



**APPROCHE METAGENOMIQUE POUR L'ETUDE DE LA  
BIODIVERSITE TOTALE DU SOL**

**MetaGENOMIC insights into the study of total biodiversity  
of SOILS**

Thibaud Decaëns  
*Laboratoire d'Ecologie (EA 1293) – Université de Rouen*  
*thibaud.Decaëns@univ-rouen.fr*

**Synthèse du rapport final**  
*05 mars 2014*

## Sommaire

2	
1.	Contexte général..... 3
2.	Objectifs généraux du projet ..... 3
3.	Quelques éléments de méthodologie ..... 4
3.1	Echantillons de sol et d'invertébrés..... 4
3.2	Constitution des bibliothèques de référence..... 4
3.3	Séquençage environnemental de l'ADN extracellulaire – Protocole MetaBar ..... 4
3.4	Séquençage environnemental de l'ADN intracellulaire ..... 5
4.	Résultats obtenus..... 5
4.1	Bibliothèques de référence ..... 5
4.2	Séquençage environnemental de l'ADN extracellulaire – Protocole MetaBar ..... 6
4.3	Séquençage environnemental de l'ADN intracellulaire ..... 9
5.	Implications pratiques, recommandations, réalisations pratiques, valorisation ..... 10
5.1	Implications pratiques ..... 10
5.2	Recommandations ..... 11
5.3	Réalisations pratiques et valorisation ..... 11
6.	Partenariats mis en place, projetés, envisagés ..... 12
7.	Liste des opérations de valorisation et de transfert issues du contrat..... 12
	Rapports de stage..... 13
8.	BIBLIOGRAPHIE..... 14

## 1. CONTEXTE GENERAL

Les sols sont parmi les habitats les plus riches en espèce au sein des écosystèmes terrestres (Swift et al. 1979). A une échelle globale, ils renferment environ 25% de la biodiversité actuellement décrite (Decaëns et al. 2006, 2008). Dans les sols tempérés, un mètre carré de sol peut abriter jusqu'à 1000 espèces d'invertébrés (Schaefer and Schauerermann 1990), et un gramme de sol peut renfermer plusieurs milliers de génotypes bactériens et fongiques (Torsvik et al. 1994, Hawksworth 2001). Cependant, les organismes du sol ont été jusqu'à présent peu pris en considération par les écologistes et les taxonomistes (Decaëns et al. 2006) et des questions fondamentales, relatives notamment aux niveaux globaux de biodiversité du sol restent encore sans réponse, particulièrement pour les espèces de petite taille (Wall et al. 2005). Les sols ont ainsi été décrits comme la dernière frontière biotique après les abysses océaniques et la canopée des forêts équatoriales (André et al. 1994, 2002, Giller 1996, Hagvar 1998, Blaxter et al. 2005). Cet état de fait est particulièrement préoccupant dans le contexte de l'étude de l'érosion de la biodiversité, car les estimations actuelles de niveaux d'extinction d'espèces se sont historiquement basées sur des résultats obtenus pour les groupes taxonomiques les mieux connus, et ont systématiquement ignoré la grande diversité des formes de vie endogées (Wilson 2002, 2008). L'existence d'un tel biais taxonomique a probablement conduit à une sous-estimation significative des niveaux actuels d'érosion de la biodiversité Wilson et al. (2007). La plupart des études suggèrent une vulnérabilité importante de cette biocénose à l'impact des contraintes environnementales liées aux activités humaines, notamment celles liées aux changements d'usage des sols (Decaëns 2010). Cependant ces études se sont généralement concentrées sur un nombre limité de taxons ou au mieux sur l'une des catégories morphologiques d'organismes du sol (microflore, mésofaune ou macrofaune). A ce jour, il n'existe pas d'études ayant abordé la question de l'érosion de la biodiversité des sols dans une approche plus holistique, ou tout au moins en considérant un spectre taxonomique plus important. Parmi les difficultés rencontrées, on peut noter (1) l'existence d'une forte variabilité intraspécifique sur le plan phénotypique; (2) des niveaux de diversité cryptique élevés; (3) des clefs d'identification morphologiques fréquemment utilisables uniquement pour stade de développement, un genre ou une caste donnés; (4) un niveau d'expertise scientifique élevé exigé pour l'utilisation de ces clefs ; (5) une diminution régulière des taxonomiques spécialistes de ces groupes au sein de la communauté scientifique.

Notre projet met en œuvre, pour la première fois, dans une étude plus globale de la biodiversité du sol, les approches de barcode classique et de metabarcoding qui sont hautement complémentaires. Le barcode ADN utilise la technologie conventionnelle de séquençage (Sanger) pour obtenir individuellement des séquences pour des spécimens précis, et elle présente donc des limites concernant le nombre d'organismes pouvant être analysés, notamment pour les petits organismes. Le séquençage environnemental permet au contraire d'obtenir des séquences à partir d'un grand nombre d'individus mélangés. D'un autre côté, les bibliothèques de référence de barcodes ADN permettent l'identification à l'espèce de l'énorme volume de séquences inconnues obtenues par le séquençage environnemental. Grâce à cette combinaison, nous levons le verrou taxonomique rendant approximative voire impossible l'estimation des niveaux de biodiversité pour un spectre taxonomique complet d'organismes édaphiques.

## 2. OBJECTIFS GENERAUX DU PROJET

Le projet GENOSOIL propose d'utiliser la génomique comme alternative ou complément aux approches traditionnelles d'étude de la biodiversité du sol qui se basent sur des identifications morphologiques des organismes. Les approches morphologiques présentent de sérieuses limites dans le cas des organismes édaphiques. Elles ne sont en effet utilisables que pour les métazoaires, requièrent fréquemment des observations microscopiques incompatibles avec l'étude d'un grand nombre d'individus, sont difficiles à mettre en œuvre pour les stades juvéniles. L'utilisation des outils moléculaires dans le cadre d'une méthode standardisée applicable indifféremment sur l'ensemble des organismes du sol permet de contourner ces difficultés.

Ce projet a donc pour objectif principal de développer une méthode de caractérisation de la biodiversité du sol en s'appuyant sur une combinaison d'approches de taxonomie moléculaire basées sur le barcode ADN et le metabarcoding (barcode environnemental à partir d'échantillons totaux d'invertébrés et/ou d'échantillons de sol). La démarche s'appuie sur (1) la constitution de bibliothèques de référence de barcodes ADN dans le cadre des campagnes internationales de barcode (projet iBOL-<http://ibol.org/>) pour une sélection de groupes d'invertébrés clefs ; (2) l'élaboration de protocoles d'analyse de l'ADN environnemental du sol à l'aide des outils de séquençage de nouvelle génération (Roche 454 et Illumina MiSeq). L'élaboration du protocole a été effectuée à l'aide d'échantillons

d'invertébrés et de sol prélevés dans différents gradients environnementaux caractéristiques des écosystèmes du Nord-Ouest de la France.

### 3. QUELQUES ELEMENTS DE METHODOLOGIE

#### 3.1 Echantillons de sol et d'invertébrés

Pour chaque type d'écosystème (prairie ou forêt), trois répétitions indépendantes ont été choisies, et trois points d'échantillonnage distants de 20m les uns des autres ont été positionnés. Sur chacun de ces points, une série d'échantillons a été prélevée de façon à permettre la collecte des vers de terre, de la macrofaune et de la mésofaune. Une carotte de sol supplémentaire a été prélevée en prévision de l'analyse environnementale sur la matrice de sol (Figure 1).

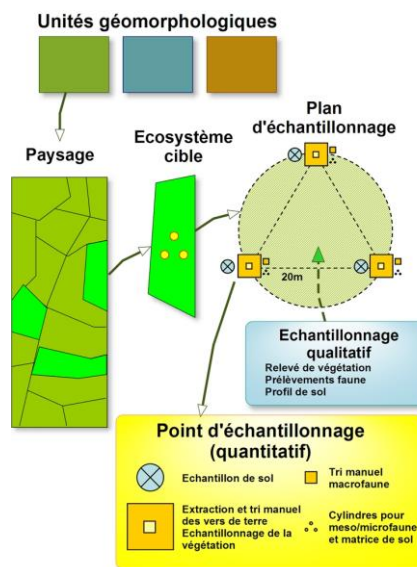


Fig. 1 Plan d'échantillonnage des organismes du sol, de la végétation et du sol. La description géomorphologique et paysagère a été utilisée pour localiser les écosystèmes cibles (répliqués trois fois chacun).

#### 3.2 Constitution des bibliothèques de référence

Elles ont été obtenues à partir de séquences de spécimens appartenant à divers groupes d'invertébrés du sol ainsi que des groupes dont les larves se développent dans ce compartiment des écosystèmes. Le fragment de gène séquencé est celui qui est utilisé en routine pour le barcode des animaux (en 5' du gène mitochondrial COI – Hebert et al. 2003). Les séquences ont été déposées sur la plateforme bioinformatique en ligne BOLD (Barcode Of Life Data system) dans le jeu de données en accès public 'Bibliothèques de référence GENOSOIL -Invertébrés du sol' [DS-BGENO].

#### 3.3 Séquençage environnemental de l'ADN extracellulaire – Protocole MetaBar

Ce protocole permet de séquencer des fragments d'ADN dit extracellulaires : il s'agit d'ADN fragmenté qui n'est pas extrait des tissus des organismes, mais qui est celui directement disséminé dans l'environnement après la mort et la décomposition de ceux-ci. Ces ADN extracellulaires sont plus fragmentés que ceux qu'on peut trouver dans une cellule encore intacte, c'est pourquoi la taille du fragment cible amplifiée est peu importante (100pb). Ainsi, ce protocole permet de détecter la présence d'un organisme sur un point d'échantillonnage même si aucun de ces représentants vivant n'est présent dans l'aliquot de sol prélevé.

Le site ciblé est une des prairies pâturée de 5 ans d'âge du Lycée agricole d'Yvetot. Vingt-quatre carottes de sols ont été prélevées sur la parcelle selon un maillage régulier de 20m de côté

environ. Tamisées, homogénéisées, puis conservées 3 ans à sec avant d'être aliquotées pour l'analyse.

L'extraction d'ADN total s'effectue à partir d'aliqots de sol de 1g sur les 5g déjà échantillonnés, suit une amplification d'un fragment cible de 100pb du gène nucléaire 18S et un séquençage sur une plateforme Illumina MiSeq. Après filtrage des séquences trop courtes et des séquences chimères, les 100 haplotypes les plus abondants ont été comparés aux séquences de 18S correspondantes à la portion amplifiée par le protocole MetaBar présentes sur Genbank. Cela a permis d'établir une correspondance au niveau des grands embranchements, classes et ordres.

### 3.4 Séquençage environnemental de l'ADN intracellulaire

Cette technique met en œuvre l'amplification de fragments du barcode standard COI de taille importante (250 et 320 pb). Avec une telle taille de marqueur, on se positionne ici non plus sur les fragments d'ADN à l'état libre (extracellulaire) comme dans le protocole précédent, mais sur de l'ADN d'animaux pris vivants ou entrés récemment en stade de décomposition dans l'échantillon prélevé (ADN intracellulaire).

Un point a été échantillonné par habitat en forêt d'Eawy, ainsi qu'en prairie et en forêt de coteau calcaire de St Adrien avec trois pseudoréplicats distant de 20m sur chaque point. Chaque pseudoréplicat est constitué d'une carotte de sol de 5cm de diamètre sur 10cm de longueur.

Après élimination des séquences chimères ou trop courtes, celles-ci ont été comparées avec la base de données nucléotides de Genbank (version Novembre 2013) et la bibliothèque de référence GENOSOIL.

## 4. RESULTATS OBTENUS

### 4.1 Bibliothèques de référence

Vingt et un groupes ont été échantillonnés à la fois dans les milieux prairiaux et forestiers pour un total de 2338 spécimens représentant environ 500 espèces ou unités taxonomiques opérationnelles moléculaires (MOTUs) définis par un seuil de variabilité génétique variable d'un groupe taxonomique à l'autre (Table 1).

Groupe taxonomique	Effectif	# MOTUs
Diplura	1	1
Lepidoptera	4	1
Orthoptera	5	3
Thysanoptera	5	3
Psocoptera	6	1
Pseudoscorpiones	9	3
Homoptera	16	5
Enchytraeidae	18	12
Opiliones	19	4
Gastropoda	26	10
Hymenoptera	43	8
Hemiptera	44	29
Isopoda	68	4
Plectidae	68	1
Myriapoda	92	35
Coleoptera	153	24
Diptera	178	49
Araneae	221	100
Acarina	225	85
Collembola	401	80
Lumbricidae	736	36

Table 1 Liste des principaux groupes taxonomiques échantillonnés, effectifs des spécimens barcodés pour alimenter les bibliothèques de référence, et nombre d'unités taxonomiques opérationnelles moléculaires (MOTUs) pour chacun d'entre eux.

L'avantage de cette approche par rapport aux estimations classiques basées sur des identifications morphologiques est qu'elle permet l'identification et donc la prise en compte des

juvéniles (Richard et al., 2010) mais aussi des spécimens fragmentés au moment de l'échantillonnage. Elle permet également de différencier les espèces cryptiques pour lesquelles les critères morphologiques sont inapplicables (morphologie semblable mais divergence génétique aussi important qu'entre espèces distinctes morphologiquement). Le cas des vers de terre a été l'un des plus fouillés et a révélé un taux de diversité cryptique de 38,8% à 45% pour la Haute Normandie, soit de 16 à 20 espèces nouvelles jusque là ignorées et confondues avec celles déjà décrites (Figure 2) (James et al., 2010, Porco et al., 2013). Si un tel biais est attendu dans des milieux tropicaux ou des groupes peu explorés, il est plus surprenant de la retrouver dans des régions tempérées maintes fois étudiées et de surcroît pour des groupes emblématiques comme les vers de terre.

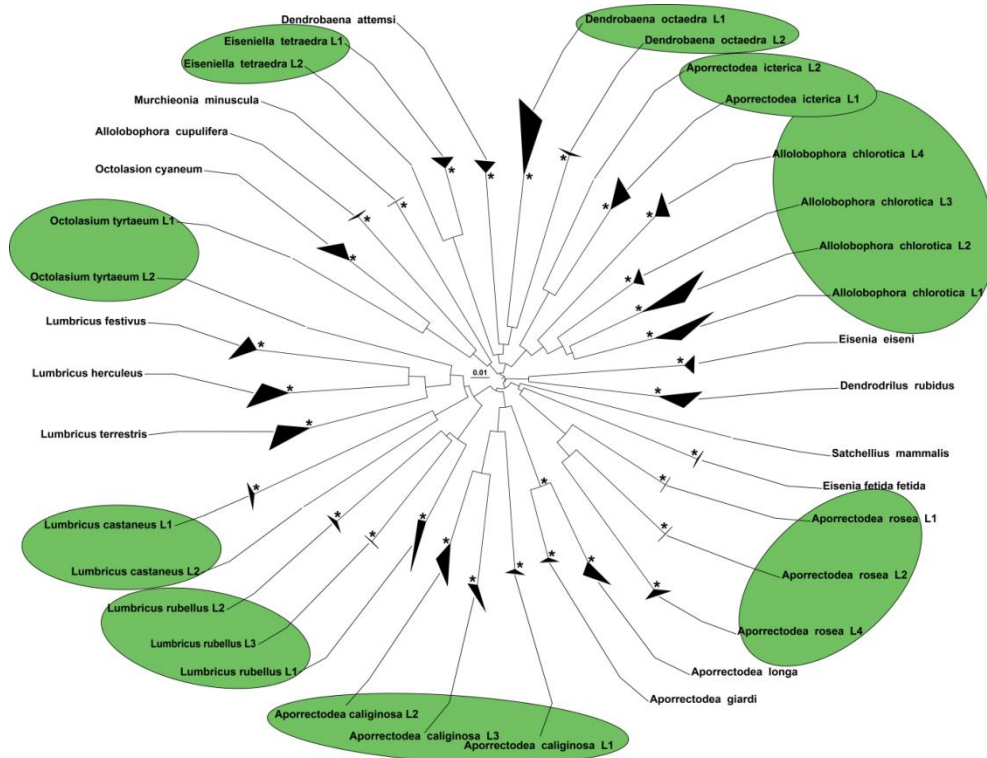


Fig. 2 Arbre de neighbor joining de distances génétiques k2p avec les 36 MOTUs détectées. Les nœuds soutenus par des valeurs de bootstrap  $\geq 99\%$  sont marqués par un astérisque. Les MOTUs cryptiques au sein d'une même espèce morphologique sont entourées d'une ellipse verte.

Ces MOTUs cryptiques sont d'autant plus importantes à répertorier et à prendre en compte qu'elles présentent pour certaines des préférences d'habitat et donc potentiellement des différences écologiques et biologiques intrinsèques. A l'inverse, dans bon nombre de localités, on a retrouvé plusieurs des MOTUs cryptiques d'une même espèce en sympatrie, suggère un évitement compétitif au niveau de l'exploitation de la ressource. Cette potentielle spécificité trophique met aussi en exergue le besoin urgent de caractérisation et de prise en compte de ces MOTUs cryptiques pour la compréhension du fonctionnement des sols.

La constitution de bibliothèques de référence exhaustives, qui sont un prérequis nécessaire à l'application de méthode de métabarcoding, donne un bon exemple de ce que peut concrètement apporter cette première phase à elle seule. De plus, cet outil, utilisé en routine, permettra de recommencer la tâche d'accumulation des données biologiques et écologiques sur chacune entité cryptique qui constituent le groupe d'espèces. On retrouvera alors l'utilité initiale du référencement transversal que constituent les noms d'espèces dans la littérature scientifique, et incidemment une meilleure compréhension du compartiment sol des écosystèmes.

## 4.2 Séquençage environnemental de l'ADN extracellulaire – Protocole MetaBar

La variation du gène nucléaire 18S est moins importante que celle observée pour les gènes mitochondriaux utilisés en routine pour le barcode. Ces données permettent néanmoins dans un premier temps d'obtenir d'importantes informations sur la présence, et à un certain niveau sur

l'abondance des grands groupes présents dans les sols (Figure 3, Table 2). On obtient donc une série d'haplotypes identifiés de l'embranchement à l'ordre, chacun étant susceptible d'appartenir à un large panel d'espèces.

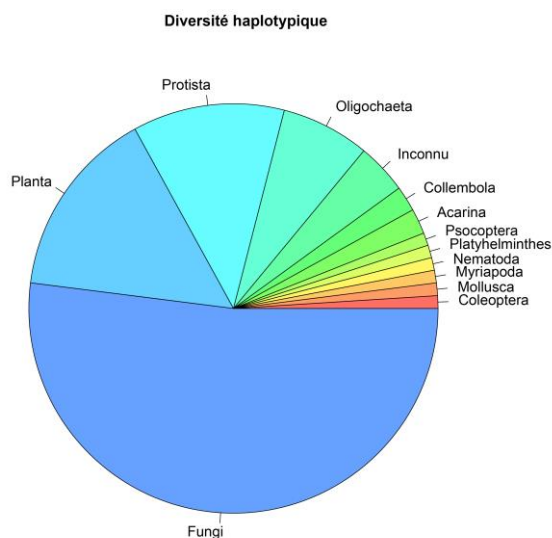


Fig. 3 Proportion des différents groupes taxonomiques détectées en nombre d'haplotypes produits.

Groupes taxonomiques	Nb haplotypes	Nb séquences
Coleoptera	1	1324
Mollusca	1	2241
Myriapoda	1	2247
Nematoda	1	1421
Platyhelminthes	1	2126
Psocoptera	1	1542
Acari	2	9927
Collembola	2	3003
Inconnu	4	7232
Oligochaeta	7	75781
Protista	12	23233
Planta	15	162585
Fungi	52	350643

Table 2 Nombre de séquences et d'haplotypes obtenus par l'analyse pour les différents groupes taxonomiques dont les correspondances ont été établies avec les séquences disponibles sur Genbank.

Les haplotypes les plus amplifiés sur les 24 échantillons appartiennent aux trois groupes attendus comme dominants en termes de biomasse dans les sols hors bactéries : les champignons, les plantes et les oligochètes. Ces résultats permettent, en outre, l'obtention d'informations sur la composition relative des communautés d'invertébrés du sol pour les principaux groupes dominants en chacun des points d'échantillonnage (Figure 4).

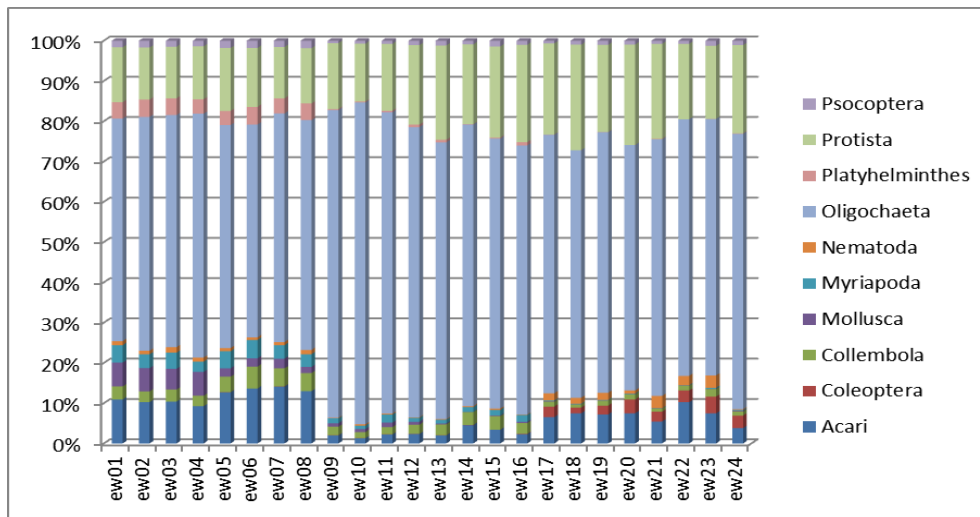


Fig. 4 Composition relative des communautés pour les 24 points d'échantillonnage.

Ceci permet également, dans une certaine mesure, d'inférer le niveau de densité de population. La visualisation des proportions de séquences aux différents points d'échantillonnage de la parcelle, donne une cartographie de la présence des groupes cibles (Figure 5). Cela met en évidence l'hétérogénéité spatiale de distribution selon les groupes, et suggère, pour certains taxa, leurs probables relations trophiques.

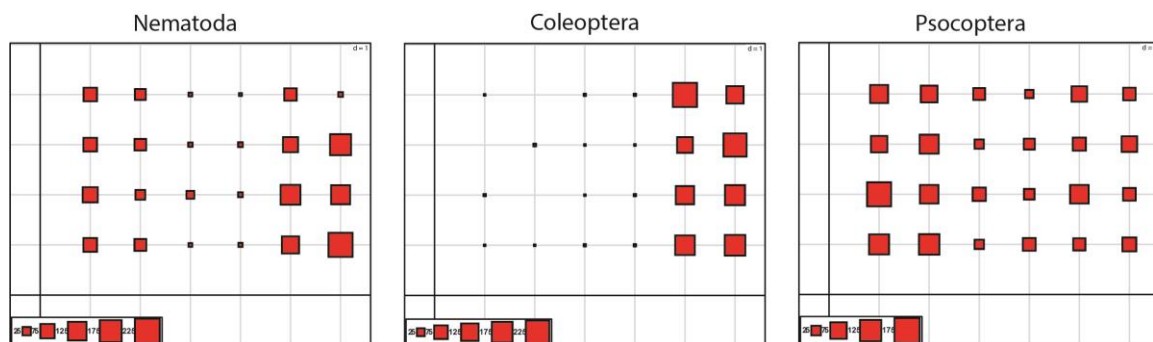


Fig. 5 Exemples de distribution spatiale du nombre de séquences selon la grille d'échantillonnage des sols pour trois des groupes taxonomiques détectés.

Les résultats obtenus avec ce dispositif expérimental montrent que l'on peut potentiellement utiliser des sols secs stockés depuis plusieurs années pour recouvrir des traces d'ADN permettant de remonter à la composition des communautés. La confrontation systématique des données de communautés avec de nombreux éléments (biotiques et abiotiques) à une échelle étendue, permettrait alors d'émettre de solides hypothèses quant aux facteurs clefs qui déterminent leur structuration. En outre, on pourra également utiliser des dispositifs expérimentaux qui mettraient en évidence la dynamique spatio-temporelle de structuration des communautés d'invertébrés du sol à une échelle courte, saisonnière, ou bien longue, pluriannuelle, permettant de mesurer entre autres l'impact de changements climatiques ou d'usage des sols.

Ici, l'essai a été effectué avec un fragment de gène nucléaire dont la variabilité ne permet qu'une discrimination au niveau taxonomique de l'embranchement à l'ordre, mais le succès de cette démarche permet d'envisager l'application de techniques impliquant des gènes mitochondriaux tel que COI (ou 16S – gène utilisé dans le protocole MetaBar en ce qui concerne les vers – Bienert et al. 2012) pour lesquels nous avons de nombreuses séquences de référence, ou en tout cas les extraits d'ADN qui permettraient de les constituer rapidement pour un autre marqueur. L'utilisation de marqueurs mitochondriaux rendra alors possible l'établissement de listes d'espèces, ou du moins l'obtention, avec le dénombrement des MOTUs, d'un proxy de la diversité spécifiques dans différents groupe cibles dont les bibliothèques de référence n'auront pas été constituées.



### 4.3 Séquençage environnemental de l'ADN intracellulaire

La double comparaison avec la bibliothèque de référence GENOSOIL et les séquences présentes pour COI sur Genbank ont permis de détecter une partie de la diversité présente dans les carottes de sol prélevées pour les principaux groupes cibles du projet au niveau spécifique (Table 3), mais aussi au niveau familial et ordinal. Ces informations ont rendu possible la comparaison des communautés présentes dans trois types d'habitats caractéristiques de Haute Normandie. On remarque, pour les trois niveaux taxonomiques, une forte hétérogénéité dans la mesure des communautés entre les pseudoréplicats de chaque habitat, avec en moyenne 78 % des espèces, 66 % des familles et 24 % des ordres ayant été retrouvés dans un seul des pseudoréplicats pour un habitat donné. Ceci montre la nécessité absolue d'effectuer plusieurs pseudoréplicats pour chaque point échantillonné, et du même coup met en exergue le besoin d'un calibrage du nombre de pseudoréplicats nécessaires pour optimiser la capture de la diversité en invertébrés du sol présente au niveau d'un point d'échantillonnage.

					EA20			SA70			SABO					
					1	2	3	1	2	3	1	2	3			
Annelida	Haplotaxida	Enchytraeidae	<i>Cognettia</i>	<i>cognetti</i>	0	0	0	0	0	0	0	65	0			
			<i>Enchytraeus</i>	<i>bulbosus</i>	0	0	0	0	0	0	2	2	0	0		
				<i>norvegicus</i>	24	0	65	24	0	0	0	0	0	0		
				<i>galba</i>	0	0	4	0	0	0	0	0	0	0		
			<i>Fridericia</i>	<i>isseli</i>	25	0	0	25	26	0	0	0	0	0		
				<i>magna</i>	0	30	0	0	0	0	0	0	0	0		
				<i>nasuta</i>	0	0	0	0	3	0	0	0	0	0		
			<i>Henlea</i>	<i>perpusilla</i>	0	0	0	0	0	0	0	225	0	0		
				<i>clavata</i>	32	0	0	32	0	0	0	0	0	0		
				<i>cambrensis</i>	1260	2034	6806	1260	0	6	0	370	0	0		
			Lumbricidae	<i>Apporectodea</i>	0	0	0	0	61	8	0	0	0	0		
				<i>Satchellius</i>	0	0	0	0	875	0	0	0	0	0		
				<i>multifasciata</i>	2	2	4	2	0	0	0	0	0	0		
			Arthropoda	Collembola	Entomobryidae	<i>Lepidocyrtus</i>	<i>lanuginosus</i>	0	0	0	0	5	0	0	0	5
						<i>Orchesella</i>	<i>cincta</i>	0	0	18	0	3	0	0	0	0
Hypogastruridae	<i>Ceratophysella</i>	<i>denticulata</i>				2	0	0	2	0	0	0	0	0		
	<i>Willemia</i>	<i>denisi</i>			0	2	0	0	0	0	0	0	0			
Isotomidae	<i>Folsomia</i>	<i>quadrioculata</i>			2	0	0	2	0	0	0	0	0			
	<i>Isotomiella</i>	<i>minor</i>			12	2	0	12	0	0	0	0	0			
	<i>Parisotoma</i>	<i>notabilis</i>			35	75	0	35	119	41	120	0	0			
Araneae	Agelenidae	<i>Tegenaria</i>			<i>agrestis</i>	0	0	2	0	0	0	0	0	0		
		<i>Dysdera</i>			<i>alegranzaensis</i>	0	0	0	0	0	0	2	0	0		
		<i>Megalephyphantes</i>			<i>nebulosus</i>	0	0	0	0	0	0	0	0	2		
Chilopoda	Lithobiidae	<i>Lithobius</i>			<i>fortificatus</i>	0	85	0	0	0	0	0	0	0		
		Coleoptera			Elateridae	<i>Agrypnus</i>	<i>murinus</i>	0	0	0	0	0	0	37	0	
						<i>Athous</i>	<i>puncticollis</i>	0	0	0	0	0	0	0	28	0
<i>Meligethes</i>	<i>erysimicola</i>					0	0	2	0	0	0	0	0	0		
Diptera	Chironomidae	<i>Ablabesmyia</i>			<i>monilis</i>	0	0	0	0	0	0	2	0	0		
		<i>Culicidae</i>	<i>Anopheles</i>	<i>pristinus</i>	2	0	0	2	0	0	0	0				
		<i>Tephritidae</i>	<i>Urophora</i>	<i>cardui</i>	0	3	0	0	0	0	0	0				
Geophilomorpha	Schendylidae	<i>Schendyla</i>	<i>nemorensis</i>	0	0	0	0	106	0	0	0					
Hemiptera	Acanthosomatidae	<i>Acanthosoma</i>	<i>haemorrhoidale</i>	0	0	0	0	0	0	0	0	3				
		<i>Miridae</i>	<i>subpulchellus</i>	0	2	0	0	0	0	0	0	0				
		<i>Charmon</i>	<i>extensor</i>	0	0	0	0	0	0	2	0	0				
Hymenoptera	Braconidae	<i>Ancistrocerus</i>	<i>parietum</i>	0	0	0	0	0	0	17	0	0				
		<i>Pandemis</i>	<i>cerasana</i>	0	13	0	0	0	0	0	0	0				
		<i>Locusta</i>	<i>migratoria</i>	0	0	0	0	0	0	5	0	2				
Lepidoptera	Achipteridae	<i>Achipteria</i>	<i>coleoprata</i>	0	0	2	0	0	0	0	0	0				
		<i>Camisiidae</i>	<i>Platynothus</i>	<i>peltifer</i>	0	6	3	0	20	0	0	0				
		<i>Eremaeidae</i>	<i>Eueremaeus</i>	<i>oblongus</i>	2	0	0	2	0	0	0	0				
Orthoptera	Hypochthoniidae	<i>Hypochthonius</i>	<i>rufulus</i>	0	0	0	0	0	0	18	0	0				
		<i>Nothrus</i>	<i>silvestris</i>	6	0	0	6	0	0	0	3	0				
		<i>Scutoverticidae</i>	<i>Scutovertex</i>	<i>sculptus</i>	5	0	0	5	0	0	0	0				

Table 3 Liste des espèces détectées dans les différents pseudoréplicats des trois habitats échantillonnés. Le nombre de séquences est figuré avec un gradient de couleur en fonction de l'abondance recouvrée pour chaque espèce.

Dans les trois types d'habitats, les groupes dominants dans les analyses sont les acariens, les collemboles, les enchytraeides et les diptères. La comparaison des communautés montre que très peu d'espèces sont partagées entre les trois habitats cibles. On a donc une bonne caractérisation au niveau des communautés d'espèces recouvrées pour les trois types d'habitat, caractérisation qui s'érode naturellement avec la dégradation du niveau taxonomique, mais qui reste néanmoins très importante pour les familles et les genres d'insectes, de myriapodes et d'acariens (groupés sous 'Autres Arthropodes') (Figure 6).

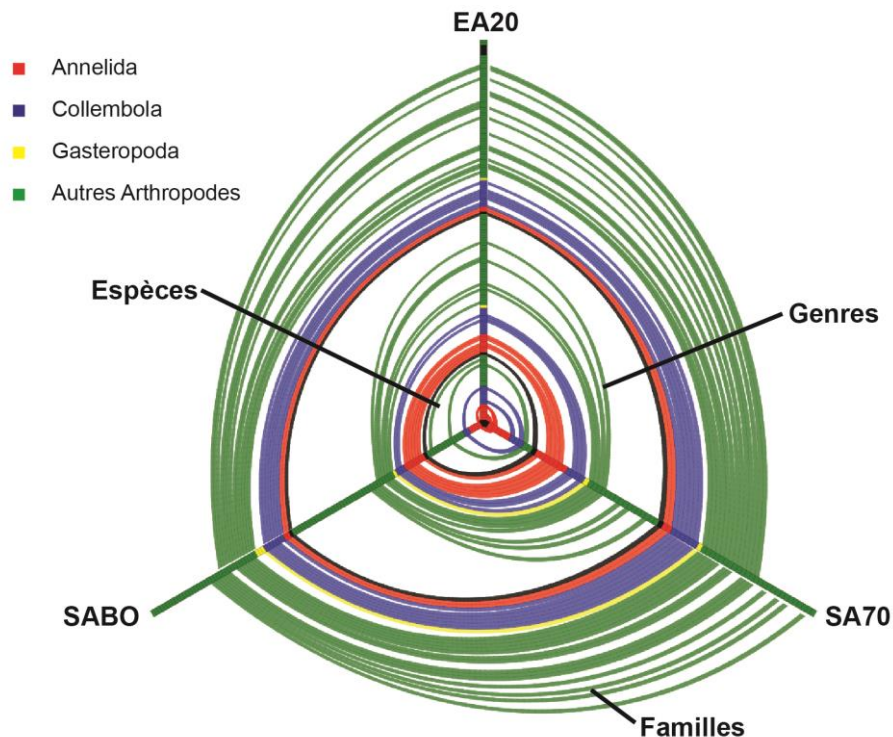


Fig. 6 Projection sur les 3 axes des communautés d'invertébrés recouvrées dans les 3 types d'habitat pour les niveaux taxonomiques espèces, genres et familles. Les arcs de cercles représentent les taxa communs entre les types d'habitat.

Il ressort de ces premiers résultats que la technique d'échantillonnage peut être encore être optimisée ainsi que calibrées suivant les groupes en termes de pression d'échantillonnage/nombre de pseudoréplicats (en particulier pour les éléments de grande taille de la macrofaune comme les vers de terre), et que la complétude des bibliothèques de références doit être augmentée de façon à obtenir une meilleure caractérisation des communautés. Il est cependant à noter que le niveau de diversité spécifique de certains groupes pourra être inféré de manière plus approfondie dans un second jeu d'analyses avec un seuil de divergence génétique fixé par rapport à ce qui a été observé précédemment dans ces mêmes groupes. On obtiendrait alors un proxy de la diversité spécifique en nombre de MOTUs. Cette technique permettra ainsi de caractériser les niveaux de diversité spécifique de nombreux groupes et de comparer la composition relative de leurs communautés sans avoir recours à une bibliothèque de référence détaillée.

## 5. IMPLICATIONS PRATIQUES, RECOMMANDATIONS, REALISATIONS PRATIQUES, VALORISATION

### 5.1 Implications pratiques

L'utilisation de ces approches en routine liée à des protocoles d'échantillonnage standardisés permettra d'établir des indicateurs de la santé des sols pour mesurer objectivement les effets de politiques de restauration, de réaménagement, de changement d'utilisation des terres ou de l'installation d'infrastructures. On pourra alors avoir une vision en termes de conservation et de bioindication liée à un compartiment des écosystèmes pour lequel de telles approches restaient difficiles voire impraticables.

L'autre application importante qui peut découler de l'utilisation de ce panel de techniques est la détection d'éléments invasifs dans les communautés des sols. L'utilisation en routine sur des sites sentinelles permettrait d'établir une veille quant à l'apparition et la progression de tels éléments. En

outre cela produirait des informations précieuses concernant l'impact de ces espèces invasives sur les communautés d'invertébrés édaphiques, que ce soit via la compétition pour la ressource, la prédation directe ou encore la transformation physico-chimique de la matrice des sols ; différents facteurs qui pourront être mesurés et corrélés à la pression des propagules d'invasifs. L'application de veille sanitaire peut aussi être envisagée de la même façon pour des ravageurs agricoles et sylvicoles (par exemple pour les nématodes phytoparasites). Cette démarche fournirait des informations de premier choix aux services régionaux de protection des végétaux et permettrait d'élaborer des stratégies de contrôle appropriées.

On voit donc que le développement de ces approches, et la standardisation des protocoles attendant, en permettant leur utilisation en routine, produiraient une masse d'informations importantes, quasiment en temps réel, qui éclairerait à l'aide d'indicateurs objectifs la prise de décisions pour une gestion plus efficace de l'usage des sols et des ressources attenantes.

## 5.2 Recommendations

Quatre points importants sont à considérer pour une utilisation et la valorisation optimale de ces approches :

- La complétion des bibliothèques de références concernant les groupes d'intérêt, cela en plus de révéler leur réel niveau de diversité (cf. le cas des vers de terre), permettra d'obtenir une liste d'espèces et donc de lier cette liste aux informations accumulées sur celles-ci dans la littérature scientifique.
- Calibrer le protocole d'échantillonnage environnemental de façon à trouver un volume d'échantillonnage adéquat pour une capture représentative de la diversité.
- Le séquençage en parallèle pour un même échantillon d'autres gènes employés spécifiquement comme barcode dans d'autres groupes essentiels dans le fonctionnement des sols: 18S pour les nématodes, 16S pour les bactéries et ITS pour les champignons. En élargissant le spectre de l'analyse, cela enrichirait considérablement les conclusions au niveau des problématiques liées aux communautés mais aussi à leur fonctionnement. Pour ces groupes édaphiques très importants, mais pour lesquels la constitution de bibliothèques de référence est difficile voire impossible car ces organismes sont pour la plupart ni décrits ni cultivables, l'approche d'estimation de la diversité spécifique par la définition de MOTUs est la meilleure stratégie.
- Le croisement systématique des données obtenues avec des données de végétations, de caractéristiques physico-chimiques du sol, ou attenantes aux modes de gestion, permettra de valoriser grandement ces données en les remettant dans le contexte d'études plus larges et pluridisciplinaires. Par exemple, ce type d'approche est particulièrement prometteur si on se place dans une optique d'agrégation avec la masse des données compilées dans le cadre du Réseau de Mesures de la Qualité des Sols (RMQS).

## 5.3 Réalisations pratiques et valorisation

Les activités menées dans le cadre de GENOSOIL ont donné lieu à plusieurs stages de licence, master et post-doctorat. David Porco a ainsi été recruté dès le début du projet en tant que post-doc pour une période de deux ans (première année financée par la région Haute Normandie, l'autre par le projet), et Samuel James a également travaillé en partie sur les bibliothèques de barcodes des vers de terre dans le cadre de sa bourse « Marie Curie France Régions ». Deux étudiantes de M2 ont participé à la description des sites d'étude par télédétection. Une étudiante de M2 et trois étudiants de L3 ont travaillé à l'élaboration des bibliothèques de barcodes ADN.

Les premiers résultats obtenus (phase préliminaire et phase financée par GESSOL 3) ont d'ores et déjà été fortement valorisés. Au total, 6 articles ont été publiés dans des journaux internationaux, et un manuscrit en cours de préparation sera soumis dans les mois prochains. La plupart de ces premières publications (6/7) portent sur la valorisation des bibliothèques de barcodes ADN. Une publication concerne le développement d'outils statistiques pour l'analyse des données du projet. Le projet dans son ensemble et ses résultats préliminaires ont également été présentés dans 5 communications orales (dont 3 conférences invitées) et 4 communications affichées dans des colloques internationaux.

L'ensemble des séquences de référence de COI obtenues a été placé dans des projets spécifiques sur la plateforme bioinformatique BOLD (jeu de données 'Bibliothèques de référence GENOSOIL -Invertébrés du sol' [DS-BGENO]). L'ensemble de ces données est en accès libre avant la fin du projet GENOSOIL, de façon à servir de base de référence pour l'utilisation des approches métagénomiques développées dans le projet GENOSOIL ou dans d'autres comme par exemple le projet MetaBar coordonné par Pierre Taberlet.

## 6. PARTENARIATS MIS EN PLACE, PROJETS, ENVISAGES

L'exploitation et la valorisation des résultats va se poursuivre avec l'équipe de Mehrdad Hajibabaei de l'université de Guelph. La réalisation de ce projet nous a également permis d'explorer différentes voies et notamment de rentrer en contact avec Pierre Taberlet (Université de Grenoble - LECA) sur des approches d'ADN extracellulaires. Plusieurs collaborations avec cette équipe sont en développement concernant la constitution de bibliothèques de référence directement par des moyens de séquençage massif parallèle (ce qui ferait grandement chuter le coût de celles-ci), la caractérisation de régimes alimentaires de certains groupes de la faune du sol, mais aussi le développement de la méthode appliquée à des pédothèques comme celle du RMQS.

## 7. LISTE DES OPERATIONS DE VALORISATION ET DE TRANSFERT ISSUES DU CONTRAT

### PUBLICATIONS SCIENTIFIQUES

#### Publications scientifiques parues

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.

Porco D., Decaëns T., Deharveng L., James S.W., Skarżyński D., Erséus C., Butt K., Richard B., Hebert P. (2013) Biological invasions in soil: DNA barcoding as a monitoring tool in a multiple taxa survey targeting European earthworms and collembolans in North America. *Biological Invasion*, 15, 899-910.

Dupont L., Lazrek F., Porco D., King R.A., Rougerie R., Symondson W.O.C., Livet A., Richard B., Decaëns T., Butt K.R., Mathieu J. (2011) New insight into the genetic structure of the *Allolobophora chlorotica* aggregate in Europe using microsatellite and mitochondrial data. *Pedobiologia*, 54(4), 217-224.

Rossi J.-P. (2011) rich: an R package to analyse species richness. *Diversity* 3, 112-120.

#### *Antérieures à 2011 (phase préliminaire de GENOSOIL)*

James S., Porco D., Decaëns T., Richard B., Rougerie R., Erséus C. (2010) DNA Barcoding Reveals Cryptic Diversity in *Lumbricus terrestris* L., 1758 (Clitellata): Resurrection of *L. herculeus* (Savigny, 1826). *PLoS One*, 5, e15629.

Richard B., Decaëns T., Rougerie R., James S., Porco D., Hebert P. (2010) Re-integrating earthworm juveniles into soil biodiversity studies: species identification through DNA barcoding. *Molecular Ecology Resources*, 10, 606-614.

#### Publications scientifiques à paraître

Porco D., Dupont L., Bodilis J., Chang C. H., Dolédec S., Dubs F., Hajibabaei M., James S., Lavelle P., Oszwald J., Richard B., Rossi J.-P., Rougerie R., Roy V., Sebag D., Shokralla S., Decaëns T. A reference library of DNA barcodes for the earthworm of upper Normandy: Biodiversity assessment and cryptic diversity. En préparation.

### COLLOQUES

### Participations passées à des colloques

Decaëns T., Porco D., James S., Smith M.A., Rougerie R., Hajibabaei M. (2012) Barcoding the underground – recent advances and new prospects for the study of soil animal biodiversity. Conférence invitée (Keynote Talk) au « XVth International Colloquium on Soil Zoology », Coimbra, Portugal, 6-10 août 2012.

Decaëns T., Porco D., Sebag D., Oszwald J., Rougerie R., Husté A., Dupont L., Dolédec S., Rossi J.-P., Hajibabaei M. (2012) The GENOSOIL-Fr project: insights from metagenomics into the study of total biodiversity of soils. Poster au « XVth International Colloquium on Soil Zoology », Coimbra, Portugal, 6-10 août 2012.

Decaëns T., Rougerie R., James S., Porco D., Hebert P. 2012 DNA barcodes as a tool for the study of invertebrate biodiversity. Conférence invitée à l'Universidade do Estado de Santa Catarina, Lages, Brésil, 03 mars 2012.

Decaëns T., Porco D., Sebag D., Oswald J., Rougerie R., Husté A., Dupont L., Doledec S., Rossi J.-P., Hajibabaei M. (2011) The GENOSOIL-Fr project : insights from metagenomics into the study of total biodiversity of soils. Poster et lightning au « Fourth International Barcode of Life Conference », Adelaide, Australie, 30 novembre – 3 décembre 2011.

### *Antérieures à 2011 (phase préliminaire de GENOSOIL)*

Decaëns T., Porco D., James S.W. (2010) El potencial de los barcodes de DNA en los estudios taxonómicos y ecológicos con lombrices de tierra. Conférence invitée au « Encuentro Latino Americano de Ecología y Taxonomía de Oligochetos », Curitiba, Brésil, 12-14 octobre 2010.

Decaëns T., Rougerie R., Deharveng L., Porco P., James S.W., Chang C.H., Richard B., Hebert P.D.N. (2010) DNA barcodes for soil animal taxonomy: transcending the final frontier. Présentation orale, colloque « ECBOL2 - 2nd Conference of the European Consortium for the Barcode of Life », Braga, Portugal, 2-4 June 2010.

James S.W., Decaëns T., Porco D., Richard B., Erseus C., Meusnier I., Rougerie R., Hebert P.D.N. (2010) Chimerical biology: a plea for good taxonomy as exemplified by the common nightcrawler (*Lumbricus terrestris* L.). Poster au colloque « ECBOL2 - 2nd Conference of the European Consortium for the Barcode of Life », Braga, Portugal, 2-4 June 2010.

Porco D., Decaëns T., Deharveng L., James S., Butt K., Mark S., Rougerie R., Richard B., Hebert P. (2010) Barcoding invasive: a new tool for invasion monitoring in soil. Présentation orale, colloque « ECBOL2 - 2nd Conference of the European Consortium for the Barcode of Life », Braga, Portugal, 2-4 June 2010.

Richard B., Decaëns T., James S., Rougerie R., Porco D., Hebert P.D.N. (2010) Re-integrating earthworm juveniles into soil biodiversity studies: species identification through DNA barcoding. Poster, colloque « ECBOL2 - 2nd Conference of the European Consortium for the Barcode of Life », Braga, Portugal, 2-4 June 2010.

### Rapports de stage

Coupiereau D. (2012) *Evaluation de la diversité des vers de terre dans deux gradients environnementaux à l'aide du barcoding ADN*. Stage de L3 EBO (Univ Rouen), réalisé au lab ECODIV (encadrant: T. Decaëns).

Salim H. (2012) *Utilisation du barcoding ADN pour l'estimation de la richesse spécifique des invertébrés du sol dans différents écosystèmes de Normandie*. Stage de M2 Environnement Continental et Hydrosociété, parcours Sol (AgroParisTech/UPMC), réalisé au lab ECODIV (encadrants: T. Decaëns & D. Porco).

James S.W. (2011) *Utilisation du barcoding ADN dans l'étude des vers de terre : mise en évidence de la diversité cryptique et amélioration de nos connaissances écologiques*. Rapport de fin de recherche, bourse « Marie Curie France Régions », laboratoire ECODIV (resp. scientifique : T. Decaëns).

Lebarque M. (2011) *Caractérisation différentielle en prairie et forêt des communautés de la mésofaune et de la macrofaune*. Stage de L3 EBO (Univ Rouen), réalisé au lab ECODIV (Encadrants : T. Decaëns & D. Porco).

Lallemand H. (2011) *Caractérisation différentielle en prairie et forêt des communautés de la mésofaune et de la macrofaune*. Stage de L3 EBO, réalisé au Lab ECODIV (Encadrants : T. Decaëns & D. Porco).

Leroux L. (2011) *Caractérisation des dynamiques paysagères entre 1947 et 2010 dans la zone de St Adrien (Normandie)*. Stage de Master 1 TASE (Université Rennes 2) réalisé au laboratoire COSTEL (encadrant : J. Oszwald).

Poirier E. (2011) *Caractérisation des dynamiques paysagères entre 1947 et 2010 dans la zone de Eawy (Normandie)*. Stage de Master 1 TASE (Université Rennes 2) réalisé au laboratoire COSTEL (encadrant : J. Oszwald).

## 8. BIBLIOGRAPHIE

- André, H. M., Ducarme, X., & Lebrun, P. (2002). Soil biodiversity: Myth, reality or conning? *Oikos*, 96, 3-24.
- André, H. M., Noti, M. I., & Lebrun, P. (1994). The soil fauna - the other last biotic frontier. *Biodiversity and Conservation*, 3, 45-56.
- Bienert, F., De Danieli, S., Miquel, C., Coissac, E., Poillot, C., Brun, J. J., et al. (2012). Tracking earthworm communities from soil DNA. *Molecular Ecology*, 21, 2017-2030.
- Blaxter, M., Mann, J., Chapman, T., Thomas, F., Whitton, C., Floyd, R., et al. (2005). Defining operational taxonomic units using DNA barcode data. *Royal Society Philosophical Transactions Biological Sciences*, 360, 1935-1943.
- Decaëns, T. (2010). Macroecological patterns in soil communities. *Global Ecology and Biogeography*, 19, 287-302.
- Decaëns, T., Jimenez, J. J., Gioia, C., Measey, G. J., & Lavelle, P. (2006). The values of soil animals for conservation biology. *European Journal of Soil Biology*, 42, S23-S38.
- Decaëns, T., Margerie, P., Aubert, M., Hedde, M., & Bureau, F. (2008). Assembly rules within earthworm communities in north-western France - a regional analysis. *Applied Soil Ecology*, 39, 321-335.
- Giller, P. S. (1996). The diversity of soil communities, the 'poor man's tropical rainforest'. *Biodiversity and Conservation*, 5, 135-168.
- Hagvar, S. (1998). The relevance of the Rio Convention on biodiversity to conserving the biodiversity of soils. *Applied Soil Ecology*, 9, 1-7.
- Hawksworth, D. L. (2001). The magnitude of fungal diversity: The 1.5 million species estimate revisited. *Mycological Research*, 105, 1422-1432.
- Schaefer, M., & Schauermann, J. (1990). The soil fauna of beech forests - comparison between a mull and a moder soil. *Pedobiologia*, 34, 299-314.
- Swift, M. J., Heal, O. W., & Anderson, J. M. (1979). Decomposition in terrestrial ecosystems. *Studies in Ecology*, 5.
- Torsvik, V., Goksoyr, J., Daae, F. L., Sorheim, R., Michalsen, J., & Salte, K. 1994. Use of DNA analysis to determine the diversity of microbial communities. In K. Ritz, J. Dighton & K. E. Giller (Eds) *Beyond the biomass: Compositional and functional analysis of soil microbial communities* pp. 39-48).
- Wall, D. H., Fitter, A. H., & Paul, E. A. 2005. Developing new perspectives from advances in soil biodiversity research. In R. D. Bardgett, D. W. Hopkins & M. B. Usher (Eds) *Biological diversity and function in soils. [ecological reviews.]* pp. 3-27).
- Wilson, E. O. (2002). *The future of life*. New York: Vintage Books.
- Wilson, E. O. (2003). The encyclopedia of life. *Trends in ecology & evolution (Personal edition)*, 18, 77-80.
- Wilson, J. R. U., Proches, S., Braschler, B., Dixon, E. S., & Richardson, D. M. (2007). The (bio)diversity of science reflects the interests of society. *Frontiers in Ecology and the Environment*, 5, 409-414.