, Kaufmann M, Patzelt G, van der Plicht J, Thurner A (2005) Holocene tree-line variability in the Kauner Valley, Central Eastern Alps, indicated by dendrochronological analysis of living trees and subfossil logs. *Vegetation History and Archaeobotany*, **14**, 221-234.
- Oksanen J, Blanchet FJ, Kindt R *et al.* (2013) vegan: Community Ecology Package. *R package version 2.0-7*. <http://CRAN.R-project.org/package=vegan>
- Parducci L, Matetovici I, Fontana SL *et al.* (2013) Molecular- and pollen-based vegetation analysis in lake sediments from central Scandinavia. *Molecular Ecology*, **22**, 3511-3524.
- Pearman PB, Randin CF, Broennimann O *et al.* (2008) Prediction of plant species distributions across six millennia. *Ecology Letters*, **11**, 357-369.
- Pedersen MW, Ginolhac A, Orlando L *et al.* (2013) A comparative study of ancient environmental DNA to pollen and macrofossils from lake sediments reveals taxonomic overlap and additional plant taxa. *Quaternary Science Review*, **75**, 161-168.
- Pietramellara G, Ascher J, Borgogni F *et al.* (2009) Extracellular DNA in soil and sediment: fate and ecological relevance. *Biology and Fertility of Soils*, **45**, 219-235.
- Prentice IC, Jolly D (2001) Mid-Holocene and glacial-maximum vegetation geography of the Northern continents and Africa. *Journal of Biogeography*, **27**, 507-519.
- Ravasi DF, Peduzzi S, Guidi V *et al.* (2012) Development of a real-time PCR method for the detection of fossil 16S rDNA fragments of phototrophic sulfur bacteria in the sediments of Lake Cadagno. *Geobiology*, **10**, 196-204.
- Rey F, Schworer C, Gobet E *et al.* (2013). Climatic and human impacts on mountain vegetation at Lauenensee (Bernese Alps, Switzerland) during the last 14,000 years. *The Holocene*, **23**, 1415-1427.
- Roepke A, Krause R (2013) High montane-subalpine soils in the Montafon Valley (Austria, Northern Alps) and their link to land-use, fire and settlement history. *Quaternary International*, **308**, 178-189.
- Röpke A, Stobbe A, Oeggel K, Kalis AJ, Tinner W (2011) Late-Holocene land-use history and environmental changes at the high altitudes of St Antonien (Switzerland,

- Northern Alps): Combined evidence from pollen, soil and tree-ring analyses. *The Holocene*, **21**, 485–498.
- Ruddiman WF (2012) The Anthropocene. *Annual Review of Earth and Planetary Sciences*, **41**, 45–68.
- Scheffer M, Carpenter S, Foley JA, Folke C, Walker B (2011) Catastrophic shifts in ecosystems. *Nature*, **413**, 591–596.
- Schmidt R, Koinig KA, Thompson R, Kamenik C (2002) A multi proxy core study of the last 7000 years of climate and alpine land-use impacts on an Austrian mountain lake (Unter Landschitzsee, Niedere Tauern). *Palaeogeography, Palaeoclimatology, Palaeoecology*, **187**, 101–120.
- Seddon AWR, Mackay AW, Baker AG *et al.* (2014) Looking forward through the past: identification of 50 priority research questions in palaeoecology. *Journal of Ecology*, **102**, 256–267.
- Seppä H, Bennett KD (2003) Quaternary pollen analysis: recent progress in palaeoecology and palaeoclimatology. *Progress in Physical Geography*, **27**, 548–579.
- Sønstebo JH, Gjelty L, Brysting AK *et al.* (2010) Using next-generation sequencing for molecular reconstruction of past Arctic vegetation and climate. *Molecular Ecology Resources*, **10**, 1009–1018.
- Steffen W, Grinevald J, Crutzen P, McNeill J (2011) The Anthropocene: conceptual and historical perspectives. *Philosophical Transactions of the Royal Society of London Series A*, **369**, 842–867.
- Stoof-Leichsenring KR, Epp LS, Trauth MH, Tiedemann R (2012) Hidden diversity in diatoms of Kenyan Lake Naivasha: a genetic approach detects temporal variation. *Molecular Ecology*, **21**, 1918–1930.
- Taberlet P, Coissac E, Pompanon F *et al.* (2007) Power and limitations of the chloroplast *trnL* (UAA) intron for plant DNA *barcoding*. *Nucleic Acids Research*, **35**, 1–8.
- Taberlet P, Coissac E, Hajibabaei M, Rieseberg LH (2012a) Environmental DNA. *Molecular Ecology*, **21**, 1789–1793.
- Taberlet P, Prud'homme SM, Campione E *et al.* (2012b) Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for *metabarcoding* studies. *Molecular Ecology*, **21**, 1816–1820.
- Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E (2012c) Towards next-generation biodiversity assessment using DNA *metabarcoding*. *Molecular Ecology*, **21**, 2045–2050.
- Tinner W, Lotter AF, Ammann B *et al.* (2003) Climatic change and contemporaneous land-use phases north and south of the Alps 2300 BC to 800 AD. *Quaternary Science Reviews*, **22**, 1447–1460.
- Valentini A, Miquel C, Nawaz, MA *et al.* (2009) New perspectives in diet analysis based

- on DNA *barcoding* and parallel pyrosequencing: the *trnL* approach. *Molecular Ecology Resources*, **9**, 51-60.
- von Grafenstein U, Erlenkeuser H, Brauer A, Jouzel J, Johnsen SJ (1999) A mid-European decadal isotope-climate record from 15,500 to 5000 years BP. *Science*, **284**, 1654-1657.
- Walsh K, Mocci F, Palet-Martinez J (2007) Nine thousand years of human/landscape dynamics in a high altitude zone in the southern French Alps (Parc National des Ecrins, Hautes-Alpes). *Preistoria Alpina*, **42**, 9-22.
- Wick L, Tinner W (1997) Vegetation changes and timberline fluctuations in the central alps as indicators of Holocene climatic oscillations. *Arctic and Alpine Research*, **29**, 445-458.
- Wick L, van Leeuwen JFN, van der Knaap WO, Lotter AF (2003) Holocene vegetation development in the catchment of Sägistalsee (1935 m asl), a small lake in the Swiss Alps. *Journal of Paleolimnology*, **30**, 261-272.
- Wilkinson BH (2005) Humans as geologic agents: A deep-time perspective. *Geology*, **33**, 161-164.
- Willerslev E, Hansen AJ, Binladen J *et al.* (2003) Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science*, **300**, 791-795.
- Willerslev E, Davison J, Moora M *et al.* (2014) Fifty thousand years of Arctic vegetation and megafaunal diet. *Nature*, **506**, 47-51.
- Wilmshurst JM, McGlone MS (2005) Corroded pollen and spores as indicators of changing lake sediment sources and catchment disturbance. *Journal of Paleolimnology*, **34**, 503-517.
- Zhou X, Li Y, Liu S *et al.* (2013) Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. *GigaScience*, **2**, 4.

b. Données complémentaires

Supplementary material

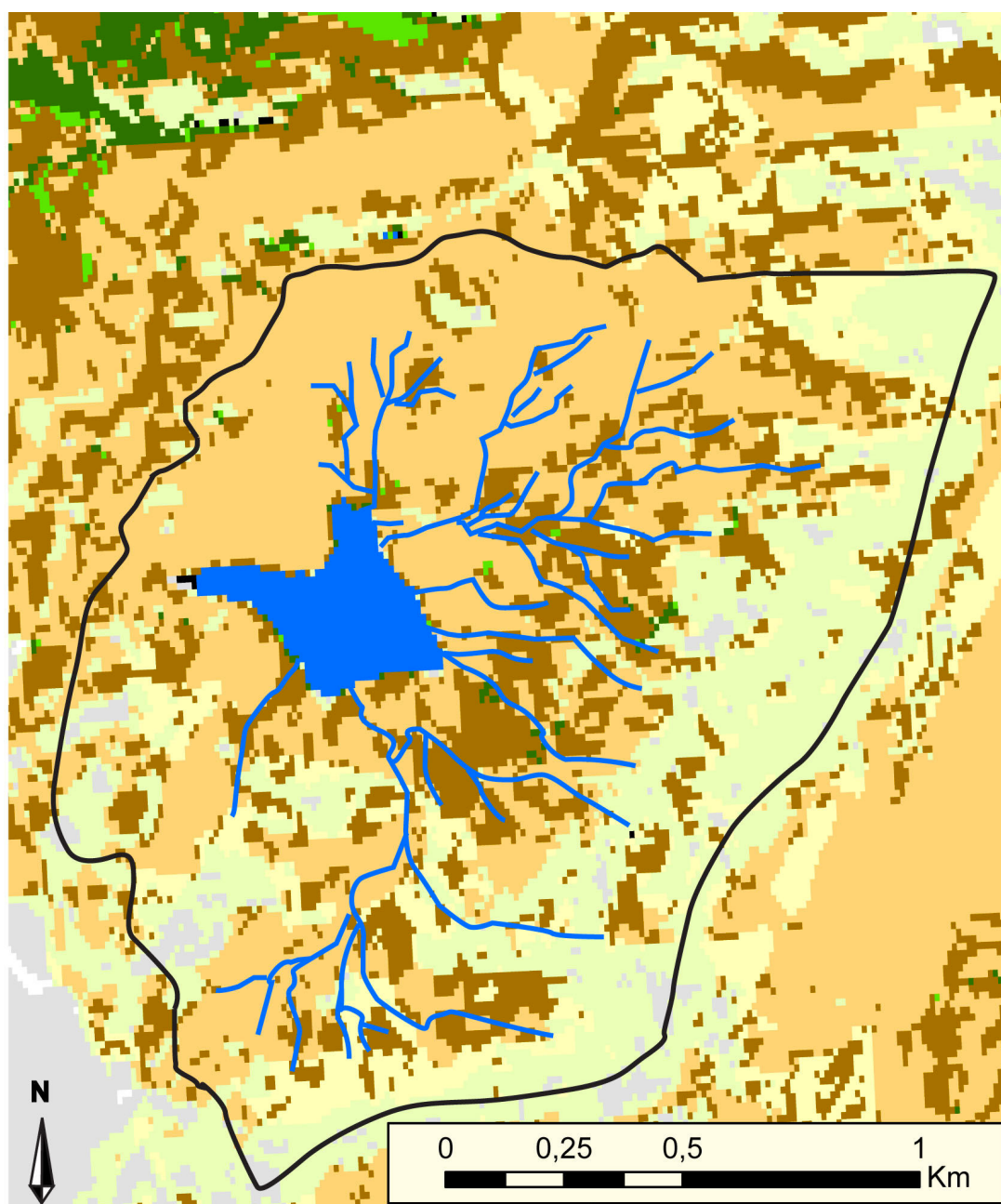
Figure S1: Land cover of the Lake Anterne catchment

Figure S2: Geochemical analysis from sediment core

Figure S3: DNA diagrams of filtered sequences over the Holocene

Figure S4: Ordination of taxa detected by *metabarcoding* according to their current co-occurrence pattern

Table S1: Detection probability and false presence probability of filtered sequences

Figure S1: Land cover of the Lake Anterne catchment**Legend**

	MOORS		ALPINE MEADOWS
	DECIDUOUS WOOD		SUBALPINES MEADOWS
	CONIFEROUS		BARE SOILS WITHOUT VEGETATION
	WATER		BARE SOILS WITH VEGETATION

Figure S2: Geochemical analysis from sediment core (from Giguët-Covex *et al.* 2011).

CaO (calcium) and SiO₂ (silicium) are expressed as oxide weight percentage (Giguët-Covex *et al.* 2011). CaO is used in this study as a proxy for past soil acidity. Hydrogen index is the amount of hydrocarbon-rich products released during the pyrolysis phase, normalized to total organic carbon content (TOC) and is expressed as mg HC/g TOC. Oxygen index (mg O₂/g TOC) is a measure of the oxygen content of the organic matter. These index are usually used as indicators of organic matter quality and reflects the origin of organic matter and its preservation state (Giguët-Covex *et al.* 2011).

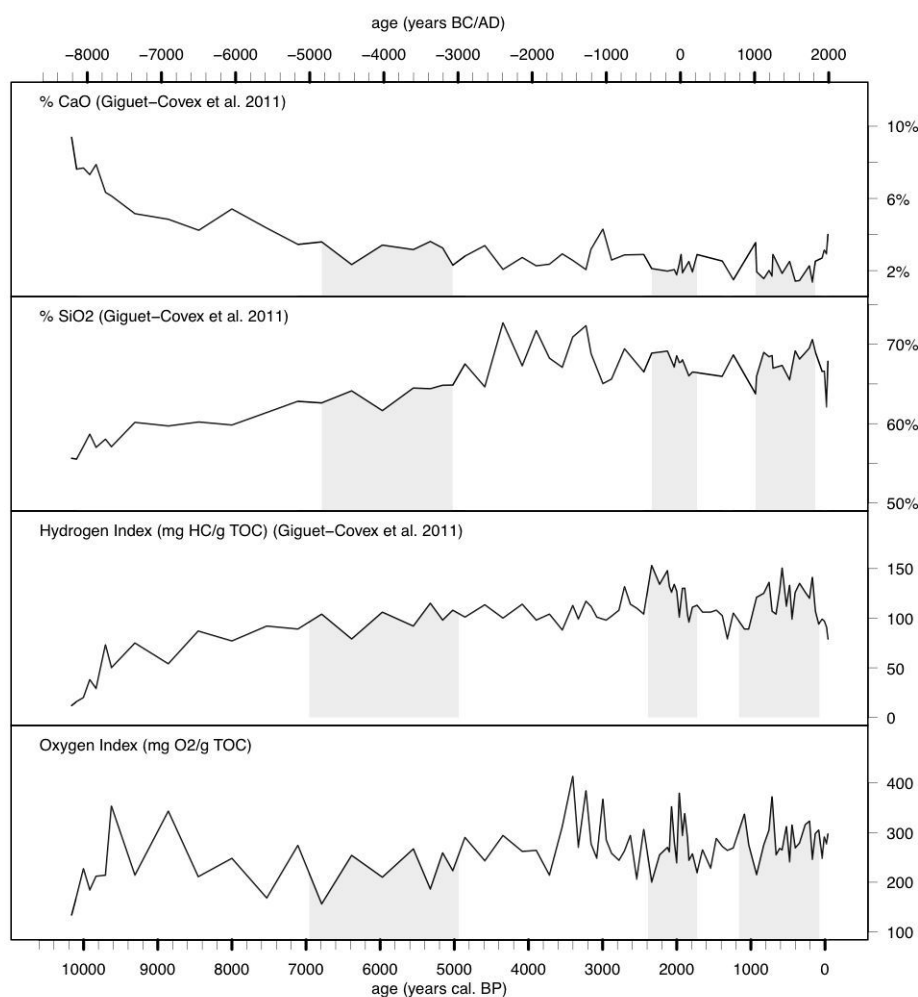
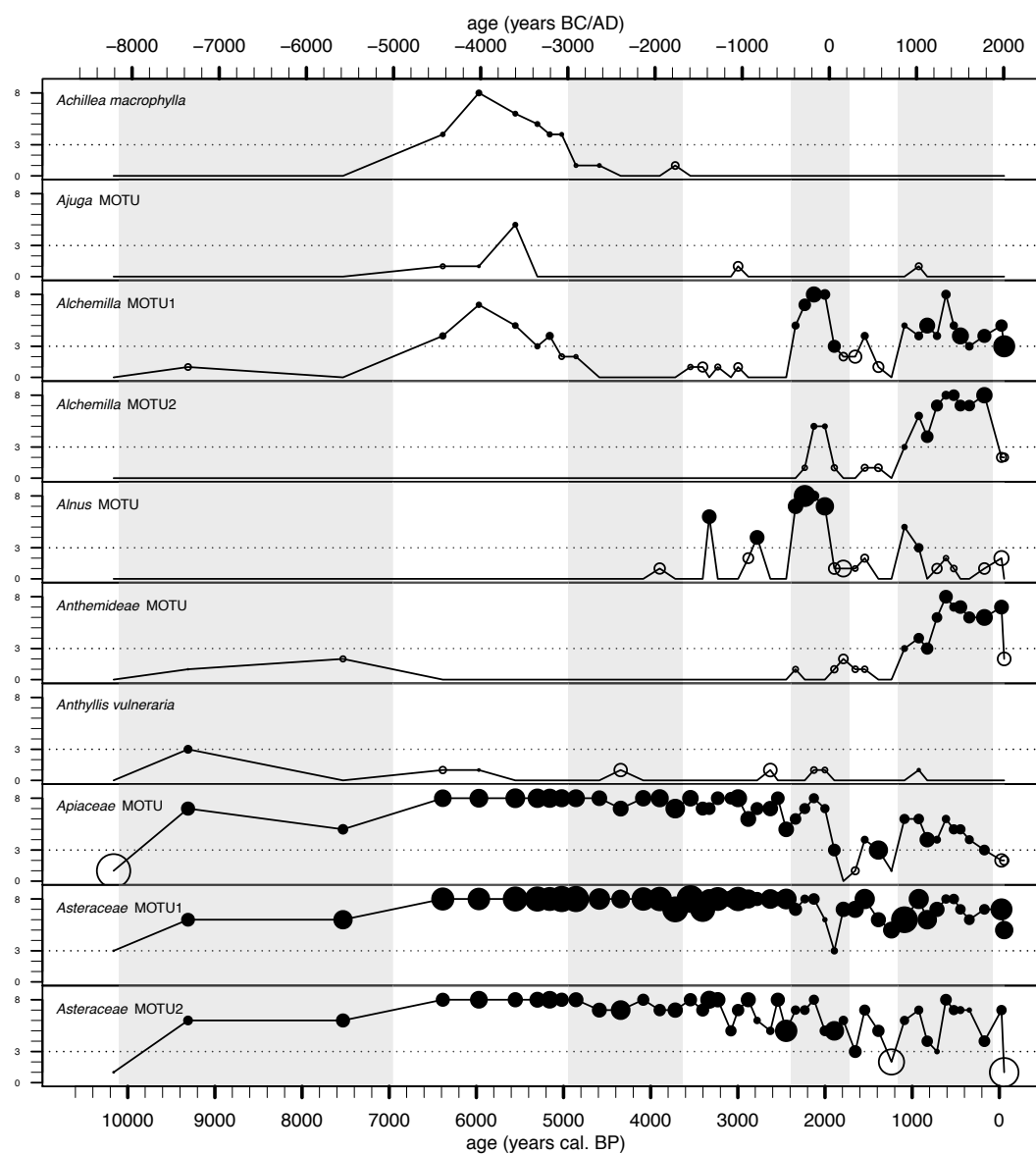
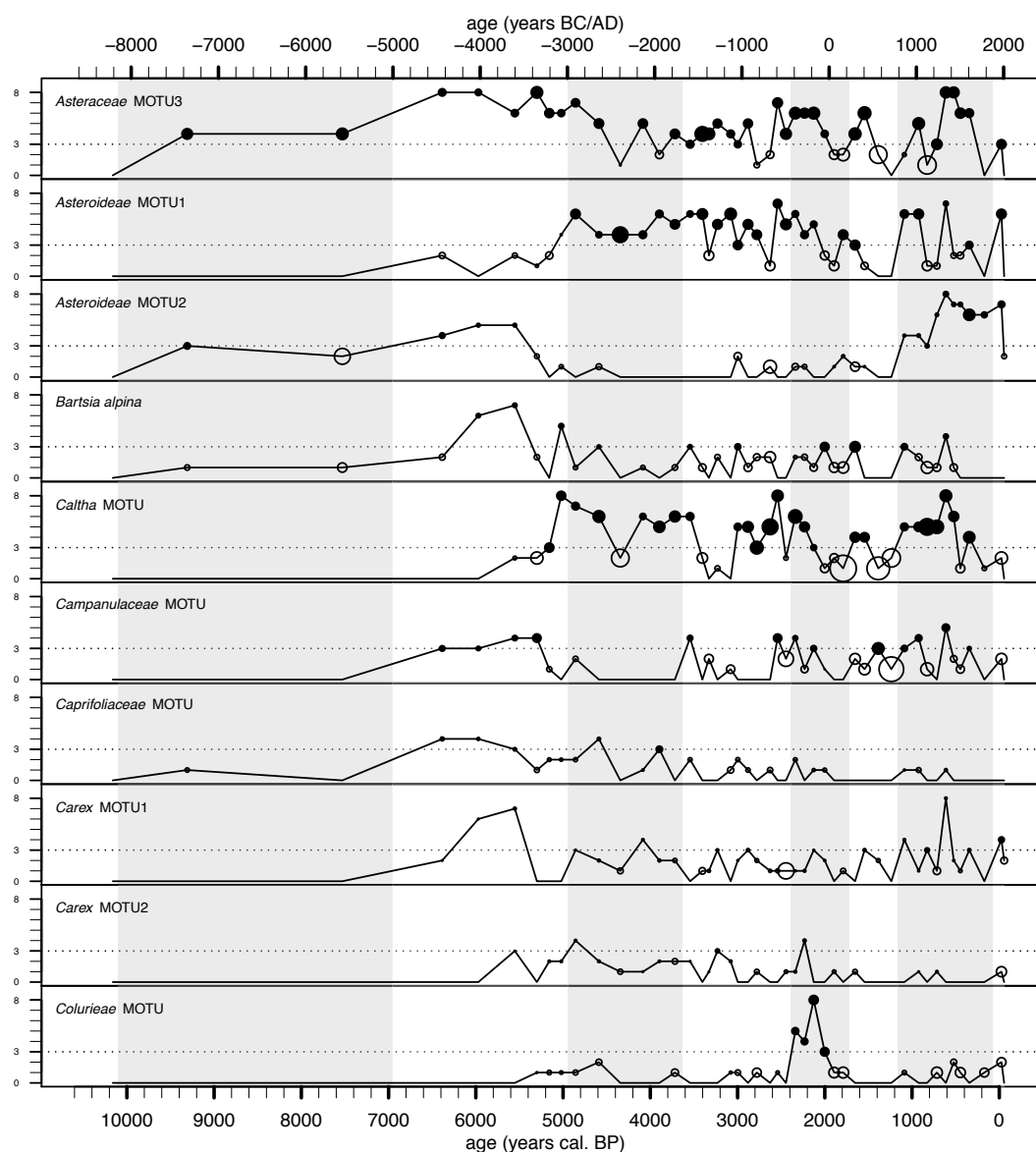


Figure S3: DNA diagrams of filtered sequences over the Holocene

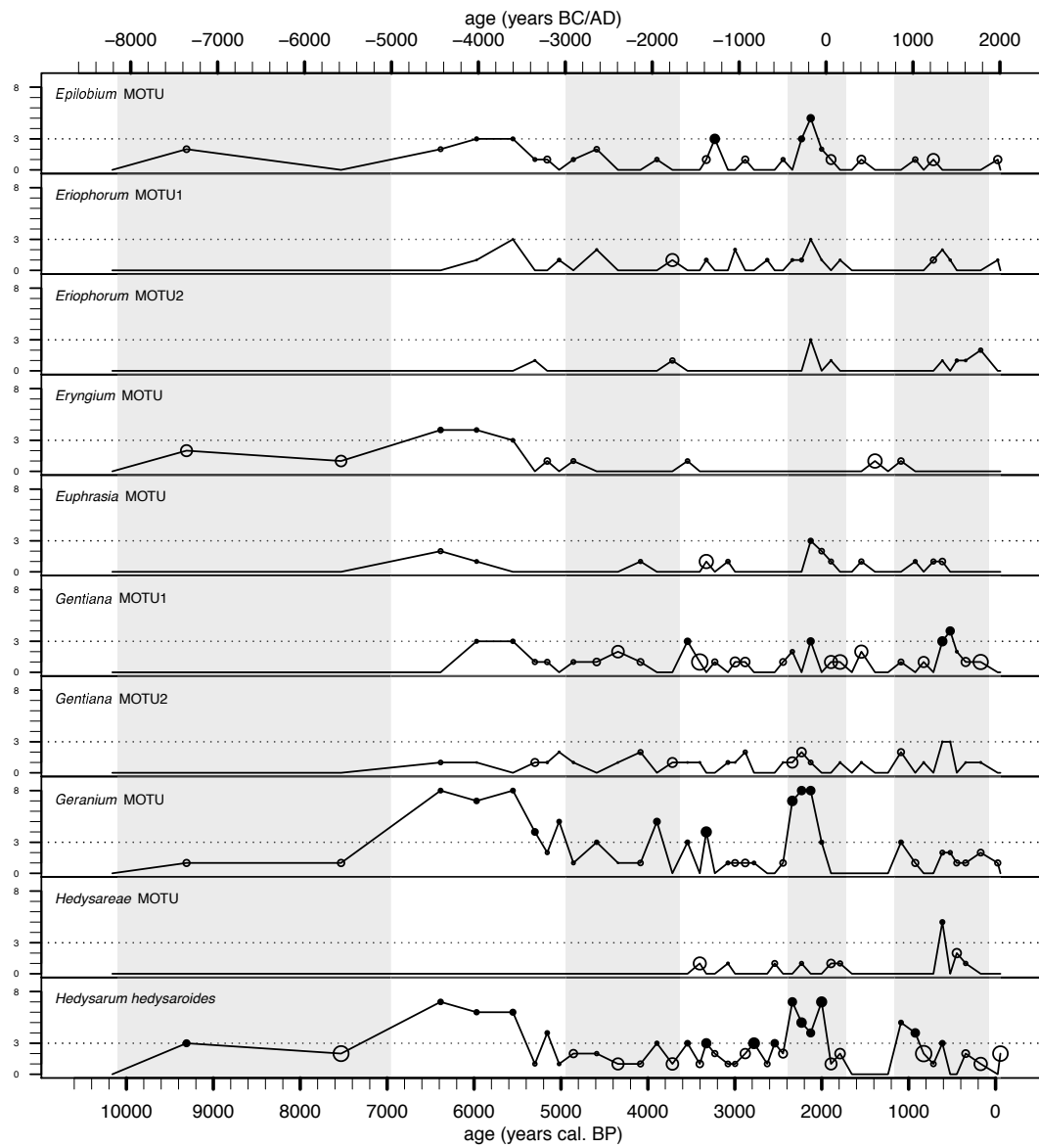
Full circles: sequences found in at least 3 PCR replicates. Open circles: sequences found in less than 3 PCR replicates. The diameter of circles is proportional to the mean relative abundance of sequences of the taxa within the sequences from a given sample.

a.

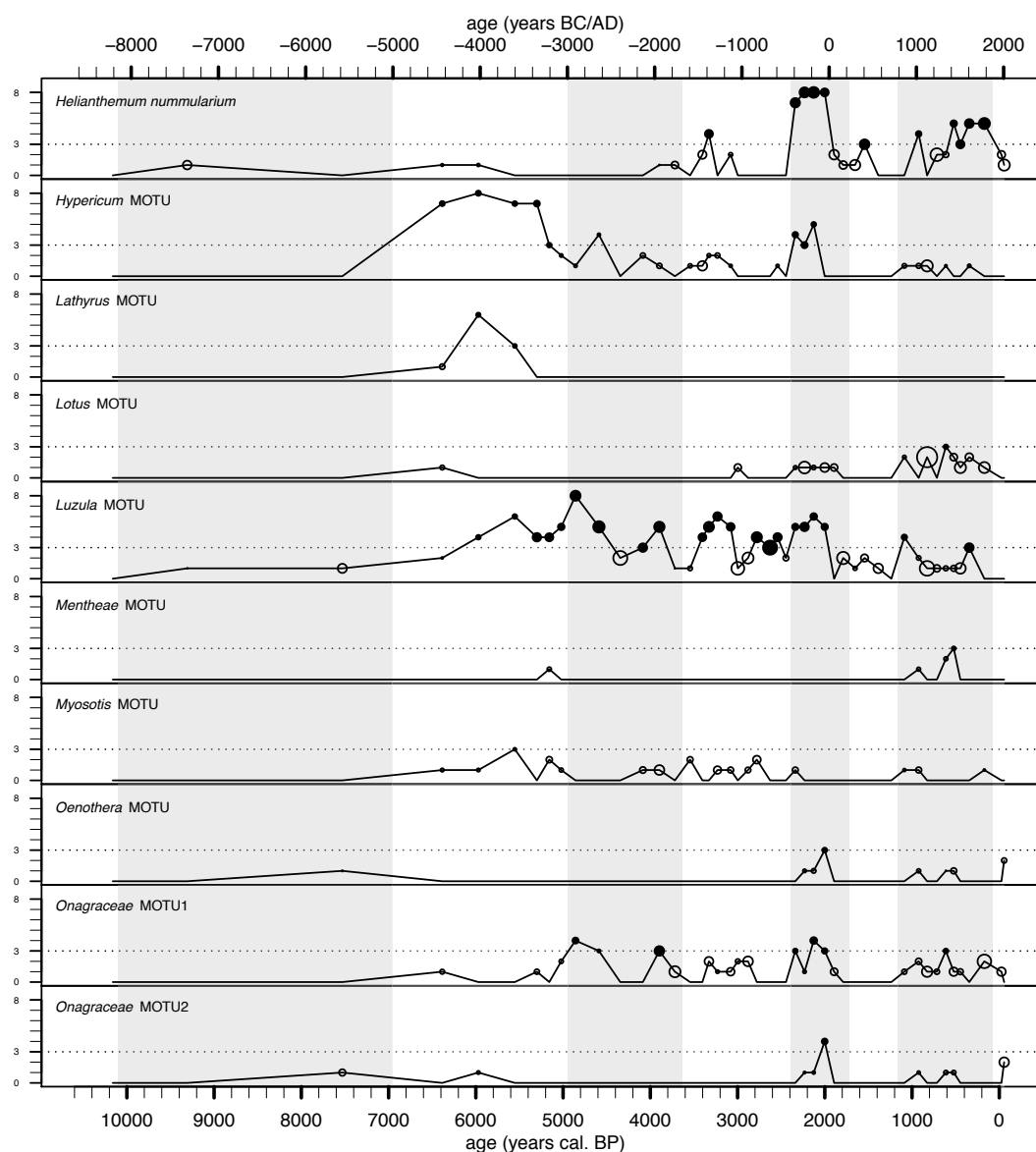
b.



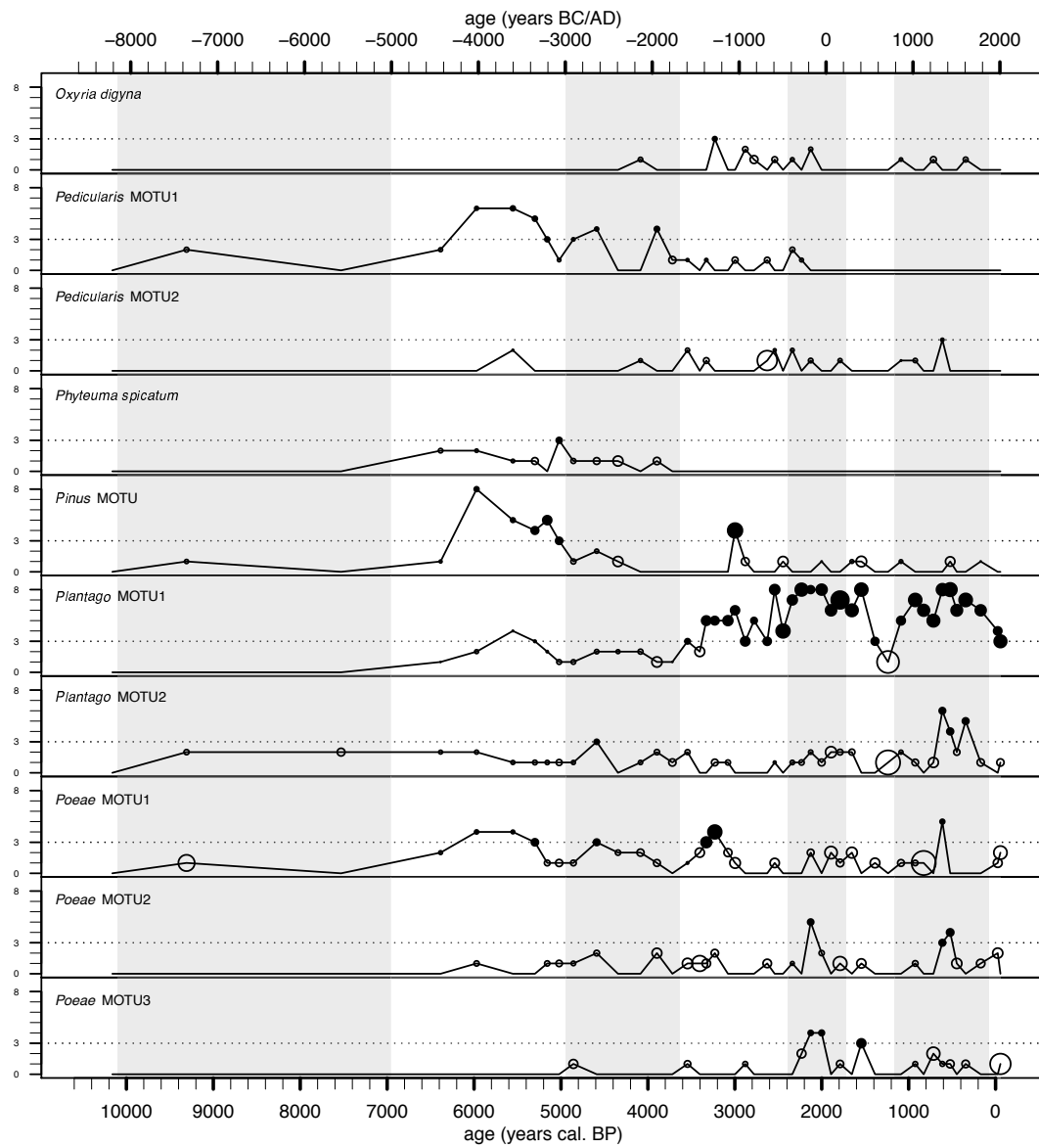
C.



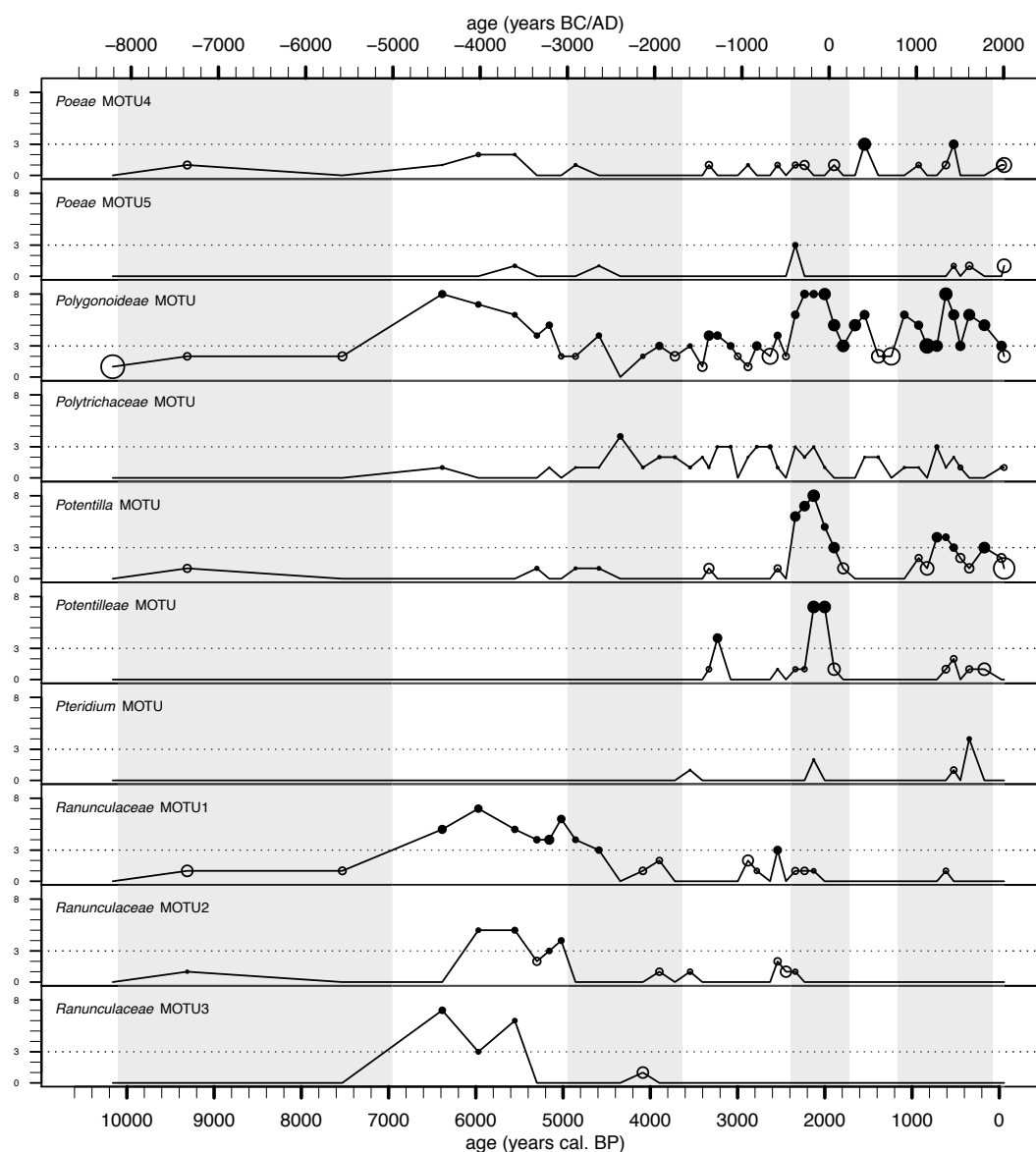
d.



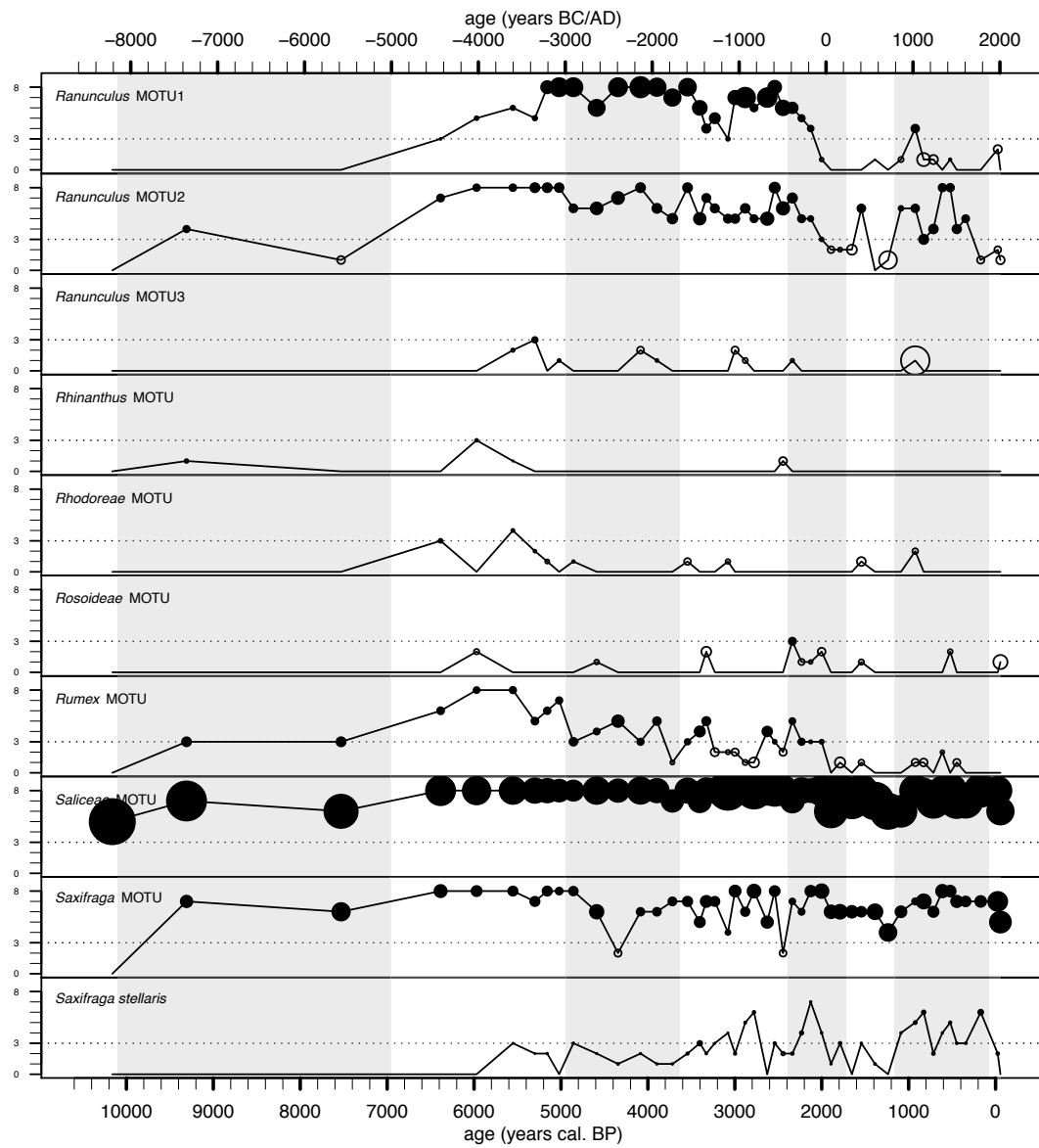
e.



f.



g.



h.

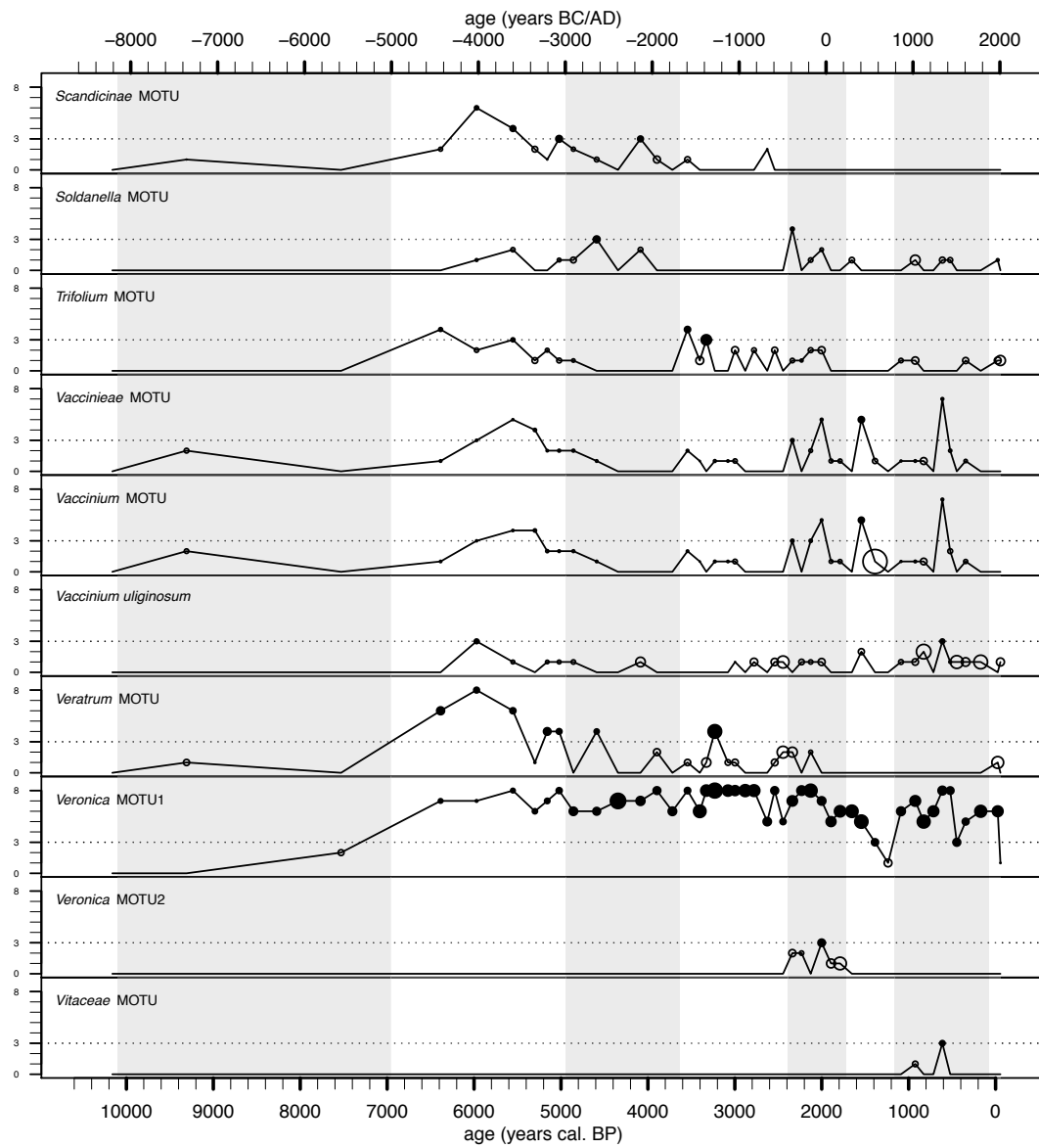
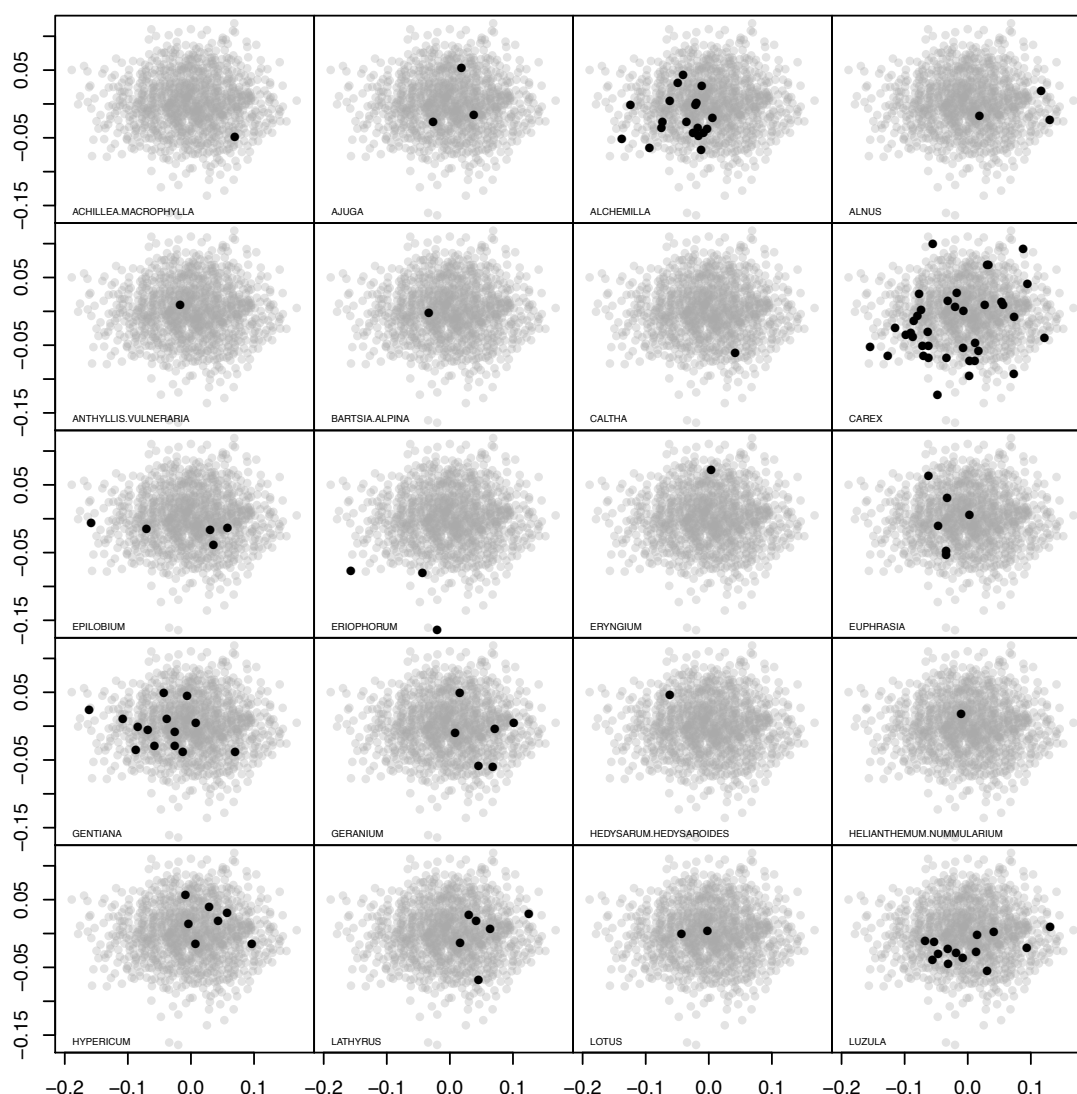


Figure S4: Ordination of taxa detected by *metabarcoding* according to their current co-occurrence pattern

A multivariate analysis (non-metric multidimensional scaling) was performed based on 3077 vegetation relevés conducted in the Northern part of the French Alps in order to ordinate 982 species according to their current co-occurrence pattern. A vector fitting approach was used to relate the ordination of current species with their abiotic preferences, measured with the Ellenberg indicator values from 223 alpine plants. Diagnostic taxa detected by *metabarcoding* analysis were replaced on this ordination to infer their ecological characteristics. Ecological gradients represented by the vector fitting analysis are shown in panel b.

a.



b.

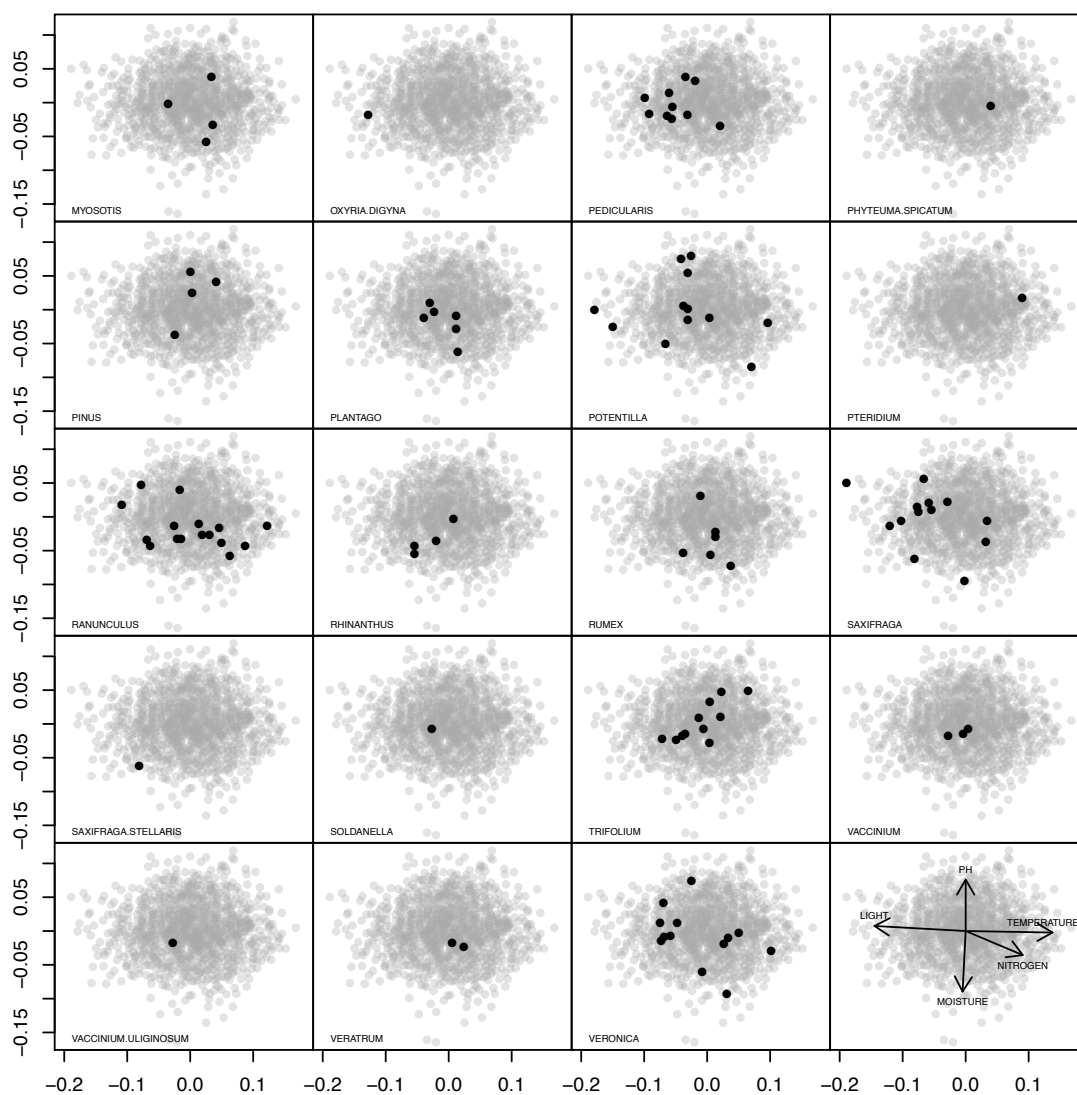


Table S1: Detection probability and false presence probability of filtered sequences

Four DNA extractions have been performed per each core slice, and two PCR amplifications have been performed on each DNA extract (i.e eight PCR replicates per sample). We used occupancy models to evaluate the detection probability of taxa and to assess the probability of false positives (Miller *et al.* 2011). We assumed that taxon presence was particularly uncertain if the taxon was detected in only one out of eight PCR replicates. Miller's *et al.* (2011) approach was used to estimate: *i*) the detection probability and *ii*) the probability of false detection of each taxon, per each PCR replicate. Simulation analyses showed this approach can accurately estimate both parameters in *metabarcoding* studies, if at least eight PCR replicates are available. On the basis of the probability of false detection per each PCR replicate, we also calculated the probability of falsely stating that a taxon is present, if detected in at least three PCR replicates from the same sample. This analysis was limited to MOTUs identified at least up to the genus level that were detected with certitude in more than one sediment core slice.

Column heading:

DP: Detection probability *e.g.* probability of detection of the species in one PCR, if it is present

FP: Probability of false detection per each PCR

FP_3: Probability of obtaining 3 false positives of the species from the same sample

Identified MOTU	DP	FP	FP_3	Identified MOTU	DP	FP	FP_3
<i>Achillea macrophylla</i>	0.645	0.010	0.000	<i>Luzula</i> MOTU	0.513	0.095	0.003
<i>Ajuga</i> MOTU	0.174	0.000	0.000	<i>Myosotis</i> MOTU	0.252	0.038	0.003
<i>Alchemilla</i> MOTU1	0.588	0.038	0.003	<i>Oxyria digyna</i>	0.107	0.000	0.000
<i>Alchemilla</i> MOTU2	0.742	0.016	0.000	<i>Pedicularis</i> MOTU1	0.470	0.025	0.001
<i>Alnus</i> MOTU	0.601	0.027	0.001	<i>Pedicularis</i> MOTU2	0.470	0.025	0.001
<i>Anthyllis vulneraria</i>	0.064	0.000	0.000	<i>Phyteuma spicatum</i>	0.104	0.000	0.000
<i>Bartsia alpina</i>	0.391	0.062	0.010	<i>Pinus</i> MOTU	0.552	0.040	0.003
<i>Caltha</i> MOTU	0.590	0.059	0.009	<i>Plantago</i> MOTU1	0.659	0.096	0.034
<i>Carex</i> MOTU1	0.410	0.069	0.014	<i>Plantago</i> MOTU2	0.345	0.078	0.019
<i>Carex</i> MOTU2	0.302	0.042	0.004	<i>Potentilla</i> MOTU	0.537	0.038	0.003
<i>Epilobium</i> MOTU	0.337	0.012	0.000	<i>Ranunculus</i> MOTU1	0.760	0.045	0.004
<i>Eriophorum</i> MOTU1	0.351	0.051	0.006	<i>Ranunculus</i> MOTU2	0.737	0.106	0.044
<i>Eryngium</i> MOTU	0.412	0.019	0.000	<i>Rumex</i> MOTU	0.513	0.057	0.008
<i>Euphrasia</i> MOTU	0.265	0.028	0.001	<i>Saxifraga</i> MOTU	0.860	0.000	0.000
<i>Gentiana</i> MOTU1	0.331	0.059	0.009	<i>Saxifraga stellaris</i>	0.460	0.050	0.006
<i>Gentiana</i> MOTU2	0.273	0.064	0.012	<i>Soldanella</i> MOTU	0.314	0.029	0.001
<i>Geranium</i> MOTU	0.595	0.066	0.012	<i>Trifolium</i> MOTU	0.298	0.042	0.004
<i>Hedysarum hedysaroides</i>	0.459	0.083	0.023	<i>Vaccinium</i> MOTU	0.466	0.082	0.022
<i>Helianthemum nummularium</i>	0.557	0.036	0.002	<i>Vaccinium uliginosum</i>	0.292	0.059	0.009
<i>Hypericum</i> MOTU	0.542	0.044	0.004	<i>Veratrum</i> MOTU	0.502	0.030	0.001
<i>Lathyrus</i> MOTU	0.409	0.000	0.000	<i>Veronica</i> MOTU1	0.848	0.083	0.024
<i>Lotus</i> MOTU	0.270	0.029	0.001				

4. Discussion chapitre IV

a. Effets à long terme des activités anthropiques en milieu alpin

Plusieurs travaux récents ont souligné l'importance des études de long terme pour comprendre les réponses des écosystèmes aux changements (*e.g.* Anderson *et al.* 2006 ; Dietl & Flessa 2011 ; Boessenkool *et al.* 2014). A court terme, les écosystèmes peuvent montrer une certaine résilience aux perturbations ; les études paléo-environnementales, elles, mettent en lumière le fait que les trajectoires à long terme des écosystèmes sont liées principalement à des changements dans les régimes de perturbations (McLauchlan *et al.* 2014). Les activités humaines engendrent ainsi des perturbations chroniques qui vont modifier profondément, et souvent irrémédiablement, les communautés biologiques et leur environnement (Bengtsson 2002).

Ces deux études mettent en évidence l'impact ancien et durable qu'ont exercé les activités humaines sur les communautés biologiques mais également sur leur environnement. Les communautés végétales alpines au cours de l'Holocène ont été dirigées, non pas uniquement par le climat, mais également par les activités humaines (Gobet *et al.* 2003). Les changements de végétation observés dans l'article D résultent ainsi de l'impact direct de l'herbivorie sur la communauté mais aussi de l'ouverture du milieu sous l'effet probable de coupes et des feux. L'analyse des charbons retrouvés dans les sédiments pourrait se révéler un outil précieux pour la reconstitution de l'histoire des feux dans le bassin versant et des changements de végétation qu'ils ont pu induire (Marlon *et al.* 2012 ; Rey *et al.* 2013). Dans le même temps, l'introduction de troupeaux a également fortement impacté les sols en augmentant les flux érosifs, ce qui a engendré un lessivage des sols et une modification de leurs caractéristiques physico-chimiques (Etude C). Or, le pastoralisme et l'acidité des sols apparaissent significativement corrélés avec les trajectoires à long terme des communautés végétales. Ces études montrent, une fois de plus, l'effet des activités humaines, *via* l'altération des sols, sur les communautés biologiques. De résultats similaires ont par ailleurs été mis en évidence concernant la fréquence des feux et la composition des communautés végétales autour de la limite supérieure des arbres (Colombaroli *et al.* 2010).

Nos résultats ne remettent pas en cause l'importance du climat sur les trajectoires de végétation, d'autant plus que la reconstruction paléo-climatique utilisée dans l'article D n'est pas idéalement adaptée au site d'étude. Les effets du climat se

reflètent néanmoins plus sûrement sur les aires de répartition des espèces et leurs variations latitudinales et/ou altitudinales (*e.g.* Giesecke *et al.* 2005 ; Tinner *et al.* 2005 ; Miller *et al.* 2008 ; Harsch *et al.* 2009). A l'échelle locale, ces effets existent ; l'expansion altitudinale du pin cembro (*Pinus cembra*) est ainsi favorisée par des températures moyennes en juillet supérieures à celles observées actuellement (Colombaroli *et al.* 2010). Néanmoins, ces effets sont souvent surpassés par les perturbations d'origine anthropique.

L'intensification des pratiques a mené à des changements drastiques de végétation et à une érosion accrue, particulièrement à la période romaine. Néanmoins, localement, les perturbations induites par le pastoralisme ont engendré une mosaïque de sols et d'habitats. L'ouverture du milieu forestier a ainsi été à l'origine d'une diversification des habitats (Fischer *et al.* 2008), les perturbations d'origine anthropique, notamment la présence de troupeaux, ayant par la suite assuré le maintien d'une matrice paysagère diversifiée (Young *et al.* 2005). Le déclin actuel des activités traditionnelles de montagne (*e.g.* brulis, pâturage, fauche) peut amener de grands changements dans les communautés végétales alpines (Colombaroli *et al.* 2010). Leur disparition se traduit notamment par une recolonisation de la forêt et des buissons (Galop *et al.* 2011). On manque ici de recul pour reconstituer le recouvrement des différentes formations sur le bassin versant à partir des données issues de l'ADN sédimentaire ancien (cf. partie V.1.d.) ; cependant plusieurs études palynologiques ont montré une diminution de la richesse végétale en lien avec la baisse de charge pastorale en milieu de montagne (Valsecchi *et al.* 2010 ; Galop *et al.* 2011). Ces résultats mettent en évidence l'importance des activités humaines passées dans la formation des paysages observés aujourd'hui en milieu de montagne, et ce même dans les zones relativement isolées. Par là même, ils soulèvent les difficultés rencontrées pour définir des états de référence en biologie de la conservation (Seddon *et al.* 2014) puisque ces milieux, considérés aujourd'hui comme très faiblement impactés par l'Homme, dérivent en réalité de plusieurs millénaires d'activités anthropiques.

b. Le metabarcoding : un outil complémentaire pour l'étude des paléoenvironnements...

Ces travaux démontrent que le *metabarcoding* environnemental peut constituer une approche complémentaire aux outils classiques pour étudier les communautés biologiques passées, et leurs trajectoires, sous l'évolution des régimes de perturbations

naturelles et anthropiques (Jørgensen 2012a). Ensemble, ces deux études illustrent la façon dont cette approche permet de retracer, en une même expérimentation, à la fois les communautés passées et une partie des facteurs biotiques guidant leur évolution. En plus de fournir des informations sur les perturbations biotiques, elle offre un accès à l'ensemble de la diversité passée. Les pollens sont les fossiles les plus aisément retrouvés dans les archives naturelles, ce qui explique que les études paléoécologiques aient été majoritairement concentrées sur la végétation. L'étude des communautés animales passées est, elle, plus limitée car elle passe par l'analyse de macrorestes, plus rares, et avec un niveau d'identification taxonomique très variable. L'apport de l'outil moléculaire pour la reconstruction des communautés animales, et *a fortiori* pour la reconstruction des perturbations biotiques induites par l'Homme, est indéniable. L'ADN environnemental a ainsi permis de détecter des taxons, dans notre cas vaches et moutons, en l'absence de fossiles visibles. Cette approche offre aujourd'hui l'opportunité d'étudier plus intensément les communautés autres que végétales, et leur évolution, sur le long terme. Les travaux présentés ici ont montré l'importance des activités humaines sur les sols et leurs biodiversité. A l'avenir, la recherche de marqueurs caractéristiques de la faune des sols anciens (*e.g.* nématodes, Willerslev *et al.* 2014) pourrait s'avérer un élément d'importance pour la compréhension de l'histoire des sols sous l'effet conjugué des activités humaines et des forçages climatiques. On peut, par exemple, imaginer étudier les trajectoires à long terme des communautés des sols.

En ce qui concerne les communautés végétales, les données recueillies par *metabarcoding* dans l'article C ne couvrent pas l'ensemble de la flore passée pour chaque période donnée. L'identification au genre ou à l'espèce est, par exemple, souvent moins fréquente que ce qui peut être obtenu typiquement avec des pollens (Parducci *et al.* 2013). Cependant, les résultats sont en accord avec les connaissances acquises depuis plusieurs décennies par les approches paléo-environnementales classiques. Dans l'article D, nous avons montré qu'il était possible d'appliquer des approches écologiques, par couplage avec des données botaniques actuelles, pour retracer les caractéristiques environnementales des milieux passés.

Plusieurs études récentes ont notamment mis en évidence les caractéristiques qui différencient le *metabarcoding* environnemental des approches traditionnelles pour la reconstruction des paléo-communautés végétales. Tout d'abord, les signaux inférés par les deux approches reflètent différentes échelles spatiales : alors que les données

palynologiques fournissent un signal régional, l'ADN retrouvé dans les archives sédimentaires lacustres, à l'instar des macrorestes, a une origine plutôt locale. Cela explique en partie les différences de composition observées dans les études visant à comparer ces méthodes (Jørgensen *et al.* 2012a; Parducci *et al.* 2013; Boessenkool *et al.* 2014). Ces différentes approches peuvent procurer des niveaux d'identification très variables en fonction des taxons et, aussi bien l'une que l'autre, peuvent être biaisées envers ou contre certaines espèces. L'outil moléculaire présente l'avantage d'être totalement indépendant du niveau d'expertise, ce qui limite les biais liés à l'expérimentateur (Hofreiter *et al.* 2012). Néanmoins, cette approche est complètement dépendante de l'exhaustivité des bases de référence (cf. partie V.2.). Compte tenu des faibles quantités d'ADN initial, de la compréhension limitée des processus de transfert de l'ADN vers le lac, de sa conservation et des biais inhérents au processus de PCR, ces données ne sont pas quantitatives (Parducci *et al.* 2013). Par exemple, le lien entre charge pastorale et quantité d'ADN est impossible à établir en l'état actuel de nos connaissances (cf. parties IV.4.c. et V.1.d.). A l'inverse, la longue histoire de la palynologie a notamment permis de développer des approches quantitatives prenant en compte les biais liés aux différences de productivité et de dispersion des pollens ou encore à la taille des bassins versants (*e.g.* Gaillard *et al.* 2008 ; Soepboer *et al.* 2010). L'ADN sédimentaire lacustre constitue ainsi un nouvel outil pour reconstituer l'histoire des environnements à l'échelle locale, néanmoins la reconstruction précise des paléo-communautés végétales passe par le couplage de l'ensemble de ces approches.

c. ... Mais un outil encore développement

De nombreux progrès restent à faire dans ce domaine. Si l'extraction de l'ADN ou encore le filtrage des données et l'analyse des séquences (cf. chapitre V) sont encore perfectibles, des études complémentaires concernant notamment la taphonomie (*i.e.* l'étude des processus de fossilisation, Andersen *et al.* 2012), l'origine de l'ADN sédimentaire et sa préservation doivent être menées. L'origine de l'ADN retrouvé dans les sédiments est encore incertaine. Si la source locale de l'ADN animal ne fait aucun doute, les modalités de son transfert dans le lac sont encore peu connues. L'utilisation de champignons coprophiles comme marqueurs d'activités pastorales (*e.g.* Etienne *et al.* 2013) indique cependant que les fèces sont un vecteur important. Comme nous l'avons vu dans l'introduction de cette partie, l'origine de l'ADN végétal est encore sujette à

débat. Dans l'article D, la détection quasi-constante de l'ADN de plantes de milieux humides (*Cattha* sp. et Juncaceae) depuis le milieu de l'Holocène suggère l'importance du réseau hydrographique du bassin versant dans le transfert de l'ADN vers les sédiments. Néanmoins, les données présentées dans cette étude montrent que l'ADN provient également des zones de prairies et de landes couvrant les pentes du bassin versant.

Les conditions environnementales sont, bien entendu, déterminantes pour la préservation de l'ADN ancien, cependant on en sait encore aujourd'hui peu sur les processus de dégradation de l'ADN dans les sédiments (Parducci *et al.* 2013). Les études réalisées sur l'ADN provenant d'os ou de fossiles ont mis en évidence l'importance de la dégradation post-mortem de l'ADN pour les études basées sur l'ADN ancien (*e.g.* Lindhal *et al.* 1993 ; Hofreiter *et al.* 2001 ; Willerslev & Cooper 2005 ; Briggs *et al.* 2007). Dans le cas présent, les deux études ont la particularité de se concentrer uniquement sur l'ADN extracellulaire. Les processus de dégradation affectant l'ADN extracellulaire contenu dans les sédiments lacustres, ainsi que les facteurs qui en sont à l'origine, sont encore relativement peu compris (Corinaldesi *et al.* 2008). Néanmoins, les études menées sur des sédiments marins suggèrent que la cinétique de dépurination de l'ADN extracellulaire dans la matrice sédimentaire est largement inférieure à celle estimée pour les écosystèmes terrestres (Corinaldesi *et al.* 2008). De plus, l'adsorption de l'ADN extracellulaire sur les particules minérales et organiques est supposée réduire fortement sa disponibilité aux enzymes de dégradation (DNase) (Coolen & Overmann 2007) alors même que l'activité de ces dernières est réduite par leur propre adsorption sur les particules sédimentaires (Demanèche *et al.* 2001). Les résultats obtenus ici sont en accord avec la littérature sur l'évolution des paysages alpins et les études préalables menées sur le lac d'Anterne. Aussi la fiabilité des résultats suggèrent-elles une préservation de qualité de l'ADN extracellulaire, une fois celui-ci fixé sur les particules sédimentaires (Pietramellara *et al.* 2009).

Comme nous l'avons vu dans la dernière étude, l'utilisation de l'ADN sédimentaire ancien est actuellement limitée par la résolution taxonomique des marqueurs ADN. Ce point est un élément crucial pour accroître leur valeur et l'utilisation des données paléo-environnementales en conservation (Willis & Birks 2006). De plus, les *metabarcodes* utilisés, dits universels, peuvent présenter des biais d'amplification liés au fait que les sites de fixation des amorces ne sont pas complètement conservés entre espèces. Travailler avec de faibles quantités d'ADN dégradé, comme c'est le cas ici,

peut conduire à passer à côté de certains taxons. Parducci *et al.* (2013) ont ainsi échoué à retrouver des séquences d'arbres (*Pinus sylvestris* et *Picea abies*) dans un lac suédois avec le marqueur *gh* (Taberlet *et al.* 2007) alors que ceux-ci avaient auparavant été détectés à l'aide d'amorces spécifiques de ces espèces (Parducci *et al.* 2012).

Ces deux problèmes majeurs sont cependant perfectibles. La taille des marqueurs utilisés (<100 pb) est telle qu'une infime fraction de l'information contenue dans les archives naturelles a été exploitée ici. Le défi est, aujourd'hui, de parvenir à extraire le maximum d'information taxonomique de ces fragments d'ADN. Une première approche serait de combiner des marqueurs ADN spécifiques de groupes taxonomiques plus restreints, ciblant par exemple une famille en particulier (Parducci *et al.* 2013). Néanmoins, son application en routine est limitée par les coûts et ne permet l'accès qu'à une petite partie de l'information disponible. L'alternative suggérée dans l'article D repose sur le séquençage direct en *shotgun* (*e.g.* sans PCR préalable) de l'ADN extrait (Taberlet *et al.* 2012c ; Ji *et al.* 2013 ; Zhou *et al.* 2013). Là encore, son application doit faire face à plusieurs défis majeurs (cf. partie V.2.).

Chapitre V. Considérations méthodologiques pour une meilleure évaluation de l'impact des activités anthropiques sur la biodiversité par *metabarcoding* environnemental

1. Discussion autour de la méthode

Ces travaux ont contribué à montrer les potentialités du *metabarcoding* environnemental en tant que nouvel outil en écologie des perturbations (Valentini *et al.* 2009a ; Bik *et al.* 2012). Néanmoins, son utilisation en routine, notamment par les gestionnaires (Kermarrec *et al.* 2014), est encore loin d'être chose aisée. La partie suivante dresse ainsi un inventaire (non-exhaustif) des éléments à considérer afin d'évaluer au mieux l'effet des activités humaines sur la biodiversité.

a. L'échantillonnage : une étape cruciale et adaptable à de nombreuses investigations

La prise en compte de l'hétérogénéité spatiale est un élément primordial de l'échantillonnage. En ce qui concerne les sédiments lacustres, des disparités locales, liées aux différences d'apports d'éléments provenant du bassin versant, peuvent exister au sein d'un même lac (Etienne *et al.* 2013). La multiplication des carottes associée à un choix rigoureux des sites est ainsi souhaitable afin d'intégrer au mieux cette hétérogénéité.

Dans les études basées sur l'analyse de l'ADN du sol présentées dans le chapitre III, 100 prélèvements de sol uniformément répartis sur la zone d'étude ont été récoltés puis mélangés afin d'obtenir un échantillon composite. Cette méthodologie s'est révélée pertinente pour obtenir des profils cohérents de diversité et caractériser les patrons de beta-diversité des communautés du sol, et les facteurs qui en sont à l'origine, aussi bien sur 100m² que sur 1ha. Elle présente également l'avantage de minimiser les coûts, notamment pour une application à grande échelle. Néanmoins, cette pratique peut poser problème dans le cas d'inventaire exhaustif de biodiversité, particulièrement en milieu riche. En effet, à chaque étape du processus d'analyse (*e.g.* extraction, amplification et séquençage), seule une fraction du produit de l'étape précédente est utilisée ; ce qui amène au final à n'obtenir, puis analyser, qu'un sous-échantillon de la diversité initiale du produit récolté (Zinger *et al.* 2014). Ainsi, en mélangeant les prélèvements de sols

avant extraction, la diversité de l'échantillon d'ADN initial est augmentée mais la probabilité de détection des espèces rares va, elle, diminuer. Imaginons le cas d'une espèce avec une distribution très localisée sur la parcelle : en mélangeant les échantillons avant l'étape d'extraction, son ADN va se retrouver bien minoritaire comparé à celui d'une espèce répartie de façon plus homogène sur la zone d'étude. Cette « dilution » va, de fait, diminuer sa probabilité de détection.

Une solution serait alors d'extraire puis d'amplifier chaque prélèvement de sol indépendamment, ce qui augmente considérablement le temps de travail et engendre des coûts financiers autrement plus importants. Une telle procédure permettrait cependant d'avoir une vue plus exhaustive de la richesse initiale de chaque échantillon de sol et de limiter l'incertitude sur les estimateurs de diversité (Zinger *et al.* 2014). Une meilleure évaluation *a posteriori* de la richesse de la parcelle est alors attendue.

Ce type d'échantillonnage, s'il est correctement spatialisé, peut permettre d'étudier les turnover de biodiversité à fine échelle, qui, compte-tenu de la faible dispersion de certains taxons, notamment dans les sols, peuvent être très importants. En modifiant la richesse et la distribution spatiale des taxons, les perturbations rendent ces mesures particulièrement sensibles. L'homogénéisation des communautés, sous l'effet de vastes perturbations (*e.g.* conversion des terres), entraîne, par exemple, une diminution des estimateurs de diversité beta. Les relations aire-taxons ou encore les relations de « *distance-decay* » (*e.g.* diminution de la similarité avec la distance, Soininen *et al.* 2007) peuvent alors être utilisées comme des indicateurs de changements aux niveaux des communautés (Seabloom *et al.* 2002 ; Tittensor *et al.* 2007). De même, un tel échantillonnage peut permettre d'étudier plus finement les réseaux de co-occurrence dans les sols. La structure et la typologie des réseaux écologiques affectent la réponse des communautés face aux perturbations. Il a été suggéré que des hauts niveaux de modularité et d'emboîtement (qui reflètent le niveau de structuration du réseau) permettent une plus grande persistance du système, notamment face à l'extinction d'espèces (Stouffer & Bascompte 2011). Ces éléments sont la conséquence des interactions biotiques. Plusieurs études ont ainsi suggéré que les patrons de co-occurrence pouvaient indiquer l'existence d'interactions biotiques, en partant du principe que, si deux espèces sont plus souvent retrouvées ensemble que ce qui peut être attendu par chance, elles sont plus à même d'interagir (Araújo *et al.* 2011 ; Gravel *et al.* 2011). Les données acquises à partir d'un échantillonnage spatialisé peuvent se

révéler utiles pour tester ces prédictions théoriques sur les communautés des sols (Poisot *et al.* 2013), ce qui ouvrirait alors la voie pour une étude plus fine des réseaux d'interactions et de leurs réponses face aux perturbations (Morris *et al.* 2010). Le rétablissement des caractéristiques des réseaux après une perturbation de grande envergure peut se faire dans des délais relativement courts. Une étude a ainsi montré le rétablissement rapide des caractéristiques des réseaux plantes-herbivores en forêts secondaires suite à l'abandon des terres, et ce malgré des différences de richesse et de composition des communautés par rapport aux forêts primaires (Villa-Galaviz *et al.* 2012).

b. Rémanence de l'ADN et évaluation à court terme des impacts

La persistance de l'ADN dans les sols offre au *metabarcoding* environnemental la perspective d'intégrer la diversité locale sur une plus longue période de temps, évitant les variations temporelles liées notamment à la saison d'échantillonnage et à l'activité des taxons (Bienert *et al.* 2012 ; Taberlet *et al.* 2012a ; Yoccoz *et al.* 2012). Cette persistance constitue clairement un avantage pour les études paléo-environnementales ; néanmoins, la présence potentielle d'ADN résiduel ancien (Yoccoz *et al.* 2012) limite l'application de cette méthode pour déterminer les variations à court terme dans la composition des communautés. En effet, la présence d'ADN résiduel peut entraîner des biais dans les estimateurs de biodiversité. Après une perturbation, la richesse locale (diversité alpha) va ainsi être artificiellement augmentée par la présence de cet ADN résiduel tandis que le turnover de biodiversité (diversité beta) va tendre à être sous-estimé. Dans les deux premières études (A et B), les changements induits par l'Homme semblent suffisamment anciens pour pouvoir être évalués par *metabarcoding*. Dans les îles Kerguelen, où le climat semble *a priori* favorable à la préservation de l'ADN, celui des plantes natives impactées par le lapin n'a ainsi jamais été retrouvé dans les sites perturbés.

Suite à la disparition d'une espèce d'un site donné, son ADN dans les sols va décroître avec le temps (Yoccoz *et al.* 2012), de même que sa probabilité de détection par *metabarcoding* environnemental. On peut alors utiliser le nombre de réplicats dans lequel un MOTU va être retrouvé comme un indicateur de sa présence à un instant donné (cf. articles A et B et annexe); une séquence retrouvée dans peu de réplicats sera ainsi considérée comme potentiellement résiduelle. Se pose alors le problème de la

différenciation entre séquences résiduelles et séquences rares dont la probabilité de détection sera également faible (cf. partie V.1.c.).

S'il est aujourd'hui bien établi que les taux de dégradation de l'ADN environnemental varient grandement selon les conditions environnementales (Dejean *et al.* 2011 ; Thomsen *et al.* 2012b), on manque toutefois de recul concernant les paramètres influençant ce processus. La caractérisation des changements à fine échelle temporelle et leur suivi régulier passent aujourd'hui nécessairement par le développement d'études concernant les processus de dégradation de l'ADN dans les sols.

c. Patrons de présence/absence et estimation de richesse

La perte de richesse est la conséquence la plus connue des activités humaines (*e.g.* Chapin *et al.* 2000). Les estimations de diversité à partir de *metabarcoding* environnemental peuvent être inférées à l'aide de deux approches différentes. La première consiste à utiliser l'assignation taxonomique des séquences, et leur diversité, comme indicateur de richesse. Cependant, la résolution taxonomique souvent limitée des marqueurs et l'incomplétude des bases de références ne permettent pas d'assigner toutes les séquences à un même niveau taxonomique. Une des solutions est alors de caractériser la diversité à des niveaux taxonomiques supérieurs lorsque cela est possible (*e.g.* famille). La seconde approche consiste à utiliser la diversité des séquences, ou de clusters de séquences (MOTUs), comme indicateur de richesse. Son utilisation permet principalement de comparer des niveaux de richesse inter-sites comme cela a été fait dans l'article A pour les communautés fongiques. Le recours à la raréfaction (ou l'utilisation de courbe de raréfaction) est alors un préalable nécessaire afin d'homogénéiser les profondeurs de séquençage entre échantillons (Gotelli & Colwell 2001 ; Gihring *et al.* 2012 ; Lundberg *et al.* 2013). Cette seconde approche est largement répandue en microbiologie (*e.g.* Lentendu *et al.* 2011 ; Zinger *et al.* 2011 ; Lindahl *et al.* 2013 ; Botnen *et al.* 2014), néanmoins elle n'est encore que très peu utilisée pour caractériser les communautés autres que fongiques ou bactériennes. Or, bien qu'imparfaite, elle est souvent la seule alternative pour évaluer la richesse des groupes taxonomiques peu ou pas connus et/ou les zones géographiques sous-étudiées (Rougerie *et al.* 2009 ; Decaëns *et al.* 2013).

Il est important de noter que, quelque soit l'approche utilisée, elle ne permet en aucun cas la détermination de la richesse réelle. En effet, dans les deux cas, se pose le problème de la représentativité de la diversité effectivement présente dans l'échantillon. Comme pour les méthodes d'inventaires classiques, des faux positifs (principalement liés à des erreurs de PCR/séquençage, à des contaminations sporadiques ou à la présence d'ADN résiduel) peuvent apparaître. L'utilisation de contrôles négatifs (*e.g.* témoins d'extraction et témoins de PCR) et positifs (*e.g.* dont l'origine et la concentration initiale de l'ADN sont connues) sont des éléments clés dans leur identification (Cooper & Poinar 2000 ; De Barba *et al.* 2014). A l'inverse, l'utilisation de faible quantité d'ADN dégradé, et la stochasticité de certains processus comme la PCR, peuvent entraîner la non détection de taxons effectivement présents (faux négatifs). La multiplication des réplicats biologiques et techniques (*e.g.* réplicats d'extraction et d'amplification) permet de limiter ces faux négatifs. Cependant, elle engendre également une augmentation du risque de faux positifs.

Distinguer les vraies séquences des erreurs, notamment à faible abondance, reste un des défis majeurs en *metabarcoding* environnemental (Coissac *et al.* 2012 ; Morgan *et al.* 2013). Des programmes dits de « *denoising* » permettant de détecter les erreurs de PCR et/ou de séquençage, sont utilisés mais nécessitent l'application de filtres sévères. La majorité d'entre eux, à l'instar d'*obiclean* le programme utilisé ici, se base, au moins en partie, sur l'abondance des séquences (*e.g.* Quince *et al.* 2011 ; Morgan *et al.* 2013). L'idée générale est de regrouper les erreurs avec les séquences réelles dont elles dérivent, sur la base de leur similarité, *via* des réseaux ou du clustering. On part alors du principe qu'un artefact sera moins fréquent que la séquence dont il dérive. L'utilisation de la distribution des séquences à travers des échantillons indépendants peut également aider à évaluer leur véracité (Morgan *et al.* 2013). Une des approches utilisées ici consiste à exclure les taxons dont la présence est incertaine (*e.g.* retrouvé dans un faible nombre de réplicats). L'étude présentée en annexe a montré que cette méthode permet d'éliminer efficacement les fausses présences (cf. Ficetola *et al.* accepté en annexe). Se pose alors la question de savoir à partir de quelle abondance une séquence est considérée comme présente dans un échantillon. A l'heure actuelle, leur détermination est principalement empirique et repose sur la distribution des abondances et les niveaux de diversité attendue. Ces seuils vont différer selon que l'on observe des communautés potentiellement riches avec beaucoup d'espèces rares (*e.g.* champignons,

bactéries) ou d'autres dont la diversité attendue est plus faible (*e.g.* vers de terre). L'élimination des séquences « rares » est ainsi fréquemment utilisée pour diminuer le « bruit de fond ». Si cette approche s'avère nécessaire pour limiter la détection de faux positifs, elle peut néanmoins produire l'effet inverse, à savoir des faux négatifs.

L'utilisation des modèles d'occupation d'espèces (*e.g.* Miller *et al.* 2011), sur la base de réplicats biologiques et/ou techniques, permet de fournir une évaluation de la pertinence des patrons de présence/absence inférés par *metabarcoding* environnemental. Néanmoins, l'utilisation de ces modèles est restreinte à un certain nombre de groupes taxonomiques, dont la diversité est relativement bien décrite (*e.g.* vertébrés, vers de terre). Leur utilisation pour caractériser la biodiversité des groupes cryptiques, comme les champignons ou les bactéries, est de fait limitée par la dominance de taxons extrêmement rares avec une probabilité de détection très faible, nécessitant un nombre inenvisageable de réplicats (cf. Ficetola *et al.* accepté en annexe)

Pour conclure, les outils développés à l'heure actuelle ne permettent pas de rendre compte de l'ensemble de la diversité des communautés hyperdiversifiées (*e.g.* forêt tropicale, champignons, bactéries) où la majorité des espèces sont rares. Le traitement des données issues du séquençage à haut-débit est une étape cruciale vers un inventaire total de la biodiversité et de nombreux progrès restent à faire. L'analyse des espèces rares reste ainsi très limitée et requière la plus grande précaution.

d. Une approche quantitative ?

La dynamique des populations est le fondement de la réponse des communautés face aux perturbations (*e.g.* Gill *et al.* 1996 ; Gardner *et al.* 2009 ; Schindler *et al.* 2010). De nombreuses métriques de diversité biologique sont basées sur l'abondance des espèces, ce dernier étant un paramètre particulièrement sensible aux changements environnementaux (Magurran 2004). Sans nécessairement engendrer la disparition d'une espèce, les perturbations vont modifier les niches écologiques disponibles dans le système et donc engendrer des fluctuations dans la taille des populations. Par exemple, la diminution brutale de la population d'une espèce, suite à une perturbation, peut permettre l'accroissement d'une autre espèce, moins performante, avec laquelle elle est en compétition. Si ces deux populations sont impliquées dans une même fonction écosystémique, ces variations de populations peuvent permettre de stabiliser le processus dans son ensemble.

S'il est possible d'estimer correctement les patrons de présence/absence par *metabarcoding* environnemental (Ficetola *et al.* accepté ; Ji *et al.* 2013 ; Etude D), l'estimation de l'abondance des taxons identifiés reste, quant à elle, encore sujette à caution. Les quantités d'ADN issues d'échantillons de sol obtenues après séquençage semblent par ailleurs, être plutôt liées à la biomasse des espèces qu'à leur abondance, le nombre de cellules, et, par conséquent, le nombre de copies du génome, étant influencé par la taille des organismes (Chariton *et al.* 2012). Des corrélations ont ainsi été observées entre la biomasse des grands vertébrés et les quantités d'ADN retrouvées dans les sols (Andersen *et al.* 2012). Plus intéressant, l'étude de Yoccoz *et al.* (2012) suggère que les quantités d'ADN dans les sols ne sont pas reliées à la biomasse en elle-même mais à son turnover. Dans l'article A présentée ici, la corrélation entre les proportions d'ADN retrouvé dans les sols de surface et les relevés botaniques apparaît très bonne. Ces résultats sont cependant à relativiser. En effet, la diversité végétale de l'archipel des Kerguelen est très faible, ce qui limite le nombre d'espèces rares, plus difficiles à détecter (cf. partie précédente et article A). De plus, la présence d'une seule strate de végétation limite les différences de turnover de biomasse pouvant exister entre les différentes formes de croissance (Yoccoz *et al.* 2012). L'évaluation de ces corrélations reste aujourd'hui limitée à la végétation et à la macro-faune, de telles comparaisons étant inapplicables pour la micro-biodiversité.

Le nombre de lectures de chaque MOTUs est couramment employé en microbiologie comme indicateur de l'abondance relative des taxons et de l'équitabilité (« eveness ») des communautés ainsi que pour décrire les patrons de diversité beta dont certains indices prennent en compte l'abondance relative des taxons (cf. encadré 3 ; Ramette 2007 ; Anderson *et al.* 2011 ; Zinger *et al.* 2012). Cette approche présente néanmoins plusieurs limites et sa transposition à l'ensemble des données issues de *metabarcoding* environnemental nécessite de prendre en compte d'importantes considérations. Premièrement, certains taxons sont plus susceptibles que d'autres de relâcher de l'ADN (*e.g.* vers de terre par ingestion de sol). Deuxièmement, il y a actuellement un manque d'informations sur la persistance de l'ADN dans les échantillons environnementaux et sur les facteurs affectant sa conservation (Bohmann *et al.* 2014). Les taux de dégradation dans les sols et les sédiments varient grandement selon les conditions environnementales (Yoccoz *et al.* 2012). Enfin, lors du processus d'analyse, des biais méthodologiques peuvent affecter la proportionnalité attendue

entre la composition en séquences initiales dans l'échantillon et celles obtenues après séquençage. Amend *et al.* (2010) ont ainsi montré que les estimations d'abondance, à partir de données issues de séquençage à haut-débit, étaient fortement influencées par le niveau de dilution des échantillons (Poisot *et al.* 2013). De même, l'utilisation d'amorces universelles est connue pour entraîner des biais d'amplification entre séquences (*e.g.* Acinas *et al.* 2004 ; Lovejoy & Potvin 2010).

Bien que critiquée, l'utilisation du nombre de lectures obtenu après séquençage est actuellement le seul moyen disponible pour estimer l'abondance des taxons microbiens (cf. Etude A, Zinger *et al.* 2012). En ce qui concerne les groupes taxonomiques pour lesquels d'autres techniques d'inventaire existent, de grandes précautions sont nécessaires, en particulier pour l'estimation de la diversité beta. Dans l'article A, compte tenu de la corrélation élevée entre les proportions d'ADN végétal retrouvées dans les sols et les relevés botaniques, les données d'abondance relative issues de *metabarcoding* ont été utilisées pour caractériser les changements de composition des communautés végétales. Cela reste toutefois une exception, une telle corrélation étant probablement limitée aux communautés pauvres et relativement simples (cf. ci-dessus). Dans les autres cas, l'utilisation d'indices de diversité beta basés sur des données de présence/absence est alors à privilégier. De même, concernant la faune, les biais liés aux différences de taille et d'activité entre taxons, ajoutés au manque d'éléments de comparaison concernant la contribution relative des taxons à l'ADN environnemental, nécessitent généralement l'utilisation de données de présence/absence (*e.g.* Yu *et al.* 2012 ; Ji *et al.* 2013 ; article B).

En conclusion, plusieurs contraintes limitent actuellement l'utilisation de l'ADN environnemental pour inférer des données quantitatives. Le développement de méthodes permettant de se dispenser du processus d'amplification par PCR (approche « sans PCR », cf. partie V.2.) constitue une première étape (Ji *et al.* 2013). Néanmoins l'extrapolation de données quantitatives requière à la fois des considérations d'ordre méthodologique mais également le développement d'approches statistiques permettant notamment de calibrer les différences de taux de dégradation (Chariton 2012). L'utilisation des rangs d'abondance, bien qu'imparfaite et moins sensible aux perturbations, apparaît, pour le moment, comme une alternative pertinente pour caractériser les variations de populations à partir de l'ADN environnemental (Kelly *et al.* 2014).

2. En perspectives : résolution taxonomique, bases de références et approche « sans PCR »

Au travers de ce document, il ressort que la principale limitation de l'approche de *metabarcoding* environnemental, autant en écologie qu'en paléoécologie, concerne la résolution taxonomique des marqueurs. Le développement d'approches robustes pour l'assignation taxonomique des MOTUs, limitant les faux positifs et les faux négatifs en donnant par exemple une probabilité d'assignation à chaque niveau taxonomique (cf. Coissac *et al.* 2012 pour une revue détaillée), est ainsi une priorité. La nécessité d'agrandir les bases de données génétiques pour un meilleur suivi de la biodiversité à partir de l'ADN environnemental ne fait également aucun doute. Ce développement est avant tout, aujourd'hui, un problème de stratégie. De fortes sommes ont été investies dans le cadre du CBoL (« *Consortium for Barcode of Life* ») pour créer des bases de référence sur des marqueurs standardisés ; néanmoins, ces derniers ne sont pas adaptés aux études basées sur l'ADN environnemental (cf. partie II.1.a. ; Yoccoz *et al.* 2012 ; Deagle *et al.* 2014). Le développement de bases de données spécifiques à chaque *metabarcod*e est également relativement limité à la fois géographiquement et surtout financièrement. Les technologies de séquençage à haut-débit permettent aujourd'hui de séquencer à un coût raisonnable et d'assembler efficacement les génomes complets des organelles ou les ARN ribosomiques, *e.g.* les portions de génome où sont élaborés la majorité des *barcodes* (*e.g.* Moore *et al.* 2006 ; Timmermans *et al.* 2010 ; Nock *et al.* 2011 ; McPherson *et al.* 2013 ; programme *orgasm* <https://www.grenoble.prabi.fr/svn/LECASofts/>, Coissac *et al.* non publié).

L'ADN extrait dans le cadre des projets « *Barcode of Life* » pourrait, par ailleurs, servir de base pour de tels projets (Taberlet *et al.* 2012c ; Deagle *et al.* 2014). Ceci permettrait d'obtenir des bases de références complètes utilisables pour la majorité des marqueurs existant actuellement mais également pour d'autres types d'études telles que les phylogénies.

Ces bases de référence ouvrent également la voie au développement d'approches basées sur le séquençage *shotgun* (Zhou *et al.* 2013). Celles-ci présentent l'avantage de ne plus se focaliser sur un fragment d'ADN particulier pour l'identification taxonomique mais de pouvoir exploiter potentiellement l'ensemble de l'information taxonomique contenu dans l'échantillon d'ADN (Taberlet *et al.* 2012c ; Zhou *et al.* 2013). Cette

méthodologie, se dispensant de l'étape d'amplification par PCR (« *approche sans PCR* »), permettrait, par ailleurs, de limiter les erreurs de réplication et de s'affranchir des biais d'amplification préférentielle de certains fragments. Les données obtenues offrent alors potentiellement une meilleure représentativité des quantités d'ADN dans l'échantillon initial (Taberlet *et al.* 2012c ; cf. parties V.1.c. et V.1.d.).

Zhou *et al.* (2013) ont montré la faisabilité technique d'une telle approche pour caractériser les communautés d'arthropodes à partir de soupes d'insectes (« *bulk sample* »). Cependant, plusieurs défis sont encore à relever pour une application à l'ADN environnemental. Premièrement, comme nous l'avons vu ci-dessus, cela nécessite le développement de bases de référence quasi-exhaustives de l'ensemble des génomes des organelles et de l'ARN ribosomique afin de pouvoir faire correspondre le maximum de fragments séquencés à un taxon (Taberlet *et al.* 2012c). Des développements statistiques et bioinformatiques sont également nécessaires afin d'adapter l'assignation taxonomique de tels jeux de données. Deuxièmement, on ignore encore la proportion de fragments informatifs, *e.g.* dans notre cas la proportion d'ADN mitochondrial, chloroplastique et ribosomique que l'on peut retrouver dans les échantillons environnementaux. Compte tenu de leurs nombreuses copies par cellule, les génomes des organelles représentent entre 1 et 10% de l'ADN total chez les plantes et les animaux tandis que les ADN ribosomiques sont massivement répétés (Taberlet *et al.* 2012c). Au vu des capacités de séquençages actuelles (plusieurs centaines millions de lectures par expérimentation), les données potentiellement recueillies apparaissent suffisantes pour recréer correctement les communautés et les paléo-communautés, si tant est que la proportion d'ADN issu des micro-organismes (archées, bactéries et champignons) ne soit pas trop élevée.

Conclusion et perspectives

Les paysages actuels, et la biodiversité qu'ils renferment, résultent en partie des activités humaines passées. Les travaux présentés dans ce document thèse rendent compte de l'impact durable, et parfois très ancien, qu'exercent les activités humaines sur les communautés biologiques. L'approche de *metabarcoding* environnemental s'est révélée être un outil prometteur pour caractériser les changements de biodiversité, actuels ou passés, suite aux perturbations d'origine anthropique. Les réponses des communautés, face à ces perturbations, sont souvent idiosyncratiques (Lewis 2009) mais le résultat principal qui ressort de ces études est que leur résilience n'est jamais complète. D'une part, il existe une part stochastique dans la résilience des communautés, liée à la limitation de la dispersion et du recrutement, qui fait que le pool d'espèces initial, capable d'occuper le milieu, est différent de celui présent avant perturbation. L'ordre d'arrivée des espèces dans le milieu, ainsi que leurs caractéristiques, jouent également un rôle important dans les processus de succession. D'autre part, les activités humaines limitent le retour aux conditions initiales, soit par leur fréquence ou par synergie entre elles, mais surtout en altérant durablement les caractéristiques biotiques et abiotiques de l'environnement.

Les résultats présentés ici supportent la thèse selon laquelle l'accent ne doit pas être mis sur le maintien ou le retour à l'état initial du système (*e.g.* en terme de composition d'espèces) mais sur la préservation des options pour le futur en maintenant la diversité, et son potentiel évolutif, ainsi que les processus qui contrôlent le système (Bengtsson 2002 ; Moritz 2002). La définition traditionnelle de la résilience se concentre sur la stabilité du système, proche d'un état d'équilibre permanent, où la résistance aux perturbations et la vitesse de retour aux conditions initiales sont utilisées comme propriétés de la perturbation (Holling 1996 ; cf. encadré 4). Cependant, l'absence de retour complet aux conditions initiales incite à se tourner vers la seconde définition de la stabilité qui met l'accent sur les conditions pouvant faire basculer le système dans un autre domaine de stabilité. La résilience de la communauté se mesure alors à la magnitude de perturbations qui peut être absorbée par le système avant que celui-ci ne modifie sa structure, en changeant les variables et processus qui contrôlent son évolution (Holling 1973 ; Holling 1996).

La résilience des caractéristiques fonctionnelles de l'écosystème (Smithwick 2011 ; Mouillot *et al.* 2013) est un aspect non exploré dans ce document. Dans les sols en particulier, les changements dans les communautés conduisent à une altération des processus écosystémiques (Allison & Martiny 2008). L'ADN environnemental, *via* la métagénomique, offre déjà l'opportunité de relier les gènes retrouvés dans l'environnement aux fonctions majeures de l'écosystème (*e.g.* Torsvik & Øvreås 2002 ; Green *et al.* 2008 ; Gilbert *et al.* 2010). Paula *et al.* (2014) ont ainsi évalué l'impact du changement d'utilisation des terres sur la diversité fonctionnelle des microbes dans les sols tropicaux. Leur étude a montré que la composition et le turnover des gènes environnementaux fonctionnels étaient différents entre des forêts primaires et secondaires, en particulier les gènes impliqués dans les cycles de l'azote et du carbone, alors que les niveaux de richesse étaient similaires.

Une des perspectives intéressantes offerte par le *metabarcoding* environnemental est de pouvoir établir un lien entre les communautés difficiles d'accès, actuelles ou passées, et les processus écosystémiques. Les données de *metabarcoding* pourraient ainsi être utilisées pour relier les taxons à des mesures de fonctionnement de l'écosystème. Combinée à des indicateurs paléoécologiques permettant d'évaluer les processus écosystémiques (*e.g.* stockage du carbone, cycle des nutriments...), cette approche permettrait notamment de corrélérer l'histoire des communautés et les perturbations avec le fonctionnement à long terme des écosystèmes (McLauchlan *et al.* 2014). Les connaissances acquises sur les grands groupes taxonomiques détectés sont parfois utilisées pour obtenir des informations sur les caractéristiques fonctionnelles des écosystèmes (*e.g.* Bellemain *et al.* 2013) mais cette méthode reste relativement limitée. Une autre approche consiste à coupler directement les caractéristiques fonctionnelles des communautés, inférées à partir de données métagénomiques, avec les données taxonomiques issues de *metabarcoding* (Barberà *et al.* 2012 ; Fierer *et al.* 2013). Bien qu'encore très peu utilisée (Bohmann *et al.* 2014), cette approche a, par exemple, permis de mettre en évidence des différences significatives entre les profils taxonomiques microbiens et les données de traits fonctionnels en milieu océanique (Barberà *et al.* 2012) ou la corrélation entre les diversités taxonomiques et fonctionnelles bactériennes dans les prairies américaines (Fierer *et al.* 2013). Un tel couplage offre d'intéressantes perspectives pour relier les changements induits par les perturbations dans les communautés au fonctionnement plus global de l'écosystème.

En conclusion, le *metabarcoding* environnemental est un outil de diagnostic précieux, mais encore perfectible, pour évaluer l'impact des perturbations sur les communautés à travers différentes échelles spatiales et temporelles. Relier ces données de biodiversité avec les traits et les fonctions auxquels ils sont associés permettra à l'avenir de mieux comprendre la résistance et la résilience des processus écosystémiques face aux perturbations humaines.

Références

- Acinas SG, Sarma-Rupavtarm R, Klepac-Ceraj V, Polz MF (2005) PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Applied and environmental microbiology*, **71**, 8966–8969.
- Adrian R, O'Reilly CM, Zagarese H (2009) Lakes as sentinels of climate change. *Limnology and Oceanography*, **54**, 2283–2297.
- Allison SD, Martiny JBH (2008) Colloquium paper: resistance, resilience, and redundancy in microbial communities. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 11512–11519.
- Amend AS, Seifert KA, Bruns TD (2010) Quantifying microbial communities with 454 pyrosequencing: does read abundance count? *Molecular ecology*, **19**, 5555–5565.
- Andersen K, Bird KL, Rasmussen M *et al.* (2012) Meta-barcoding of “dirt” DNA from soil reflects vertebrate biodiversity. *Molecular ecology*, **21**, 1966–1979.
- Anderson NJ, Bugmann H, Dearing JA, Gaillard M-J (2006) Linking palaeoenvironmental data and models to understand the past and to predict the future. *Trends in ecology & evolution*, **21**, 696–704.
- Anderson MJ, Crist TO, Chase JM *et al.* (2011) Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist. *Ecology letters*, **14**, 19–28.
- Anderson-Carpenter LL, McLachlan JS, Jackson ST *et al.* (2011) Ancient DNA from lake sediments: bridging the gap between paleoecology and genetics. *BMC evolutionary biology*, **11**, 30.
- André HM, Ducarme X, Lebrun P (2002) Soil biodiversity: myth, reality or conning? *Oikos*, **96**, 3–24.
- Araújo MB, Rozenfeld A, Rahbek C, Marquet PA (2011) Using species co-occurrence networks to assess the impacts of climate change. *Ecography*, **34**, 897–908.
- Arnaud F, Lignier V, Revel M *et al.* (2002) Flood and earthquake disturbance of 210Pb geochronology (Lake Anterne, NW Alps). *Terra Nova*, **14**, 225–232.
- Asner GP, Knapp DE, Broadbent EN *et al.* (2005) Selective logging in the Brazilian Amazon. *Science*, **310**, 480–482.
- Baird DJ, Hajibabaei M (2012) Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. *Molecular Ecology*, **21**, 2039–2044.
- Baldwin DS, Colloff MJ, Rees GN *et al.* (2013) Impacts of inundation and drought on eukaryote biodiversity in semi-arid floodplain soils. *Molecular ecology*, **22**, 1746–1758.

- De Barba M, Miquel C, Boyer F *et al.* (2014) DNA *metabarcoding* multiplexing and validation of data accuracy for diet assessment: application to omnivorous diet. *Molecular ecology resources*, **14**, 306–323.
- Barberán A, Fernández-Guerra A, Bohannan BJM, Casamayor EO (2012) Exploration of community traits as ecological markers in microbial metagenomes. *Molecular ecology*, **21**, 1909–1917.
- Bardgett RD, Bowman WD, Kaufmann R, Schmidt SK (2005) A temporal approach to linking aboveground and belowground ecology. *Trends in ecology & evolution*, **20**, 634–641.
- Bardgett RD, Wardle DA (2003) Herbivore-mediated linkages between aboveground and belowground communities. *Ecology*, **84**, 2258–2268.
- Barlow J, Gardner TA, Araujo IS *et al.* (2007) Quantifying the biodiversity value of tropical primary, secondary, and plantation forests. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 18555–18560.
- Barlow J, Gardner TA, Lees AC, Parry L, Peres CA (2012) How pristine are tropical forests? An ecological perspective on the pre-Columbian human footprint in Amazonia and implications for contemporary conservation. *Biological Conservation*, **151**, 45–49.
- Barlow J, Peres CA (2004) Ecological responses to el Niño-induced surface fires in central Brazilian Amazonia: management implications for flammable tropical forests. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **359**, 367–380.
- de Beaulieu J-L, Andrieu-Ponel V, Cheddadi R *et al.* (2006) Apport des longues séquences lacustres à la connaissance des variations des climats et des paysages pléistocènes. *Comptes Rendus Palevol*, **5**, 65–72.
- Behre K-E (1988) The rôle of man in European vegetation history. In: *Vegetation history* (eds Huntley B, Webb T), pp. 633–672. Springer Netherlands, Dordrecht.
- Beisner B, Haydon D, Cuddington K (2003) Alternative stable states in ecology. *Frontiers in Ecology and the Environment*, **1**, 376–382.
- Bellemain E, Carlsen T, Brochmann C *et al.* (2010) ITS as an environmental DNA barcode for fungi: an in silico approach reveals potential PCR biases. *BMC microbiology*, **10**, 189.
- Bellemain E, Davey ML, Kauserud H *et al.* (2013) Fungal palaeodiversity revealed using high-throughput *metabarcoding* of ancient DNA from arctic permafrost. *Environmental microbiology*, **15**, 1176–1189.
- Bengtsson J (2002) Disturbance and resilience in soil animal communities. *European Journal of Soil Biology*, **38**, 119–125.
- Bent SJ, Pierson JD, Forney LJ *et al.* (2007) Measuring species richness based on microbial community fingerprints: the emperor has no clothes. *Applied and environmental microbiology*, **73**, 2399–2401.

- Berry D, Ben Mahfoudh K, Wagner M, Loy A (2011) Barcoded primers used in multiplex amplicon pyrosequencing bias amplification. *Applied and environmental microbiology*, **77**, 7846–7849.
- Bienert F, De Danieli S, Miquel C *et al.* (2012) Tracking earthworm communities from soil DNA. *Molecular ecology*, **21**, 2017–2030.
- Bik HM, Porazinska DL, Creer S *et al.* (2012) Sequencing our way towards understanding global eukaryotic biodiversity. *Trends in ecology & evolution*, **27**, 233–243.
- Binladen J, Gilbert MTP, Bollback JP *et al.* (2007) The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PloS one*, **2**, e197.
- Birks HH, Birks HJB (2000) Future uses of pollen analysis must include plant macrofossils. *Journal of Biogeography*, **27**, 31–35.
- Björk RG, Björkman MP, Andersson MX, Klemedtsson L (2008) Temporal variation in soil microbial communities in Alpine tundra. *Soil Biology and Biochemistry*, **40**, 266–268.
- Blaalid R, Carlsen T, Kumar S *et al.* (2012) Changes in the root-associated fungal communities along a primary succession gradient analysed by 454 pyrosequencing. *Molecular Ecology*, **21**, 1897–1908.
- Blaalid R, Davey ML, Kauserud H *et al.* (2014) Arctic root-associated fungal community composition reflects environmental filtering. *Molecular ecology*, **23**, 649–659.
- Blarquez O, Carcaillet C (2010) Fire, fuel composition and resilience threshold in subalpine ecosystem. *PloS one*, **5**, e12480.
- Blaxter M (2003) Molecular systematics: Counting angels with DNA. *Nature*, **421**, 122–124.
- Blaxter M, Mann J, Chapman T *et al.* (2005) Defining operational taxonomic units using DNA barcode data. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **360**, 1935–1943.
- Boessenkool S, Epp LS, Haile J *et al.* (2012) Blocking human contaminant DNA during PCR allows amplification of rare mammal species from sedimentary ancient DNA. *Molecular ecology*, **21**, 1806–1815.
- Boessenkool S, McGlynn G, Epp LS *et al.* (2014) Use of ancient sedimentary DNA as a novel conservation tool for high-altitude tropical biodiversity. *Conservation biology*, **28**, 446–455.
- Bohmann K, Evans A, Gilbert MTP *et al.* (2014) Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in ecology & evolution*, **29**, 358–367.
- Bohmann K, Monadjem A, Lehmkuhl Noer C *et al.* (2011) Molecular diet analysis of two african free-tailed bats (molossidae) using high throughput sequencing. *PloS one*, **6**, e21441.

- Bongers F, Poorter L, Hawthorne WD, Sheil D (2009) The intermediate disturbance hypothesis applies to tropical forests, but disturbance contributes little to tree diversity. *Ecology letters*, **12**, 798–805.
- Botnen S, Vik U, Carlsen T *et al.* (2014) Low host specificity of root-associated fungi at an Arctic site. *Molecular ecology*, **23**, 975–985.
- Briggs AW, Stenzel U, Johnson PLF *et al.* (2007) Patterns of damage in genomic DNA sequences from a Neandertal. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 14616–14621.
- Brooks TM, Mittermeier RA, Mittermeier CG *et al.* (2002) Habitat loss and extinction in the hotspots of biodiversity. *Conservation Biology*, **16**, 909–923.
- Brown KA, Gurevitch J (2004) Long-term impacts of logging on forest diversity in Madagascar. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 6045–6049.
- Calderon-Aguilera LE, Rivera-Monroy VH, Porter-Bolland L *et al.* (2012) An assessment of natural and human disturbance effects on Mexican ecosystems: current trends and research gaps. *Biodiversity and Conservation*, **21**, 589–617.
- Caporaso JG, Kuczynski J, Stombaugh J *et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nature methods*, **7**, 335–336.
- Carcaillet C, Ali AA, Blarquez O *et al.* (2009) Spatial variability of fire history in subalpine forests: from natural to cultural regimes. *Ecoscience*, **16**, 1–12.
- Cardinale BJ, Srivastava DS, Duffy JE *et al.* (2006) Effects of biodiversity on the functioning of trophic groups and ecosystems. *Nature*, **443**, 989–992.
- CBOL Plant Working Group (2009) A DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 12794–12797.
- Chapin FS, Zavaleta ES, Eviner VT *et al.* (2000) Consequences of changing biodiversity. *Nature*, **405**, 234–242.
- Chapuis J-L, Bousses P, Barnaud G (1994) Alien mammals, impact and management in the french subantarctic islands. *Biological Conservation*, **67**, 97–104.
- Chapuis J, Frenot Y, Lebouvier M (2004) Recovery of native plant communities after eradication of rabbits from the subantarctic Kerguelen Islands, and influence of climate change. *Biological Conservation*, **117**, 167–179.
- Chapuis J-L, Le Roux V, Asseline J, Lefèvre L, Kerleau F (2001) Eradication of rabbits (*Oryctolagus cuniculus*) by poisoning on three islands of the subantarctic Kerguelen Archipelago. *Wildlife Research*, **28**, 323–331.
- Chariton AA (2012) Short and informative DNA products to indirectly measure vascular plant biodiversity. *Molecular ecology*, **21**, 3637–3739.
- Chariton AA, Court LN, Hartley DM, Colloff MJ, Hardy CM (2010) Ecological assessment of estuarine sediments by pyrosequencing eukaryotic ribosomal DNA. *Frontiers in Ecology and the Environment*, **8**, 233–238.

- Chariton AA, Ho KT, Proestou D *et al.* (2014) A molecular-based approach for examining responses of eukaryotes in microcosms to contaminant-spiked estuarine sediments. *Environmental toxicology and chemistry*, **33**, 359–369.
- Chazdon RL (2003) Tropical forest recovery: legacies of human impact and natural disturbances. *Perspectives in Plant Ecology, Evolution and Systematics*, **6**, 51–71.
- Chazdon RL, Letcher SG, van Breugel M *et al.* (2007) Rates of change in tree communities of secondary Neotropical forests following major disturbances. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **362**, 273–289.
- Chiarucci A, Bacaro G, Scheiner SM (2011) Old and new challenges in using species diversity for assessing biodiversity. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **366**, 2426–2437.
- Chown SL (2010) Temporal biodiversity change in transformed landscapes: a southern African perspective. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **365**, 3729–3742.
- Clark JA, May RM (2002) Taxonomic bias in conservation research. *Science*, **297**, 191–192.
- Coissac E, Riaz T, Puillandre N (2012) Bioinformatic challenges for DNA *metabarcoding* of plants and animals. *Molecular ecology*, **21**, 1834–1847.
- Colombaroli D, Beckmann M, van der Knaap WO, Curdy P, Tinner W (2013) Changes in biodiversity and vegetation composition in the central Swiss Alps during the transition from pristine forest to first farming. *Diversity and Distributions*, **19**, 157–170.
- Colombaroli D, Henne PD, Kaltenrieder P, Gobet E, Tinner W (2010) Species responses to fire, climate and human impact at tree line in the Alps as evidenced by palaeo-environmental records and a dynamic simulation model. *Journal of Ecology*, **98**, 1346–1357.
- Compton JE, Boone RD (2000) Long-Term Impacts of Agriculture on Soil Carbon and Nitrogen in New England Forests. *Ecology*, **81**, 2314–2330.
- Connell JH (1978) Diversity in tropical rain forests and coral reefs. *Science*, **199**, 1302–1310.
- Coolen M, Gibson J (2009) Ancient DNA in lake sediment records. *PAGES news*, **17**, 104–106.
- Coolen MJL, Overmann J (2007) 217 000-year-old DNA sequences of green sulfur bacteria in Mediterranean sapropels and their implications for the reconstruction of the paleoenvironment. *Environmental microbiology*, **9**, 238–249.
- Cooper A, Poinar HN (2000) Ancient DNA: Do it right or not at all. *Science*, **289**, 1139–1139.

- Cordonnier T (2004) Perturbations, diversité et permanence des structures dans les écosystèmes forestiers. Ecole Nationale du Génie Rural des Eaux et des Forêts.
- Corinaldesi C, Barucca M, Luna GM, Dell'Anno A (2011) Preservation, origin and genetic imprint of extracellular DNA in permanently anoxic deep-sea sediments. *Molecular ecology*, **20**, 642–654.
- Corinaldesi C, Beolchini F, Dell'Anno A (2008) Damage and degradation rates of extracellular DNA in marine sediments: implications for the preservation of gene sequences. *Molecular ecology*, **17**, 3939–3951.
- Courchamp F, Chapuis J-L, Pascal M (2003) Mammal invaders on islands: impact, control and control impact. *Biological reviews of the Cambridge Philosophical Society*, **78**, 347–383.
- Crutzen PJ (2002) Geology of mankind. *Nature*, **415**, 23–23.
- Cugny C, Mazier F, Galop D (2010) Modern and fossil non-pollen palynomorphs from the Basque mountains (western Pyrenees, France): the use of coprophilous fungi to reconstruct pastoral activity. *Vegetation History and Archaeobotany*, **19**, 391–408.
- Curry JP (2004) Factors affecting the abundance of earthworms in soils. In: *Earthworm Ecology* (ed Edwards CA), pp. 91–113. CRC Press, Boca Raton.
- D'Anjou RM, Bradley RS, Balascio NL, Finkelstein DB (2012) Climate impacts on human settlement and agricultural activities in northern Norway revealed through sediment biogeochemistry. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 20332–20337.
- Dale VH, Joyce LA, McNulty S *et al.* (2001) Climate Change and Forest Disturbances. *BioScience*, **51**, 723–734.
- Dambrine E, Dupouey J-L, Laüt L *et al.* (2007) Present forest biodiversity patterns in France related to former Roman agriculture. *Ecology*, **88**, 1430–1439.
- David F (2010a) Expansion of green alder (*Alnus alnobetula* [Ehrh] K. Koch) in the northern French Alps: a palaeoecological point of view. *Comptes rendus biologies*, **333**, 424–428.
- David F (2010b) An example of the consequences of human activities on the evolution of subalpine landscapes. *Comptes Rendus Palevol*, **9**, 229–235.
- Davis OK, Shafer DS (2006) Sporormiella fungal spores, a palynological means of detecting herbivore density. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **237**, 40–50.
- Deagle BE, Chiaradia A, McInnes J, Jarman SN (2010) Pyrosequencing faecal DNA to determine diet of little penguins: is what goes in what comes out? *Conservation Genetics*, **11**, 2039–2048.
- Deagle BE, Jarman SN, Coissac E, Pompanon F, Taberlet P (2014) DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. *Biology Letters*, **10**, In Press.

- Dearing JA, Braimoh AK, Reenberg A, Turner BL, van der Leeuw S (2010) Complex land systems : the need for long time perspectives to assess their future. *Ecology and Society*, **15**, 21.
- Decaëns T (2010) Macroecological patterns in soil communities. *Global Ecology and Biogeography*, **19**, 287–302.
- Decaëns T, Jiménez JJ, Gioia C, Measey GJ, Lavelle P (2006) The values of soil animals for conservation biology. *European Journal of Soil Biology*, **42**, S23–S38.
- Decaëns T, Porco D, Rougerie R, Brown GG, James SW (2013) Potential of DNA *barcoding* for earthworm research in taxonomy and ecology. *Applied Soil Ecology*, **65**, 35–42.
- Dejean T, Valentini A, Duparc A *et al.* (2011) Persistence of environmental DNA in freshwater ecosystems. *PloS one*, **6**, e23398.
- Delsuc F, Brinkmann H, Philippe H (2005) Phylogenomics and the reconstruction of the tree of life. *Nature Reviews Genetics*, **6**, 361–375.
- DeLuca TH, MacKenzie MD, Gundale MJ, Holben WE (2006) Wildfire-produced charcoal directly influences nitrogen cycling in ponderosa pine forests. *Soil Science Society of America Journal*, **70**, 448–453.
- Demanèche S, Jocteur-Monrozier L, Quiquampoix H, Simonet P (2001) Evaluation of biological and physical protection against nuclease degradation of clay-bound plasmid DNA. *Applied and environmental microbiology*, **67**, 293–9.
- Dennis PG, Rushton SP, Newsham KK *et al.* (2012) Soil fungal community composition does not alter along a latitudinal gradient through the maritime and sub-Antarctic. *Fungal Ecology*, **5**, 403–408.
- De Deyn GB, Van der Putten WH (2005) Linking aboveground and belowground diversity. *Trends in ecology & evolution*, **20**, 625–633.
- Diedhiou AG, Dupouey J-L, Buée M *et al.* (2009) Response of ectomycorrhizal communities to past Roman occupation in an oak forest. *Soil Biology and Biochemistry*, **41**, 2206–2213.
- Dietl GP, Flessa KW (2011) Conservation paleobiology: putting the dead to work. *Trends in ecology & evolution*, **26**, 30–37.
- Domaizon I, Savichtcheva O, Debroas D *et al.* (2013) DNA from lake sediments reveals the long-term dynamics and diversity of *Synechococcus* assemblages. *Biogeosciences Discussions*, **10**, 2515–2564.
- Dornelas M (2010) Disturbance and change in biodiversity. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **365**, 3719–3727.
- Drake JA (1991) Community-assembly mechanics and the structure of an experimental species ensemble. *The American Naturalist*, **137**, 1–26.
- Duffy JE (2003) Biodiversity loss, trophic skew and ecosystem functioning. *Ecology Letters*, **6**, 680–687.

- Dupouey J-L, Dambrine E, Laffite JD, Moares C (2002) Irreversible impact of past land-use on forest soil and biodiversity. *Ecology*, **83**, 2978–2984.
- Ellis EC (2011) Anthropogenic transformation of the terrestrial biosphere. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, **369**, 1010–1035.
- Ellis EC, Ramankutty N (2008) Putting people in the map: anthropogenic biomes of the world. *Frontiers in Ecology and the Environment*, **6**, 439–447.
- Epp LS, Boessenkool S, Bellemain EP *et al.* (2012) New environmental metabarcodes for analysing soil DNA: potential for studying past and present ecosystems. *Molecular ecology*, **21**, 1821–1833.
- Etienne D, Wilhelm B, Sabatier P, Reyss J-L, Arnaud F (2013) Influence of sample location and livestock numbers on *Sporormiella* concentrations and accumulation rates in surface sediments of Lake Allos, French Alps. *Journal of Paleolimnology*, **49**, 117–127.
- Ettema CH, Wardle DA (2002) Spatial soil ecology. *Trends in Ecology & Evolution*, **17**, 177–183.
- Everham EM, Brokaw NVL (1996) Forest damage and recovery from catastrophic wind. *The Botanical Review*, **62**, 113–185.
- Ficetola GF, Coissac E, Zundel S *et al.* (2010) An in silico approach for the evaluation of DNA barcodes. *BMC genomics*, **11**, 434.
- Ficetola GF, Miaud C, Pompanon F, Taberlet P (2008) Species detection using environmental DNA from water samples. *Biology letters*, **4**, 423–425.
- Ficetola GF, Pansu J, Bonin A *et al.* (2014) Replication levels, false presences, and the estimation of presence / absence from eDNA metabarcoding data. *Accepté*.
- Fierer N, Jackson RB (2006) The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 626–631.
- Fierer N, Ladau J, Clemente JC *et al.* (2013) Reconstructing the microbial diversity and function of pre-agricultural tallgrass prairie soils in the United States. *Science*, **342**, 621–624.
- Fischer M, Rudmann-Maurer K, Weyand A, Stöcklin J (2008) Agricultural Land Use and Biodiversity in the Alps. *Mountain Research and Development*, **28**, 148–155.
- Fonseca VG, Carvalho GR, Sung W *et al.* (2010) Second-generation environmental sequencing unmask marine metazoan biodiversity. *Nature communications*, **1**, 98.
- Fonseca VG, Nichols B, Lallias D *et al.* (2012) Sample richness and genetic diversity as drivers of chimera formation in nSSU metagenetic analyses. *Nucleic acids research*, **40**, e66.

- Foster DR, Motzkin G, Slater B (1998) Land-use history as long-term broad-scale disturbance : regional forest dynamics in Central New England. *Ecosystems*, **1**, 96–119.
- Foster D, Swanson F, Aber J *et al.* (2003) The importance of land-use legacies to ecology and conservation. *BioScience*, **53**, 77–88.
- Frenot Y, Chown SL, Whinam J *et al.* (2005) Biological invasions in the Antarctic: extent, impacts and implications. *Biological reviews of the Cambridge Philosophical Society*, **80**, 45–72.
- Frenot Y, Gloaguen J-C, Massé L, Lebouvier M (2001) Human activities, ecosystem disturbance and plant invasions in subantarctic Crozet, Kerguelen and Amsterdam Islands. *Biological Conservation*, **101**, 33–50.
- Gaillard M-J, Sugita S, Bunting MJ *et al.* (2008) The use of modelling and simulation approach in reconstructing past landscapes from fossil pollen data: a review and results from the POLLANDCAL network. *Vegetation History and Archaeobotany*, **17**, 419–443.
- Galop D, Houet T, Mazier F, Leroux G, Rius D (2011) Grazing activities and biodiversity history in the Pyrenees : New insights on high altitude ecosystems in the framework of a Human-Environment Observatory. *PAGES news*, **19**, 53–55.
- Gardner TA, Barlow J, Chazdon R *et al.* (2009) Prospects for tropical forest biodiversity in a human-modified world. *Ecology Letters*, **12**, 561–582.
- Gardner TA, Barlow J, Sodhi NS, Peres CA (2010) A multi-region assessment of tropical forest biodiversity in a human-modified world. *Biological Conservation*, **143**, 2293–2300.
- Gause GF (1934) *The struggle for existence*. Williams and Wilkins, Baltimore, MD.
- Gibson L, Lee TM, Koh LP *et al.* (2011) Primary forests are irreplaceable for sustaining tropical biodiversity. *Nature*, **478**, 378–381.
- Giesecke T (2005) Holocene dynamics of the southern boreal forest in Sweden. *The Holocene*, **15**, 858–872.
- Giguet-Covex C, Arnaud F, Enters D *et al.* (2012) Frequency and intensity of high-altitude floods over the last 3.5ka in northwestern French Alps (Lake Anterne). *Quaternary Research*, **77**, 12–22.
- Giguet-Covex C, Arnaud F, Poulenard J *et al.* (2011) Changes in erosion patterns during the Holocene in a currently treeless subalpine catchment inferred from lake sediment geochemistry (Lake Anterne, 2063 m a.s.l., NW French Alps): The role of climate and human activities. *The Holocene*, **21**, 651–665.
- Giguet-Covex C, Pansu J, Arnaud F *et al.* (2014) Long livestock farming history and human landscape shaping revealed by lake sediment DNA. *Nature communications*, **5**, 3211.

- Giguët-Covex C, Poulenard J, Chalmin E *et al.* (2013) XANES spectroscopy as a tool to trace phosphorus transformation during soil genesis and mountain ecosystem development from lake sediments. *Geochimica et Cosmochimica Acta*, **118**, 129–147.
- Giguët-Covex C, Rey P-J (2013) Evolution du paysage, du climat et des pratiques pastorales autour de la montagne d'Arterne depuis 10000 ans. *Nature & Patrimoines en pays de Savoie*, 15–24.
- Gihring TM, Green SJ, Schadt CW (2012) Massively parallel rRNA gene sequencing exacerbates the potential for biased community diversity comparisons due to variable library sizes. *Environmental microbiology*, **14**, 285–290.
- Gilbert JA, Field D, Swift P *et al.* (2010) The taxonomic and functional diversity of microbes at a temperate coastal site: a “multi-omic” study of seasonal and diel temporal variation. *PloS one*, **5**, e15545.
- Gill JA, Sutherland WJ, Watkinson AR (1996) A method to quantify the effects of human disturbance on animal populations. *The Journal of Applied Ecology*, **33**, 786–792.
- Gobet E, Tinner W, Hochuli PA, van Leeuwen JFN, Ammann B (2003) Middle to Late Holocene vegetation history of the Upper Engadine (Swiss Alps): the role of man and fire. *Vegetation History and Archaeobotany*, **12**, 143–163.
- Gotelli NJ, Colwell RK (2001) Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, **4**, 379–391.
- Gravel D, Canard E, Guichard F, Mouquet N (2011) Persistence increases with diversity and connectance in trophic metacommunities. *PloS one*, **6**, e19374.
- Green JL, Bohannan BJM, Whitaker RJ (2008) Microbial biogeography: from taxonomy to traits. *Science*, **320**, 1039–1043.
- Green JL, Holmes AJ, Westoby M *et al.* (2004) Spatial scaling of microbial eukaryote diversity. *Nature*, **432**, 747–750.
- Gunderson LH (2000) Ecological resilience - in theory and application. *Annual Review of Ecology and Systematics*, **31**, 425–429.
- Haas BJ, Gevers D, Earl AM *et al.* (2011) Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome research*, **21**, 494–504.
- Haile J, Froese DG, Macphee RDE *et al.* (2009) Ancient DNA reveals late survival of mammoth and horse in interior Alaska. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 22352–22357.
- Hajibabaei M, Shokralla S, Zhou X, Singer G a C, Baird DJ (2011) Environmental *barcoding*: a next-generation sequencing approach for biomonitoring applications using river benthos. *PloS one*, **6**, e17497.
- Haouchar D, Haile J, McDowell MC *et al.* (2014) Thorough assessment of DNA preservation from fossil bone and sediments excavated from a late Pleistocene–

- Holocene cave deposit on Kangaroo Island, South Australia. *Quaternary Science Reviews*, **84**, 56–64.
- Harrison RD, Tan S, Plotkin JB *et al.* (2013) Consequences of defaunation for a tropical tree community. *Ecology letters*, **16**, 687–694.
- Harsch MA, Hulme PE, McGlone MS, Duncan RP (2009) Are treelines advancing? A global meta-analysis of treeline response to climate warming. *Ecology letters*, **12**, 1040–1049.
- Hausmann S, Lotter AF, van Leeuwen JFN *et al.* (2002) Interactions of climate and land use documented in the varved sediments of Seebergsee in the Swiss Alps. *The Holocene*, **12**, 279–289.
- Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003a) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **270**, 313–321.
- Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) Ten species in one: DNA *barcoding* reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 14812–14817.
- Hebert PDN, Ratnasingham S, DeWaard JR (2003b) *Barcoding* animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **270**, S96–S99.
- Hebsgaard MB, Gilbert MTP, Arneborg J *et al.* (2009) “The Farm Beneath the Sand” – an archaeological case study on ancient “dirt” DNA. *Antiquity*, **83**, 430–444.
- Heckenberger MJ, Kuikuro A, Kuikuro UT *et al.* (2003) Amazonia 1492: pristine forest or cultural parkland? *Science*, **301**, 1710–1714.
- Heckenberger MJ, Russell JC, Fausto C *et al.* (2008) Pre-Columbian urbanism, anthropogenic landscapes, and the future of the Amazon. *Science*, **321**, 1214–1217.
- Heckenberger MJ, Russell JC, Toney JR, Schmidt MJ (2007) The legacy of cultural landscapes in the Brazilian Amazon: implications for biodiversity. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **362**, 197–208.
- Hector A (1999) Plant Diversity and Productivity Experiments in European Grasslands. *Science*, **286**, 1123–1127.
- Hiiesalu I, Opik M, Metsis M *et al.* (2012) Plant species richness belowground: higher richness and new patterns revealed by next-generation sequencing. *Molecular ecology*, **21**, 2004–2016.
- Hofreiter M (2001) DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Research*, **29**, 4793–4799.
- Hofreiter M, Collins M, Stewart JR (2012) Ancient biomolecules in Quaternary palaeoecology. *Quaternary Science Reviews*, **33**, 1–13.

- Holling CS (1973) Resilience and stability of ecological systems. *Annual Review of Ecology and Systematics*, **4**, 1–23.
- Holling CS (1996) Engineering resilience versus ecological resilience. In: *Engineering Within Ecological Constraints* (ed National Academy of Engineering), pp. 31–43.
- Hooper DU, Chapin FS, Ewel JJ *et al.* (2005) Effects of biodiversity on ecosystem functioning: a consensus of current knowledge. *Ecological Monographs*, **75**, 3–35.
- Hu FS, Hampe A, Petit RJ (2009) Paleoecology meets genetics: deciphering past vegetational dynamics. *Frontiers in Ecology and the Environment*, **7**, 371–379.
- Hubbell SP (1979) Tree dispersion, abundance, and diversity in tropical dry forest. *Science*, **203**, 1299–1309.
- Hubbell SP, Foster RB, O'Brien ST *et al.* (1999) Light-gap disturbances, recruitment limitation, and tree diversity in a neotropical forest. *Science*, **283**, 554–557.
- Irwin MT, Wright PC, Birkinshaw C *et al.* (2010) Patterns of species change in anthropogenically disturbed forests of Madagascar. *Biological Conservation*, **143**, 2351–2362.
- Iversen J (1964) Plant indicators of climate, soil, and other factors during the Quaternary. In: pp. 421–428. Congress, Report of the VIth International on Quaternary Warsaw 1961. Vol. II: Palaeobotanical Section.
- Ives AR, Carpenter SR (2007) Stability and diversity of ecosystems. *Science*, **317**, 58–62.
- Jaeger CH, Monson RK, Fisk MC, Schmidt SK (1999) Seasonal partitioning of nitrogen by plants and soil microorganisms in an alpine ecosystem. *Ecology*, **80**, 1883–1891.
- Jeffery S, Gardi C (2010) Soil biodiversity under threat - a review. *Acta Societatis Zoologicae Bohemicae*, **74**, 7–12.
- Ji Y, Ashton L, Pedley SM *et al.* (2013) Reliable, verifiable and efficient monitoring of biodiversity via *metabarcoding*. *Ecology letters*, **16**, 1245–1257.
- Jørgensen T, Haile J, Möller P *et al.* (2012a) A comparative study of ancient sedimentary DNA, pollen and macrofossils from permafrost sediments of northern Siberia reveals long-term vegetational stability. *Molecular Ecology*, **21**, 1989–2003.
- Jørgensen T, Kjaer KH, Haile J *et al.* (2012b) Islands in the ice: detecting past vegetation on Greenlandic nunataks using historical records and sedimentary ancient DNA *meta-barcoding*. *Molecular ecology*, **21**, 1980–1988.
- Jussy JH, Koerner W, Dambrine E, Dupouey JL, Benoit M (2002) Influence of former agricultural land use on net nitrate production in forest soils. *European Journal of Soil Science*, **53**, 367–374.
- Keesing F, Holt RD, Ostfeld RS (2006) Effects of species diversity on disease risk. *Ecology letters*, **9**, 485–498.
- Kelly RP, Port JA, Yamahara KM, Crowder LB (2014) Using environmental DNA to census marine fishes in a large mesocosm. *PloS one*, **9**, e86175.

- Kennedy PG, Bruns TD (2005) Priority effects determine the outcome of ectomycorrhizal competition between two *Rhizopogon* species colonizing *Pinus muricata* seedlings. *The New phytologist*, **166**, 631–638.
- Kennedy TA, Naeem S, Howe KM *et al.* (2002) Biodiversity as a barrier to ecological invasion. *Nature*, **417**, 636–668.
- Kennedy PG, Peay KG, Bruns TD (2009) Root tip competition among ectomycorrhizal fungi: are priority effects a rule or an exception? *Ecology*, **90**, 2098–2107.
- Kermarrec L, Franc A, Rimet F *et al.* (2014) A next-generation sequencing approach to river biomonitoring using benthic diatoms. *Freshwater Science*, **33**, 349–363.
- Kidder JH (1876) Contributions to the Natural History of Kerguelen Island, Made in Connection with the United States Transit-of-Venus Expedition, 1874-75. *Bulletin of the United States Natural Museum*, **3**, 1–122.
- Kircher M, Kelso J (2010) High-throughput DNA sequencing - concepts and limitations. *BioEssays*, **32**, 524–536.
- Knops JMH, Tilman D, Haddad NM *et al.* (1999) Effects of plant species richness on invasion dynamics, disease outbreaks, insect abundances and diversity. *Ecology Letters*, **2**, 286–293.
- Koleff P, Gaston KJ, Lennon JJ (2003) Measuring beta diversity for presence-absence data. *Journal of Animal Ecology*, **72**, 367–382.
- Kueffer C, Kaiser-Bunbury CN (2014) Reconciling conflicting perspectives for biodiversity conservation in the Anthropocene. *Frontiers in Ecology and the Environment*, **12**, 131–137.
- Lavelle P, Bignell D, Lepage M *et al.* (1997) Soil function in a changing world: the role of invertebrate ecosystem engineers. *European Journal of Soil Biology*, **33**, 159–193.
- Lawler JJ, Aukema JE, Grant JB *et al.* (2006) Conservation science: a 20-year report card. *Frontiers in Ecology and the Environment*, **4**, 473–480.
- Lawton JH, Bignell DE, Bolton B *et al.* (1998) Biodiversity inventories, indicator taxa and effects of habitat modification in tropical forest. *Nature*, **391**, 72–76.
- Lebouvier M, Frenot Y (2007) Conservation and management in the French subantarctic Islands. *Papers and Proceedings of the Royal Society of Tasmania*, **141**, 23–28.
- Lebouvier M, Frenot Y, Gloaguen JC, Hennion F, Vernon P (2002) Changements climatiques et évolution de la biodiversité dans les îles subantarctiques françaises. *Rapport d'activité IPEV 2002*, 31–36.
- van Leeuwen JFN, Froyd CA, van der Knaap WO *et al.* (2008) Fossil pollen as a guide to conservation in the Galapagos. *Science*, **322**, 1206.
- Lejzerowicz F, Esling P, Majewski W *et al.* (2013) Ancient DNA complements microfossil record in deep-sea subsurface sediments. *Biology letters*, **9**, 20130283.

- Lentendu G, Zinger L, Manel S *et al.* (2011) Assessment of soil fungal diversity in different alpine tundra habitats by means of pyrosequencing. *Fungal Diversity*, **49**, 113–123.
- Lenzen M, Moran D, Kanemoto K *et al.* (2012) International trade drives biodiversity threats in developing nations. *Nature*, **486**, 109–112.
- Levine JM (2000) Species diversity and biological invasions: relating local process to community pattern. *Science*, **288**, 852–854.
- Levy-Booth DJ, Campbell RG, Gulden RH *et al.* (2007) Cycling of extracellular DNA in the soil environment. *Soil Biology and Biochemistry*, **39**, 2977–2991.
- Lewis OT (2009) Biodiversity change and ecosystem function in tropical forests. *Basic and Applied Ecology*, **10**, 97–102.
- Lindahl T (1993) Instability and decay of the primary structure of DNA. *Nature*, **362**, 709–715.
- Lindahl BD, Nilsson RH, Tedersoo L *et al.* (2013) Fungal community analysis by high-throughput sequencing of amplified markers - a user's guide. *The New phytologist*, **199**, 288–299.
- Lindenmayer DB, Likens GE (2011) Direct measurement versus surrogate indicator species for evaluating environmental change and biodiversity loss. *Ecosystems*, **14**, 47–59.
- Loreau M, Mouquet N, Gonzalez A (2003) Biodiversity as spatial insurance in heterogeneous landscapes. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 12765–12770.
- Lovejoy C, Potvin M (2010) Microbial eukaryotic distribution in a dynamic Beaufort Sea and the Arctic Ocean. *Journal of Plankton Research*, **33**, 431–444.
- Lundberg DS, Yourstone S, Mieczkowski P, Jones CD, Dangl JL (2013) Practical innovations for high-throughput amplicon sequencing. *Nature methods*, **10**, 999–1002.
- Lydolph MC, Jacobsen J, Arctander P *et al.* (2005) Beringian paleoecology inferred from permafrost-preserved fungal DNA. *Applied and environmental microbiology*, **71**, 1012–1017.
- MacDougall a S, McCann KS, Gellner G, Turkington R (2013) Diversity loss with persistent human disturbance increases vulnerability to ecosystem collapse. *Nature*, **494**, 86–89.
- Mack MC, Antonio CMD (1998) Impacts of biological invasions on disturbance regimes. *Trends in Ecology & Evolution*, **13**, 195–198.
- Mack MC, Bret-Harte MS, Hollingsworth TN *et al.* (2011) Carbon loss from an unprecedented Arctic tundra wildfire. *Nature*, 489–492.

- Mack RN, Simberloff D, Mark Lonsdale W *et al.* (2000) Biotic invasions: causes, epidemiology, global consequences and control. *Ecological applications*, **10**, 689–710.
- Mackey RL, Currie DJ (2001) The diversity-disturbance relationship: is it generally strong and peaked? *Ecology*, **82**, 3479–3492.
- MacLean D, Jones JDG, Studholme DJ (2009) Application of “next-generation” sequencing technologies to microbial genetics. *Nature Reviews Microbiology*, **7**, 96–97.
- Madeja J, Wacnik A, Wypasek E, Chandran A, Stankiewicz E (2010) Integrated palynological and molecular analyses of late Holocene deposits from Lake Miłkowskie (NE Poland): Verification of local human impact on environment. *Quaternary International*, **220**, 147–152.
- Magurran AE (2004) *Measuring biological diversity*. Wiley.
- Magurran AE, Dornelas M (2010) Biological diversity in a changing world. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **365**, 3593–3597.
- Malik I, Wistuba M (2012) Dendrochronological methods for reconstructing mass movements — An example of landslide activity analysis using tree-ring eccentricity. *Geochronometria*, **39**, 180–196.
- Marlon JR, Bartlein PJ, Gavin DG *et al.* (2012) Long-term perspective on wildfires in the western USA. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, E535–E543.
- Marquer L, Gaillard M-J, Sugita S *et al.* (2014) Holocene changes in vegetation composition in northern Europe: why quantitative pollen-based vegetation reconstructions matter. *Quaternary Science Reviews*, **90**, 199–216.
- Mazier F, Galop D, Gaillard MJ *et al.* (2009) Multidisciplinary approach to reconstructing local pastoral activities: an example from the Pyrenean Mountains (Pays Basque). *The Holocene*, **19**, 171–188.
- McCann KS (2000) The diversity-stability debate. *Nature*, **405**, 228–233.
- McLauchlan KK, Higuera PE, Gavin DG *et al.* (2014) Reconstructing Disturbances and Their Biogeochemical Consequences over Multiple Timescales. *BioScience*, **64**, 105–116.
- McNeill JR, Winiwarter V (2004) Breaking the sod: humankind, history, and soil. *Science*, **304**, 1627–1629.
- McPherson H, Merwe M Van Der, Delaney SK *et al.* (2013) Capturing chloroplast variation for molecular ecology studies: a simple next generation sequencing approach applied to a rainforest tree. *BMC Ecology*, **13**, 8.
- Meusnier I, Singer GAC, Landry J-F *et al.* (2008) A universal DNA mini-barcode for biodiversity analysis. *BMC genomics*, **9**, 214.

- Miehe G, Miehe S, Böhner J *et al.* (2014) How old is the human footprint in the world's largest alpine ecosystem? A review of multiproxy records from the Tibetan Plateau from the ecologists' viewpoint. *Quaternary Science Reviews*, **86**, 190–209.
- Millenium Ecosystem Assessment (2005) *Ecosystems and human well-being: biodiversity synthesis*. World Resources Institute, Washington, DC.
- Miller PA, Giesecke T, Hickler T *et al.* (2008) Exploring climatic and biotic controls on Holocene vegetation change in Fennoscandia. *Journal of Ecology*, **96**, 247–259.
- Miller AD, Roxburgh SH, Shea K (2011) How frequency and intensity shape diversity-disturbance relationships. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 5643–5648.
- Millet L, Arnaud F, Heiri O *et al.* (2009) Late-Holocene summer temperature reconstruction from chironomid assemblages of Lake Anterne, northern French Alps. *The Holocene*, **19**, 317–328.
- Mocci F, Walsh K, Richer S *et al.* (2008) Archéologie et paléoenvironnement dans les Alpes méridionales françaises. Hauts massifs de l'Argentiérois, du Champsaur et de l'Ubaye (Hautes-Alpes et Alpes de Haute Provence) (Néolithique final – début de l'Antiquité). *Cahiers de Paléoenvironnement*, 253–272.
- Molino JF, Sabatier D (2001) Tree diversity in tropical rain forests: a validation of the intermediate disturbance hypothesis. *Science*, **294**, 1702–1704.
- Montoya JM, Pimm SL, Solé R V (2006) Ecological networks and their fragility. *Nature*, **442**, 259–264.
- Moore MJ, Dhingra A, Soltis PS *et al.* (2006) Rapid and accurate pyrosequencing of angiosperm plastid genomes. *BMC Plant Biology*, **6**, 17.
- Morgan MJ, Chariton AA, Hartley DM, Court LN, Hardy CM (2013) Improved inference of taxonomic richness from environmental DNA. *PloS one*, **8**, e71974.
- Moritz G (2002) Strategies to protect biological diversity and the evolutionary processes that sustain it. *Systematic Biology*, **51**, 238–254.
- Morris RJ (2010) Anthropogenic impacts on tropical forest biodiversity: a network structure and ecosystem functioning perspective. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **365**, 3709–3718.
- Mouillot D, Graham NAJ, Villéger S, Mason NWH, Bellwood DR (2013) A functional approach reveals community responses to disturbances. *Trends in ecology & evolution*, **28**, 167–177.
- Mullis K, Faloona F, Scharf S *et al.* (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbor symposia on quantitative biology*, **51 Pt 1**, 263–73.
- Murray DC, Pearson SG, Fullagar R *et al.* (2012) High-throughput sequencing of ancient plant and mammal DNA preserved in herbivore middens. *Quaternary Science Reviews*, **58**, 135–145.

- Myers JI, Chase JM, Jiménez I *et al.* (2013) Beta-diversity in temperate and tropical forests reflects dissimilar mechanisms of community assembly. *Ecology letters*, **16**, 151–157.
- Myers N, Knoll AH (2001) The biotic crisis and the future of evolution. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 5389–5392.
- Naeem S, Duffy JE, Zavaleta E (2012) The functions of biological diversity in an age of extinction. *Science*, **336**, 1401–1406.
- Naeem S, Knops JMH, Tilman D *et al.* (2000) Plant diversity increases resistance to invasion in the absence of covarying extrinsic factors. *Oikos*, **91**, 97–108.
- Naeem S, Thompson LJ, Lawler SP, Lawton JH, Woodfin RM (1994) Declining biodiversity can alter the performance of ecosystems. *Nature*, **368**, 734–737.
- Nepstad DC, Verissimo A, Alencar A *et al.* (1999) Large-scale impoverishment of Amazonian forests by logging and fire. *Nature*, **398**, 505–508.
- Nielsen R, Matz M (2006) Statistical approaches for DNA *barcoding*. *Systematic biology*, **55**, 162–169.
- Nock CJ, Waters DLE, Edwards MA *et al.* (2011) Chloroplast genome sequences from total DNA for plant identification. *Plant Biotechnology Journal*, **9**, 328–333.
- Nyström M, Folke C, Moberg F (2000) Coral reef disturbance and resilience in a human-dominated environment. *Trends in ecology & evolution*, **15**, 413–417.
- Ogram A, Sayler GS, Barkay T (1987) The extraction and purification of microbial DNA from sediments. *Journal of Microbiological Methods*, **7**, 57–66.
- Overpeck JT, Rind D, Goldberg R (1990) Climate-induced changes in forest disturbance and vegetation. *Nature*, **343**, 51–53.
- Paine RT, Tegner MJ, Johnson EA (1998) Compounded perturbations yield ecological Surprises. *Ecosystems*, **1**, 535–545.
- Parducci L, Jørgensen T, Tollefsrud MM *et al.* (2012) Glacial survival of boreal trees in northern Scandinavia. *Science*, **335**, 1083–1086.
- Parducci L, Matetovici I, Fontana SL *et al.* (2013) Molecular- and pollen-based vegetation analysis in lake sediments from central Scandinavia. *Molecular ecology*, **22**, 3511–3524.
- Parducci L, Suyama Y, Lascoux M, Bennett KD (2005) Ancient DNA from pollen: a genetic record of population history in Scots pine. *Molecular ecology*, **14**, 2873–2882.
- Paula FS, Rodrigues JLM, Zhou J *et al.* (2014) Land use change alters functional gene diversity, composition and abundance in Amazon forest soil microbial communities. *Molecular ecology*, **23**, 2988–2999.
- Pedersen MW, Ginolhac A, Orlando L *et al.* (2013) A comparative study of ancient environmental DNA to pollen and macrofossils from lake sediments reveals

- taxonomic overlap and additional plant taxa. *Quaternary Science Reviews*, **75**, 161–168.
- Peters DPC, Lugo AE, Chapin FS *et al.* (2011) Cross-system comparisons elucidate disturbance complexities and generalities. *Ecosphere*, **2**, 81.
- Peterson G, Allen CR, Holling CS (1998) Ecological resilience, biodiversity, and scale. *Ecosystems*, **1**, 6–18.
- Pietramellara G, Ascher J, Borgogni F *et al.* (2008) Extracellular DNA in soil and sediment: fate and ecological relevance. *Biology and Fertility of Soils*, **45**, 219–235.
- Pimm SL (1984) The complexity and stability of ecosystems. *Nature*, **307**, 321–326.
- Platt WJ (1975) The colonization and formation of equilibrium plant species associations on badger disturbances in a tall-grass prairie. *Ecological Monographs*, **45**, 285–305.
- Poinar HN, Hofreiter M, Spaulding WG *et al.* (1998) Molecular coproscopy: dung and diet of the extinct ground sloth *Nothrotheriops shastensis*. *Science*, **281**, 402–406.
- Poisot T, Péquin B, Gravel D (2013) High-throughput sequencing: a roadmap toward community ecology. *Ecology and evolution*, **3**, 1125–1139.
- Pompanon F, Deagle BE, Symondson WOC *et al.* (2012) Who is eating what: diet assessment using next generation sequencing. *Molecular ecology*, **21**, 1931–1950.
- Porazinska DL, Giblin-Davis RM, Esquivel A *et al.* (2010) Ecometagenetics confirm high tropical rainforest nematode diversity. *Molecular ecology*, **19**, 5521–5530.
- Puillandre N, Lambert A, Brouillet S, Achaz G (2012) ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Molecular ecology*, **21**, 1864–1877.
- Puissant J, Cécillon L, Mills RTE *et al.* Seasonal effects of climate manipulation on microbial community structure and function in mountain soils. *Soumis*.
- Purschke O, Schmid BC, Sykes MT *et al.* (2013) Contrasting changes in taxonomic, phylogenetic and functional diversity during a long-term succession: insights into assembly processes (J Fridley, Ed.). *Journal of Ecology*, **101**, 857–866.
- Quince C, Lanzen A, Davenport RJ, Turnbaugh PJ (2011) Removing noise from pyrosequenced amplicons. *BMC bioinformatics*, **12**, 38.
- Ramette A (2007) Multivariate analyses in microbial ecology. *FEMS microbiology ecology*, **62**, 142–160.
- Ratnasingham S, Hebert PDN (2007) BOLD: The Barcode of Life Data System (<http://www.barcodinglife.org>). *Molecular ecology notes*, **7**, 355–364.
- Redon M, Luque S, Gosselin F, Cordonnier T (2014) Is generalisation of uneven-aged management in mountain forests the key to improve biodiversity conservation within forest landscape mosaics? *Annals of Forest Science*, **71**, 751–760.

- Reeder J, Knight R (2010) Rapid denoising of pyrosequencing amplicon data: exploiting the rank-abundance distribution. *Nature methods*, **7**, 668–669.
- Rey PJ (2008) Premières occupations de la montagne sur les versants du col d'Anterne (2257m). *ADLFI. Archéologie de la France - Informations*.
- Rey PJ (2012) Expériences d'altitude : premières occupations humaines dans les Alpes savoyardes et valdôtaines. Deux opérations récentes de prospections et sondages entre 700 et 2500 m d'altitude : versants du col du Petit Saint Bernard (Savoie – Val d'Aoste) et envir. In: *Actes des Journées archéologiques régionales, Yenne 2011 et Valence 2012. Bilan Scientifique Rhône-Alpes*
- Rey F, Schworer C, Gobet E *et al.* (2013) Climatic and human impacts on mountain vegetation at Lauenensee (Bernese Alps, Switzerland) during the last 14,000 years. *The Holocene*, **23**, 1415–1427.
- Riaz T, Shehzad W, Viari A *et al.* (2011) ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic acids research*, **39**, e145.
- Robin M, Chapuis J-L, Lebouvier M (2011) Remote sensing of vegetation cover change in islands of the Kerguelen archipelago. *Polar Biology*, **34**, 1689–1700.
- Robinson JF, Dickerson JE (1987) Does invasion sequence affect community structure? *Ecology*, **68**, 587–595.
- Robinson J V., Edgemon MA (1988) An experimental evaluation of the effect of invasion history on community structure. *Ecology*, **69**, 1410–1417.
- Rodrigues JLM, Pellizari VH, Mueller R *et al.* (2013) Conversion of the Amazon rainforest to agriculture results in biotic homogenization of soil bacterial communities. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 988–993.
- Ropke A, Stobbe A, Oeggl K, Kalis AJ, Tinner W (2011) Late-Holocene land-use history and environmental changes at the high altitudes of St Antonien (Switzerland, Northern Alps): Combined evidence from pollen, soil and tree-ring analyses. *The Holocene*, **21**, 485–498.
- Rösch M, Jacomet S, Karg S (1992) The history of cereals in the region of the former Duchy of Swabia (Herzogtum Schwaben) from the Roman to the Post-medieval period: results of archeobotanical research. *Vegetation History and Archeobotany*, **1**, 193–231.
- Rougerie R, Decaëns T, Deharveng L *et al.* (2009) DNA barcodes for soil animal taxonomy. *Pesquisa Agropecuaria Brasileira*, **44**, 789–801.
- Roy J, Albert CH, Ibanez S *et al.* (2013) Microbes on the cliff: alpine cushion plants structure bacterial and fungal communities. *Frontiers in microbiology*, **4**, 64.
- Ruddiman WF (2013) The Anthropocene. *Annual Review of Earth and Planetary Sciences*, **41**, 45–68.

- Rydgren K, ØKland RH, Hestmark G (2004) Disturbance severity and community resilience in a boreal forest. *Ecology*, **85**, 1906–1915.
- Rykiel EJJ (1985) Towards a definition of ecological disturbance. *Australian Journal of Ecology*, **10**, 361–365.
- Savichtcheva O, Debroas D, Kurmayer R *et al.* (2011) Quantitative PCR enumeration of total/toxic *Planktothrix rubescens* and total cyanobacteria in preserved DNA isolated from lake sediments. *Applied and environmental microbiology*, **77**, 8744–8753.
- Scheffer M, Carpenter SR (2003) Catastrophic regime shifts in ecosystems: linking theory to observation. *Trends in Ecology & Evolution*, **18**, 648–656.
- Scheffer M, Carpenter S, Foley JA, Folke C, Walker B (2001) Catastrophic shifts in ecosystems. *Nature*, **413**, 591–6.
- Schimel DS, Asner GP, Moorcroft P (2013) Observing changing ecological diversity in the Anthropocene. *Frontiers in Ecology and the Environment*, **11**, 129–137.
- Schindler DE, Hilborn R, Chasco B *et al.* (2010) Population diversity and the portfolio effect in an exploited species. *Nature*, **465**, 609–612.
- Schoch CL, Seifert KA, Huhndorf S *et al.* (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 6241–6246.
- Seabloom EW, Dobson AP, Stoms DM (2002) Extinction rates under nonrandom patterns of habitat loss. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 11229–11234.
- Seastedt TR, Adams GA (2001) Effects of mobile tree islands on alpine tundra soils. *Ecology*, **82**, 8–17.
- Seddon AWR, Mackay AW, Baker AG *et al.* (2014) Looking forward through the past: identification of 50 priority research questions in palaeoecology. *Journal of Ecology*, **102**, 256–267.
- Seppä H, Bennett KD (2003) Quaternary pollen analysis: recent progress in palaeoecology and palaeoclimatology. *Progress in Physical Geography*, **27**, 548–579.
- Shea K, Roxburgh SH, Rauschert ESJ (2004) Moving from pattern to process: coexistence mechanisms under intermediate disturbance regimes. *Ecology Letters*, **7**, 11229–11234.
- Shehzad W (2011) DNA-based diet analyses in carnivores. Université de Grenoble.
- Shehzad W, McCarthy TM, Pompanon F *et al.* (2012a) Prey preference of snow leopard (*Panthera uncia*) in South Gobi, Mongolia. (R DeSalle, Ed.). *PloS one*, **7**, e32104.
- Shehzad W, Riaz T, Nawaz M a *et al.* (2012b) Carnivore diet analysis based on next-generation sequencing: application to the leopard cat (*Prionailurus bengalensis*) in Pakistan. *Molecular ecology*, **21**, 1951–1965.

- Shokralla S, Spall JL, Gibson JF, Hajibabaei M (2012) Next-generation sequencing technologies for environmental DNA research. *Molecular ecology*, **21**, 1794–1805.
- Smithwick EAH (2011) Pyrogeography and biogeochemical resilience. In: *The Landscape Ecology of Fire Ecological Studies*. (eds McKenzie D, Miller C, Falk DA), pp. 143–163. Springer Netherlands, Dordrecht.
- Smol JP, Douglas MS (2007) From controversy to consensus: making the case for recent climate change in the Arctic using lake sediments. *Frontiers in Ecology and the Environment*, **5**, 466–474.
- Soepboer W, Sugita S, Lotter AF (2010) Regional vegetation-cover changes on the Swiss Plateau during the past two millennia: A pollen-based reconstruction using the REVEALS model. *Quaternary Science Reviews*, **29**, 472–483.
- Soininen J, McDonald R, Hillebrand H (2007) The distance decay of similarity in ecological communities. *Ecography*, **30**, 3–12.
- Soininen EM, Ravolainen VT, Bråthen KA *et al.* (2013) Arctic Small Rodents Have Diverse Diets and Flexible Food Selection. *PloS one*, **8**, e68128.
- Sønstebo JH, Gielly L, Brysting a K *et al.* (2010) Using next-generation sequencing for molecular reconstruction of past Arctic vegetation and climate. *Molecular ecology resources*, **10**, 1009–1018.
- Sousa WP (1984) The role of disturbance in natural communities. *Annual Review of Ecology and Systematics*, **15**, 353–391.
- Spehn EM, Hector A, Joshi J *et al.* (2005) Ecosystem effects of biodiversity manipulations in European grasslands. *Ecological Monographs*, **75**, 37–63.
- Stachowicz JJ, Fried H, Osman RW, Whitlatch RB (2002) Biodiversity, invasion resistance, and marine ecosystem function: reconciling pattern and process. *Ecology*, **83**, 2575–2590.
- Staley JT, Konopka A (1985) Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annual review of microbiology*, **39**, 321–346.
- Stauffer M, Leyval C, Brun J-J, Leportier P, Berthelin J (2014) Effect of willow short rotation coppice on soil properties after three years of growth as compared to forest, grassland and arable land uses. *Plant and Soil*, **377**, 423–438.
- Steffen W, Grinevald J, Crutzen P, McNeill J (2011) The Anthropocene: conceptual and historical perspectives. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, **369**, 842–867.
- Stoof-Leichsenring KR, Epp LS, Trauth MH, Tiedemann R (2012) Hidden diversity in diatoms of Kenyan Lake Naivasha: a genetic approach detects temporal variation. *Molecular ecology*, **21**, 1918–1930.

- Stouffer DB, Bascompte J (2011) Compartmentalization increases food-web persistence. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 3648–3652.
- Taberlet P, Coissac E, Hajibabaei M, Rieseberg LH (2012a) Environmental DNA. *Molecular ecology*, **21**, 1789–1793.
- Taberlet P, Coissac E, Pompanon F *et al.* (2007) Power and limitations of the chloroplast trnL (UAA) intron for plant DNA *barcoding*. *Nucleic acids research*, **35**, e14.
- Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E (2012b) Towards next-generation biodiversity assessment using DNA *metabarcoding*. *Molecular ecology*, **21**, 2045–2050.
- Taberlet P, Prud’Homme SM, Campione E *et al.* (2012c) Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for *metabarcoding* studies. *Molecular ecology*, **21**, 1816–1820.
- Thomsen PF, Elias S, Gilbert MTP *et al.* (2009) Non-destructive sampling of ancient insect DNA. *PloS one*, **4**, e5048.
- Thomsen PF, Kielgast J, Iversen LL *et al.* (2012a) Detection of a diverse marine fish fauna using environmental DNA from seawater samples. *PloS one*, **7**, e41732.
- Thomsen PF, Kielgast J, Iversen LL *et al.* (2012b) Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular ecology*, **21**, 2565–2573.
- Thuiller W (2007) Climate change and the ecologist. *Nature*, **448**, 550–552.
- Tilman D (2004) Niche tradeoffs, neutrality, and community structure: a stochastic theory of resource competition, invasion, and community assembly. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 10854–10861.
- Tilman D, Knops J, Wedin D *et al.* (1997) The influence of functional diversity and composition on ecosystem processes. *Science*, **277**, 1300–1302.
- Tilman D, Reich PB, Knops J *et al.* (2001) Diversity and productivity in a long-term grassland experiment. *Science*, **294**, 843–845.
- Timmermans MJTN, Dodsworth S, Culverwell CL *et al.* (2010) Why barcode? High-throughput multiplex sequencing of mitochondrial genomes for molecular systematics. *Nucleic acids research*, **38**, e197.
- Tinner W, Ammann B, Germann P (1996) Treeline fluctuations recorded for 12,500 years by soil profiles, pollen, and plant macrofossils in the Central Swiss Alps. *Arctic and Alpine Research*, **28**, 131–147.
- Tinner W, Kaltenrieder P (2005) Rapid responses of high-mountain vegetation to early Holocene environmental changes in the Swiss Alps. *Journal of Ecology*, **93**, 936–947.
- Tittensor DP, Micheli F, Nyström M, Worm B (2007) Human impacts on the species-area relationship in reef fish assemblages. *Ecology letters*, **10**, 760–772.

- Torsvik V, Øvreås L (2002) Microbial diversity and function in soil: from genes to ecosystems. *Current Opinion in Microbiology*, **5**, 240–245.
- Tringe SG, Rubin EM (2005) Metagenomics: DNA sequencing of environmental samples. *Nature Reviews Genetics*, **6**, 805–814.
- Turbé A, De Toni A, Benito P *et al.* (2010) *Soil biodiversity : functions, threats and tools for policy makers. Bio Intelligence Service, IRD, and NIOO, Report for European Commission (DG Environment).*
- Turner MG (2010) Disturbance and landscape dynamics in a chnaging world. *Ecology*, **91**, 2833–2849.
- Turner MG, Collins SL, Lugo AE *et al.* (2003) Disturbance dynamics and ecological response: the contribution of long-term ecological research. *BioScience*, **53**, 46–56.
- Turner MG, Dale VH (1998) Comparing large, infrequent disturbances : what have we learned? *Ecosystems*, **1**, 493–496.
- Turner MG, Romme WH, Gardner RH, Neill RVO, Kratz TK (1993) A revised concept of landscape equilibrium : disturbance and stability on scaled landscapes. *Landscape Ecology*, **8**, 213–227.
- Valentini A, Miquel C, Nawaz MA *et al.* (2009a) New perspectives in diet analysis based on DNA *barcoding* and parallel pyrosequencing: the trnL approach. *Molecular ecology resources*, **9**, 51–60.
- Valentini A, Pompanon F, Taberlet P (2009b) DNA *barcoding* for ecologists. *Trends in ecology & evolution*, **24**, 110–117.
- Valsecchi V, Carraro G, Conedera M, Tinner W (2010) Late-Holocene vegetation and land-use dynamics in the Southern Alps (Switzerland) as a basis for nature protection and forest management. *The Holocene*, **20**, 483–495.
- Vandermeer J (2000) Hurricane disturbance and tropical tree species diversity. *Science*, **290**, 788–791.
- Villa-Galaviz E, Boege K, Del-Val E (2012) Resilience in plant-herbivore networks during secondary succession. *PloS one*, **7**, e53009.
- Vitousek PM, A MH, Lubchenco J, Melillo JM (1997) Human Domination of Earth's Ecosystems. *Science*, **277**, 494–499.
- Wagg C, Bender SF, Widmer F, van der Heijden MG a (2014) Soil biodiversity and soil community composition determine ecosystem multifunctionality. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 5266–5270.
- Walker LR, Wardle DA, Bardgett RD, Clarkson BD (2010) The use of chronosequences in studies of ecological succession and soil development. *Journal of Ecology*, **98**, 725–736.
- Wall DH (2005) Biodiversity and ecosystem functioning in terrestrial habitats of Antarctica. *Antarctic Science*, **17**, 523–531.

- Wall DH, Bardgett RD, Kelly E (2010) Biodiversity in the dark. *Nature Geoscience*, **3**, 297–298.
- Wardle DA (2006) The influence of biotic interactions on soil biodiversity. *Ecology letters*, **9**, 870–886.
- Wardle DA, Bardgett RD, Klironomos JN *et al.* (2004) Ecological linkages between aboveground and belowground biota. *Science*, **304**, 1629–1633.
- White PS, Jentsch A (2001) The search for generality in studies of disturbance and ecosystem dynamics. *Progress in Botany*, **62**, 399–449.
- White PS, Pickett STA (1985) Natural disturbance and patch dynamics: an introduction. In: *The Ecology of Natural Disturbance and Patch Dynamics* (eds Pickett STA, White PS), pp. 3–13. Academic Press.
- Whittaker RH (1965) dominance and diversity in land plant communities: numerical relations of species express the importance of competition in community function and evolution. *Science*, **147**, 250–260.
- Whittaker RH (1972) Evolution and Measurement of Species Diversity. *Taxon*, **21**, 213.
- Wick L, Tinner W (1997) Vegetation changes and timberline fluctuations in the Central Alps as indicators of Holocene climatic oscillations. *Arctic and Alpine Research*, **29**, 445–458.
- Willard DA, Cronin TM (2007) Paleoecology and ecosystem restoration: case studies from Chesapeake Bay and the Florida Everglades. *Frontiers in Ecology and the Environment*, **5**, 491–498.
- Willerslev E, Cappellini E, Boomsma W *et al.* (2007) Ancient biomolecules from deep ice cores reveal a forested southern Greenland. *Science*, **317**, 111–114.
- Willerslev E, Cooper A (2005) Ancient DNA. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **272**, 3–16.
- Willerslev E, Davison J, Moora M *et al.* (2014) Fifty thousand years of Arctic vegetation and megafaunal diet. *Nature*, **506**, 47–51.
- Willerslev E, Hansen AJ, Binladen J *et al.* (2003) Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science*, **300**, 791–795.
- Willis KJ, Birks HJB (2006) What is natural? The need for a long-term perspective in biodiversity conservation. *Science*, **314**, 1261–1265.
- Willis KJ, Gillson L, Brncic TM (2004) Ecology. How “virgin” is virgin rainforest? *Science*, **304**, 402–403.
- Winter M, Schweiger O, Klotz S *et al.* (2009) Plant extinctions and introductions lead to phylogenetic and taxonomic homogenization of the European flora. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 21721–21725.

- Wolters V, Silver WL, Bignell DE *et al.* (2000) Effects of global changes on above- and belowground biodiversity in terrestrial ecosystems: implications for ecosystem functioning. *BioScience*, **50**, 1089–1098.
- Wood JR, Wilmshurst JM, Wagstaff SJ *et al.* (2012) High-resolution coproecology: using coprolites to reconstruct the habits and habitats of New Zealand's extinct upland moa (*Megalapteryx didinus*). *PloS one*, e40025.
- Wright SJ (2005) Tropical forests in a changing environment. *Trends in ecology & evolution*, **20**, 553–560.
- Wu J, Loucks OE (1995) From balance of nature to hierarchical patch dynamics: a paradigm shift in ecology. *The Quarterly Review of Biology*, **70**, 439–466.
- Yergeau E, Bokhorst S, Huiskes AHL *et al.* (2007) Size and structure of bacterial, fungal and nematode communities along an Antarctic environmental gradient. *FEMS microbiology ecology*, **59**, 436–451.
- Yoccoz NG, Bråthen KA, Gielly L *et al.* (2012) DNA from soil mirrors plant taxonomic and growth form diversity. *Molecular ecology*, **21**, 3647–55.
- Young J, Watt A, Nowicki P *et al.* (2005) Towards sustainable land use: identifying and managing the conflicts between human activities and biodiversity conservation in Europe. *Biodiversity and Conservation*, **14**, 1641–1661.
- Yu DW, Ji Y, Emerson BC *et al.* (2012) Biodiversity soup: *metabarcoding* of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution*, **3**, 613–623.
- Zalasiewicz J, Williams M, Haywood A, Ellis M (2011) The Anthropocene: a new epoch of geological time? *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, **369**, 835–841.
- Zhou X, Li Y, Liu S *et al.* (2013) Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. *GigaScience*, **2**, 4.
- Zinger L (2009) Variations spatio-temporelles de la microflore des sols alpins. Université Grenoble Joseph Fourier.
- Zinger L, Boetius A, Ramette A (2014) Bacterial taxa-area and distance-decay relationships in marine environments. *Molecular ecology*, **23**, 954–964.
- Zinger L, Gobet A, Pommier T (2012) Two decades of describing the unseen majority of aquatic microbial diversity. *Molecular ecology*, **21**, 1878–1896.
- Zinger L, Lejon DPH, Baptist F *et al.* (2011) Contrasting diversity patterns of crenarchaeal, bacterial and fungal soil communities in an alpine landscape. *PloS one*, **6**, e19950.
- Zinger L, Shahnava B, Baptist F, Geremia R, Choler P (2009) Microbial diversity in alpine tundra soils correlates with snow cover dynamics. *The ISME journal*, **3**, 850–859.

Annexe

Replication levels, false presences, and the estimation of presence/absence from eDNA *metabarcoding* data

Accepté dans Molecular Ecology Resources

Ficetola, G.F.¹, Pansu, J.¹, Bonin, A.¹, Coissac, E.¹, Giguët-Covex, C.¹, De Barba, M.¹, Gielly, L.¹, Martins Lopes, C.¹, Boyer, F.¹, Pompanon, F.¹, Rayé, G.¹, Taberlet, P.¹.

¹ LECA, CNRS UMR 5553, Université Joseph Fourier, 38041 Grenoble Cedex 9, France

Abstract

Environmental DNA (eDNA) *metabarcoding* is increasingly used to study present and past biodiversity. eDNA analyses often rely on amplification of very small quantities or degraded DNA. To avoid missing detection of taxa that are actually present (false negatives), multiple extractions and amplifications of the same samples are often performed. However, the level of replication needed for reliable estimates of presence / absence patterns remains an unaddressed topic. Furthermore, degraded DNA and PCR/sequencing errors might produce false positives. We used simulations and empirical data to evaluate the level of replication required for accurate detection of targeted taxa in different contexts, and to assess the performance of methods used to reduce the risk of false detections. Furthermore, we evaluated whether statistical approaches developed to estimate occupancy in presence of observational errors can successfully estimate true prevalence, detection probability, and false positive rates. Replications reduced the rate of false negatives; the optimal level of replication was strongly dependent on the detection probability of taxa. Occupancy models successfully estimated true prevalence, detection probability, and false positive rates, but their performance increased with the number of replicates. At least eight PCR replicates should be performed if detection probability is not high, such as in ancient DNA studies. Multiple DNA extractions from the same sample yielded consistent results; in some cases collecting multiple samples from the same locality allowed detecting more species. The optimal level of replication for accurate species detection strongly varies among studies, and could be explicitly estimated to improve the reliability of results.

Introduction

Environmental DNA (eDNA) *metabarcoding*, *e.g.* the simultaneous identification of multiple taxa from the DNA extracted from an environmental sample (*e.g.* soil, water, faeces samples), is an emerging approach for the study of present and past biodiversity (Valentini *et al.* 2009b; Taberlet *et al.* 2012). The analysis of eDNA has an increasing number of applications, such as the description of biodiversity of microbes, plants and animals from a wide range of environments (*e.g.* Bienert *et al.* 2012; Yoccoz *et al.* 2012; Zinger *et al.* 2012), the analysis of diet (*e.g.* Deagle *et al.* 2005; Deagle *et al.* 2009; Valentini *et al.* 2009a; Pompanon *et al.* 2012; De Barba *et al.* 2014), the reconstruction of past biodiversity and/or environmental changes (Jorgensen *et al.* 2012; Parducci *et al.*

2013; Boessenkool *et al.* 2014; Giguët-Covex *et al.* 2014; Willerslev *et al.* 2014) and environmental monitoring (Jerde *et al.* 2011; Hajibabaei *et al.* 2012; Jerde *et al.* 2013; Darling 2014; Mahon *et al.* 2014; Nathan *et al.* 2014).

Analysis and identification of eDNA is used mainly for two purposes in ecological studies. First, considering a sample we want to infer the list of all the taxa present in the sampled environment, like in many studies on microbial diversity (*e.g.* Zinger *et al.* 2012). Second, considering a set of samples, we want to infer in which of them a species or a set of species are present or absent (*e.g.* Giguët-Covex *et al.* 2014). These types of analysis are not faced with the same challenges. In the first case, a major problem is to determine within the full list of observed sequences which one corresponds to true species and which one corresponds to experimental artefacts (PCR or sequencing induced errors). Conversely, the second case corresponds to studies in which we are interested in deciphering the distribution of a defined set of species, or in asserting the presence / absence of key species in a given environment. Like with other conventional inventory techniques, it is extremely important to take into account the possibility that present taxa remain undetected (false negatives), but also the occurrence of false observations of absent taxa (false positives) (Darling & Mahon 2011). Despite the high sensitivity of techniques based on eDNA, researchers are well aware that species detection using this approach is imperfect. Indeed, these analyses are usually performed with very little starting material and stochastic processes are often implicated as to whether a given DNA molecule is amplified. For instance, Ficetola *et al.* (2008) reported several PCRs in which the DNA of a target species was not detected, despite being present in the environmental sample. To cope with this issue, eDNA studies often performed replicated analyses: they can rely on multiple environmental samples, perform multiple extractions from the same environmental sample, and/or multiple amplifications of the same extracted DNA. The number of replicates per environmental sample can be highly variable among studies, ranging from two to >10 (*e.g.* Ficetola *et al.* 2008; Jerde *et al.* 2011; Jerde *et al.* 2013; Parducci *et al.* 2013; Giguët-Covex *et al.* 2014; Willerslev *et al.* 2014). Increasing the number of replicates certainly reduces the risk of missing present taxa, but inflates costs and workload. Furthermore, false positives are always possible, and the risk of false positives might increase with the replication level.

False positives are a particularly crucial issue in ancient DNA *metabarcoding* studies performed on environmental samples because the low amount of highly

degraded template DNA requires many PCR cycles. False positives may for instance arise because of PCR or sequencing errors not detected by dedicated software, or through sporadic contamination (Willerslev *et al.* 2014). To limit false positives, studies performed on degraded DNA use multiple quality assurance practices (*e.g.* sampling of localities where target taxa are notoriously absent, extraction blanks, equipment controls) and, additionally, sometimes consider a sequence only if it was confirmed by at least two independent PCRs, while those detected in one replicate only may be discarded or considered dubious (Giguët-Covex *et al.* 2014; Willerslev *et al.* 2014). Nevertheless, this approach has drawbacks, as it might overlook taxa that are actually present at low density or very difficult to detect.

Imperfect detection is an unavoidable feature of most data on species presence/absence and abundance: even during traditional ecological field studies, individuals and species that are present at one site are not always all detected, and failure in accounting for imperfect detection may result in biased inference (Yoccoz *et al.* 2001; MacKenzie *et al.* 2006; Kéry & Schmidt 2008; Lahoz-Monfort *et al.* 2014). Several models have been developed by ecologists to limit these issues. In the last decade, species-occupancy models (SOMs) have been developed to analyse species distribution when detection probability is lower than one. In short, SOMs can use data on detection and non-detection at multiple occasions in a number of sites to evaluate the detection probability of species when they are present. SOMs can allow estimating the number of sites in which a target species is present but remain undetected, and can be integrated with other analytical methods to better understand population dynamics and relationships between species and habitat (MacKenzie *et al.* 2006; Kéry & Schaub 2012). SOMs have been developed to analyse species distribution data obtained in the context of traditional field studies (MacKenzie *et al.* 2002; MacKenzie *et al.* 2006), but recent work showed that SOMs can be successfully applied to the analysis of eDNA data (Pilliod *et al.* 2013; Schmidt *et al.* 2013). SOMs may therefore allow the evaluation of the probability of taxa detection through eDNA, and the estimation of the number of replicates required for reliable inference of taxon absence (Schmidt *et al.* 2013).

SOMs were first developed to analyse data in which a species may remain undetected (*e.g.* false absences are possible) but, when detected, a presence is always considered as genuine (*e.g.* false presences are impossible) (MacKenzie *et al.* 2002; MacKenzie *et al.* 2006). However, misidentification of species is possible even with

traditional ecological data, and SOMs have thus been expanded to account for potential false presences (Royle & Link 2006; Miller *et al.* 2011). These approaches have never been applied in the context of eDNA *metabarcoding* studies, but can be extremely valuable to obtain measures of confidence on taxa absences, to evaluate the possibility of false presences, and to estimate the optimal number of replicates required for robust results.

In this study, we applied the use of SOMs to the analysis of eDNA *metabarcoding* results and estimated the number of technical replicates required to confidently estimate the presence / absence of a taxon in a given environment. We also evaluated the performance of approaches currently used to control for false presence. First, we analysed simulated data with known properties to answer the following questions: 1) what is the optimal number of PCR replicates for present and ancient eDNA *metabarcoding* analyses?; 2) what is the impact of imperfect detection, false presences, and number of replicates on the results of eDNA *metabarcoding* studies?; 3) what is the impact of dismissing taxa detected in only one of the replicated PCRs?; 4) do occupancy models allow estimating the frequency of false presences / false absences in eDNA *metabarcoding* data?

Second, we applied occupancy models to empirical eDNA *metabarcoding* data, to evaluate their detection probability and rate of false presences. We compared two different typologies of eDNA *metabarcoding* data (earthworm DNA from present-day soils and ancient DNA of mammals from lake sediment cores) to highlight the range of differences occurring among eDNA *metabarcoding* datasets. Finally, we evaluated whether different replication approaches (multiple samples per site, multiple DNA extractions per sample, multiple amplifications per DNA extract) may influence the study outcome.

Materials and methods

Simulations

We simulated data with known properties, mimicking patterns of taxa detection / non-detection in eDNA *metabarcoding* studies. We generated datasets representing 100 environmental samples for which one taxon was analysed, assuming that the taxon has a probability of presence in each sample = 0.3 (latent occupancy, *e.g.* the true, unobserved occupancy of the taxon; MacKenzie *et al.* 2006). For each dataset, we repeated 100

simulations per combination of parameter sets (p , fp and Nr , see below). Different values of taxon detection probability p ($p = 0.25, 0.5$ and 0.75), and different occurrences of false presences fp ($fp = 0.002, 0.01$ and 0.03) were used in simulations. These values of p represent taxa ranging from very low to very high detection probability, which may represent differences of abundance/biomass among taxa, or differences between present and ancient DNA studies (see Results, section "Analysis of empirical data"). Similarly, the fp values represent error rates ranging from very low to moderate. Per each combination of the parameters p and fp , we generated patterns of taxon detection / non-detection, with different numbers of replicates Nr ($Nr = 4, 6, 8, 12$). This number of replicates reflects the range commonly observed in eDNA studies (*e.g.* De Barba *et al.* 2014; Giguet-Covex *et al.* 2014; Willerslev *et al.* 2014). In our analysis, the number of replicates represents the overall number of PCRs performed on each environmental sample. In empirical studies, a given level of replication may be reached through different ways (*e.g.* eight replicates may be obtained by performing one DNA extraction per sample and then eight PCRs on the extract, or by performing two extractions per sample and four PCRs per extraction). For simplicity, in our simulations, the number of replicates just represents the total number of PCRs, independently on how they were obtained, as analyses of empirical data suggest that differences between replicating at the amplification level or also at the extraction level are small (see Results section).

We then analysed the simulated data using different approaches:

a) *Naive approach* assuming perfect detection and no false positives, and presence of a taxon at one site if it was detected at least once. We estimated the frequency of false absences (*e.g.* the number of samples in which the taxon was present but remained undetected) and the frequency of false presences (*e.g.* the number of samples in which the taxon was absent, but was erroneously detected because of false positives) by comparing the patterns of detection / non-detection with the true simulated dataset. The frequency of false absences and false presences was also calculated analytically using conditional probability, on the basis of known values of p , fp and Nr . Specifically, the overall probability of false presences across the Nr replicates was calculated as $1 - \text{the probability of not finding any false presence [e.g. } 1-(1-fp)^{Nr}]$, while the probability of false absences was calculated as the probability of obtaining Nr false absences, conditional to the probability of not finding any false presence [*e.g.* $(1-p)^{Nr} \times (1-(1-fp)^{Nr})$].

b) Conservative approach assuming perfect detection but possibility of false presences. Taxon presence was considered 'uncertain' if it was detected only once out of the Nr replicates. Uncertain presences were discarded, and we then estimated the frequency of remaining false presences and the number of true presences incorrectly removed.

c) MacKenzie occupancy model. This approach analyses replicated data of taxon detection / non-detection assuming that detection probability is < 1 , *e.g.* false absences are possible, while supposing that all detections are correct (*e.g.* no false presences) (MacKenzie *et al.* 2002). Using this approach, we estimated the detection probability of the taxon and its occupancy (the frequency of samples where the taxon is actually present). As this approach hypothesizes that all detections are correct, we assumed that a taxon was absent if taxon presence was uncertain (*e.g.* if it was detected in one replicate only).

d) Miller occupancy model. This approach analyses data of taxon detection / non-detection assuming that detection probability is < 1 , and that false presences are possible (Miller *et al.* 2011). We classified taxon presences as certain if the taxon was detected in at least two replicates, and uncertain if it was detected in one replicate only. Using this approach, we estimated the taxon detection probability, the rate of false presences, and the occupancy.

Both occupancy models were run using the package *unmarked* in R 3.0 (Fiske & Chandler 2011; R Development Core Team 2013). The R code used for simulation is available as Supporting Information (Appendix S1).

Empirical data – molecular methods

Earthworm data. We sampled 12 sites of 1ha in the Vercors massif (Northern French Alps), of which five were situated in pasture, four in deciduous forests and three in coniferous forests. In each site, we collected 100 core samples of mineral soil and litter every 10 m over a regular grid of 100 x 100 m, and pooled them together. The sampling procedure was repeated twice for each site. Extracellular DNA extraction was performed using the protocol described by Taberlet *et al.* (2012b). For each soil sample, we carried out two extractions (*e.g.* 4 extractions per site) and two different PCRs per extraction (total: 8 PCR replicates per site). DNA was amplified using the ewD/ewE primers that target short sequences (about 70 bp) on the mitochondrial 16S gene

(Bienert *et al.* 2012). Amplicons were then sequenced on a high-throughput sequencer (Illumina HiSeq 2000, 2 x 100-bp, pair-end reads). Finally, sequences were filtered to remove PCR/sequencing errors and chimeras using the OBITOOLS software suite (<http://www.grenoble.prabi.fr/trac/OBITools>) and assigned to the relevant taxon by comparison to a reference database (Pansu *et al.* unpublished). Only sequences assigned up to the species level with >95% similarity were used (Pansu *et al.* unpublished; see Appendix S2 for additional methodological details).

Ancient DNA from lake sediment cores. We analysed ancient DNA data previously published by Giguet-Covex *et al.* (2014). Giguet-Covex *et al.* (2014) extracted ancient DNA from a 20.2 m long sediment core from Lake Anterne (2063m asl, Northern French Alps); 47 slices of ~1cm thickness, were sampled, corresponding to 10160 cal. before present to nowadays. Each sample was divided in two subsamples, with two DNA extractions per subsample (*e.g.* four per sample) and two independent PCRs per extraction, resulting in eight amplification replicates per core slice. Mammal DNA was amplified using the MamP007 primers pair targeting a 60-84-bp fragment on the mitochondrial 16S gene (Giguet-Covex *et al.* 2014). Sequencing was performed using the Illumina HiSeq 2000 platform (2 x 100-bp, pair-end reads). PCR/sequencing errors and chimeras were also filtered out from the obtained dataset and sequences were assigned to the relevant taxa (see Giguet-Covex *et al.* 2014 for details).

Empirical data – data analysis

To evaluate detection probability, false presences and occupancy, we analysed the empirical data using the MacKenzie *et al.* (2002) (earthworm dataset) and the Miller *et al.* (2011) (ancient DNA dataset) occupancy models, following the same procedure of simulated data. We used two different approaches for the two datasets because uncertain presences only occurred in the ancient DNA datasets (see result).

For empirical data, three different approaches were used to obtain replicates: multiple samples from the same site/core sample, multiple DNA extractions from the same sample, and multiple amplifications on the same DNA extract. We therefore used the analysis of similarity (ANOSIM) to evaluate whether there are significant differences among communities obtained from different environmental samples from the same site, or from different DNA extractions of the same sample (Legendre & Legendre 2012). Similarity among communities was evaluated using the Bray-Curtis distance (Legendre

& Legendre 2012); significance of ANOSIM was assessed through 999 simulations using *vegan* (Oksanen *et al.* 2013; R Development Core Team 2013). This analysis was not performed for the ancient DNA data, because many PCRs did not detect the DNA of any taxon, therefore hampering most of pairwise comparisons among replicates.

Results

Analysis of simulated data

Naive approach. When assuming perfect detection, taxon occupancy was severely underestimated if the number of replicates was low, and if detection probability was low; results of simulations and analytical solutions showed high concordance (Table 1, Fig. 1a-d). For instance, if $p = 0.25$ and $Nr = 4$, the taxon remained undetected in about one third of samples where it was actually present (Table 1, Fig. 1a). Six replicates were needed to avoid false absences if detection probability = 0.5, and 12 replicates were needed if detection probability was very low (Fig. 1b-d). However, with this approach a high number of false positives exists in the data, particularly if fp was high and many replicates were run (Fig. 1e-h, Table 1). With many replicates, the number of false positives remained limited only if fp was very low.

Conservative approach. If detection in one replicate only was considered uncertain and discarded, the false positive rate was very low (Fig. 2 e-h), with false presences remaining in the data only if fp was high and many replicates were run (Fig. 2h). However, this conservative approach incorrectly removed many true presences (Fig. 2 a-d). For instance, if only four replicates were run and detection probability was low, this approach removed two third of true presences. The number of incorrectly removed presences was limited only if detection probability was very high ($p = 0.75$), or if many replicates were run (8-12 replicates, depending on detection probability) (Fig. 2 a-d).

MacKenzie occupancy model. The accuracy of the MacKenzie *et al.* (2002) occupancy model in the estimation of detection probability strongly depended on the number of replicates and on the actual values of p (Fig. 3a-d). With just four replicates, detection probability was correctly estimated only if it was high (0.75, Fig. 3a). Six replicates were needed to correctly estimate a detection probability = 0.5, and 8-12 replicates were needed if detection probability was low (0.25, Fig. 3c-d). If the number of replicates was not large enough, the MacKenzie model tended to overestimate

detection probability. Similarly, this approach correctly estimated true occupancy if detection probability was high, or if many replicates were performed. Four replicates were sufficient to estimate occupancy if detection probability was high (0.75), while 12 replicates were needed if p was very low (0.25). If the number of replicates was not large enough, the MacKenzie model underestimated occupancy (Fig. 3e-h).

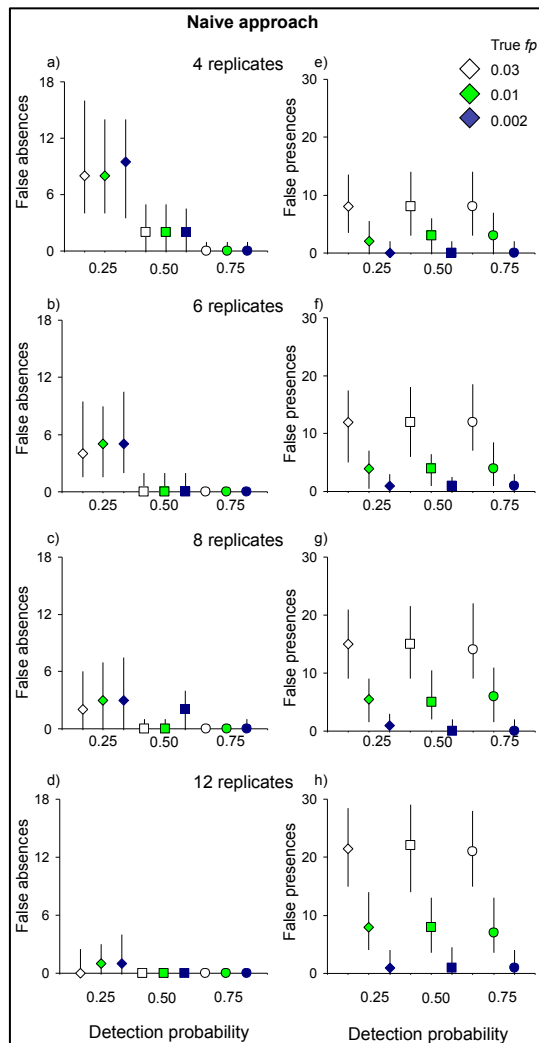


Figure 1: Analysis of simulated data using the Naive approach

a-d: number of simulated environmental samples in which the species was incorrectly assumed to be absent, depending on the number of replicates performed; e-h: number of simulated environmental samples in which the species was incorrectly assumed to be present, depending on the number of replicates performed. Symbols represents medians $\pm$ 95% CI; fp: frequency of false presences. Diamonds: detection probability = 0.25; squares: detection probability = 0.5; circles: detection probability = 0.75.

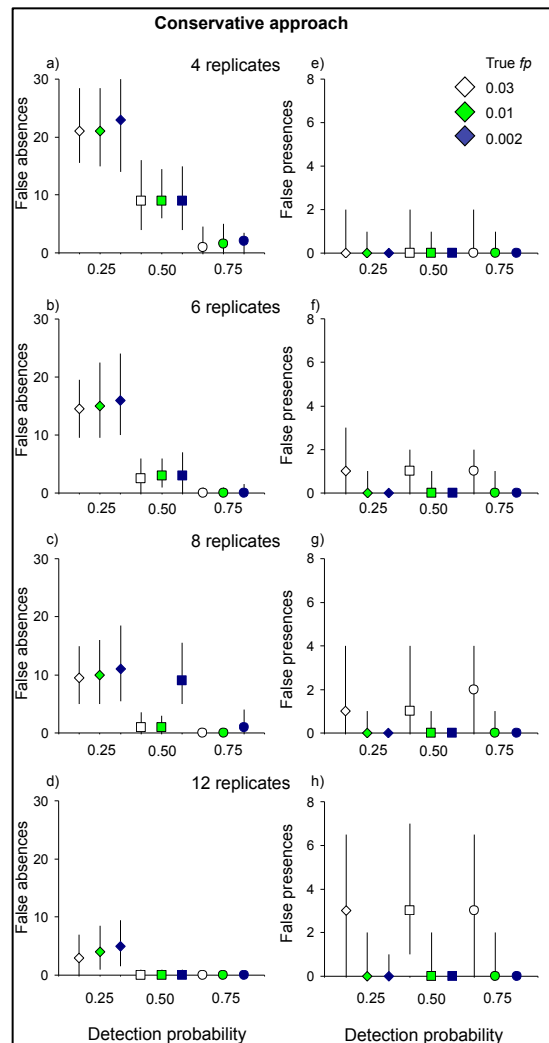


Figure 2: Analysis of simulated data using the conservative approach

a-d: number of simulated environmental samples in which the species was incorrectly assumed to be absent, depending on the number of replicates performed; e-h: number of simulated environmental samples in which the species was incorrectly assumed to be present, depending on the number of replicates performed. Symbols represents medians $\pm$ 95% CI; fp: frequency of false presences. Diamonds: detection probability = 0.25; squares: detection probability = 0.5; circles: detection probability = 0.75.

Miller's occupancy model. Estimates of detection probability and occupancy using the Miller *et al.* (2011) model were highly consistent with those of the MacKenzie model. If detection probability was high, four replicates were sufficient to estimate occupancy and detection probability, while 12 replicates were needed if p was very low (Fig. 4a-h). Furthermore, if a sufficient number of replicates (8-12) were run, the Miller's model estimated false presences with a good accuracy. Estimates of false presence were overestimated if detection probability was low, or if too few replicates were run (Fig. 4i-l)

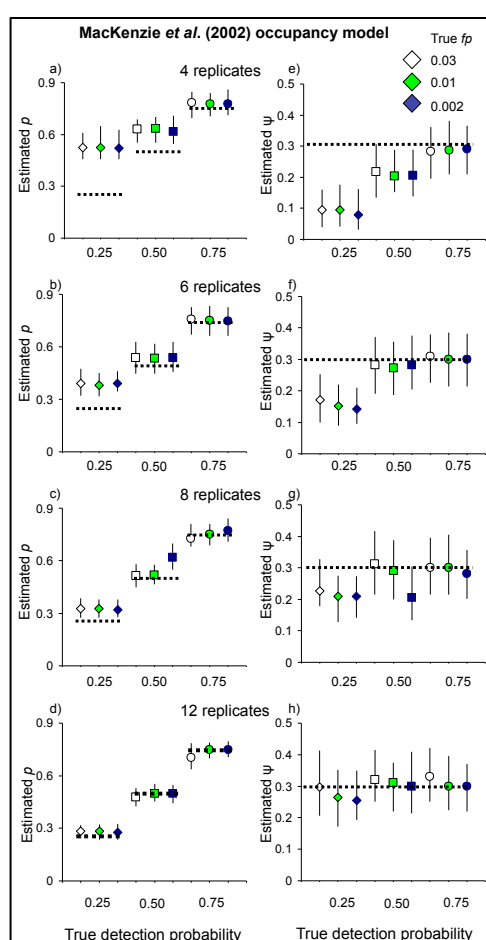


Figure 3: Analysis of simulated data using the MacKenzie *et al.* (2002) occupancy model.

a-d: estimated detection probability (p); e-h: estimated occupancy (ψ). Symbols represent medians $\pm$ 95% CI; fp : frequency of false presences. Bold dotted lines represent the true values of the parameters, used to generate the simulated data. Diamonds: detection probability = 0.25; squares: detection probability = 0.5; circles: detection probability = 0.75.

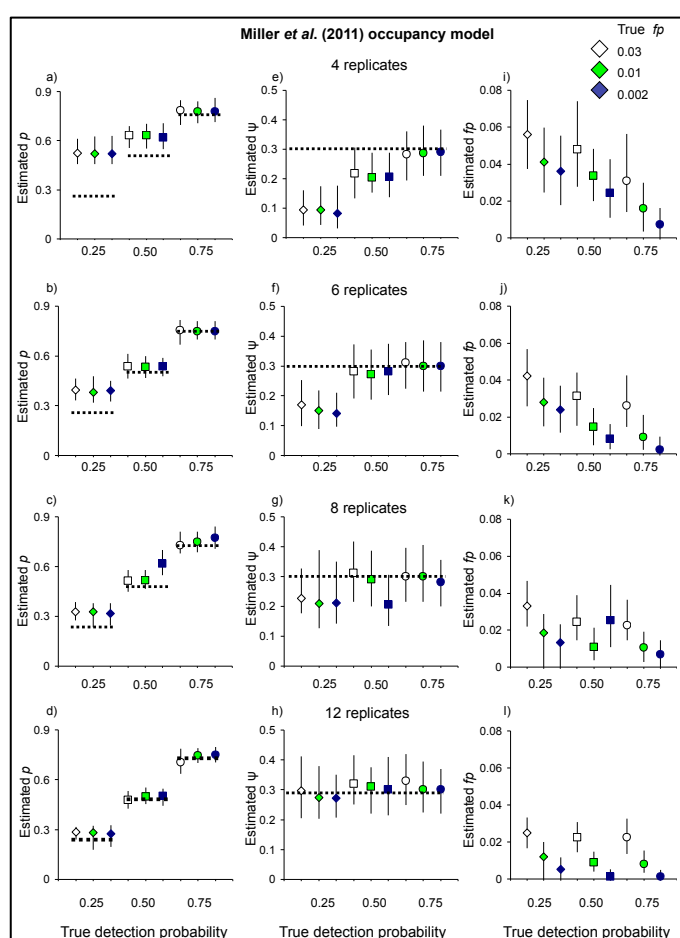


Figure 4: Analysis of simulated data using the Miller *et al.* (2011) occupancy model.

a-d: estimated detection probability (p); e-h: estimated occupancy (ψ). i-l: estimated rate of false presences (fp). Symbols represent medians $\pm$ 95% CI. Bold dotted lines represent the true values of the parameters, used to generate the simulated data. Diamonds: detection probability = 0.25; squares: detection probability = 0.5; circles: detection probability = 0.75.

Detection probability	N replicates	False presences	False absences	
			Analytical	Simulations
0.25	4	0.03	0.280	0.267
0.25	4	0.02	0.304	0.267
0.25	4	0.002	0.314	0.317
0.25	6	0.03	0.148	0.133
0.25	6	0.02	0.168	0.167
0.25	6	0.002	0.176	0.167
0.25	8	0.03	0.078	0.067
0.25	8	0.02	0.092	0.100
0.25	8	0.002	0.099	0.100
0.25	12	0.03	0.022	0.000
0.25	12	0.02	0.028	0.033
0.25	12	0.002	0.031	0.033
0.50	4	0.03	0.055	0.067
0.50	4	0.02	0.060	0.067
0.50	4	0.002	0.062	0.067
0.50	6	0.03	0.013	0.000
0.50	6	0.02	0.015	0.000
0.50	6	0.002	0.015	0.000
0.50	8	0.03	0.003	0.000
0.50	8	0.02	0.004	0.000
0.50	8	0.002	0.004	0.067
0.50	12	0.03	<0.001	0.000
0.50	12	0.02	<0.001	0.000
0.50	12	0.002	<0.001	0.000
0.75	4	0.03	0.003	0.000
0.75	4	0.02	0.004	0.000
0.75	4	0.002	0.004	0.000
0.75	6	0.03	<0.001	0.000
0.75	6	0.02	<0.001	0.000
0.75	6	0.002	<0.001	0.000
0.75	8	0.03	<0.001	0.000
0.75	8	0.02	<0.001	0.000
0.75	8	0.002	<0.001	0.000
0.75	12	0.03	<0.001	0.000
0.75	12	0.02	<0.001	0.000
0.75	12	0.002	<0.001	0.000
*	N replicates	False presences	False absences	
			Analytical	Simulations
	4	0.03	0.115	0.114
	4	0.02	0.039	0.038
	4	0.002	0.008	0.000
	6	0.03	0.167	0.171
	6	0.02	0.059	0.057
	6	0.002	0.012	0.014
	8	0.03	0.216	0.210
	8	0.02	0.077	0.079
	8	0.002	0.016	0.005
	12	0.03	0.306	0.307
	12	0.02	0.114	0.110
	12	0.002	0.024	0.014

* Results are unaffected by detection probability, as the false absences are estimated on the sites where the species is actually absent.

Table 1. Frequency of false absences and false presences obtained using the naïve approach, estimated analytically and on the basis of simulations.

Analysis of empirical data

Earthworm data. Nine earthworm species were identified in the dataset (Table 2). In all the samples and for all the species, three or more replicated PCRs confirmed species presence, therefore no presences were considered uncertain, and data were thus analysed using the MacKenzie *et al.* (2002) model. For all the earthworm species, the estimated detection probability was high, ranging from 0.52 to >0.9, and observed frequency of species corresponded very well to their estimated occupancy (Table 1a).

Distinct soil samples from the same locality sometimes yielded slightly different communities. Specifically, ANOSIM detected significant differences between soil samples from the same site (at $P < 0.05$) in six out of the twelve study sites. In the cases where ANOSIM detected differences between soil samples, one or two species were detected in only one of the two samples. Conversely, we never detected significant differences between extractions performed on the same soil sample (for all comparisons, $P \geq 0.33$).

Ancient DNA from lake sediments. Two mammal taxa (*Bos* and *Ovis*) were identified; sequences perfectly matched published sequences of domestic cow and sheep (Giguët-Covex *et al.* 2014). For the two taxa, in a few cases we obtained a single positive amplification not confirmed by PCR replication (from three samples for *Bos* and four samples for *Ovis*). Consequently, we used the Miller *et al.* (2011) approach that takes into account uncertain presence. For both taxa, the estimated detection probability was moderate (0.36-0.38). Nevertheless, the observed and estimated occupancies were extremely similar, because the number of replicates was high (eight replicates per sample). The estimated rate of false presences remained limited (about 1.3%) (Table 2).

Discussion

Careful planning is a key phase for all ecological studies. The optimal level of replication is an important parameter that often results from complex trade-offs. Inadequate replication may yield inconclusive and inconsistent results, while performing many replicates is costly and time-consuming. DNA *metabarcoding* is an emerging area of research where the optimization of a replication strategy is particularly important. Species detection from small amounts of degraded DNA is clearly imperfect, still laboratory and sequencing costs limit replication. In recent years, researchers are increasingly combining analyses of simulated and empirical data to assess the effect of incomplete sampling and for the evaluation of the performance of

analytical methods (Guillera-Arroita *et al.* 2010; Zurell *et al.* 2010; Ficetola *et al.* 2014). In this study, the analysis of simulated data allowed the identification of optimal strategies that limit the frequency of false presences and false absences in eDNA *metabarcoding* data, and we found that species occupancy models can help the analysis of eDNA data by providing accurate information on detection probability, false presences and true occupancy.

Taxon	Observed frequency	ψ	p	fp
a) Earthworms				
<i>Aporrectodea caliginosa</i>	0.750	0.750	0.939	-
<i>Aporrectodea icterica</i>	0.167	0.167	0.867	-
<i>Aporrectodea longa</i>	0.583	0.583	0.849	-
<i>Aporrectodea rosea</i>	0.833	0.833	0.892	-
<i>Dendrodrilus rubidus</i>	0.250	0.274	0.519	-
<i>Lumbricus castaneus</i>	0.750	0.750	0.826	-
<i>Lumbricus rubellus</i>	0.333	0.334	0.565	-
<i>Lumbricus terrestris</i>	0.417	0.417	0.919	-
<i>Octolasion cyaneum</i>	0.917	0.917	0.963	-
b) Ancient DNA from lake sediments				
<i>Bos</i> sp.	0.273*	0.283	0.377	0.013
<i>Ovis</i> sp.	0.113*	0.118	0.363	0.014

* calculated using samples with two detections or more.

Table 2. Results of occupancy models applied to present-day soil DNA *metabarcoding* data (earthworms), and to ancient DNA of mammals from lake sediments.

ψ : estimated occupancy; p : estimated detection probability; fp : estimated rate of false presences.

How many replicates should we perform? Increasing the number of replicated PCRs quickly reduces the risk of false negatives, improving the reliability of results. In practice, if only four replicates are run, the number of false absences can be very high, particularly if detection probability is low, and species occupancy may be severely underestimated (Fig. 1, Table 1). To limit false absences, at least six replicates are needed when detection probability is about 0.5, and at least eight are needed if detection probability is lower. In many cases, *metabarcoding* studies target multiple organisms, with wide variation of detection probability. For instance, p varied between 0.5 and > 0.9 for the earthworm species analyses, while it was clearly lower for the mammals of the ancient DNA dataset. If we want a reliable detection of the community, replication level should match the requirements of species with both low and high p . Eight

replicates may represent an appropriate level of replication if no a-priori information is available on the detection probability of species, as they may be suitable even for detecting species under difficult conditions, such as with ancient DNA. Nevertheless, it should be remarked that the optimal level of replication needed will also depend on the particular study and research objectives. For example, studies aiming at screening the possible biodiversity present in the samples (*e.g.* Bienert *et al.* 2012; Thomsen *et al.* 2012; Yoccoz *et al.* 2012; Parducci *et al.* 2013; Boessenkool *et al.* 2014), as the two real datasets analysed here, would differ from diet studies, where the main goal is to detect food items of biological importance for the animals (Pompanon *et al.* 2012; De Barba *et al.* 2014). In this latter case, non-detecting food items consumed only in traces or small quantities, and therefore present with low detection probabilities in some samples, would not affect the study conclusions, and a limited replication will most likely be sufficient. Sequencing depth, choice of primer and sequencing platforms are additional parameters that influence taxa detection, and should be also considered during the planning of studies (Tang *et al.* 2012; Zinger *et al.* 2012; Deagle *et al.* 2013; Smith & Peay 2014)

Researchers working with eDNA *metabarcoding* are well aware that false positives are always a risk, because of contamination or because of errors during PCR and sequencing, and it is not unusual to detect sequences of taxa that are actually exotic to the study sites. Blank and positive controls are key tools to identify these taxa (Cooper & Poinar 2000; De Barba *et al.* 2014). An additional approach is to exclude as “uncertain” the taxa that have been detected only once out the *Nr* replicates (Giguët-Covex *et al.* 2014; Willerslev *et al.* 2014). Our simulations showed that, with moderate levels of false positives, this method can successfully remove all the false presences (Fig. 2), but again it requires a sufficient number of replicates: if replicates are not enough, this approach may be too conservative and would remove taxa with low detection probability.

Species occupancy methods have excellent performance in the estimation of detection probability, true occupancy and even false presences, and can be successfully applied to eDNA data (Schmidt *et al.* 2013). SOMs can be a key resource for *metabarcoding* studies, but unfortunately they are only seldom used. In our simulations, two major approaches to SOMs (MacKenzie *et al.* 2002; Miller *et al.* 2011) yielded similar results in the estimation of occupancy and detection probability. Their

performance was generally good, and improved with the number of replicates and if target taxa showed high detection probability (Mackenzie & Royle 2005; Guillera-Arroita *et al.* 2010): if the number of replicates and detection probability were too low, these approaches tended to overestimate detection probability and underestimated occupancy, while they required at least eight replicates for a robust inference over a wide range of p values (Fig. 3-4). Furthermore, SOMs can be successfully applied to eDNA data for the estimation of error rate (Miller *et al.* 2011), if taxa that have been detected only once out of the Nr replicates are considered “uncertain”. This is excellent news for eDNA *metabarcoding* studies, in which dubious presences occur, as it may be possible to provide a measure of the reliability of such dubious presences. Importantly, if we know the error rate, we can eventually change the minimum number of positives required to consider taxon presence as “genuine”, and therefore identify taxon- or study-specific thresholds for the filtering of uncertain presences. For instance, at least three positive PCRs might be required to confirm the presence of species for which the risk of false detection is high.

Nevertheless, the approach described here has some limitations. First, PCR and sequencing errors may result in highly reproducible sequences. These errors may be found in multiple replicates, and therefore incorrectly assumed to be genuine. We stress the importance of using the appropriate procedures of bioinformatic filtering, to limit the occurrence of these artefacts; the approach described here does not ensure removal of this kind of errors. Second, this approach may be particularly suited for *metabarcoding* studies focusing on a restricted number of potential taxonomic units, or for which it is particularly relevant evaluating whether a given taxon is present or absent in a sample. Examples of this application of *metabarcoding* include studies on bioindicators, or targeting taxa for which a relatively good proportion of species is well known. However, many *metabarcoding* studies try to describe the biodiversity of cryptic, poorly known taxa such as soil microorganisms. In this case, thousands of potential operational taxonomic units are potentially present, many of which have very low occurrence. In this case, other approaches should be developed, for instance for the estimation of the richness of present taxa using accumulation curves (Lundberg *et al.* 2013).

In addition, occupancy modelling works well only if detection probability is reasonably high. Occupancy modelling is thus applicable to taxa with moderately low

detection probability (about 0.25) only if at least eight replicates are performed. However, biodiversity is dominated by many extremely rare species that likely have very low detection probability. An unrealistically high number of replicates would be needed to apply occupancy modelling to these taxa. Inappropriately applying occupancy modally to taxa that are very rare and difficult to detect may lead to inaccurate conclusions (*e.g.* Fig. 3e).

DNA *metabarcoding* data are increasingly used for ecological inference (Ji *et al.* 2013). Objective measures of data reliability, such as rate of false presences and false absences, can also be used to improve the outcome of ecological analysis. For instance, strategies exist to integrate measures of detection probability within models relating taxa occurrence to environmental variables (MacKenzie *et al.* 2006; Fiske & Chandler 2011; Gómez-Rodríguez *et al.* 2012). If detection probability is not perfect, these approaches (*e.g.* mixture models) allow better ecological inference and help identifying the factors determining biodiversity (MacKenzie *et al.* 2006; Kéry *et al.* 2009). Furthermore, in our study we simplistically assumed that detection probability and false presences are constant across the samples. However, these parameters may be influenced by environmental, biological (*e.g.* sample age, substrate, environmental temperature, differences among species, etc.) and technical factors (*e.g.* differences among operators and laboratories, preferential detection of shorter sequences, bias linked to the specific PCR and sequencing protocol, etc.). Occurrence of potential sources of biases may be tested and, if needed, integrated into models to limit their impact on the conclusion of studies. Finally, some studies are comparing the effectiveness of eDNA sampling with more traditional methods (*e.g.* amphibian calls, electroshocking, pit traps, etc) (Dejean *et al.* 2012; Tréguier *et al.* 2014). These comparisons should take into account that both eDNA and traditional approaches are imperfect methods affected by false presences and false absences, and require the application of appropriate occupancy models.

Through this study, we broadly used the term “replication levels” to indicate the total number of PCR replicates for each sample, but a given number of replicates may be reached through multiple ways (multiple samples per localities, multiple extractions per sample, multiple PCRs per extraction). Replication strategies are extremely different among studies: some researchers favour multiple PCRs on the same sample, while others favour multiple samples per site. Are these different approaches to replication

equivalent? In the analysis of earthworm communities, we never found differences between distinct DNA extractions performed on the same environmental sample. This suggests that technical reproducibility is high when studying present-day samples (De Barba *et al.* 2014). Conversely, in some cases, two samples from the same locality yielded slightly different results, suggesting that microhabitat heterogeneity may be strong, and stressing that DNA *metabarcoding* results often represent very well local communities. Collecting multiple environmental samples per site may help limiting the effects of microhabitat/spatial heterogeneity, allowing more exhaustive results.

Replication level may have a strong impact on eDNA *metabarcoding* studies. Researchers should adjust replication level, depending on their aims and on the features of their study system. Before performing biological inference from eDNA *metabarcoding* data, we suggest 1) running occupancy models to evaluate detection probability and rate of false presences; 2) evaluating whether the current level of replication is appropriate to control for false negatives; 3) if needed, removing “uncertain presences”, *e.g.* positives not confirmed by multiple PCRs. These three steps may allow improving the robustness of conclusions based on eDNA *metabarcoding* data.

Acknowledgments

The comments of K. Deiner, S. Creer and two anonymous reviewers greatly improved an early version of the manuscript. We thank L. Chalmandrier for useful discussions on statistics. This project was funded by: FRB, Fondation pour la Recherche sur la Biodiversité, in collaboration with Irstea Grenoble (earthworm dataset); Retromont program (Université Grenoble Alpes Savoie and CNRS in the framework of the DiPEE Chambéry-Grenoble) (ancient DNA dataset) and by French National Research Agency (Metabar research program; ANR 11 BSV7 020 01). We specifically thank J.J. Brun and S. De Danieli to coordinate soil sampling; P. Choler, J. Poulenard and F. Arnaud for coordinating the Retromont program with P. Taberlet.

References

- Bienert F, De Danieli S, Miquel C, *et al.* (2012) Tracking earthworm communities from soil DNA. *Molecular Ecology*, **21**, 2017-2030.
- Boessenkool S, McGlynn G, Epp LS, *et al.* (2014) Use of Ancient Sedimentary DNA as a Novel Conservation Tool for High-Altitude Tropical Biodiversity. *Conservation Biology*, **28**, 446-455.
- Cooper A, Poinar HN (2000) Ancient DNA: Do it right or not at ALL *Science*, **289**, 1139-1139.
- Darling JA (2014) Genetic studies of aquatic biological invasions: closing the gap between research and management. *Biological Invasions*, **in press**.
- Darling JA, Mahon AR (2011) From molecules to management: Adopting DNA-based methods for monitoring biological invasions in aquatic environments. *Environmental Research*, **111**, 978-988.
- De Barba M, Miquel C, Boyer F, *et al.* (2014) DNA *metabarcoding* multiplexing and validation of data accuracy for diet assessment: application to omnivorous diet. *Molecular Ecology Resources*, **14**, 306-323.
- Deagle BE, Kirkwood R, Jarman SN (2009) Analysis of Australian fur seal diet by pyrosequencing prey DNA in faeces. *Molecular Ecology*, **18**, 2022-2038.
- Deagle BE, Thomas AC, Shaffer AK, Trites AW, Jarman SN (2013) Quantifying sequence proportions in a DNA-based diet study using Ion Torrent amplicon sequencing: which counts count? *Molecular Ecology Resources*, **13**, 620-633.
- Deagle BE, Tollit DJ, Jarman SN, *et al.* (2005) Molecular scatology as a tool to study diet: analysis of prey DNA in scats from captive Steller sea lions. *Molecular Ecology*, **14**, 1831-1842.
- Dejean T, Valentini A, Miquel C, *et al.* (2012) Improved detection of an alien invasive species through environmental DNA *barcoding*: the example of the American bullfrog *Lithobates catesbeianus*. *Journal of Applied Ecology*, **49**, 953-959.
- Ficetola GF, Cagnetta M, Padoa-Schioppa E, *et al.* (2014) Sampling bias inverts ecogeographical relationships in island reptiles. *Global Ecology and Biogeography*, **in press**, DOI: 10.1111/geb.12201.
- Ficetola GF, Miaud C, Pompanon F, Taberlet P (2008) Species detection using environmental DNA from water samples. *Biology Letters*, **4**, 423-425.
- Fiske I, Chandler R (2011) unmarked: An R package for fitting hierarchical models of wildlife occurrence and abundance. *Journal of Statistical Software*, **43**, 1-23.
- Giguët-Covex C, Pansu J, Arnaud F, *et al.* (2014) Long livestock farming history and human landscape shaping revealed by lake sediment DNA. *Nature Communications*, **5**, 3211.

- Gómez-Rodríguez C, Bustamante J, Díaz-Paniagua C, Guisan A (2012) Integrating detection probabilities in species distribution models of amphibians breeding in Mediterranean temporary ponds. *Diversity and Distributions*, **18**, 260-272.
- Guillera-Arroita G, Ridout M-S, Morgan BJT (2010) Design of occupancy studies with imperfect detection. *Methods in Ecology and Evolution*, **1**, 131-139.
- Hajibabaei M, Spall JL, Shokralla S, van Konynenburg S (2012) Assessing biodiversity of a freshwater benthic macroinvertebrate community through non-destructive environmental *barcoding* of DNA from preservative ethanol. *Bmc Ecology*, **12**, 10.
- Jerde CL, Chadderton WL, Mahon AR, *et al.* (2013) Detection of Asian carp DNA as part of a Great Lakes basin-wide surveillance program. *Canadian Journal of Fisheries and Aquatic Sciences*, **70**, 522-526.
- Jerde CL, Mahon AR, Chadderton WL, Lodge DM (2011) "Sight-unseen" detection of rare aquatic species using environmental DNA. *Conservation Letters*, **4**, 150-157.
- Ji Y, Ashton L, Pedley SM, *et al.* (2013) Reliable, verifiable and efficient monitoring of biodiversity via *metabarcoding*. *Ecology Letters*, **16**, 1245-1257.
- Jorgensen T, Kjaer KH, Haile J, *et al.* (2012) Islands in the ice: detecting past vegetation on Greenlandic nunataks using historical records and sedimentary ancient DNA *Meta-barcoding*. *Molecular Ecology*, **21**, 1980-1988.
- Kéry M, Dorazio RM, Soldaat L, *et al.* (2009) Trend estimation in populations with imperfect detection. *Journal of Applied Ecology*, **46**, 1163-1172.
- Kéry M, Schaub M (2012) *Bayesian Population Analysis using WinBUGS: A hierarchical perspective* Academic Press, Waltham, MA.
- Kery M, Schmidt BR (2008) Imperfect detection and its consequences for monitoring for conservation. *Community Ecology*, **9**, 207-216.
- Lahoz-Monfort JJ, Guillera-Arroita G, Wintle BA (2014) Imperfect detection impacts the performance of species distribution models. *Global Ecology and Biogeography*, **23**, 504-515.
- Legendre P, Legendre L (2012) *Numerical Ecology*, 3rd Edition edn. Elsevier, Amsterdam.
- Lundberg DS, Yourstone S, Mieczkowski P, Jones CD, Dangl JL (2013) Practical innovations for high-throughput amplicon sequencing. *Nature Methods*, **10**, 999-1002.
- MacKenzie DI, Nichols JD, Lachman GB, *et al.* (2002) Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, **83**, 2248-2255.
- MacKenzie DI, Nichols JD, Royle JA, *et al.* (2006) *Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence* Academic Press, Burlington, MA.
- Mackenzie DI, Royle JA (2005) Designing occupancy studies: general advice and allocating survey effort. *Journal of Applied Ecology*, **42**, 1105-1114.

- Mahon AR, Nathan LR, Jerde CL (2014) Meta-genomic surveillance of invasive species in the bait trade. *Conservation Genetics Resources*, **6**, 563-567.
- Miller DA, Nichols JD, McClintock BT, *et al.* (2011) Improving occupancy estimation when two types of observational error occur: non-detection and species misidentification. *Ecology*, **92**, 1422-1428.
- Nathan LR, Jerde CL, Budny ML, Mahon AR (2014) The Use of Environmental DNA in Invasive Species Surveillance of the Great Lakes Commercial Bait Trade. *Conservation Biology*, n/a-n/a.
- Oksanen J, Blanchet FG, Kindt R, *et al.* (2013) *vegan: Community Ecology Package version 2.0-8*, www.r-project.org.
- Parducci L, Matetovici I, Fontana SL, *et al.* (2013) Molecular- and pollen-based vegetation analysis in lake sediments from central Scandinavia. *Molecular Ecology*, **22**, 3511-3524.
- Pilliod DS, Goldberg CS, Arkle RS, Waits LP (2013) Estimating occupancy and abundance of stream amphibians using environmental DNA from filtered water samples. *Canadian Journal of Fisheries and Aquatic Sciences*, **70**, 1123-1130.
- Pompanon F, Deagle BE, Symondson WOC, *et al.* (2012) Who is eating what: diet assessment using next generation sequencing. *Molecular Ecology*, **21**, 1931-1950.
- R Development Core Team (2013) *R: A language and environment for statistical computing* R Foundation for Statistical Computing, Vienna.
- Royle JA, Link WA (2006) Generalized site occupancy models allowing for false positive and false negative errors. *Ecology*, **87**, 835-841.
- Schmidt BR, Kéry M, Ursenbacher S, Hyman OJ, Collins JP (2013) Site occupancy models in the analysis of environmental DNA presence/absence surveys: a case study of an emerging amphibian pathogen. *Methods in Ecology and Evolution*, **4**, 646-653.
- Smith DP, Peay KG (2014) Sequence depth, not PCR replication, improves ecological inference from next generation DNA sequencing. *PLoS ONE*, **9**, e90234.
- Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, **21**, 2045-2050.
- Tang CQ, Leasi F, Obertegger U, *et al.* (2012) The widely used small subunit 18S rDNA molecule greatly underestimates true diversity in biodiversity surveys of the meiofauna. *Proceedings of the National Academy of Sciences*, **109**, 16208-16212.
- Thomsen PF, Kielgast JOS, Iversen LL, *et al.* (2012) Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular Ecology*, **21**, 2565-2573.
- Tréguier A, Paillisson J-M, Dejean T, *et al.* (2014) Environmental DNA surveillance for invertebrate species: advantages and technical limitations to detect invasive crayfish *Procambarus clarkii* in freshwater ponds. *Journal of Applied Ecology*, **51**, 871-879.

- Valentini A, Miquel C, Nawaz N, *et al.* (2009a) New perspectives in diet analysis based on DNA *barcoding* and parallel pyrosequencing: the *trnL* approach. *Molecular Ecology Resources*, **9**, 51-60.
- Valentini A, Pompanon F, Taberlet P (2009b) DNA *barcoding* for ecologists. *Trends in Ecology & Evolution*, **24**, 110-117.
- Willerslev E, Davison J, Moora M, *et al.* (2014) Fifty thousand years of Arctic vegetation and megafaunal diet. *Nature*, **506**, 47-51.
- Yoccoz NG, Brathen KA, Gielly L, *et al.* (2012) DNA from soil mirrors plant taxonomic and growth form diversity. *Molecular Ecology*, **21**, 3647-3655.
- Yoccoz NG, Nichols JD, Boulinier T (2001) Monitoring of biological diversity in space and time. *Trends in Ecology & Evolution*, **16**, 446-453.
- Zinger L, Gobet A, Pommier T (2012) Two decades of describing the unseen majority of aquatic microbial diversity. *Molecular Ecology*, **21**, 1878-1896.
- Zurell D, Berger U, Cabral JS, *et al.* (2010) The virtual ecologist approach: simulating data and observers. *Oikos*, **119**, 622-635.

Résumé

La plupart des écosystèmes sont aujourd'hui soumis à une pression anthropique croissante. Les études portant sur l'effet des activités humaines sur la biodiversité se multiplient mais elles se focalisent, principalement pour des raisons méthodologiques, sur un nombre restreint de groupes taxonomiques. L'originalité de ces travaux repose sur l'application d'une approche basée sur l'ADN environnemental permettant d'accéder à l'ensemble de la diversité biologique. Elle a ici été utilisée pour étudier l'impact des activités anthropiques sur les communautés biologiques et la résilience de ces dernières, à travers différentes échelles spatiales et temporelles. Dans une première partie, les communautés édaphiques, sous l'influence de diverses perturbations anthropiques, ont été caractérisées grâce à l'ADN contenu dans les sols. Ces études, réalisées dans différents milieux, mettent en évidence l'impact direct des activités humaines et leur influence sur les paramètres biotiques et abiotiques déterminant la distribution spatiale de la biodiversité des sols. La seconde partie se propose d'examiner l'impact à long terme des activités anthropiques au travers d'un exemple particulier en milieu de montagne. L'ADN contenu dans les archives sédimentaires lacustres a permis de retracer l'histoire des activités pastorales au cours de l'Holocène mais également les changements environnementaux que ceux-ci ont engendrés dans le bassin versant. L'approche mise en œuvre se révèle pertinente pour étudier à la fois les changements induits par les perturbations humaines sur les communautés et les facteurs influençant leurs dynamiques. Les résultats obtenus mettent notamment en lumière l'impact à long terme qu'exercent les activités anthropiques sur les communautés biologiques *via* les modifications profondes qu'elles génèrent sur les caractéristiques abiotiques du milieu.

Abstract

Most ecosystems undergo an increasing anthropogenic pressure. Studies about the effects of human activities on biodiversity are proliferating but they focus on few taxonomical groups, mainly for methodological reasons. The originality of this work is based on the application of an approach based on environmental DNA that allows access to the biodiversity as a whole. It was used here to study the impact of anthropogenic activities on biological communities and the resilience of these latter, across different spatial and temporal scales. In the first part, edaphic communities under the influence of several human disturbances were characterized from soil DNA. These studies, performed in various environments, bring out the direct impact of human activities and their influence on biotic and abiotic parameters driving spatial distribution of soil biodiversity. In the second part, long-term impact of human activities was investigated through the analysis of lake sediments in the Alps. DNA from lacustrine sediments allowed to reconstruct livestock farming history over the Holocene and environmental changes that they induced in the catchment. We showed that this approach was relevant to study both changes generated by human disturbances on communities and factors driving their dynamics. Our results highlight the long-term impact of anthropogenic activities on biological communities, in relation to the alteration of environmental characteristics.