

LES OUTILS DE BIOLOGIE MOLECULAIRE ET LEUR UTILISATION DANS LE DOMAINE DE LA GESTION DES SITES POLLUES

Janvier 2015

Étude réalisée pour le compte de l'ADEME par :

Jean-Michel Monier¹ et Sébastien Cécillon²

¹ ENOVEO, 7 place Antonin Poncet, 69002 Lyon.

² Environmental Microbial Genomics Group, Ecole Centrale de Lyon, 36 avenue Guy de Collongue, 69134 Ecully.

N° de contrat : 1472C0042

Coordination technique ADEME : Yves Duclos et Cécile Grand –
Service Friches Urbaines et Sites Pollués - SFUSP – ADEME (Angers)



SYNTHESE D'ETUDE

En partenariat avec :



CITATION DE CETTE SYNTHÈSE

ADEME. Jean-Michel Monier et Sébastien Cécillon. 2015. Les outils de biologie moléculaire et leur utilisation dans le domaine de la gestion des sites pollués – Synthèse. 39 pages.

Cet ouvrage est disponible en ligne www.ademe.fr, rubrique Médiathèque.

En français :

Toute représentation ou reproduction intégrale ou partielle faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause est illicite selon le Code de la propriété intellectuelle (art. L 122-4) et constitue une contrefaçon réprimée par le Code pénal. Seules sont autorisées (art. 122-5) les copies ou reproductions strictement réservées à l'usage privé de copiste et non destinées à une utilisation collective, ainsi que les analyses et courtes citations justifiées par le caractère critique, pédagogique ou d'information de l'œuvre à laquelle elles sont incorporées, sous réserve, toutefois, du respect des dispositions des articles L 122-10 à L 122-12 du même Code, relatives à la reproduction par reprographie.

Table des matières

1	INTRODUCTION.....	4
1.1	Objectif de cet ouvrage.....	4
1.2	Utilisation des outils de biologie moléculaire dans la gestion des sites et sols pollués	5
2	LES OUTILS DE BIOLOGIE MOLECULAIRE	6
2.1	Introduction	6
2.2	Synthèse : comparaison des OBMs	9
3	APPLICATIONS DES OBMs POUR LA GESTION DES SITES ET SOLS POLLUES	12
3.1	Application des OBMs en SSP	12
3.1.1	Applications et outils déjà utilisés	12
3.1.2	Intégration aux données de caractérisations classiques	15
3.2	Limites actuelles des OBMs	17
3.3	Acceptation et perception des OBMs	18
3.3.1	Perception des OBMs	18
3.3.2	Identification de la demande	19
3.3.3	Enquête auprès des acteurs	19
4	CONCLUSION	20
	FICHES TECHNIQUES : OUTILS DE BIOLOGIE MOLECULAIRE	20
	Fiche Technique 1 : Amplification par PCR	21
	Fiche Technique 2 : Empreintes Génétiques	24
	Fiche Technique 3 : Puces Phylogénétiques et Fonctionnelles.....	27
	Fiche Technique 4 : Séquençage Massif	29
	Fiche Technique 5 : Analyses Isotopiques (CSIA et SIP)	31
	Fiche Technique 6 : Analyses Cellulaires	33
	LEXIQUE	35
	Index des tableaux et figures	37

1 INTRODUCTION

1.1 Objectif de cet ouvrage

Les outils de biologie moléculaire (OBMs) regroupent un ensemble de techniques de pointes utilisées pour les analyses biologiques d'organismes vivants et la compréhension des mécanismes de la cellule à l'échelle des molécules. Initialement développés et appliqués en laboratoire, ces outils ont trouvé de nombreuses applications dans différents secteurs (médecine, défense, génétique humaine, environnement...). Depuis quelques années, ces outils ont été appliqués avec succès à différentes problématiques environnementales et sont notamment de plus en plus souvent intégrés aux plans de gestions de sites pollués. Leur succès croissant dans ce domaine réside dans le fait que ces outils fournissent des informations complémentaires et souvent uniques permettant d'optimiser les stratégies de remédiation en améliorant l'efficacité des traitements de remédiation et/ou en diminuant la durée et le coût des traitements. Leur intégration aux plans de gestion, en complément des outils conventionnels (données physico-chimiques, hydrogéologiques ou historiques) reste cependant encore limitée, en grande partie due à une méconnaissance de leur existence et de leur valeur ajoutée et parfois de la difficulté apparente de leur interprétation.

Cet article propose de faire un tour d'horizon des différents outils de biologie moléculaire disponibles en mettant en avant les avantages et limites de chacun, leurs applications, leur valeur ajoutée, ainsi que le type d'informations fournies et leur interprétation. Bien que non traitées dans cet ouvrage, les étapes amont d'échantillonnage, de conservation de la matrice analysée (sol, eau, sédiment, biofilm...) et d'extraction des biomarqueurs (ADN, ARN, Protéines...) font partie intégrante de la chaîne analytique. La qualité des résultats générés par les analyses de biologie moléculaire dépendant fortement de ces étapes, les protocoles d'échantillonnage et d'extraction doivent être établis avec rigueur. Les méthodes d'échantillonnage (direct vs passif) et d'extraction des différentes molécules utilisées pour les analyses (ADN, ARN, protéines et lipides) étant en constante évolution, une concertation des différents acteurs (gestionnaire, bureau d'étude, laboratoire d'analyses...) est préconisée afin de sélectionner les approches les plus pertinentes pour la matrice considérée (nature de la matrice, quantité de biomasse, présence potentielle d'inhibiteurs, types d'analyses...).

1.2 Utilisation des outils de biologie moléculaire dans la gestion des sites et sols pollués

Les OBMs peuvent fournir des informations uniques et complémentaires des données conventionnelles durant toutes les étapes de gestion d'un projet de réhabilitation d'un site pollué. Bien que leur réalisation n'a pas vocation à suppléer les outils conventionnels, ils offrent de nombreux bénéfices et fournissent de nouvelles données permettant d'optimiser les stratégies de dépollution. Ils peuvent être utilisés durant toutes les étapes de gestion, du diagnostic initial à la restitution en passant par la définition de la stratégie de dépollution et son suivi. Ils permettent par exemple de déterminer si un traitement par bioremédiation est envisageable ou si la dégradation biotique est effective, de redéfinir les stratégies, de suivre l'efficacité d'un traitement et en combinaison avec les autres données géochimiques, de déterminer si la restitution est envisageable (**Figure 1**). Leur intérêt réside également dans le fait qu'ils permettent de comprendre les mécanismes microbiens en jeu et ainsi de pouvoir anticiper le devenir des polluants et d'identifier les causes d'un éventuel dysfonctionnement ou d'une perte d'efficacité d'un traitement mis en place. Les microorganismes responsables de la biodégradation sont-ils présents, se maintiennent-elles au cours du traitement, sont-elles toujours actives, les produits de dégradation risquent-ils de s'accumuler... ?



Figure 1 : Exemples de réponses apportées par les outils de biologie moléculaire au cours des différentes étapes de traitement de sites et sols pollués.

Lors de la caractérisation d'un site (identification des contaminants présents, importance et étendue de la pollution, estimation du risque, évaluation de l'atténuation naturelle...), les OBMs pourront contribuer à cette caractérisation en apportant des données sur sa composante biologique et à l'élaboration des différentes options de traitement possibles.

Lors de l'étape de dépollution (sélection de la stratégie, pilote, mise en place), les OBMs peuvent être intégrés à l'échelle pilote (laboratoire ou terrain) pour la réalisation de tests de faisabilité ou traitabilité, peuvent permettre de comparer l'efficacité d'un traitement par rapport à une atténuation naturelle, de déterminer l'impact d'un amendement ou d'un traitement sur les communautés et processus microbiens impliqués. En conjonction avec les données hydrogéologiques et physico-chimiques, ils permettront de définir les stratégies les plus appropriées tenant compte des différentes contraintes du site.

Lors du suivi du traitement, qui par définition nécessite plusieurs échantillonnages dans le temps, les OBMs vont fournir des données complémentaires aux données conventionnelles. Leur avantage réside principalement dans le fait qu'ils permettent d'anticiper l'efficacité du traitement puisqu'ils vont fournir une information sur la dynamique des communautés microbiennes responsables de la biodégradation des contaminants. Est-ce que les microorganismes et gènes de dégradation maintiennent, accroissent leur activité ou au contraire est-ce que leur nombre et activité décroissent

au cours du traitement. Ces données sur la dynamique des acteurs permettent au gestionnaire d'anticiper l'évolution des contaminants sur le site en fournissant une information prédictive.

La réception est possible lorsque les objectifs de dépollution sont atteints et les risques environnementaux acceptables. L'atteinte des objectifs est classiquement déterminée par des analyses chimiques (baisse tendancielle des concentrations, réduction des panaches...). Les OBMs peuvent venir conforter ces tendances en mettant en évidence que la biodégradation est toujours effective sur le site et conforter les décisions concernant l'atteinte des objectifs et le devenir de la pollution résiduelle ou l'accumulation de produits de dégradation.

2 LES OUTILS DE BIOLOGIE MOLECULAIRE

2.1 Introduction

Les outils de biologie moléculaire, au sens large, sont les outils permettant d'isoler, de manipuler ou encore de caractériser les composants moléculaires d'une cellule ou d'un ensemble de cellules. Les composants moléculaires classiquement étudiés sont l'ADN et l'ARN mais peuvent également inclure les protéines de structure (eg, constituant cellulaire) ou de fonction (eg, enzymes responsables de la dégradation des contaminants) et les acides gras (constituant de la membrane des cellules) (**Figure 2**). Ces outils offrent un intérêt tout particulier pour les microorganismes environnementaux car ces derniers sont rarement isolables et cultivables sur des milieux synthétiques. On observe en effet que près de 99% des microorganismes présents dans un échantillon environnemental (eg, sol) ne sont généralement pas cultivables et sont donc complètement ignorés lors d'étude utilisant des méthodes de microbiologie classiques basées sur la culture de ces microorganismes sur milieux synthétiques. Les outils de biologie moléculaire vont permettre de passer outre cette limitation puisqu'ils ne nécessitent pas de cultiver et d'isoler les microorganismes pour les étudier. Leur force réside également dans le fait que l'on puisse étudier et analyser lors d'une même étape, l'ensemble des microorganismes présents dans un échantillon. L'analyse de la littérature scientifique montre cependant que ces outils sont encore sous exploités (**Figure 6**).

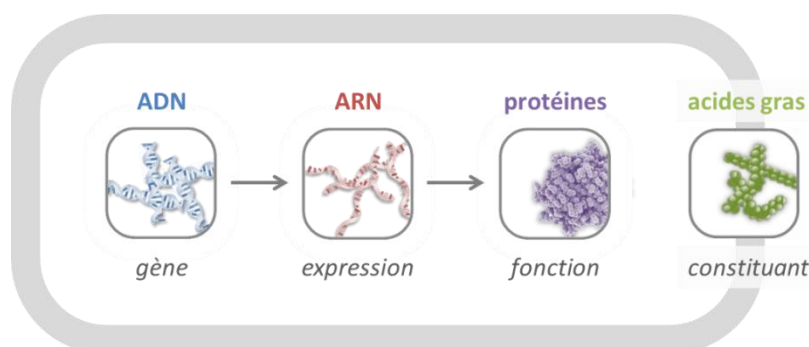


Figure 2 : Composants de la cellule analysés à l'aide des outils de biologie moléculaire. L'ADN (Acide DésoxyriboNucléique) est le support de l'information génétique d'un individu et constitue son génome (ensemble des gènes). Lorsqu'un individu exprime un gène (activité de ce gène), l'ADN est transcrit en ARN. L'ARN (Acide RiboNucléique) est la molécule servant à transférer les instructions génétiques inscrites dans l'ADN. L'ARN, qui constitue un reflet de l'activité de la bactérie, est ensuite traduit pour former les protéines. Les protéines peuvent avoir un rôle structurel (constituant de la cellule) ou fonctionnel (enzymes). Les acides gras sont des constituants de la cellule, pouvant être utilisés pour différencier les organismes (voir PLFA).

Les analyses s'effectuent classiquement après une étape d'extraction et de purification des composants moléculaires des cellules dans la chaîne analytique de traitement des échantillons (**Figure 3**). Notons que certains outils permettent d'étudier ces molécules (ADN principalement) directement dans la cellule (**Figure 4** ; **Fiche Technique 6 : Analyses Cellulaires**). Parmi la diversité des outils disponibles (**Figure 4**), ceux trouvant leur champ d'application dans le domaine des sites et sols pollués, s'appliquent aujourd'hui essentiellement aux ADN et ARN. Les acides gras, constituant de l'enveloppe cellulaire, peuvent être utilisés pour comparer la structure des communautés microbiennes. Les analyses portant sur les protéines restent encore très anecdotiques dans le domaine des Sites et Sols Pollués (SSP) et ne seront pas détaillées dans ce document.

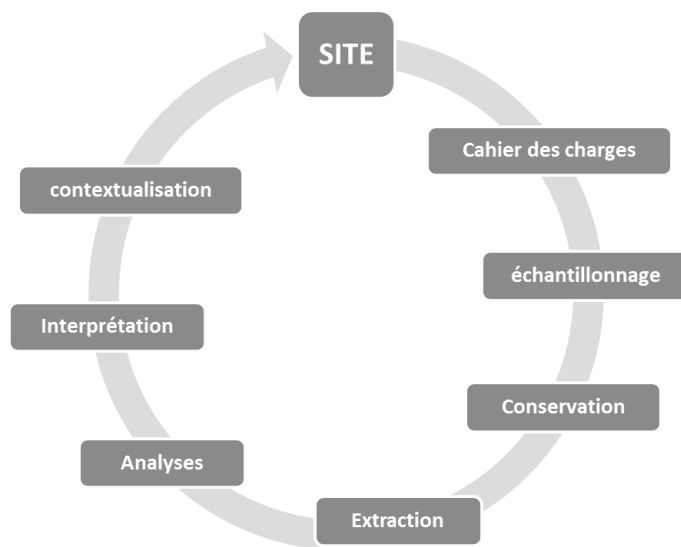


Figure 3 : Représentation schématique des différentes étapes à considérer lors de la réalisation d'analyses d'échantillons environnementaux à l'aide des outils de biologie moléculaire.

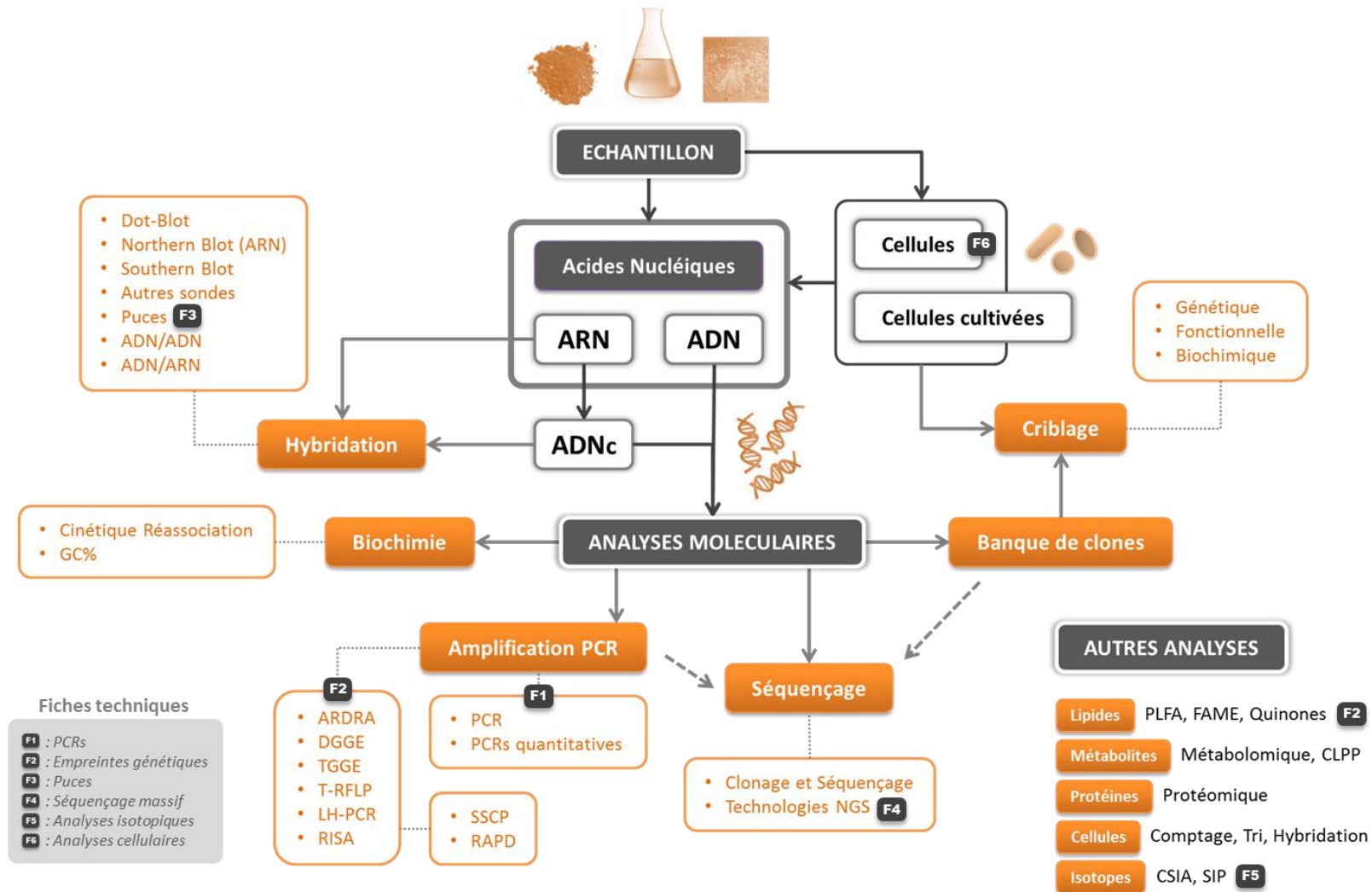


Figure 4 : Palette d'outils de biologie moléculaire disponibles. Les outils classiquement utilisés en SSP sont signalés par les codes F1 à F6, correspondants aux fiches techniques de chacun des outils décrits en annexe du document.

L'ADN (Acide DésoxyriboNucléique) est le support de l'information génétique d'un individu et constitue son génome, c'est-à-dire l'ensemble des gènes codant chacun une fonction précise, de biodégradation par exemple. Dans le cas d'un échantillon complexe (sol, sédiment, eau de nappe...) où plusieurs individus sont présents, on parle de métagénome (ensemble des génomes et donc de tous les gènes des organismes présents dans l'échantillon étudié). Son analyse va permettre notamment d'identifier les microorganismes présents ou par exemple, dans le cas de fonctions précises, de rechercher quelle est la diversité et l'abondance des gènes de biodégradation. **L'analyse basée sur l'ADN renseigne sur le potentiel d'une communauté microbienne.**

Si l'on prend l'exemple de l'ADN, chaque organisme contient des milliers de gènes (génome) dont la taille et l'ordre des bases A, C, G et T les constituants (la séquence) leur sont propres. De manière schématique, les OBM, vont permettre à partir des molécules d'ADN extraites, d'analyser les tailles, les nombres, la diversité ou les séquences et l'ordre des bases A, C, G, et T dans les molécules extraites (**Figure 5**) afin de déterminer par exemple la diversité, l'abondance ou encore le potentiel d'activité microbienne dans les échantillons analysés.

```
ATGAAAGAGTTTCATTCCACACTCTCGCGACGTGATTTTATGAAAAGCCTTGGGGTAGTCGGAGCAGGCTTGGGTACTATGAGCGCTGCTGCC
CTGTTTTCCACGATTTGGATGAGGTACAGTCTTCAACCCTTGGTATCAATAAAAAATCCTTGGTGGGTAAGGAAAGGATTTCAAGAATCCTAC
AGTTCCTATAGACTGGTCAAAAGTTACCCGTCAGCCGGGTGATTTTCAGGGTCTGCCAGACCCACTGTAGCTGATTTCAAAAAGCCGGCTT
GTGGCGGTACTTCCACTGACCTTGAACCCCTGAAATGGCCCTGACTCTTTATGATGCCATGGCCAAGGAGTTTCCGGGCTGGACTCCGGGCT
ATGCCGGTATGGGTGACGTGAGGACTACCTCACTCTGCAATGCTTCCAAATTTATGATGATGGGTGCCTGGCCCGGCAATATGGAATGGCGG
TAAAAGGATTAATGTTATCGGCGCTATTATGGCGGCGCGGCGGAGCCCTACCTTTACTCCGTGGTTGGGGCTCAGCTGGATACTACTACCCGT
CCCCAGGATTTCCGGGCTCCGGTCTGGCAGGGTACTCCCGAAGAAAACCTTAAAACCTGCCGCTCTGCCTTCAGATTTCTTGGCGGCTCGGATG
TGGGTGCTATTGAAATGACGACGATGTTCTTAAATTTATCCATTCCAGATTGGCGGTAAGCTGTAGTCGTTGAAGACGTAGAGGAAGCTTA
TGAGACCACAACCAAGATGGTTATCCCCGCAAGTGCAAGTGGATACTTATGTGGAGCGCCAGACAGTCACTCGAAGCTACCAGACGCCAGGCA
GGCATAACCGAAAGTCAATGCTGTTTGGTATTTCATATTCGCCCTTCCCAAGGTAGGTGCCAGTTCAGGAATTTATCCGCGGTTTGGGTTATC
AGGCCTTGAATCCGGGCATGATGGGCTTTTGGCTAATCCTTTAGCAGCTTTAGCCGGTATGGGCGAACATGGCCGCATGTCTTCAACCCACAT
CACTCCCAATACGGCACAACCTAACCGGCTATGTGGCGGTTAATCACCGACCTGCCATTGCTGCCACTCCGCCATAGACTTTGGTGCCTAC
AAATTCGCAAAACCTGCGGTATTTGCGCGGATGCTTGTCTTTCGGGCTTATCCAGAAGGGTGACCCACCTGGGAAAACCCGGCTTCAGCCA
AGTCCGGCATTCAGCAGGGCACATTTGAGGGCTGGCGAACCAATACCGCAGATTGTCTCACTGCCACCTGTGAGGGCACTTGCCCTTCAA
CAGCAAAACCGGATTCATTCCTGCATGCTGTGGTTAAGGGCACAGTTGCCAATACCCCGCTTTTAAATAGCTTCTTACAAATATGAAAAAGCT
ATGGATTACGGGGCAAAAGACCCCGAAGAGTGGTGGGATATGGATGACTTTACCTACGGTATTGACACCAGTTACTAATAGCTTAAGGAATTGG
AGGTAGGAAGCATG
```

Figure 5 : Exemple de séquence génétique (gène de la bactérie *Dehalococcoides* sp. CBDB1 codant pour une déhalogénase responsable de la déchloration réductrice anaérobie des chloroéthènes).

Pour savoir si ce potentiel s'exprime, on s'intéresse alors aux ARN (Acide RiboNucléique). Lorsqu'un individu exprime un gène (activité de ce gène), l'ADN est transcrit en ARN. L'ARN est la molécule servant à transférer les instructions génétiques inscrites dans l'ADN. L'ARN (qui constitue un reflet de l'activité de la bactérie) est ensuite traduit pour former les protéines de structure (eg, constituant de la cellule) ou de fonction (eg, enzymes). La formation d'une enzyme traduit donc l'expression (ou l'activité) d'un gène pour une fonction donnée, dite d'intérêt, si celle-ci est, par exemple responsable de la biodégradation d'un polluant. **L'analyse basée sur l'ARN renseigne ainsi sur l'activité d'une communauté microbienne.**

Les paragraphes suivant décrivent l'utilisation des différents outils en SSP. Pour plus d'information les outils actuellement utilisés dans le domaine des SSP, des fiches techniques détaillant le principe et les informations fournies par les principaux OBM sont reportées à la fin de ce document. Les avantages et limites des différents outils sont synthétisés dans le **Tableau 1** ci-dessous.

2.2 Synthèse : comparaison des OBM

La description, les avantages et limites des différents OBM décrits dans les fiches techniques sont synthétisés ci-dessous (**Tableau 1**). Notons que les outils décrits indépendamment sont souvent utilisés de manière combinée pour appréhender les différentes questions inhérentes aux problématiques spécifiques de chaque site et en association aux données géochimiques classiques.

Tableau 1 (partie 1/2) : Description, avantages et limites potentielles des différents outils de biologie moléculaire.

Outils (1/2)		Description / type de données	Avantages	Limites potentielles
Amplification par PCRs	<ul style="list-style-type: none"> • PCR classique 	<ul style="list-style-type: none"> • Amplification spécifique de séquences génétiques d'intérêt (cibles). <p>⇒ Présence ou absence de séquences génétiques spécifiques.</p>	<ul style="list-style-type: none"> • Disponible pour un grand nombre de microorganismes et fonctions de biodégradation. • Confirme la présence des séquences cibles. • Rapidité et coût. 	<ul style="list-style-type: none"> • Ne permet pas de quantifier le nombre et l'activité des séquences ciblées. • Nécessite de connaître au moins partiellement la séquence de la cible.
	<ul style="list-style-type: none"> • qPCR • RT-qPCR • dPCR 	<ul style="list-style-type: none"> • Amplification spécifique de séquences génétiques d'intérêt permettant la quantification des acides nucléiques. <p>⇒ Nombre de copies de séquences génétiques spécifiques ciblant un groupe ou une fonction (ADN : présence, ARN : expression).</p>	<ul style="list-style-type: none"> • Disponible pour un grand nombre de microorganismes et fonctions de biodégradation. • Confirme la présence des séquences cibles. • Permet de quantifier le nombre (qPCR) et l'activité (RT-qPCR) des microorganismes ou fonctions ciblées. • Permet de déterminer les dynamiques de populations (croissance, distribution spatiale). 	<ul style="list-style-type: none"> • Seule la RT-qPCR (ARN) permet de mettre en évidence l'activité des microorganismes ou fonctions ciblées. • RT-qPCR moins facilement disponible commercialement du fait de la technicité de l'outil et de l'instabilité des ARN. • Nécessite de connaître au moins partiellement la séquence de la cible.
Empreintes génétiques	<ul style="list-style-type: none"> • DGGE • RFLP • RISA • PLFA 	<ul style="list-style-type: none"> • Ensemble de méthodes permettant de mettre en évidence la structure et la diversité des communautés microbiennes. <p>⇒ Empreintes génétiques (codes-barres) d'un échantillon.</p>	<ul style="list-style-type: none"> • Fournit une information de base sur la structure et l'évolution des communautés microbiennes. • Permet de suivre des organismes ou groupes de microorganismes dans le temps et l'espace. • Rapidité et coût pour certaines. 	<ul style="list-style-type: none"> • La plupart de ces techniques ne permettent pas, sans analyses complémentaires, d'identifier les microorganismes ou groupes de microorganismes. • Méthodes fournissant une information relative, non quantitative. • Microorganismes abondants peuvent dominer les profils et masquer la diversité apparente. • Ne permettent pas de mettre en évidence les activités de biodégradation.
Puces	<ul style="list-style-type: none"> • Phylogénétiques • Fonctionnelles 	<ul style="list-style-type: none"> • Détermination rapide de la présence ou de l'absence d'un grand nombre de gènes (fonctions ou identité). <p>⇒ Empreintes génétiques (codes-barres) d'un échantillon et présence/absence des séquences génétiques ciblées.</p>	<ul style="list-style-type: none"> • Fournit une information plus complète que les empreintes génétiques classiques sur la structure et l'évolution des communautés microbiennes. • Permet de suivre des organismes ou groupes de microorganismes dans le temps et l'espace. 	<ul style="list-style-type: none"> • Disponibilité commerciale limitée. • Nécessite une haute technicité pour la mise en œuvre et l'interprétation. • Détection non exhaustive, limitée à la spécificité des sondes présentes sur la puce.

Tableau 1 (partie 2/2) : Description, avantages et limites potentielles des différents outils de biologie moléculaire.

Outils (2/2)		Description / type de données	Avantages	Limites potentielles
Séquençage massif (NGS)	<ul style="list-style-type: none"> • Roche (454) • SOLiD • Ion Torrent • Illumina 	<ul style="list-style-type: none"> • Techniques de séquençage massif des acides nucléiques permettant de déterminer les séquences génétiques d'un gène ou d'un ensemble de gènes (ADN, ARN) présents dans la totalité d'un échantillon. 	<ul style="list-style-type: none"> • technique dite de haut débit permettant d'obtenir une information très précise sur la structure des communautés. • Détection et abondance relative des fonctions même faiblement représentées. • Applicable pour différentes matrices (ADN, ARN, produits amplifiés par PCR) • Réduction importante des coûts de séquençage rend ces technologies abordables pour des applications en routine. 	<ul style="list-style-type: none"> • Peut nécessiter des compétences et des moyens informatiques importants pour le traitement et l'analyse des données. • Quantification relative et non absolue du nombre de séquences d'intérêt.
	⇒ <i>Diversité et abondance relative des microorganismes ou fonctions recherchés.</i>			
Analyses isotopiques	<ul style="list-style-type: none"> • CSIA 	<ul style="list-style-type: none"> • Quantification de l'abondance relative des isotopes stables dans les contaminants (eg, $^{13}\text{C}/^{12}\text{C}$). 	<ul style="list-style-type: none"> • Mise en évidence directe de la dégradation biotique ou abiotique des contaminants. • Utile pour identifier les mécanismes et acteurs de la biodégradation. • Contribue à l'identification de la source d'une pollution. 	<ul style="list-style-type: none"> • Facteurs de fractionnement isotopique nécessaire. • Un nombre limité de laboratoires peuvent effectuer ces analyses • Coût, sachant que plusieurs analyses sont nécessaires pour évaluer les cinétiques de dégradation. • Réalisation / résolution délicates pour des certaines pollutions.
	⇒ <i>Abondances relatives des isotopes.</i>			
Analyses cellulaires	<ul style="list-style-type: none"> • FISH • Viabilité • Cytométrie 	<ul style="list-style-type: none"> • Visualisation, quantification et spatialisation des microorganismes et de leur activité. 	<ul style="list-style-type: none"> • Mise en évidence directe de la présence de microorganismes, de leur nombre et de leur activité métabolique. • Information sur la répartition spatiale et la structure des communautés. • Quantité de biomasse et activité globale de la communauté. 	<ul style="list-style-type: none"> • Peu disponible commercialement. • FISH : Nombre et diversités des sondes nécessaires à la visualisation des cellules encore limités ou nécessitant un développement en interne.
	⇒ <i>Images en microscopie, nombres de cellules (cytométrie).</i>			

3 APPLICATIONS DES OBM_s POUR LA GESTION DES SITES ET SOLS POLLUES

Les OBM_s peuvent fournir des informations uniques et complémentaires des données conventionnelles durant toutes les étapes de gestion d'un projet de dépollution. Les OBM_s ne peuvent en aucun cas remplacer les données conventionnelles mais apportent de nouvelles informations permettant d'optimiser les stratégies à mettre en place. Les données conventionnelles permettent d'établir des stratégies d'intervention classiques mais par manque d'informations sur la composante biologique, ne permettent souvent pas d'envisager des alternatives pouvant conduire à des traitements plus efficaces, moins coûteux ou nécessitant moins de temps.

Cette section se propose de synthétiser comment les OBM_s sont et peuvent être utilisés pour la gestion des projets de dépollution, de montrer leur complémentarité avec les outils classiques ainsi que leurs spécificités, limites et avantages.

3.1 Application des OBM_s en SSP

3.1.1 Applications et outils déjà utilisés

Bien qu'il existe de nombreux exemples d'applications des OBM_s dans les projets de R&D appliqués à des réhabilitations de sites et sols pollués, leur utilisation reste globalement encore très minoritaire dans la plupart des projets industriels de bioremédiation. Une analyse de la littérature scientifique montre que l'exploitation de ces outils est relativement récente et encore très limitée (**Figure 6**).

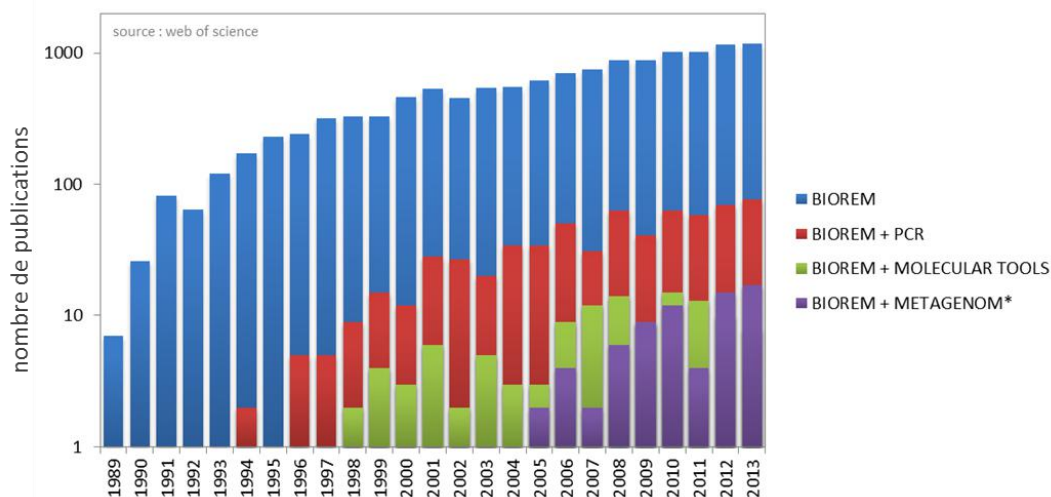


Figure 6 : Nombre de publications référencées dans le domaine de la bioremédiation intégrant les outils de biologie moléculaire. Les recherches ont été effectuées sur la base de données Web Of Science à partir des mots clés Bioremédiation ; Bioremédiation + PCR ; Bioremédiation + Outils de biologie moléculaire ; Bioremédiation + Métagénomique (webofknowledge.com).

En moyenne pour l'année 2013, seulement 6% des études publiées intégraient effectivement la PCR (technique de base de biologie moléculaire ciblant spécifiquement un gène ou ensemble de gènes) et seulement 1% utilisaient des approches dites de métagénomiques (telles que puces ou séquençages massifs) basées sur l'extraction et l'analyse de l'ensemble des gènes présents dans un échantillon (**Figure 6**). Bien que ces pourcentages soient en augmentation depuis la première utilisation de ces

outils (milieu des années 90 pour la PCR et milieu des années 2000 pour la métagénomique), le constat est que ces outils ne sont pas encore exploités en routine malgré leur forte valeur ajoutée potentielle dans la réhabilitation des sites et sols pollués.

Les raisons peuvent être multiples selon les acteurs. Une enquête réalisée dans le cadre de cette étude et visant à élaborer un portrait de l'utilisation des OBMs et des besoins dans la gestion des SSPs auprès de différents acteurs a permis de tracer une image plus précise de l'exploitation actuelle de ces outils. Les principales conclusions de cette enquête sont reportées dans le paragraphe **3.3.1 Perception des OBMs**.

Des exemples concrets de questions auxquelles les différents OBMs permettent de répondre sont listés dans le **Tableau 2**. Ces questions s'appliquent notamment à l'ensemble des polluants organiques dont ceux couramment rencontrés tels que les hydrocarbures, BTEX, HAP, PCB, solvants halogénés. Des exemples d'applications des différents outils sont listés ci-dessous :

Méthodes

Dominguez et al. (2008) Aerobic bioremediation of chlorobenzene source-zone soil in flow-through columns: Performance assessment using quantitative PCR. *Biodegradation* 19:545-553. (PCR)

Hendrickx et al. (2005) Dynamics of an oligotrophic bacterial aquifer community during contact with a groundwater plume contaminated with benzene, toluene, ethylbenzene, and xylenes: an in situ mesocosm study. *Applied and Environmental Microbiology*. 71: 3815-25. (Empreintes génétiques)

Rowe et al. (2008) Characterization of the community structure of a dechlorinating mixed culture and comparisons of gene expression in planktonic and biofilm-associated *Dehalococcoides* and *Methanospirillum* species. *Applied and Environmental Microbiology* 74:6709-6719. (FISH)

Uhlir et al. (2014) Stable isotope probing in the metagenomics era: A bridge towards improved bioremediation. *Biotechnology Advances* 31:154-165. (SIP)

Meckenstock et al. (1999) $^{13}\text{C} / ^{12}\text{C}$ isotope fractionation of aromatic hydrocarbons during microbial degradation. *Environmental Microbiology* 1:409-414. (CSIA)

He et al. (2012) Applications of functional gene microarrays for profiling microbial communities. *Current Opinion in Biotechnology* 23:460-466. (Puces fonctionnelles)

Lee et al. (2012) Phylogenetic microarray analysis of a microbial community performing reductive dechlorination at a TCE-contaminated site. *Environmental Science and Technology* 46:1044-1054. (Puces taxinomiques)

Shah et al. (2013) Taxonomic profiling and metagenome analysis of a microbial community from a habitat contaminated with industrial discharges. *Microbial Ecology* 66:533-550. (NGS)

Articles de synthèse

Chikere (2013) Application of molecular microbiology techniques in bioremediation of hydrocarbons and other pollutants. *British Biotechnology Journal* 3:90-115.

Deeb et al. (2012) Advanced Diagnostic Tools. Chapter 8 in Kueper B, HF Stroo and CH Ward (eds.). *Chlorinated Solvent Source Zone Remediation*. Springer, New York, NY.

Desai et al. (2010) Advances in molecular and "-omics" technologies to gauge microbial communities and bioremediation at xenobiotic/anthropogen contaminated sites. *Bioresource Technology* 101:1558-1569.

Maphosa et al. (2012) Ecogenomics of microbial communities in bioremediation of chlorinated contaminated sites. *Frontiers in Microbiology* 3:351.

Petrovskis et al. (2013) Microbial monitoring during bioaugmentation with *Dehalococcoides*." In Leeson A, H. Stroo and CH Ward (eds), *Bioaugmentation for Groundwater Remediation*, Springer, New York, NY.

Srivastava et al. (2014) Advances in microbial bioremediation and the factors influencing the process. *International Journal of Environmental Science and Technology* 11:1787-1800.

Widada et al. (2002) Recent developments in molecular techniques for identification and monitoring of xenobiotic-degrading bacteria and their catabolic genes in bioremediation. *Applied Microbiology and Biotechnology* 60:45-59.

Tableau 2 : Exemples de questions auxquelles peuvent répondre les différents OBMs* dans le cadre des SSP.

Questions	Amplification PCR		Empreintes génétiques	Puces	NGS	Analyses Isotopiques		Analyses Cellulaires
	qPCR	RT-qPCR				CSIA	SIP	
Quels sont les microorganismes présents ?			+	+	++			
Quels sont les microorganismes actifs ?				+	++			
Microorganismes sont présents en quantité importante ?	++							+
Présence de microorganismes capables de biodégradation ?	++	++				+	+	
Microorganismes capables de dégrader le polluant sont actifs ?		++					++	
La structure des communautés évolue-elle ?			+	+	++			
Les microorganismes introduits (bioaugmentation) se maintiennent-ils ?	+	+						
Abondance et l'activité des microorganismes évoluent ?	++	++	+	+	++			+
Quelle est la cinétique de biodégradation ?						+	+	
Les produits de dégradation risquent-ils de s'accumuler ?	+	+		+	+	+		
Est-ce que les produits de dégradation s'accumulent ?						+		
Y-a-t-il dégradation abiotique ?						+		

* Les outils mentionnés pour chacune des questions sont les outils les plus appropriés. Certains des autres outils peuvent également être associés pour adresser certaines de ces questions.

3.1.2 Intégration aux données de caractérisations classiques

Comme mentionné préalablement les OBMs fournissent des informations complémentaires à celles fournies par les outils conventionnels. Ces outils ont donc vocation à être intégrés dans les projets de réhabilitation (**Figure 7**). Le constat actuel est que très souvent, les OBMs sont un recours lorsque, durant les étapes de caractérisation et de gestion, les analyses conventionnelles (historiques, hydrogéologiques, physico-chimiques...) soulèvent des interrogations. Les apports des données de biologie moléculaire aux données classiques de gestion des sites et sols pollués sont synthétisés dans le **Tableau 3**.

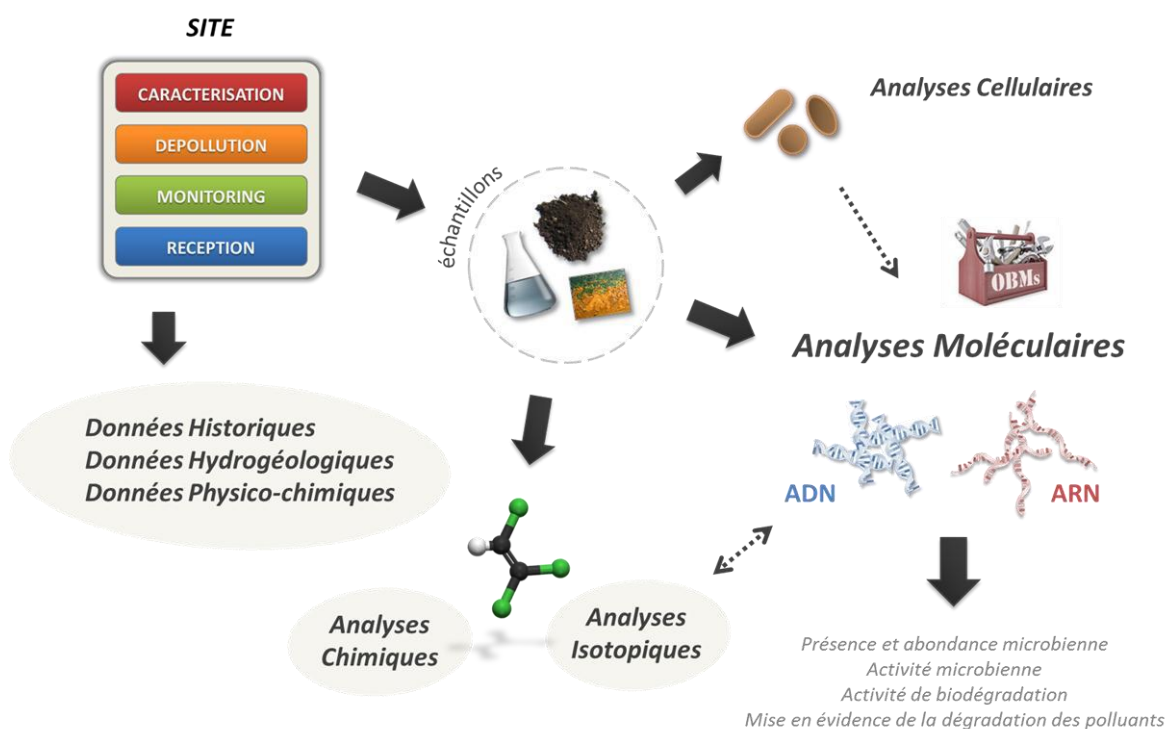


Figure 7 : Intégration des données de biologie moléculaire aux données classiques de caractérisation des sites et sols pollués.

Tableau 3 : Apport des données de biologie moléculaire aux données classiques de caractérisation des sites et sols pollués.

Données conventionnelles	Limites potentielles	Apport des OBMs
Caractérisation du site		
<ul style="list-style-type: none"> • Evaluation des concentrations et spatialisation des contaminants. • Compréhension des paramètres influençant la stratégie de remédiation. 	<ul style="list-style-type: none"> • Incapacité à estimer le potentiel de biodégradation microbien. • Absence d'information sur les mécanismes de biodégradation en place ou potentiel. 	<ul style="list-style-type: none"> • Identification des communautés microbiennes et de leurs capacités à dégrader les contaminants. • Estimation de la dégradation biotique et abiotique des contaminants. • Consolidation des données conventionnelles.
Dépollution		
<ul style="list-style-type: none"> • Caractérisation des conditions physico-chimiques et adéquation des processus microbiens. • Evaluation des cinétiques de dégradation. • Identification des produits de dégradation. • Evaluation de l'atténuation naturelle. • Décision des stratégies de traitement à mettre en place. 	<ul style="list-style-type: none"> • Difficulté à établir que les mécanismes de dégradation désirés sont en place et actifs. • Capacité limitée à anticiper les cinétiques de dégradation biotique et abiotique et les risques d'accumulation des produits de dégradation. 	<ul style="list-style-type: none"> • Identification des acteurs et des mécanismes en place. • Détermination précise des cinétiques de dégradation biotique et abiotique. • Confirmation de la pertinence de la stratégie mise en place. • Robustesse des informations sur la nécessité de maintenir ou de changer de stratégie.
Monitoring		
<ul style="list-style-type: none"> • Evaluation des concentrations en contaminants et produits de dégradation afin d'évaluer la pertinence et les performances de la stratégie mise en place. • Détermination de l'évolution des paramètres géochimiques pendant et après la mise en place du traitement. 	<ul style="list-style-type: none"> • Difficulté à établir la contribution relative des processus biotique et abiotique pour certains contaminants. • Données contradictoires ou difficilement interprétables. • Capacité limitée à anticiper les cinétiques de dégradation biotique et abiotique. 	<ul style="list-style-type: none"> • Information précise sur la réponse des microorganismes au traitement mis en place. • Information sur le maintien et l'activité des processus de dégradation en place. • Evolution des cinétiques de dégradation biotique et abiotique.
Réception		
<ul style="list-style-type: none"> • Détermination des concentrations en contaminants et produits de dégradation permettant d'établir si les objectifs sont atteints. 	<ul style="list-style-type: none"> • Incapacité à établir la pertinence et la contribution de la stratégie de remédiation mise en place. • Difficulté à prédire les cinétiques de dégradation et l'évolution de la pollution sur le site. 	<ul style="list-style-type: none"> • Données complémentaires permettant de confirmer les processus de dégradation en jeu et l'atteinte des objectifs. • Information sur l'évolution et le devenir des contaminants et produits sur site après réception.

3.2 Limites actuelles des OBMs

L'utilisation des OBMs dans le domaine des SSPs peut présenter un certain nombre de biais et de limites. Les sources principales de biais sont liées aux aspects non abordés dans cet article, à savoir l'échantillonnage et l'extraction des acides nucléiques. Il est évident que les analyses moléculaires, aussi informatives soient-elles, fournissent une information utile au gestionnaire à partir du moment où l'échantillonnage est représentatif du site (**Figure 3**). Le coût de certaines analyses moléculaires peut représenter un frein et en conséquence limiter le nombre d'échantillons considérés. Si le nombre d'échantillons est effectivement restreint, la sélection des échantillons à analyser doit donc être consolidée et définie sur la base d'une caractérisation précise du site (*ie*, analyses conventionnelles). L'efficacité de l'extraction des acides nucléiques (*ie*, rendement d'extraction et diversité microbienne représentative) peut également être une source de biais. En effet, cette efficacité dépend très souvent de la nature de l'échantillon (*eg*, sol, eau, biofilm...) et des contaminants présents qui peuvent conduire à une réduction importante des rendements d'extraction et/ou à la co-extraction d'inhibiteurs des réactions enzymatiques utilisées en biologie moléculaire. Ces sources de biais amènent donc très souvent à adapter les protocoles d'extraction en fonction des différentes situations rencontrées. L'adaptation, parfois nécessaire des protocoles, requérant une certaine technicité et une connaissance approfondie des OBMs, il est difficilement envisageable, dans le cas des SSPs, d'élaborer des méthodes d'extraction standardisées.

Les limites actuelles des OBMs dans le domaine des SSPs, mais également de manière plus générale en microbiologie environnementale, sont liées à notre connaissance partielle de la diversité taxinomique et fonctionnelle des microorganismes dans l'environnement. En effet, la recherche ou la quantification de gènes taxinomiques ou fonctionnels d'intérêt ne peuvent s'effectuer et ne présentent un intérêt qu'à partir du moment où les séquences génétiques et leur fonction sont clairement établies. Bien que les OBMs et le séquençage massif, en particulier, ont permis ces dernières années de générer une quantité importante d'informations et l'identification de nouveaux gènes, leur fonction reste encore très souvent putative ou inconnue. De ce fait, les analyses de biologie moléculaire appliquées aux SSP portent uniquement sur les connaissances actuelles. En conséquence, certains OBMs (*eg*, PCR quantitatives, puces...) ne peuvent être appliqués aux contaminants pour lesquels les mécanismes de biodégradation n'ont pas encore été clairement caractérisés. Inversement, pour les composés dont certains mécanismes de biodégradation ont été identifiés, il est envisageable que d'autres mécanismes existent et conduisent donc à une recherche partielle et une sous-estimation du potentiel microbien réel. Les limites des OBMs ne sont donc, dans ce cas, pas techniques mais simplement liées à une méconnaissance partielle ou totale de certains mécanismes de biodégradation. Les solutions proposées au gestionnaire sont donc en constante progression mais reste tributaire d'une recherche plus fondamentale réalisée au sein des laboratoires de recherche publics ou privés.

3.3 Acceptation et perception des OBMs

3.3.1 Perception des OBMs

Une enquête, mandatée par l'ADEME, a été réalisée afin d'élaborer un portrait de l'utilisation actuelle des Outils de Biologie Moléculaires (OBMs) et des besoins industriels pour la gestion des Sites et Sols Pollués. Le questionnaire, disponible en ligne, a été envoyé à plus de 400 professionnels travaillant dans le domaine des sites pollués. Cette enquête a été structurée en 4 parties permettant d'identifier le type de société et ses activités et d'appréhender leur connaissance, leur utilisation et leur perception des Outils de Biologie Moléculaire. La synthèse des résultats est présentée ci-dessous.

Les résultats portent sur 57 réponses correspondant approximativement à 14% des personnes informées de l'enquête. 90% des sondés sont issus du secteur privé contre 10% du milieu académique (Universitaire, EPIC). La majorité des réponses ont été fournies par des acteurs français (96%). Les activités des sondés sont majoritairement liées au conseil et ingénierie environnementale et la maîtrise d'œuvre, principalement dans le secteur des sites et sols pollués (85%).

Concernant la connaissance des OBMs, la vaste majorité des sondés a entendu parler de ces outils (91%) essentiellement par le biais de leur réseau, participation à des congrès et l'intégration de ces outils dans certaines de leurs études. La technique la plus connue reste l'amplification par PCR (76%). Les autres outils (empreintes génétiques, Puces à ADN et séquençage massif) restent moins connus (25 à 40% des sondés).

Concernant l'utilisation des OBMs, 30% des personnes font appel à ces outils dans leur étude et 50% rarement. 15% des sondés n'utilisent pas ces outils mais considèrent cette option envisageable. Parmi les OBMs utilisés, la PCR (qPCR et RT-qPCR) reste la plus populaire (65%). Les autres outils restent beaucoup moins exploités (7 à 15%). Parmi l'ensemble des projets traités par les sondés, approximativement 15% pourraient bénéficier des OBMs et 5% les exploitent réellement. La majorité des analyses de biologie moléculaire sont réalisées par des prestataires externes (76%) et principalement en France (92%). Les analyses portent sur les différents types de matrice mais majoritairement sur les sols (84%) et eaux souterraines (65%) et en grande majorité pour des problématiques hydrocarbures (80%), PCB et composés chlorés (57%). Ces analyses sont utilisées majoritairement lors de la mise en place de la stratégie de dépollution (77%), dans une moindre mesure pour les autres étapes du projet (40%) et très rarement à réception des projets (4,5%). Ces analyses sont également exploitées en forensie environnementale et pour l'identification des sources de dysfonctionnement (14%).

Concernant la perception des OBMs, les barrières principales limitant leur exploitation sont, outre une méconnaissance de ces outils (32%), l'absence de méthodes standardisées, la difficulté à interpréter les données fournies, ainsi que le coût des analyses. Les axes proposés pour améliorer la perception des OBMs sont ceux mentionnés ci-dessus (rendre résultats plus facilement interprétables et exploitables, développer des méthodes standardisées et diminuer les coûts de réalisation). Diversifier l'offre technique, proposer des programmes de formation et fournir des guides d'échantillonnage sont également considérés comme important pour la majorité des sondés (85 à 88%).

Les OBMs sont donc globalement bien perçus par les acteurs SSP qui sont informés. Toutefois quoique présentant un intérêt grandissant de par leur pertinence concernant la réhabilitation des SSP, les OBMs sont largement sous exploités car représentant un nouveau métier dans le paysage de la dépollution environnementale. En France, les OBMs semblent encore être l'apanage du monde académique et universitaire alors qu'ils sont exploités de manière plus importante aux Etats-Unis par exemple. Les OBMs apparaissent aux acteurs privés comme étant souvent d'une grande complexité scientifique, nécessitant des études en laboratoire longues et fastidieuses dans lesquelles se mêlent recherche et développement. L'impression d'un manque de standardisation, de procédures normalisées pour l'étude d'organismes vivants induit une notion d'incertitude difficile à appréhender dans des problématiques nécessitant souvent des réponses rapides et fiables. Aussi, le fait de se trouver face à des techniques performantes mais sans savoir clairement quelles seront les informations obtenues conduit à une certaine défiance. Les deux freins essentiels restent cependant une méconnaissance de ces outils et des prestataires capables de fournir ces analyses et leur

interprétation. Bien que les OBMs paraissent extrêmement pertinents auprès des acteurs connaissant ces approches car pouvant permettre un gain de temps important, en permettant le ciblage direct de processus naturels, leur méconnaissance limite très probablement leur exploitation. Aux Etats-Unis où l'utilisation des OBMs dans la réhabilitation des SSPs est plus fréquente, la méconnaissance de ces outils reste encore importante. Un sondage mené en 2013 aux Etats-Unis par l'ITRC (*Interstate Technology & Regulatory Council*) auprès de 90 acteurs de la gestion des SSP révèlent notamment que 66% des acteurs n'avaient aucune expérience des OBMs et que 40% d'entre eux n'avaient aucune connaissance des OBMs. Un effort de pédagogie et de vulgarisation, émanant des laboratoires de recherche mais également des prestataires industriels fournissant de telles analyses, semble donc essentiel pour démocratiser l'utilisation des OBMs en SSP. Ces efforts devront porter sur le fait que l'utilisation de ces outils de pointe n'est pas restreinte à de la recherche fondamentale menée en laboratoire mais peuvent apporter des réponses très concrètes à forte valeur ajoutée pour la réhabilitation des SSP.

3.3.2 Identification de la demande

La demande principale se place très probablement dans la nécessité d'établir un catalogue de résultats possibles, répondant à des questions concrètes (**Tableau 2**) plutôt qu'un catalogue de techniques disponibles en développement constant que les acteurs des SSP ne connaissent souvent que très peu. Lors du colloque AdebioTech Biominnov (OBMs appliqués à l'environnement) organisé en 2013, il a clairement été mis en évidence la nécessité d'améliorer les transferts technologiques et de communiquer sur les outils. Cet article constitue modestement un premier effort de démocratisation des OBMs, essayant de mettre en lien les outils disponibles et les solutions concrètes pouvant être apportées.

La demande des utilisateurs porte également sur l'interprétation des données et leur lien avec les données conventionnelles. A la question « Quels sont vos remarques ou besoins spécifiques pour lesquels vous souhaiteriez avoir des outils disponibles ? », les réponses obtenues ont permis d'identifier différents besoins :

- *Documentation* : Fournir une documentation simple sur les OBMs et leur utilisation ; Inversement, mettre en adéquation les OBMs vis-à-vis des besoins en SSP.
- *Formation* : Besoin de formation théorique.
- *Procédure* : Etablir des procédures standardisées permettant de comparer les résultats issus de différents essais

En outre il apparait clairement que les résultats des analyses de biologie moléculaire doivent être accompagnés d'une interprétation et d'une transposition au site afin de fournir aux utilisateurs (ie, BE, gestionnaires) une réponse pragmatique répondant clairement à leurs besoins. Cet effort est, ou devra être, assuré par les prestataires afin de satisfaire la demande mais également de promouvoir l'intérêt et les avantages des OBMs en SSP.

3.3.3 Enquête auprès des acteurs

Les raisons d'une utilisation encore limitée des OBMs en SSP sont multiples. Nous avons mentionné la méconnaissance et la perception de ces outils mais nous pourrions également citer l'absence de compétences en interne ou la méconnaissance des prestataires académiques et industriels. Concernant les personnes ayant connaissance des OBMs, les barrières limitant leur exploitation en SSP sont principalement associées à l'absence de méthodes standardisées permettant de comparer différents sites, les coûts des analyses ainsi que la difficulté d'interprétation des données brutes.

Cette enquête met en évidence que les principaux freins à l'utilisation des OBMs en SSPs sont liés à :

- Une méconnaissance de ces outils et de leur intérêt
- Une difficulté d'interprétation des données fournies
- L'absence de méthodes standardisées
- Une méconnaissance des prestataires pouvant réaliser de telles analyses

4 CONCLUSION

La pollution des sols et milieux aquatiques par des composés xénobiotiques est un problème d'ampleur croissante et l'utilisation de microorganismes pour le traitement de ces sites a offert la possibilité de mettre en place de nouvelles stratégies de gestion des sites et de dépollution. Les stratégies de bioremédiation qui peuvent être appliquées sont classiquement définies comme la bioatténuation naturelle (dégradation des contaminants par les microorganismes naturellement présents et sans intervention humaine), la biostimulation (stimulation des microorganismes naturellement présents par addition d'accepteurs ou donneurs d'électrons, de nutriments...) et la bioaugmentation (addition de microorganismes cultivés en laboratoire possédant les capacités de biodégradation recherchées). Le choix d'une part, et l'optimisation d'autre part, des stratégies à mettre en place dépendent de d'une bonne compréhension des communautés microbiennes indigènes du site. L'étude de ces dernières par des approches de microbiologie classiques (*ie*, culture des microorganismes sur milieux synthétiques), en plus d'être longue, biaise de manière très importante l'estimation de l'abondance et de la diversité des microorganismes présents. L'utilisation des outils de biologie moléculaire permet de passer outre ces limitations et apportent des informations essentielles, jusque-là inaccessibles, sur la structure et le fonctionnement des communautés microbiennes impliquées dans la biodégradation des contaminants.

Le constat actuel est que l'application des OBMs dans le domaine des SSP est encore limitée, principalement parce que ces outils sont méconnus mais également parce que leur interprétation et valeur ajoutée n'est pas toujours évidente pour le gestionnaire. Un effort de communication et de facilitation de transfert technologique doit donc être fourni afin de promouvoir ces outils. De manière très triviale, il est nécessaire de mettre en évidence l'adéquation qui existe entre d'une part, des outils (qui le plus souvent n'évoquent rien pour le non initié) offrant des solutions et d'autre part, les besoins concrets des gestionnaires. Une meilleure connaissance de ces outils et de leurs applications permettra d'identifier les questions auxquelles il est désormais possible de répondre et *in fine* d'intégrer de nouveaux paramètres pour optimiser la gestion des sites et sols pollués. Cependant, jusqu'à présent, l'effort de communication a peut-être trop porté sur la pléiade d'outils disponibles et peut-être pas suffisamment sur les solutions concrètes qu'ils apportent. Cette amélioration de la perception des OBMs peut être assurée par un partenariat laboratoire de recherche / industriels ou par l'intermédiaire des prestataires fournissant ces analyses de biologie moléculaire. Afin d'améliorer la perception des OBMs et de favoriser leur utilisation dans le domaine des Sites et Sols Pollués, il est essentiel de ne pas (plus) mettre les outils au centre des débats mais de communiquer de manière plus efficace sur les nouvelles solutions concrètes que peuvent offrir ces outils, utilisés le plus souvent de manière combinée et en complément des analyses classiques.

FICHES TECHNIQUES : OUTILS DE BIOLOGIE MOLECULAIRE

Les fiches techniques des outils de biologie moléculaire suivants sont présentées dans les pages suivantes :

- F1** : Amplification par PCR
- F2** : Empreintes génétiques
- F3** : Pucés phylogénétiques et fonctionnelles
- F4** : Séquençage massif (NGS)
- F5** : Analyses isotopiques (CSIA et SIP)
- F6** : Analyses cellulaires

Amplification par PCR

F1

Principe

La PCR (*Polymerase Chain Reaction*, Réaction de Polymérisation en Chaîne) est une méthode d'amplification génique, mise au point par le prix Nobel de chimie Kary B. Mullis (1986), qui permet de dupliquer de manière exponentielle une séquence d'ADN ou d'ARN connue à partir d'une faible quantité. C'est une méthode essentielle dans l'analyse des acides nucléiques (*ie*, ADN et ARN) qui est utilisée pour détecter, quantifier ou simplement amplifier des séquences ou gènes d'intérêt à partir d'un mélange complexe de molécules d'acides nucléiques. Le principe de la méthode est décrit dans la **Figure 8**.

Les produits générés par PCR dépendent de la spécificité de la séquence des amorces oligonucléotidiques choisies pour amplifier la région d'intérêt. Ces amorces spécifiques sont des séquences de nucléotides synthétisés à façon. Cette technique permet d'amplifier des morceaux de gènes, des gènes entiers ou encore un ensemble de plusieurs gènes.

La PCR peut être appliquée à

- L'amplification de marqueurs taxinomiques (identité des cellules : eg, 16S, IGS) pour déterminer la structure d'une communauté.
- L'amplification de gènes de fonctions (eg, dégradation) pour déterminer ou quantifier leur présence ou leur activité.

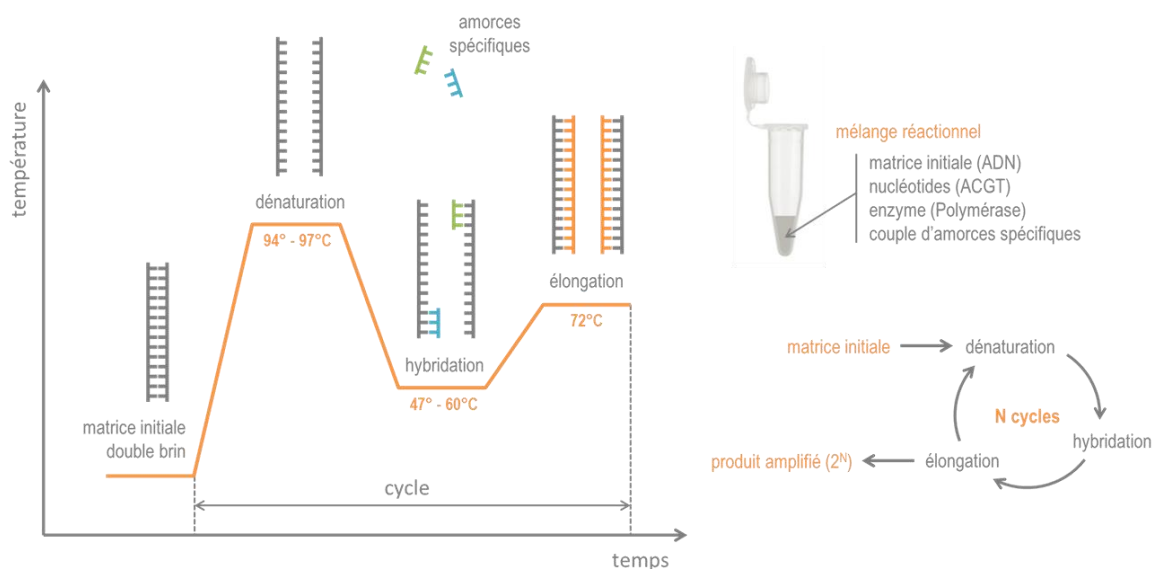


Figure 8 : Principe de la PCR. La technique est basée sur la répétition de cycles de transition de température qui alternent après une phase de dénaturation initiale des molécules d'ADN (1'). Le mélange réactionnel contient l'ADN ou l'ADNc, la polymérase, les nucléotides, et les amorces spécifiques permettant d'amplifier la région d'intérêt. Le nombre de cycles (n), constitués d'étapes de dénaturation (1), hybridation (2), élongation (3) est généralement supérieur à 30 et conduit à l'obtention théorique de 2^n molécules.

Les techniques de PCR classique (dite à point final) ne permettent pas de quantifier précisément le nombre de séquences d'intérêt. Pour cela, des techniques de PCR quantitatives ont été développées (qPCR, RT-qPCR, dPCR).

Les techniques de PCR dites quantitatives (eg, qPCR, RT-qPCR, dPCR) sont utilisées pour quantifier le nombre de copies d'un gène ou d'une région d'intérêt et apportent une information plus complète que la PCR en point final (présence/absence d'un gène). La dPCR (digital PCR) est une technique plus récente fournissant les mêmes informations que la qPCR. Le principe des techniques de PCR quantitative (dPCR et qPCR), est décrit dans les figures ci-dessous (**Figure 9** et **Figure 10**, respectivement). La qPCR et la RT-qPCR permettent respectivement de quantifier le nombre de copies de la région d'intérêt à partir de l'ADN (présence) et de l'ARN (expression). Ces outils sont couramment utilisés pour déterminer par exemple la dynamique des populations microbiennes (présence, augmentation, maintien de l'abondance microbienne) ou du maintien de l'activité des gènes de dégradation.

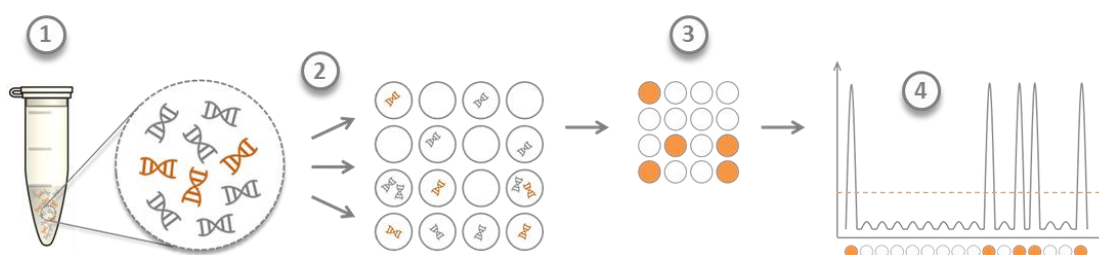


Figure 9 : Principe de la dPCR (digital PCR ou PCR numérique). Dans la PCR numérique, la réaction PCR peut prendre deux valeurs, 0 (absence) ou 1 (présence). Le principe est basé sur la multiplication du nombre de réactions ou l'échantillon est partitionné dans des milliers de compartiments distincts, chaque compartiment étant ensuite considéré comme un réacteur indépendant. L'application de la loi de Poisson permet de quantifier la quantité de séquence cible dans l'échantillon de départ. La précision de la mesure est fonction du nombre de compartiments. Cette technique présente l'avantage de permettre une quantification absolue, sans nécessité d'établir une courbe standard et reste moins sensible aux inhibiteurs du fait de la dilution de l'échantillon.

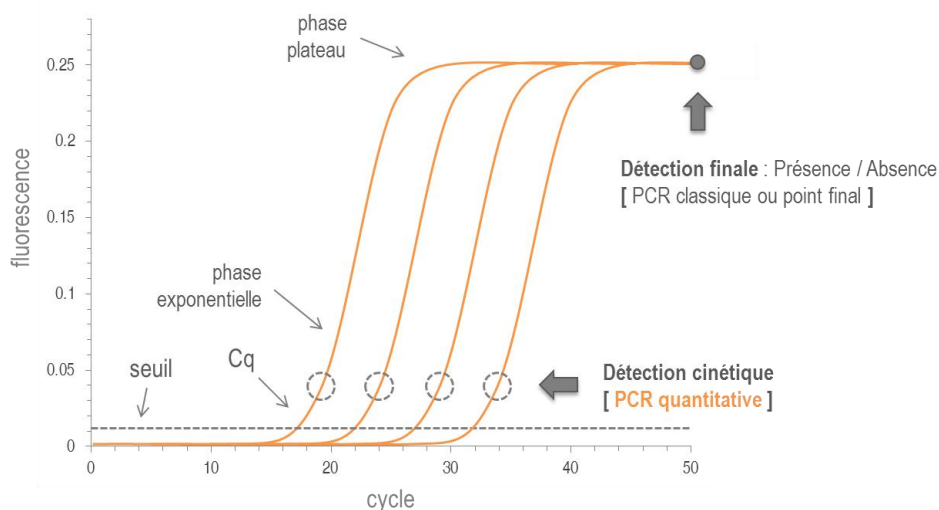


Figure 10 : Principe de la qPCR. Contrairement à la PCR en point final, qui se termine par un dépôt sur gel d'agarose permettant de déterminer si la séquence recherchée est présente ou absente, la PCR quantitative permet de quantifier le nombre de séquences présentes dans la matrice initiale en suivant l'évolution de la libération de molécules fluorescentes dont l'apparition dépend de la quantité de séquences cibles dans la matrice initiale. La RT-qPCR repose sur le même principe mais utilise l'ARN, converti en ADN complémentaire, comme matrice de départ. Elle permet de quantifier l'expression des séquences recherchées.

Informations fournies

La PCR classique, en point final, fournit simplement une information sur la présence ou absence d'un gène d'intérêt. Les techniques de PCR quantitatives (dPCR, qPCR, RT-qPCR) permettent de quantifier l'abondance des séquences recherchées.

Les applications de la PCR sont multiples. Les outils les plus couramment utilisés sont la qPCR et la RT-qPCR. Ces outils permettent la détermination de la présence, l'abondance et l'activité de microorganismes capables de biodégrader les contaminants. La possibilité de quantifier les marqueurs taxinomiques ou fonctionnels fournit des informations essentielles sur la dynamique des populations et processus en jeu (maintien, augmentation, diminution) dans un contexte environnemental et de manière très résolutive.

Les résultats fournis sont exprimés en nombre de copies du gène par millilitre ou litre (échantillon d'eau) ou par gramme pour des échantillons solides tels que les sols ou sédiments. Dans le cas de la RT-qPCR (ARN), le nombre de transcrits variant en fonction de l'activité des cellules, celui-ci peut être rapporté au nombre de copies de la même séquence déterminé par qPCR (ADN). Bien que les données de PCR quantitative fournissent des valeurs absolues, leur interprétation doit prendre en compte le type de contaminant, les voies de biodégradation existantes et ciblées ainsi que des paramètres géochimiques du site au moment de l'analyse.

Empreintes Génétiques

F2

Principe

La réalisation d'empreintes génétiques des communautés, équivalent des codes-barres, permet d'obtenir une image globale de la structure, de la diversité ou encore de l'évolution des communautés. L'avantage de cette approche est qu'il n'est pas nécessaire de savoir quels sont les organismes présents. Ces approches permettent de différencier les microorganismes ou groupes de microorganismes et dans certains cas d'identifier les membres dominants de la communauté.

Le terme générique « empreintes génétiques » regroupe différentes techniques qui permettent d'obtenir des profils génétiques (*ie*, codes-barres) de communautés microbiennes :

Empreintes génétiques à partir d'ADN métagénomique (basées sur l'amplification par PCR)

L'approche consiste, après extraction des ADN, à réaliser une amplification par PCR d'une région du génome commune à tous les microorganismes. Les produits obtenus par amplification PCR sont ensuite séparés par différentes techniques d'électrophorèse en fonction de leur taille, de leur composition en bases (ACGT) ou de leur structure (**Figure 11**). La région cible majoritairement utilisée pour ce genre d'analyses est la région des gènes codant l'ARN ribosomique.

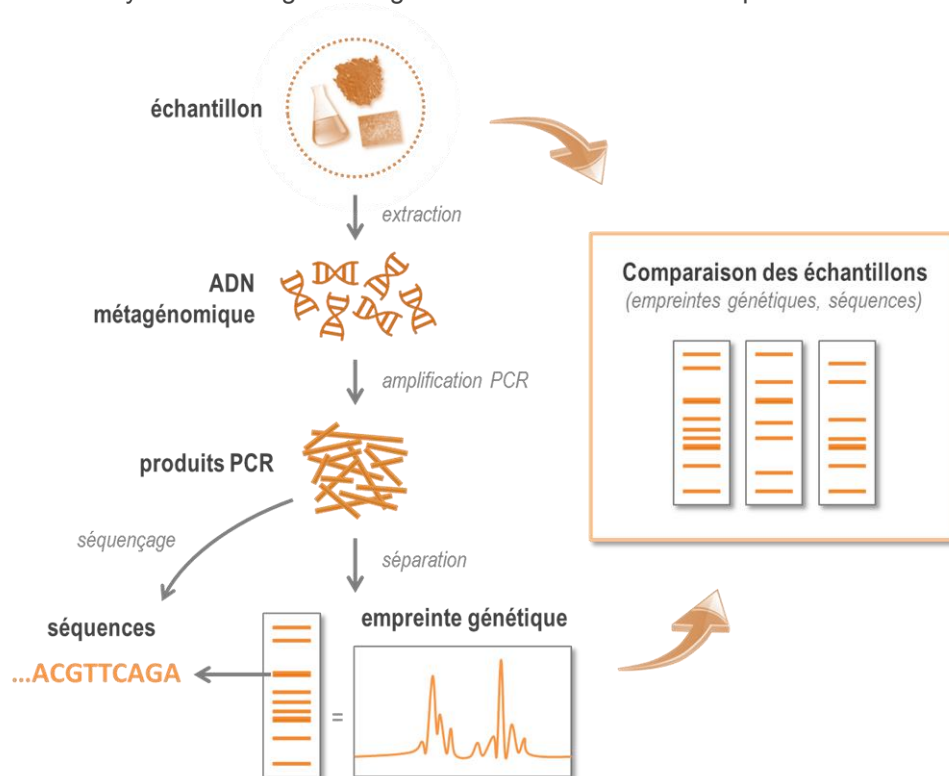


Figure 11 : Description de l'approche permettant d'obtenir une empreinte génétique d'une communauté microbienne. Les profils génétiques générés, sortes de codes-barres des communautés, peuvent être comparés afin de déterminer l'évolution de la structure de la communauté ou de l'abondance relative de certains groupes. Par défaut ces approches ne permettent pas d'identifier les groupes bactériens mais peuvent être complétés par des techniques de clonages, séquençages ou hybridation afin d'identifier les membres dominants de la communauté.

L'ADN ribosomique (ADNr) est une séquence d'ADN codant pour l'ARN ribosomique (ARNr) constituant les ribosomes. Le faible polymorphisme de cette région en fait un marqueur universel utilisé pour la classification des microorganismes (Woese, 1987).

L'ADNr est devenu le gène de référence pour l'étude de la diversité microbienne. Bien qu'il existe différents types d'ARNr (procaryote : 5S, 16S, 23S et eucaryote : 5.8S, 18S et 28S) le 16S est le plus fréquemment utilisé pour ces applications. Ces ARNr sont souvent comparés à des chronomètres moléculaires en raison de leur ubiquité et de leur forte conservation de séquence. Ils présentent aussi l'avantage d'être extrêmement abondants dans la cellule du fait de leur rôle dans la structure des ribosomes nécessaires à la synthèse des protéines.

Par conséquent, il est possible de définir des amorces ciblant différents niveaux de spécificité taxinomique (eg, amorces universelles ou spécifiques d'une espèce donnée). Les produits PCR obtenus peuvent par la suite faire l'objet de divers traitements et analyses complémentaires (eg, clonage, séquençage ou hybridation) afin de déterminer la diversité taxinomique de la communauté (par exemple, l'identification des membres dominants de la communauté).

Parmi les méthodes d'empreintes génétiques à partir d'ADN métagénomique principalement utilisées, citons :

- la DGGE (*Denaturing Gradient Gel Electrophoresis*) qui est basée sur la séparation de produits PCR de l'ADNr par électrophorèse sur gel de polyacrylamide en fonction de leur composition en bases GC.
- La RISA (*Ribosomal Intergenic Spacer Analysis*) qui est basée sur l'amplification de la région intergénique comprise entre les gènes ribosomiques, créant un profil génétique basé sur le polymorphisme de longueur de cette région.
- D'autres techniques (T-RFLP, ARDRA, SSCP, RAPD...) qui peuvent également être utilisées pour obtenir des empreintes génétiques par amplification PCR des régions d'intérêt, suivie ou non d'une étape de digestion enzymatique partielle à l'aide d'enzymes de restriction, et d'une séparation sur gel d'électrophorèse.

Empreintes génétiques basées sur l'analyse des acides gras (PLFA)

La PLFA (*PhosphoLipid-derived Fatty Acids*) diffère des approches précédentes car basée, non pas sur l'analyse de l'ADN mais, sur l'obtention des profils en acides gras (constituants de la membrane des cellules) (**Figure 12**). Cette technique fournit des données qualitatives et semi-quantitatives sur la structure des communautés. L'avantage principal de cet outil est qu'il est basé sur l'analyse des acides gras qui sont rapidement synthétisés dans les cellules et sont rapidement dégradés lorsque les cellules meurent, fournissant ainsi un instantané précis de la structure de la communauté. Cette technique n'est cependant pas applicable à tous les microorganismes (eg, méthanogènes et archaebactéries).

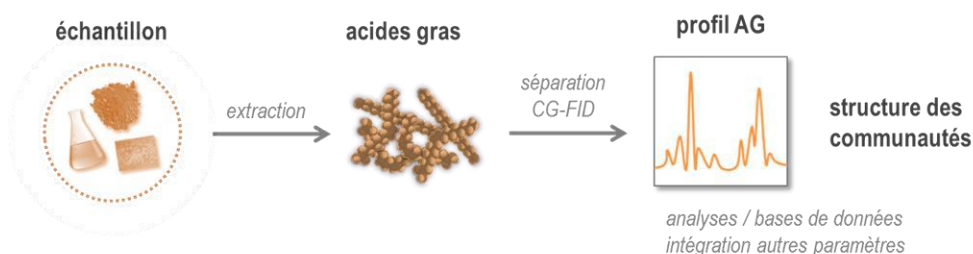


Figure 12 : Principe de l'obtention d'empreintes génétiques par la technique PLFA.

Informations fournies

Les techniques d'empreintes génétiques sont utilisées par exemple pour suivre l'évolution de la structure d'une communauté ou évaluer la structure et composition des communautés et le type de processus microbiens potentiellement présents lors de la caractérisation initiale du site afin de contribuer à l'élaboration d'un modèle conceptuel. Ces approches restent cependant globales et peuvent être couplées à d'autres outils pour une information plus ciblée (eg, PCR quantitative, séquençage).

Les résultats fournis (images des gels d'électrophorèse, analyses statistiques type ACP, indices de diversité...) permettent d'avoir une information qualitative sur la structure et la diversité des communautés. Des échantillonnages multiples (temps et espaces) vont permettre de comprendre la dynamique des populations sur le site et de faire le lien avec les autres paramètres du site.

Puces Phylogénétiques et Fonctionnelles

F3

Principe

Les puces phylogénétiques et les puces fonctionnelles représentent une collection de fragments d'ADN (sondes nucléotidiques) synthétisés artificiellement et fixés sur une surface (eg, lame de verre). Ces sondes sont spécifiques par complémentarité de certaines séquences d'ADN ou d'ARN environnementales, dénommées les cibles. Les ADN ou ARN cibles extraits de l'environnement vont être marqués par un fluorophore (molécule fluorescente permettant de détecter les fragments nucléiques). Ces cibles seront ensuite exposées aux sondes présentes sur le support où seules les séquences cibles complémentaires des sondes pourront s'hybrider (**Figure 13**). Après lavage des cibles non hybridées, le signal de fluorescence des cibles fixées aux sondes est lu. L'information obtenue est non seulement une information de présence (sondes allumées) / absence (sondes non allumées) mais aussi une information quantitative relative, dans le sens où la quantité de fluorescence détectée est proportionnellement relative à la quantité d'ADN ou d'ARN cibles hybridée.

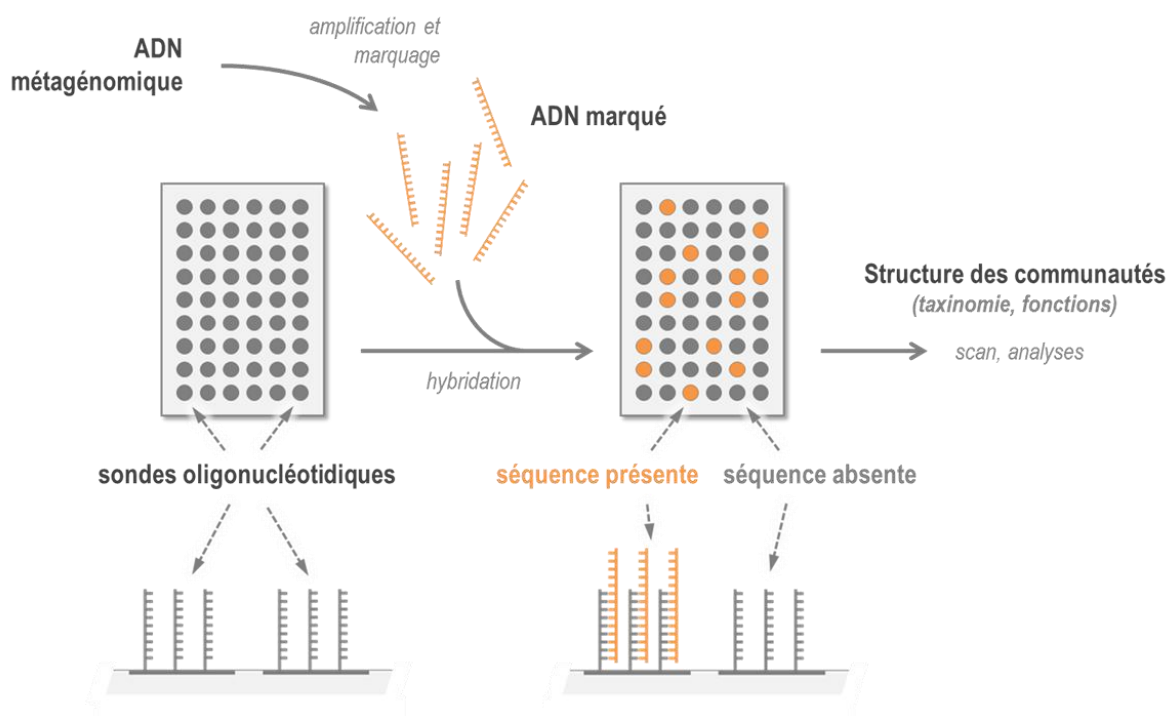


Figure 13 : Principe des puces à ADN. Les molécules cibles sont marquées avec un fluorophore (molécule fluorescente permettant de détecter les fragments nucléiques) puis hybridées sur un support sur lequel sont placées des sondes nucléotidiques (courtes séquences spécifiques des gènes ciblés) recherchées.

Informations fournies

Cette méthode permet de discriminer des biomarqueurs spécifiques dans des échantillons environnementaux qui peuvent contenir des milliers de microorganismes.

Les puces phylogénétiques visent des marqueurs taxinomiques de microorganismes (eg, le gène 16S rRNA) et permettent ainsi de déterminer quels sont les microorganismes présents et leur proportion relative dans la communauté étudiée. Appliquée avec l'ARN, cette méthode permet de savoir quels sont les organismes actifs.

Les puces fonctionnelles fournissent des informations sur le potentiel génétique, de biodégradation par exemple, d'un environnement. Ces puces ciblent des gènes impliqués dans des procédés spécifiques, par exemple des gènes codant des enzymes d'une voie de dégradation. En appliquant les puces fonctionnelles à l'ARN, il est possible de déterminer l'activité de gènes fonctionnels spécifiques de certaines voies métaboliques.

Dans les deux cas, puces phylogénétiques ou fonctionnelles, les données qui en sont extraites sont de deux types : présence/ absence (ou détecté/non détecté) ou abondance relative (en fonction des signaux mesurés). Les résultats fournis sont sous forme de tableaux avec la liste des sondes détectées et les gènes ou organismes correspondant. Une interprétation plus poussée peut permettre d'identifier les conditions rencontrées sur site et les types de métabolismes en place.

Dans le cadre des SSP, cette approche peut amener différentes informations. Par exemple, dans le cadre de l'étude de la bioremédiation d'un site contaminé, elles pourront permettre :

- de connaître les espèces présentes,
- de savoir si le potentiel fonctionnel relatif à la dégradation voulue est là,
- de suivre l'évolution des communautés et l'activité des gènes pendant le processus de bioremédiation
- de vérifier le retour à un état stable et « naturel » de l'environnement.

La limite principale de cette technique, par rapport aux approches basées sur les empreintes génétiques qui ne nécessitent pas d'*a priori*, est qu'elle permet de détecter uniquement les organismes ou gènes pour lesquelles des sondes ont été définies et placées sur la puce.

Séquençage Massif

F4

Principe

Le séquençage est un procédé utilisé pour déterminer l'ordre (la séquence) des bases dans les acides nucléiques (ADN et ARN). Le séquençage massif consiste à appliquer ce procédé sur l'ensemble de l'ADN (ou de l'ARN après transformation en ADNc) de la totalité d'un échantillon, sans *a priori*.

Le métagénome (ADN total de l'échantillon) ou le métatranscriptome (ARN total de l'échantillon) des organismes présents dans l'environnement d'intérêt sont extraits et fragmentés. Ce mélange de fragments génomiques est par la suite séquençé, c'est-à-dire que la séquence complémentaire d'un brin correspondant à un fragment est amplifiée et le signal d'amplification spécifique de chaque acide nucléique incorporé dans la séquence est lu. Cette succession de bases nucléiques (ACGT) représente la séquence du fragment d'intérêt (**Figure 14**).

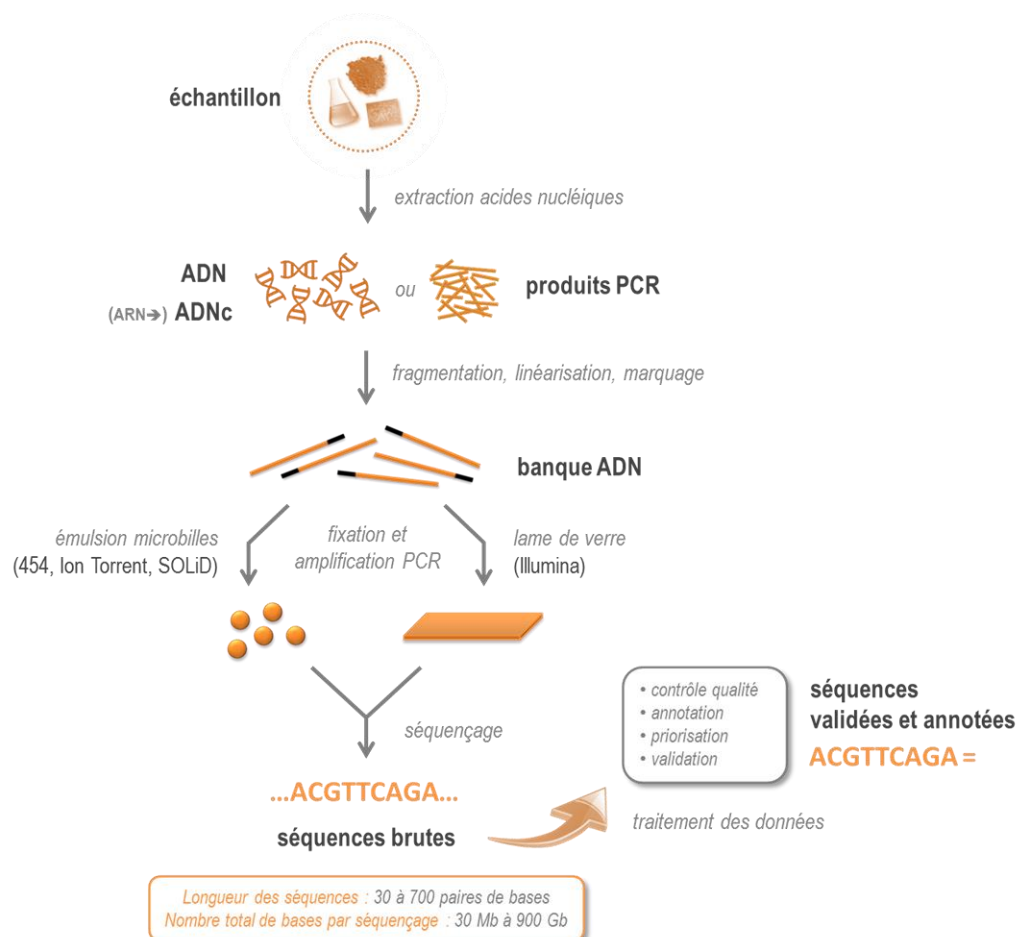


Figure 14 : Principe du séquençage massif.

Les technologies NGS (*Next Generation Sequencing*), ou nouvelles générations de séquenceurs, permettent aujourd'hui de séquencer des millions de fragments sur des longueurs allant jusqu'à 750 paires de bases. Parmi ces NGS nous pouvons citer les technologies commerciales Ion Torrent, SOLiD et Illumina (HiSeq, MiSeq) et PacBio. Un travail de bioinformatique ultérieur est nécessaire afin de reconstruire des fragments plus longs, de valider, d'annoter, d'identifier des séquences par comparaison avec des bases de données de séquences connues.

Ce type d'analyse de séquençage massif permet non seulement d'identifier des organismes présents, mais aussi d'identifier des gènes codant des enzymes impliquées dans des voies métaboliques relatives aux processus environnementaux étudiés. Sa force réside dans la quantité de séquences générée et la précision de l'information. Le biais principal de cette technologie, qui est également un biais pour les autres outils de biologie moléculaire, est qu'une grande partie des séquences obtenues sont inconnues et ne sont donc pas référencées dans les bases de données. L'analyse porte donc seulement sur les organismes et fonctions référencés dans les bases de données et pour lesquelles l'annotation des nouvelles séquences a pu être réalisée. Plus concrètement, étant donné que l'ensemble des gènes de biodégradation n'est pas encore connu, les analyses de biologie moléculaire peuvent donc sous-estimer le potentiel de biodégradation d'un site donné.

Informations fournies

Le séquençage massif sur ADN fournit une image des séquences nucléiques présentes dans l'échantillon, il s'agit du potentiel génétique. En appliquant ces technologies sur l'ARN, il est possible de déterminer si ce potentiel s'exprime et d'identifier les microorganismes et fonctions actifs. Le séquençage peut être réalisé sur l'ensemble des acides nucléiques extraits ou sur une région précise impliquant, en amont, une amplification par PCR de cette région (eg, gène universel 16S pour l'identification des organismes présents). Les données fournies par le séquençage seul ne permettent pas de faire une quantification absolue des fonctions ou bactéries présentes mais fournissent une information relative (eg, % d'une espèce ou d'une fonction donnée) permettant d'identifier les groupes dominants ou minoritaires.

Analyses Isotopiques (CSIA et SIP)

F5

Principe

L'utilisation des isotopes dans le domaine des sites et sols pollués n'est pas à proprement parlé un outil de biologie moléculaire mais exploite cependant les OBM lors des analyses et est couramment utilisé en complément des outils traditionnels. Nous décrivons ici les deux outils principalement utilisés, la technologie SIP (*Stable Isotope Probing*) et CSIA (*Compound-Specific Isotope Analysis*).

La technologie SIP (Traçage Isotopique)

La technologie SIP (Traçage Isotopique) permet de tracer spécifiquement les membres d'une communauté microbienne impliqués dans le processus de dégradation d'intérêt.

Elle repose sur la capacité de ces microorganismes à cataboliser (ou assimiler) un substrat organique préalablement marqué avec un isotope stable, lourd (eg, ^{13}C) et à incorporer cet isotope marqué dans les constituants de la cellule (**Figure 15**). Après incubation de l'échantillon en présence du substrat marqué, les acides nucléiques (ADN et ou ARN) sont extraits et les fractions lourdes et légères séparés par ultracentrifugation. Les fractions collectées peuvent ensuite être analysées et caractérisées en utilisant les différents OBM.

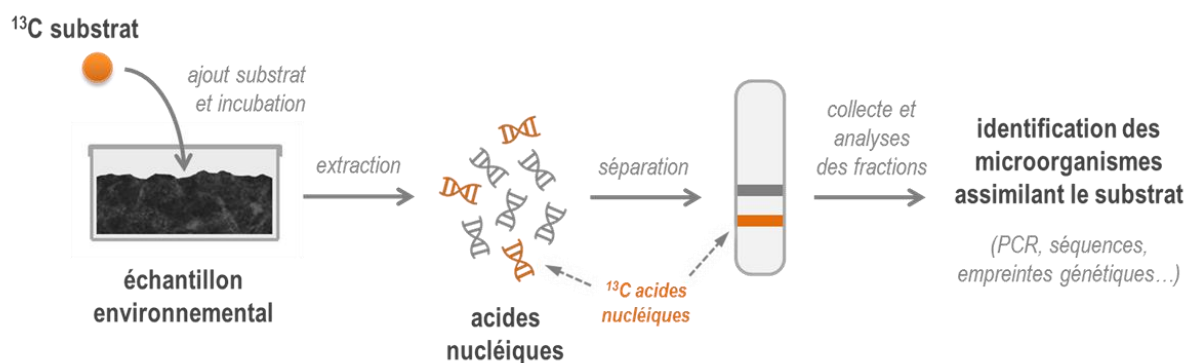


Figure 15 : Principe du traçage isotopique par SIP (*Stable Isotope Probing*).

La technologie CSIA (Analyse des rapports isotopiques des composés spécifiques)

La technologie CSIA (Analyse des rapports isotopiques des composés spécifiques) exploite le fait que les molécules organiques sont naturellement constituées d'atomes présents sous la forme de différents isotopes stables. La séparation et l'analyse des composés sont basées sur une séparation des composés par chromatographie (liquide ou gaz), une conversion chimique des molécules en gaz simples constitués d'un ou deux atomes et d'une séparation des fractions lourdes et légères par spectrométrie de masse (**Figure 16**). L'évolution des ratios entre les différentes fractions peut, dans certaines conditions, mettre en évidence des processus de dégradation biotique et abiotique.

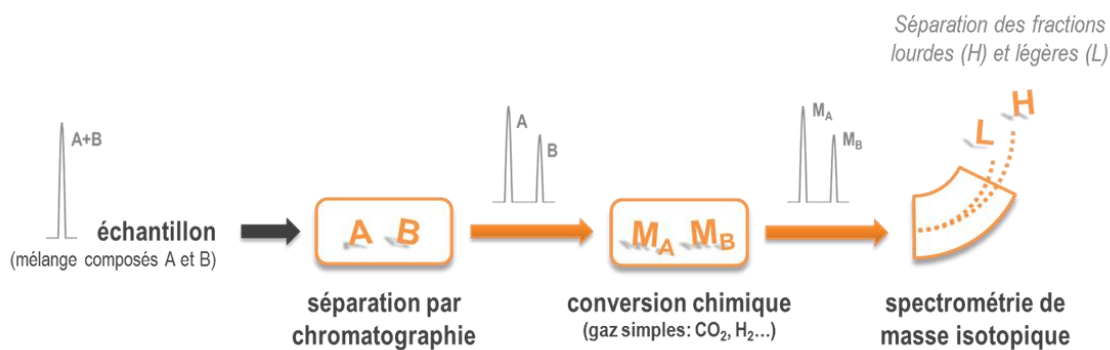


Figure 16 : Principe de l'analyse des rapports isotopiques des composés spécifiques (CSIA).

Informations fournies

La technologie SIP permet d'identifier les organismes capables de biodégrader les composés organiques ciblés (typiquement un contaminant dont on cherche à savoir s'il peut être biodégradé par les communautés microbiennes en place). Elle peut être utilisée pour tout type de molécule organique même si le coût de synthèse et de marquage de certains composés peut être très élevé. Les informations recueillies dépendent du type d'analyses réalisées à partir des acides nucléiques marqués.

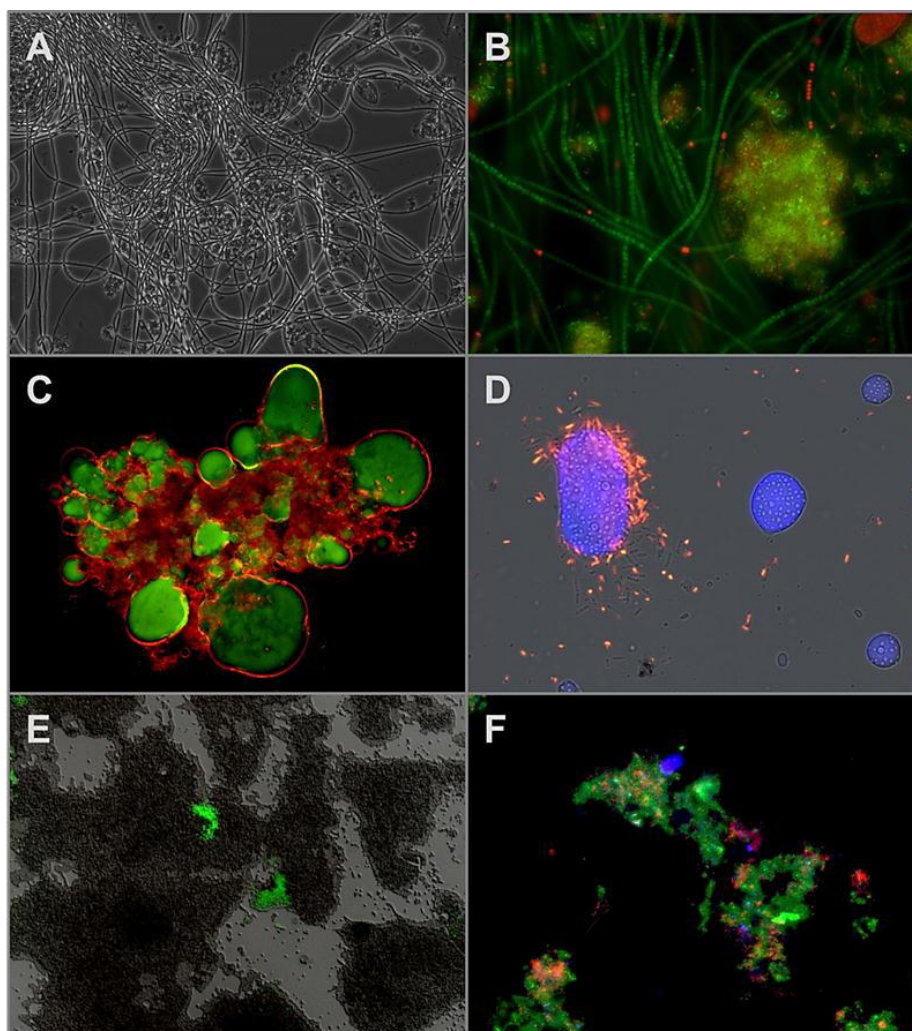
La technologie CSIA vient compléter les OBMs en permettant par exemple de retracer l'histoire des polluants, d'identifier plusieurs sources potentielles ou encore de déterminer leur origine et les voies de dégradation de certains composés. Les pollutions complexes peuvent être parfois difficiles à caractériser et les outils de biologie moléculaires (empreintes génétiques, puces, NGS) viennent compléter ces analyses CSIA notamment dans le domaine de la forensie environnementale.

Analyses Cellulaires

F6

Principe

Les analyses cellulaires sont principalement basées sur l'observation des échantillons bruts ou des microorganismes extraits des échantillons par microscopie. Notons l'utilisation de la cytométrie en flux permettant de trier et d'isoler des cellules en fonction de certains critères (eg, taille, identité, fonction...) et de réaliser des analyses de biologie moléculaire à partir des sous-populations récupérées.



Images: Jean-Michel Monnier

Figure 17 : Exemples d'utilisation de la microscopie pour la caractérisation des communautés microbiennes dans le domaine des sites et sols pollués. (A) Identification de l'origine biologique d'un échantillon et présence de bactéries filamenteuses ; (B) Visualisation des cellules actives (vertes) et non actives (rouges) ; (C) et (D) Visualisation de bactéries actives dans un échantillon pollué aux hydrocarbures ; (E) et (F) exemples de FISH permettant de visualiser un groupe ou plusieurs groupes de bactéries d'intérêt (vert, bleu, rouge) dans un échantillon complexe.

Différentes techniques de microscopie peuvent être utilisées pour étudier les communautés.

Microscopie couplée à l'utilisation de fluorochromes

Les plus informatives étant celles permettant l'utilisation de fluorochromes (molécules fluorescentes permettant de marquer spécifiquement différents composants de la cellule, tels que les acides nucléiques, membranes et protéines) comme la microscopie à épifluorescence ou la microscopie confocale. L'utilisation des fluorochromes est plus récente que celles des isotopes mais présente moins de contraintes au niveau de la gestion des échantillons et des équipements nécessaires en laboratoire et permet surtout d'obtenir une information immédiate. Utilisées dans les outils de biologie moléculaire pour le marquage des acides nucléiques (eg, PCR quantitative, puces, séquençage), elles peuvent être appliquées directement au niveau de la cellule (*Figure 17*).

La microscopie couplée à l'utilisation de fluorochromes va permettre par exemple de déterminer si des microorganismes sont présents dans l'échantillon et d'estimer leur densité et activité. Cette approche ne nécessite aucune préparation des échantillons et peut être appliquée en ajoutant directement le fluorochrome à l'échantillon, qu'il soit liquide ou solide (*Figure 17, A-D*).

Technique de FISH (Fluorescent In Situ Hybridization),

L'utilisation de sondes nucléotidiques marquées (technique de FISH) est appliquée directement sur la cellule bactérienne. Cette technique qui nécessite des étapes d'hybridation et de lavage, va permettre d'identifier, de quantifier et de localiser certains groupes bactériens d'intérêt à différents niveaux taxinomiques (*Figure 17, E-F*) dans une suspension bactérienne ou directement dans l'échantillon.

Informations fournies

L'information fournie par les analyses au niveau cellulaire est complémentaire des autres OBMs puisqu'elle renseigne sur la localisation et l'organisation des cellules dans un échantillon. Elle permet d'estimer la biomasse active et de quantifier le nombre de microorganismes.

Sa limite principale réside dans le fait que les échantillons observables en microscopie sont de petites tailles et une bonne représentativité nécessite souvent le traitement de plusieurs échantillons.

L'observation directe des échantillons en microscopie permet également d'identifier de manière catégorique l'origine biologique d'un problème de colmatage ou biofouling en amont de l'utilisation des autres OBMs permettant de caractériser les organismes présents dans les échantillons.

LEXIQUE

Acide nucléique : terme désignant une molécule constituée d'un enchaînement linéaire de nucléotides. Il en existe deux catégories : les acides désoxyribonucléiques (ADN) et les acides ribonucléiques (ARN).

ADN : abréviation d'Acide DésoxyriboNucléique. Cette molécule, de structure hélicoïdale, est plus ou moins longue selon les espèces (cyclique chez les bactéries, finie chez les organismes pluricellulaires). Elle est classiquement composée de paires de bases azotées : adénine associée à la thymine, et cytosine associée à la guanine. Sa longueur est exprimée en nombre de paires de bases (pb).

ADNc : ADN "complémentaire", copie d'un gène généré à partir de l'ARN par transcription inverse.

Amorce : courte séquence nucléotidique complémentaire de la matrice à amplifier et servant de point d'ancrage lors de la synthèse du brin complémentaire par une ADN polymérase.

ARN : Acide RiboNucléique, molécule servant à transférer les instructions génétiques inscrites dans l'ADN. L'ARN est également formé par un enchaînement précis de nucléotides.

ARNm : acide ribonucléique messager, forme sous laquelle le message génétique codé dans les gènes est transféré. Cet ARN est une sorte de "copie" d'un gène, destinée à être utilisée comme modèle (par les ribosomes) pour la synthèse d'une protéine.

ARNr : acide ribonucléique ribosomique, l'un des deux constituants des ribosomes.

ARNt : acide ribonucléique de transfert. Courte molécule d'ARN de structure complexe, qui joue un rôle fondamental dans la synthèse des protéines. Il apporte les acides aminés au ribosome.

Bactérie : micro-organisme unicellulaire sans noyau (procaryote) dont le génome est constitué d'ADN circulaire. La bactérie contient un seul chromosome et éventuellement des plasmides.

Bases azotées : molécules entrant dans la composition des acides nucléiques (ADN et ARN), ce sont les éléments porteurs de l'information des nucléotides. L'ADN est constitué des 4 bases suivantes : A = adénine ; G = guanine ; C = cytosine ; T = thymine. L'ARN est constitué des 4 bases suivantes : A = adénine ; G = guanine ; C = cytosine ; U = uracile.

Bioaugmentation : méthode consistant à ajouter des microorganismes cultivés pour accélérer la biodégradation d'un contaminant.

Biodégradation : décomposition de matières organiques par des microorganismes.

Biomarqueur : gène ou séquence génétique définie utilisé comme marqueur et pouvant être détecté ou quantifié facilement lors des analyses moléculaires.

Clonage : multiplication in vitro d'un organisme ou d'un gène, en grand nombre d'exemplaires identiques.

Communauté : Ensemble d'organismes interagissant et partageant un même environnement.

CSIA (compound specific isotope analysis) : Analyses des rapports isotopiques de composés organiques.

Dénaturation : la dénaturation (de l'ADN) est un processus qui consiste à séparer les deux brins complémentaires d'une molécule d'ADN en l'exposant à une température ou un pH élevé. C'est la rupture des liaisons hydrogènes entre les bases qui provoque cette séparation.

Désoxyribose : sucre, pentose, provenant du ribose et qui a subi la perte d'un atome d'oxygène (désoxygénation). Le désoxyribose entre dans la composition des nucléotides qui constituent l'ADN.

dPCR (digital PCR) : PCR numérique. Technique permettant de quantifier le nombre de séquences cibles dans un échantillon.

Electrophorèse : cette technique permet de séparer des molécules en fonction de leur taille et de leur charge en utilisant un courant

électrique. On peut ainsi analyser et purifier dans un milieu gélifié (gel d'agarose, gel de polyacrylamide..) l'ADN, l'ARN, les protéines.

Empreinte génétique microbienne : Technique de biologie moléculaire permettant d'obtenir un code-barres d'une communauté microbienne, renseignant sur l'abondance relative et la diversité de celle-ci.

Enzymes : Macromolécule protéique (protéines et ribozymes) jouant le rôle de catalyseur dans la synthèse ou dégradation de composés organiques. Comme toutes les protéines, les enzymes sont synthétisées à partir des informations codées dans l'ADN via la traduction de l'ARN.

Espèce : Unité de base taxinomique dans la classification des organismes, généralement associée au genre (eg, Genre *Escherichia* espèce *coli*).

Eubactérie : les eubactéries sont les bactéries dites « vraies », contrairement aux Archaeobactéries. Plus de 10,000 espèces ont déjà été repertoriées.

Expression : on dit d'un gène qu'il "s'exprime" quand il est actif, c'est à dire, quand il est transcrit sous forme d'ARN messager, puis traduit sous forme de protéine.

FISH (fluorescent *in situ* hybridization) : technique de biologie moléculaire utilisant des sondes marquées d'un composé fluorescent et utilisés en microscopie ou imagerie moléculaire.

Forensie environnementale : Discipline recouvrant les notions d'expertise et d'enquête scientifique visant à déterminer l'origine, les causes ou l'âge d'une pollution.

Gène : unité d'hérédité contrôlant un caractère particulier. Cet élément génétique correspondant à un segment d'ADN ou d'ARN (virus), situé à un endroit bien précis (locus) sur un chromosome. Chaque région de l'ADN qui produit une molécule d'ARN fonctionnelle est un gène. Le gène est responsable d'une fonction spécifique, correspondant le plus souvent à la synthèse d'une protéine. Chez les eucaryotes les gènes sont portés par les chromosomes mais aussi par l'ADN extranucléaire, cas des mitochondries et des chloroplastes. Chez les procaryotes, les gènes sont localisés dans un chromosome circulaire et éventuellement dans des plasmides.

Génétique : science de l'hérédité. La génétique étudie les caractères héréditaires des individus, leur transmission au fil des générations et leurs variations (mutations).

Génome : ensemble du matériel génétique d'un individu. Patrimoine héréditaire d'un individu.

Génomique : science des génomes. La génomique regroupe un ensemble d'analyses qui vont de l'établissement de cartes du génome (cartographie) à l'identification de nouveaux gènes, à l'étude de leurs fonctions et au séquençage des molécules d'ADN.

Genre : En systématique, le genre est un rang taxinomique regroupant un ensemble d'espèces (eg, Genre *Pseudomonas* sp., *Dehalococcoides* sp.).

Hybridation moléculaire : technique permettant de mettre en évidence au sein d'une cellule ou d'un tissu, une séquence d'acide nucléique. Elle est basée sur le principe de complémentarité des bases nucléiques.

Isotope : nucléide d'un élément chimique caractérisé par un nombre de neutrons distinct (eg, isotopes du carbone ¹²C, ¹³C, ¹⁴C).

Métabolisme : C'est l'ensemble de toutes les réactions chimiques s'effectuant dans la cellule vivante. Le catabolisme (ensemble des réactions de dégradations moléculaires de l'organisme considéré) et l'anabolisme (ensemble des réactions de synthèse) sont les deux composantes du métabolisme.

Métagénome : Dans le cas d'un échantillon complexe ou plusieurs individus sont présents, on ne parle plus de génome mais de métagénome. Le métagénome est défini comme l'ensemble des génomes, et donc de tous les gènes des organismes, présents dans un environnement donné (ADN total de l'échantillon).

Métatranscriptome : Ensemble des molécules d'ARN messagers transcrites extraits d'un environnement complexe. Dans le cas d'un échantillon complexe ou plusieurs individus sont présents, son analyse permet de déterminer les fonctions exprimées par les organismes présents et d'étudier collectivement les gènes exprimés (ARN total de l'échantillon)..

Microscopie à épifluorescence et confocale : Techniques de microscopie optique et d'imagerie des objets de petites dimensions (μm) pouvant atteindre des grossissements de 2000x et tirant profit des phénomènes de fluorescence.

NGS : Technologies de séquençage de nouvelle génération (Next Generation Sequencing) permettant de déterminer l'ordre des nucléotides (ACGT) dans une molécule d'ADN. Ces techniques permettent de séquencer massivement et à moindre coût des milliers ou des millions de séquences en même temps à partir d'un métagénome. Citons par exemple les technologies Sanger, 454, SOLiD, Illumina et Ion Torrent.

Nucléoside : élément constitutif des acides nucléiques (ADN et ARN) constitué d'une base azotée (purine ou pyrimidine), associée à un sucre (pentose : ribose pour l'ARN et désoxyribose pour l'ADN). Les quatre unités nucléosidiques de l'ADN sont : désoxyadénosine, désoxyguanosine, désoxythymidine et désoxycytidine.

Nucléotide : unité de construction des acides nucléiques, résultant de l'addition d'un sucre (ribose pour l'ARN et désoxyribose pour l'ADN), d'un groupement phosphate et d'une base azotée à l'origine de l'information. Il existe quatre nucléotides différents pour l'ADN : adénine (A), thymine (T), guanine (G), cytosine (C) et quatre nucléotides différents pour l'ARN : uracile (U), guanine (G), cytosine (C), adénine (A). C'est la succession des bases résultant de l'enchaînement des nucléotides dans l'acide nucléique qui constitue le message génétique.

PCR : Amplification d'un fragment d'ADN réalisée par une réaction en chaîne par polymérisation. Outil de biologie moléculaire d'amplification génique *in vitro* permettant de dupliquer en grand nombre une séquence d'ADN ou d'ARN connue, à partir d'une faible quantité (de l'ordre de quelques picogrammes) d'acide nucléique (voir également dPCR, qPCR et RT-qPCR).

Phylogénie : La phylogénie moléculaire est l'utilisation de séquences de macromolécules biologiques (marqueurs phylogénétiques) pour obtenir des informations sur l'histoire évolutive des êtres vivants, et notamment sur leurs liens de parenté.

Plasmide : Désigne en microbiologie ou en biologie moléculaire une molécule d'ADN surnuméraire distincte de l'ADN chromosomique, capable de répllication autonome et non essentielle à la survie de la cellule. Les plasmides sont généralement circulaires et se trouvent quasi-exclusivement chez les bactéries.

PLFA (phospholipid fatty acids) : méthode d'empreintes moléculaire basée sur l'extraction et l'analyse des acides gras phospholipidiques.

Polymérase : enzyme capable d'enchaîner des nucléotides en polymères d'ADN ou d'ARN (ADN polymérase ou ARN polymérase).

Protéine : macromolécule composée d'une ou plusieurs chaînes d'acides aminés. Les protéines peuvent jouer un rôle structural ou fonctionnel et assurent la vaste majorité des fonctions cellulaires.

Puces (phylogénétiques, fonctionnelles) : ensemble de sondes nucléotidiques fixées et ordonnées sur un petit support en verre, silice ou plastique permettant d'analyser les gènes ou transcrits présents dans un échantillon.

Pyroséquençage : technique de séquençage rapide ne nécessitant pas de clonage.

qPCR et **RT-qPCR** : outils de biologie moléculaire permettant de mesurer la quantité d'une séquence génétique dans un échantillon. Le qPCR utilise l'ADN comme matrice. Le RT-qPCR utilise l'ARN comme matrice.

Ribosome : organite du cytoplasme de la cellule où se déroule le processus de traduction de l'ARN messager en protéine. Cette fonction est commune à toutes les cellules vivantes.

Séquençage : procédé utilisé pour déterminer l'ordre (la séquence) des acides aminés d'une protéine ou des bases dans les acides nucléiques (ADN et ARN).

Séquençage massif : voir NGS.

SIP (stable isotope probing) : technique de traçage isotopique permettant d'identifier les acteurs responsables de la dégradation d'un composé marqué.

Sonde nucléotidique : segment de nucléotides synthétisés à façon et fixés sur une surface permettant de rechercher de manière spécifique un fragment d'acide nucléique complémentaire par hybridation (voir puces, FISH).

Traduction : processus permettant la synthèse d'une chaîne polypeptidique (protéine) à partir d'un brin d'ARN messager. La traduction a lieu au niveau des ribosomes.

Transcription : processus permettant la copie de l'ADN en ARN. L'ARN synthétisé peut être de type ARN messager (ARNm), de transfert (ARNt) ou ribosomique (ARNr). C'est la première étape du processus qui permet de passer de l'ADN à la protéine, ou plus concrètement du gène à son produit.

Index des tableaux et figures

Tableaux ⁽ⁱ⁾

Tableau 1 (partie 1/2) : Description, avantages et limites potentielles des différents outils de biologie moléculaire.....	10
Tableau 2 : Exemples de questions auxquelles peuvent répondre les différents OBMs* dans le cadre des SSP.....	14
Tableau 3 : Apport des données de biologie moléculaire aux données classiques de caractérisation des sites et sols pollués.....	16

Figures ⁽ⁱ⁾

Figure 1 : Exemples de réponses apportées par les outils de biologie moléculaire au cours des différentes étapes de traitement de sites et sols pollués.....	5
Figure 2 : Composants de la cellule analysés à l'aide des outils de biologie moléculaire. L'ADN (Acide DésoxyriboNucléique) est le support de l'information génétique d'un individu et constitue son génome (ensemble des gènes). Lorsqu'un individu exprime un gène (activité de ce gène), l'ADN est transcrit en ARN. L'ARN (Acide RiboNucléique) est la molécule servant à transférer les instructions génétiques inscrites dans l'ADN. L'ARN, qui constitue un reflet de l'activité de la bactérie, est ensuite traduit pour former les protéines. Les protéines peuvent avoir un rôle structurel (constituant de la cellule) ou fonctionnel (enzymes). Les acides gras sont des constituants de la cellule, pouvant être utilisés pour différencier les organismes (voir PLFA).....	6
Figure 3 : Représentation schématique des différentes étapes à considérer lors de la réalisation d'analyses d'échantillons environnementaux à l'aide des outils de biologie moléculaire.....	7
Figure 4 : Palette d'outils de biologie moléculaire disponibles. Les outils classiquement utilisés en SSP sont signalés par les codes F1 à F6, correspondants aux fiches techniques de chacun des outils décrits en annexe du document.....	8
Figure 5 : Exemple de séquence génétique (gène de la bactérie <i>Dehalococcoides</i> sp. CBDB1 codant pour une déhalogénase responsable de la déchloration réductrice anaérobie des chloroéthènes).....	9
Figure 6 : Nombre de publications référencées dans le domaine de la bioremédiation intégrant les outils de biologie moléculaire. Les recherches ont été effectuées sur la base de données Web Of Science à partir des mots clés Bioremédiation ; Bioremédiation + PCR ; Bioremédiation + Outils de biologie moléculaire ; Bioremédiation + Métagénomique (webofknowledge.com).....	12
Figure 7 : Intégration des données de biologie moléculaire aux données classiques de caractérisation des sites et sols pollués.....	15
Figure 8 : Principe de la PCR. La technique est basée sur la répétition de cycles de transition de température qui alternent après une phase de dénaturation initiale des molécules d'ADN (1'). Le mélange réactionnel contient l'ADN ou l'ADNc, la polymérase, les nucléotides, et les amorces spécifiques permettant d'amplifier la région d'intérêt. Le nombre de cycles (n), constitués d'étapes de dénaturation (1), hybridation (2), élongation (3) est généralement supérieur à 30 et conduit à l'obtention théorique de 2 ⁿ molécules.....	21
Figure 9 : Principe de la dPCR (digital PCR ou PCR numérique). Dans la PCR numérique, la réaction PCR peut prendre deux valeurs, 0 (absence) ou 1 (présence). Le principe est basé sur la multiplication du nombre de	

réactions ou l'échantillon est partitionné dans des milliers de compartiments distincts, chaque compartiment étant ensuite considéré comme un réacteur indépendant. L'application de la loi de Poisson permet de quantifier la quantité de séquence cible dans l'échantillon de départ. La précision de la mesure est fonction du nombre de compartiments. Cette technique présente l'avantage de permettre une quantification absolue, sans nécessité d'établir une courbe standard et reste moins sensible aux inhibiteurs du fait de la dilution de l'échantillon. 22

Figure 10 : Principe de la qPCR. Contrairement à la PCR en point final, qui se termine par un dépôt sur gel d'agarose permettant de déterminer si la séquence recherchée est présente ou absente, la PCR quantitative permet de quantifier le nombre de séquences présentes dans la matrice initiale en suivant l'évolution de la libération de molécules fluorescentes dont l'apparition dépend de la quantité de séquences cibles dans la matrice initiale. La RT-qPCR repose sur le même principe mais utilise l'ARN, converti en ADN complémentaire, comme matrice de départ. Elle permet de quantifier l'expression des séquences recherchées. 22

Figure 11 : Description de l'approche permettant d'obtenir une empreinte génétique d'une communauté microbienne. Les profils génétiques générés, sortes de codes-barres des communautés, peuvent être comparés afin de déterminer l'évolution de la structure de la communauté ou de l'abondance relative de certains groupes. Par défaut ces approches ne permettent pas d'identifier les groupes bactériens mais peuvent être complétés par des techniques de clonages, séquençages ou hybridation afin d'identifier les membres dominants de la communauté. 24

Figure 12 : Principe de l'obtention d'empreintes génétiques par la technique PLFA. 25

Figure 13 : Principe des puces à ADN. Les molécules cibles sont marquées avec un fluorophore (molécule fluorescente permettant de détecter les fragments nucléiques) puis hybridées sur un support sur lequel sont placées des sondes nucléotidiques (courtes séquences spécifiques des gènes ciblés) recherchées. 27

Figure 14 : Principe du séquençage massif. 29

Figure 15 : Principe du traçage isotopique par SIP (Stable Isotope Probing). 31

Figure 16 : Principe de l'analyse des rapports isotopiques des composés spécifiques (CSIA). 32

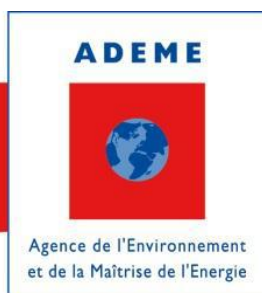
Figure 17 : Exemples d'utilisation de la microscopie pour la caractérisation des communautés microbiennes dans le domaine des sites et sols pollués. (A) Identification de l'origine biologique d'un échantillon et présence de bactéries filamenteuses ; (B) Visualisation des cellules actives (vertes) et non actives (rouges) ; (C) et (D) Visualisation de bactéries actives dans un échantillon pollué aux hydrocarbures ; (E) et (F) exemples de FISH permettant de visualiser un groupe ou plusieurs groupes de bactéries d'intérêt (vert, bleu, rouge) dans un échantillon complexe. 33

L'ADEME EN BREF

L'Agence de l'Environnement et de la Maîtrise de l'Energie (ADEME) participe à la mise en œuvre des politiques publiques dans les domaines de l'environnement, de l'énergie et du développement durable. Elle met ses capacités d'expertise et de conseil à disposition des entreprises, des collectivités locales, des pouvoirs publics et du grand public, afin de leur permettre de progresser dans leur démarche environnementale. L'Agence aide en outre au financement de projets, de la recherche à la mise en œuvre et ce, dans les domaines suivants : la gestion des déchets, la préservation des sols, l'efficacité énergétique et les énergies renouvelables, la qualité de l'air et la lutte contre le bruit.

L'ADEME est un établissement public sous la tutelle conjointe du ministère de l'Ecologie, du Développement durable et de l'Energie, et du ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche.

www.ademe.fr



ADEME
20, avenue du Grésillé
BP 90406 | 49004 Angers Cedex 01

www.ademe.fr