

---

# MicroSol database<sup>©</sup>, le Premier Système d'Information Environnemental sur la Microbiologie des Sols

F.E.R. Morin<sup>(1)</sup>, S. Dequiedt<sup>(1)</sup>, V. Koyao-Darinest<sup>(2)</sup>, B. Toutain<sup>(2)</sup>, S. Terrat<sup>(1)</sup>, M. Lelièvre<sup>(1)</sup>, V. Nowak<sup>(1,3)</sup>, C. Faivre-Primot<sup>(1)</sup>, P. Lemanceau<sup>(3)</sup>, P-A. Maron<sup>(1,3)</sup> et L. Ranjard<sup>(1,3\*)</sup>

1) INRA-Université Bourgogne, AgroSup, Plateforme GenoSol, UMR 1347 Agroécologie, F-21000 Dijon, France

2) INRA, US 1106 InfoSol, F-45075 Orléans, France

3) INRA-Université Bourgogne, AgroSup, UMR 1347 Agroécologie-pôle EcolDur, F-21000 Dijon, France

\* : Auteur correspondant: [lionel.ranjard@dijon.inra.fr](mailto:lionel.ranjard@dijon.inra.fr) - INRA-Université Bourgogne, AgroSup, UMR 1347, Agroécologie - plateforme GenoSol, 17 rue Sully, BP 86510, F-21000 Dijon, France

## RÉSUMÉ

Un effort scientifique considérable est déployé à l'heure actuelle pour parvenir à améliorer la description et la compréhension du fonctionnement des agro-écosystèmes. Cet effort est motivé par une question centrale : l'action de l'Homme et les modifications climatiques risquent-elles d'engendrer des modifications du fonctionnement de ces systèmes de nature à compromettre leur rôle vis-à-vis des sociétés humaines ? Dans ce contexte, la mise au point de descripteurs permettant de rendre compte de l'état biologique et du fonctionnement des sols constitue un enjeu majeur.

Nombre des services assurés par les sols reposent sur l'activité, l'abondance et la diversité des communautés microbiennes telluriques, ce qui positionne ces paramètres microbiens comme des indicateurs de choix pour rendre compte de l'état biologique des sols. La base de données MicroSol database mise en place au sein de la plateforme GenoSol (INRA Dijon, [http://www.dijon.inra.fr/plateforme\\_genosol](http://www.dijon.inra.fr/plateforme_genosol)) constitue un outil opérationnel pour développer ces indicateurs microbiens ainsi que les référentiels indispensables à leur interprétation. En effet, cette base de données unique en son genre renferme les résultats des analyses moléculaires de l'abondance et de la diversité des communautés microbiennes acquis dans un cadre méthodologique standardisé à partir d'échantillons couvrant la totalité du territoire national.

Cet article retrace le contexte de la création de MicroSol database ainsi que les aspects techniques relatifs à l'architecture, à la typologie des données et aux fonctionnalités opérationnelles de cette base de données.

## Mots clés

Sol, microbiologie, biodiversité, référentiel, base de données.

**SUMMARY****MICROSOL DATABASE ©, THE FIRST ENVIRONMENTAL INFORMATION SYSTEM IN SOIL MICROBIOLOGY**

Considerable scientific effort is currently provided to improve description and understanding of the functioning of agro-ecosystems. This effort is motivated by a central question: the action of Human and climate changes are they likely to cause changes in the functioning of these systems that may affect their role towards human societies? In this context, development of descriptors for reporting the biological status and functioning of soil is a major issue.

Most of the services provided by soils are based on activity, abundance and diversity of soil microbial communities, which positions these microbial parameters as promising indicators to account for the biological status of soils. The database MicroSol database developed within the platform GenoSol (INRA Dijon, [http://www.dijon.inra.fr/plateforme\\_genosol](http://www.dijon.inra.fr/plateforme_genosol)) is an operational tool to develop such microbial indicators, but also the associated standards essential for their interpretation. Indeed, this database contains results of molecular analyzes of the abundance and diversity of microbial communities acquired in a standardized methodological framework from samples covering the entire France territory.

This article outlines the context of the creation of MicroSol database. It also addresses the technical aspects of architecture, type of data and functionality of the database.

**Key-words**

Soil, microbiology, biodiversity, normal operating range, database.

**RESUMEN****MICROSOL DATABASE©, EL PRIMERO SISTEMA DE INFORMACIÓN AMBIENTAL SOBRE LA MICROBIOLOGÍA DE LOS SUELOS.**

Se desarrolla actualmente un esfuerzo científico considerable para lograr al mejoramiento de la descripción y de la comprensión del funcionamiento de los agro-ecosistemas. Este esfuerzo está motivado por una cuestión central: ¿existe riesgos que la acción central del Hombre y las modificaciones climáticas engendran modificaciones del funcionamiento de estos sistemas de manera a comprometer sus papeles frente de las sociedades humanas? En este contexto, el ajuste de descriptores que permiten dar cuenta del estado biológico y del funcionamiento de los suelos constituye un desafío mayor.

Numero de servicios proporcionados por los suelos se apoyan sobre la actividad, la abundancia y la diversidad de las comunidades microbianas telúricas, lo que posiciona estos parámetros microbianos como indicadores de elección para dar cuenta del estado biológico de los suelos. La base de datos MicroSol database instalada al seno de la plataforma GenoSol (INRA Dijon,) constituye una herramienta operacional para desarrollar estos indicadores microbianos así que los referenciales indispensables a su interpretación. En efecto, esta base de datos única a su tipo contiene los resultados de los análisis moleculares de la abundancia y de la diversidad de las comunidades microbianas adquiridos en un cuadro metodológico estandarizado a partir de muestras que cubren la totalidad del territorio nacional.

En este artículo se describe el contexto de la creación de MicroSol database así que los aspectos técnicos relativos a la arquitectura, a la tipología de los datos y a las funcionalidades operacionales de esta base de datos.

**Palabras clave**

Suelo, microbiología, referencial, base de datos.

D'un point de vue écologique, le sol représente l'un des plus importants réservoirs de diversité biologique de notre planète et surtout l'un des derniers remparts pour cette biodiversité. Il est aussi le lieu de nombreux processus biologiques et écologiques et fournit donc un nombre important de services écosystémiques résultant de la complexité des assemblages taxonomiques et fonctionnels des communautés indigènes et des interactions entre les organismes (Torsvik & Øvreås, 2002). Parmi les services écologiques, sociaux et économiques identifiés, le rôle du sol en tant que réservoir de ressource génétique est maintenant bien identifié au même titre que son rôle dans l'épuration des eaux de surface, le recyclage des éléments minéraux (fertilité du sol), le stockage du carbone (en tant que puits de CO<sub>2</sub> atmosphérique), ce dernier point étant en relation directe avec les changements climatiques et la notion de productivité végétale (Gardi *et al.*, 2009).

Le sol représente avant tout pour l'homme un support de construction et de production agricole. Sa surexploitation depuis le développement de l'agriculture intensive, de l'urbanisation et de l'industrialisation a mené à une érosion significative de sa biodiversité et par là même des fonctions et services remplis par cette biodiversité. En effet, Loreau a défini en 2000 le concept d'« assurance écologique » qui démontre que la capacité d'un sol à remplir de façon pérenne ses fonctions environnementales ou agronomiques est fortement reliée à son niveau de biodiversité. Il en résulte une prise de conscience où la notion de sols valorisés essentiellement en termes de productivité maximum est peu à peu remplacée par celle d'un écosystème « piloté de manière à fournir durablement diverses catégories de biens et de services précisément qualifiés ».

Dans ce contexte, il devient donc urgent de fournir aux utilisateurs et gestionnaires des sols les moyens d'évaluer l'état biologique de leur support de production et l'impact des pratiques associées (agricoles, urbaines, industrielles). Parmi les organismes indigènes des sols, les communautés microbiennes (bactéries et champignons) sont les plus importantes en densité (10<sup>6</sup>-10<sup>9</sup> individus/g de sol), en diversité (10<sup>3</sup> à 10<sup>6</sup> espèces/g de sols) et occupent un rôle central dans le fonctionnement biologique des sols (Curtis & Sloan, 2005). De par leur temps de génération court, leur capacité adaptative importante et leur sensibilité à de nombreuses perturbations environnementales, les communautés microbiennes ont été démontrées comme de très bons indicateurs précoces, sensibles et robustes de l'impact des perturbations d'origines naturelles ou anthropiques que peut subir un sol. Par conséquent, la mise au point de bioindicateurs basés sur l'abondance et la diversité des communautés microbiennes du sol permettra de renseigner sur l'état biologique d'un sol mais aussi sur sa capacité à fournir des services écosystémiques nécessaires au développement et au bien-être des sociétés humaines.

Les communautés microbiennes du sol sont difficiles à caractériser. Ceci s'explique par une accessibilité plus

ou moins importante des populations au sein d'une matrice hétérogène et structurée mais aussi par leurs densité et diversité très importantes qui rendent difficile à caractériser et résoudre l'information obtenue. Toutefois, grâce aux avancées méthodologiques qui ont eu lieu depuis une quinzaine d'années, les études d'écologie microbienne bénéficient maintenant d'une automatisation des outils moléculaires (extraction d'ADN, caractérisation du polymorphisme de l'ADN, séquençage haut débit, métagénomique, métaprotéomique, etc.) permettant la caractérisation des ressources génétiques microbiennes (densité, diversité taxonomique, potentiel génétique et fonctionnel) des sols en moyen débit (Maron *et al.*, 2011). Néanmoins, le développement de ces nouveaux outils moléculaires s'est effectué de façon anarchique et désordonnée à l'échelle de la communauté scientifique internationale et une multitude d'outils divers et variés ont vu le jour sans aucune standardisation et donc possibilité de comparer et généraliser les résultats entre les études. Il résulte aussi de ce constat une absence de « centralisation » des données sans laquelle il est impossible de mettre en place des « référentiels » robustes sur l'abondance et la diversité microbienne des sols servant directement à l'interprétation de leur état biologique.

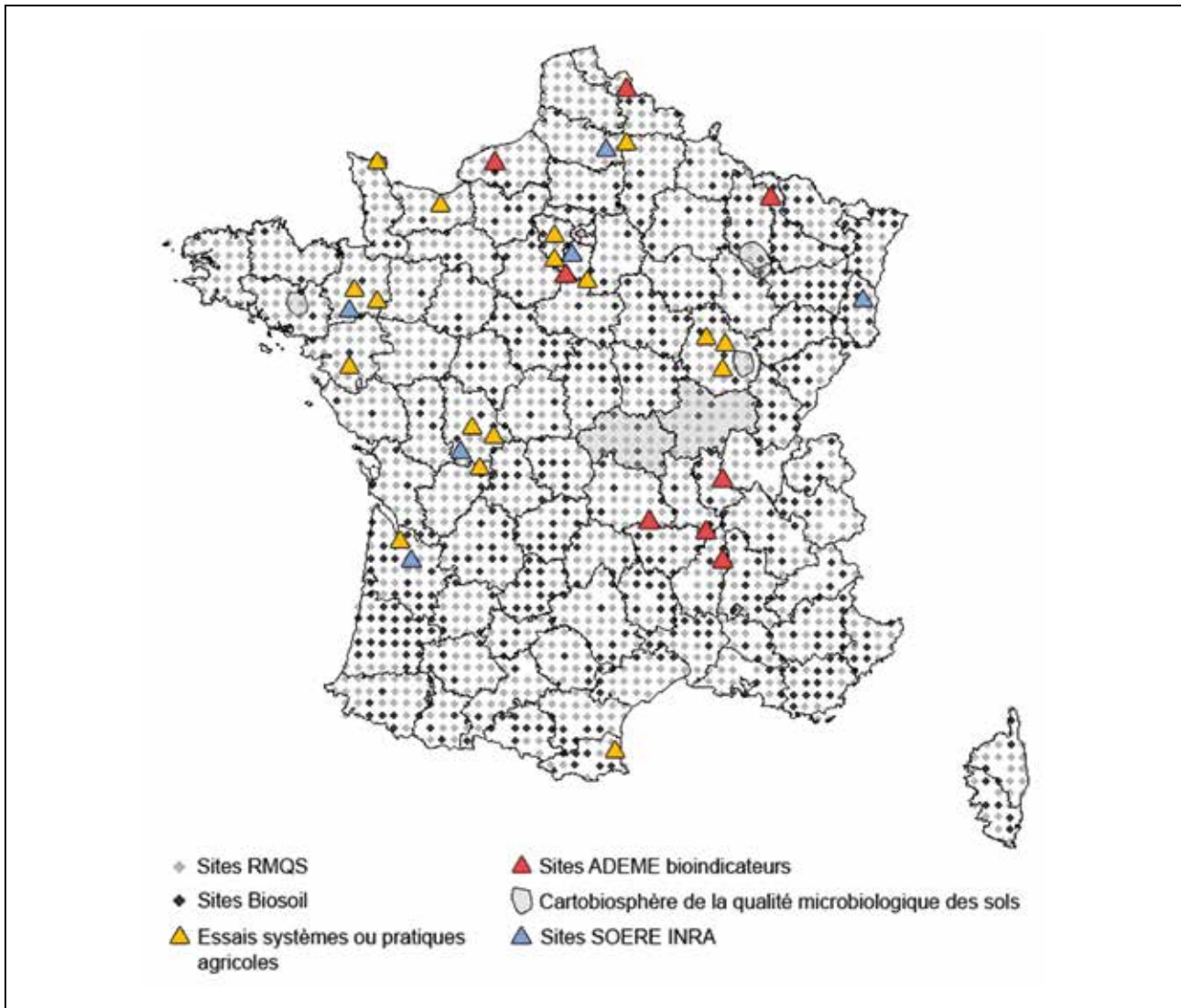
Dans ce contexte, l'UMR Agroécologie (ex Microbiologie du Sol et de l'Environnement) a créé en 2008 la plateforme « GenoSol », dont l'objectif est de fournir une structure logistique et technique assurant l'acquisition, la conservation, la caractérisation et la mise à disposition des ressources génétiques (ADN) des sols issus d'échantillonnages de grande envergure (plusieurs centaines à plusieurs milliers correspondant à de grandes échelles spatiales et/ou temporelles) (Ranjard *et al.*, 2009).

Cette plateforme est le fruit de l'expertise et du savoir-faire accumulés en microbiologie environnementale par la standardisation des outils moléculaires de caractérisation des communautés microbiennes du sol, mais aussi par la mise en place d'une librairie et d'une base de données des ressources génétiques sur les sols servant à l'interprétation de ces analyses à grande échelle (temps, espace). Ces analyses doivent permettre de conduire des travaux de recherche sur la base d'approches génériques et sans *a priori* pour évaluer et hiérarchiser les facteurs environnementaux et agronomiques intervenant dans l'évolution de la qualité biologique des sols à long terme. La finalité est une meilleure connaissance, gestion et valorisation de la biodiversité, notamment pour appréhender la relation biodiversité-fonction-service.

La plateforme GenoSol s'articule autour de trois activités principales :

- un **Centre de Ressources Génétiques** (CRG) national sur les sols (unique en Europe) qui a pour but de stocker et conserver les ressources génétiques (ADN) et de les mettre à disposition de la communauté scientifique. Les échantillons de sols conservés proviennent en majorité de réseaux de surveillance (RMQS,

**Figure 1** - Carte de France des partenaires et sites expérimentaux ou d'observation dont les sols sont en conservatoire GenoSol.  
**Figure 1** - French map of partners and both experimental and observation sites which soils are in the GenoSol conservatory.



Massif Central, etc.), et d'observations (Transect européen des sols), d'observatoires de recherche en environnement (SOERE ACBB, PRO, OPE ANDRA), de sites expérimentaux pérennes (LTO EcoFINDERS, réseau PIC, site de Fénay, réseau ADEME, sites CETIOM et Arvalis, sites CA 21 et 71), et de réseaux d'exploitations agricoles (réseau OFSV). Tous ces sites sont impliqués dans des questionnements en lien avec l'évaluation environnementale des modes d'usage des sols et la plateforme GenoSol constitue une logistique stratégique pour renforcer les systèmes d'observation de la biodiversité tellurique dans ces dispositifs. Fin 2012, plus de 7000 sols sont stockés dans le CRG et 1500 nouveaux sols sont en prévision d'acquisition pour les deux années à venir (figure 1) ;

- une **plateforme technique** permettant le développement et la veille technologique sur les méthodes d'extraction des acides nucléiques des sols et les outils de caractérisation des ressources génétiques microbiennes (génotypage, pyroséquençage, métagénome, métagprotéome, mesure d'activité, etc.). Cette plateforme permet d'améliorer la standardisation des procédures et des outils moléculaires avec des objectifs de normalisation (AFNOR et ISO). Des partenariats avec le Genoscope, les sociétés Genoscreen et Roche ont été développés en 2010 et permettent de développer les techniques de séquençage haut débit pour aborder la diversité microbienne des sols ;

- un **Système d'Information Environnementale** (SIE) centré sur le développement de la base de données « MicroSol database » qui permet d'une part de gérer le CRG et la traçabilité de ses échantillons et, d'autre part, de stocker et d'analyser les données de caractérisation génétique des sols (taxonomique et fonctionnelle). A ce jour, les données environnementales (climat, caractéristiques physico-chimiques des sols, pratiques agricoles, etc.) sont gérées par les partenaires gestionnaires des sites ou réseaux de surveillance.

*In fine*, la base de données permettra d'élaborer un véritable « référentiel d'interprétation et de traduction » de la diversité microbienne des sols qui servira à définir un « Normal Operating Range » de la biodiversité des sols.

Dans cet article, nous allons présenter plus précisément les originalités techniques et scientifiques de cette base de données qui est unique à ce jour dans le domaine de la microbiologie environnementale et qui doit stimuler des initiatives similaires au sein de la communauté scientifique internationale.

## MICROSOL AU SERVICE DE LA MICROBIOLOGIE DU SOL

MicroSol database est un outil informatique (une base de données couplée à une interface web) conçu afin de servir les activités de la plateforme GenoSol d'un point de vue technique et scientifique. Son développement a été initié dans le but de doter la plateforme d'un système performant et robuste lui permettant de gérer et pérenniser le grand nombre de données générées lors de ses programmes de recherche. En ce sens, l'axe principal de MicroSol database a été défini comme étant un système d'information représentant au mieux les processus métiers inhérents à la plateforme GenoSol. Par définition, les processus métiers sont un ensemble coordonné d'activités qui visent à produire un résultat ; ils sont exécutés par des acteurs qui, pour atteindre le résultat escompté, utilisent différents types de ressources.

Dans le contexte de la plateforme GenoSol, ces processus peuvent être décrits comme un enchaînement linéaire d'actions prenant pour origine le sol et dont le but final est la caractérisation des gènes des microorganismes de ce sol. Le scénario global de cette approche processus métier peut être exposé comme suit : la première étape est le prélèvement d'un échantillon de sol qui par la suite est stocké dans le conservatoire de sol GenoSol ; une partie de ce sol est ensuite utilisée pour subir une lyse à partir de laquelle l'ADN des microorganismes est extrait puis purifié (tout comme le sol, cet ADN est stocké dans ce qu'on appelle une DNA-thèque) ; enfin, différentes analyses moléculaires sont effectuées sur ces ADN (PCR, analyse arisa, PCR quantitative, etc.). Pour les besoins du document, il est important ici d'introduire la notion d'entité.

De manière générale, une étape spécifique du processus sera désormais nommée « entité ».

De l'analyse de ces processus (*figure 2*) a découlé la conception du modèle relationnel de MicroSol database et aussi la définition des besoins fonctionnels de l'application.

A chaque étape du processus, des données relatives à une entité sont produites (ex : masse de sol, quantité d'ADN, etc.), données brutes qui sont l'essence même du travail de GenoSol et donc stockées dans MicroSol database. Aussi, dans un souci de qualité et de traçabilité maximales, il est important de donner plus d'information à la donnée, c'est-à-dire d'enrichir cette donnée brute par des métadonnées qui permettent de répondre notamment aux questions suivantes : où, quand, qui et comment (ex : date, opérateur, mode opératoire, etc.). De ce fait, chaque entité est composée de données résultats auxquelles sont associées des métadonnées. Concrètement, un couple donnée/métadonnées est représenté dans la base par ce qu'on appelle communément un identifiant unique (entier n'apparaissant qu'une fois et une seule pour une entité).

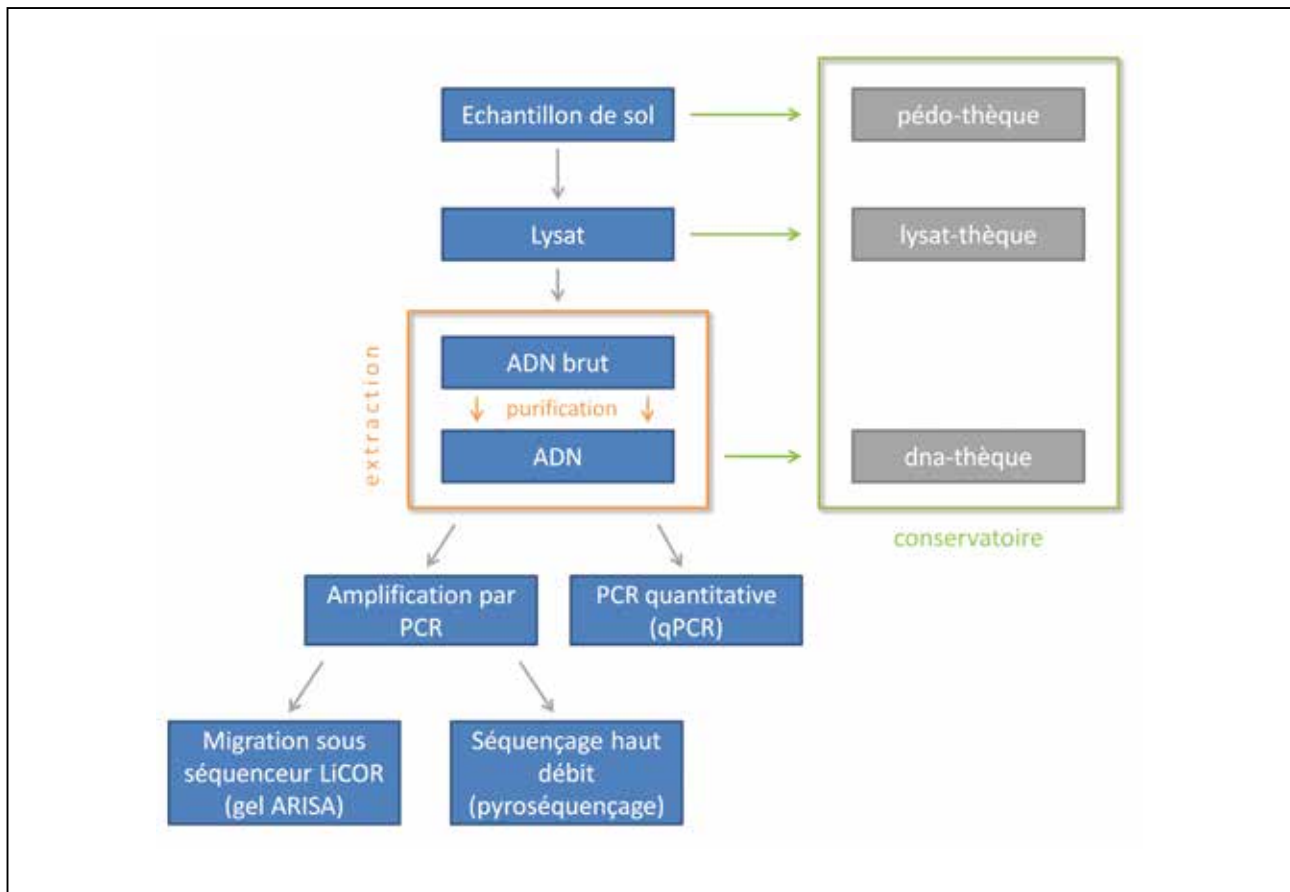
Les informations contenues dans MicroSol database peuvent être de types suivants : entier, réel, date, fichier et texte (alphanumérique). De plus, toutes les données relatives aux entités sont décrites formellement dans l'application web grâce à un dictionnaire de données qui permet à la fois de définir le nom de l'information apportée, son type (ex : entier), son format (ex : jj-mm-aaaa), son unité (ex : gramme) et son descriptif explicatif.

## DESCRIPTIF DES FONCTIONNALITÉS DE MICROSOL

Comme il a été décrit précédemment, MicroSol database se compose d'une base de données et d'une interface web. Ce choix de structure confère à l'application deux niveaux d'utilisation. Tout d'abord la base de données seule, en tant qu'entité de stockage pur, qui peut être utilisée pour la manipulation de données *via* des requêtes SQL (statistiques, requêtes spécifiques, confrontation de données...). L'autre niveau d'utilisation se fait à travers l'interface utilisateur pour l'acquisition, la consultation, la modification, etc. C'est cette seconde partie qui va être traitée ici : les fonctionnalités de l'interface web.

### Sécurité et confidentialité

L'url d'accès à MicroSol database est « <http://www.genosolsi.inra.fr> ». L'accès à cette application n'est pas directement ouvert à tous. En effet, il est nécessaire de posséder un compte GenoSol pour avoir un login et un mot de passe valide afin d'accéder à l'application. Dans le cas où la personne n'a pas de compte, elle peut faire une demande de création

**Figure 2** - Processus métier de la plateforme GenoSol : du sol au gène.**Figure 2** - Business processes of the GenoSol platform: from soil to gene.

après de GenoSol en utilisant le lien « demande d'ouverture de compte » présent sur la page d'accueil préalablement citée.

Un compte GenoSol est obligatoirement associé à un groupe utilisateur (administrateur, utilisateur, visiteur, etc.). Ces groupes servent à rassembler des utilisateurs afin de leur attribuer des privilèges communs. Par privilèges on entend le droit ou non d'effectuer certaines actions dans l'application. Par exemple une personne se trouvant dans le groupe administrateur peut accéder à toutes les fonctionnalités de MicroSol database alors qu'un visiteur sera plus restreint, dans le sens où il ne pourra pas faire d'ajouts ou de modifications de données. La gestion de ces privilèges est implémentée par l'usage d'ACL (Access Control List).

L'autre point relatif à la sécurité concerne la confidentialité des données à proprement parler. Celles-ci ne sont pas en libre accès dans l'application. Par conséquent, un système contrôlé par les administrateurs a été mis en place afin de définir « qui peut voir quoi ». Ce système s'appuie sur le fonctionnement dit en « mode projet » de la plateforme GenoSol et répond à la démarche suivante : d'un côté une donnée est associée à un ou

plusieurs projets et de l'autre un utilisateur appartient à un ou plusieurs projets. Ainsi, un utilisateur n'a accès qu'aux données inscrites dans un projet auquel il est affilié. Au final, ce contrôle, effectué à l'aide d'un filtre basé sur l'implication ou non d'un utilisateur dans un projet, assure la confidentialité des données.

## Ajout et modification des données

L'acquisition de données, pour une entité, s'effectue de deux manières distinctes, soit de façon unitaire, c'est-à-dire par l'intermédiaire d'un formulaire prévu à cet effet, soit massivement grâce à un fichier CSV. Les trames des fichiers à fournir, contenant les en-têtes des colonnes à renseigner, sont standardisées et prédéfinies pour chaque entité dans l'application. L'utilisateur doit donc respecter les contraintes imposées par ces trames pour pouvoir importer ses données dans MicroSol database. De la même manière, la modification de données peut se faire *via* un formulaire ou par un fichier CSV. Toutefois, dans un souci de respect de l'intégrité des données, la modification en masse n'est réservée qu'aux administrateurs.

**Figure 3** - Rapport d'état d'un historique d'un échantillon de sol.**Figure 3** - Status report of a soil sample historic.

Stock Sol - Historique de modification pour l'identifiant : 2389  
Retour liste

| Version actuelle (Nombre de versions : 2)                         |                 |                 | Version antérieure  |                 |
|---|-----------------|-----------------|---------------------|-----------------|
| Version du 19-04-2013 à 16:09 par Fabien Morin (id modif = 27334) |                 |                 | Version originale   |                 |
|   |                 | ancienne valeur |                     |                 |
| Id Stock Sol  | 2389            |                 | Id Stock Sol        | 2389            |
| Code échantillon  | Res2011_P8_7_24 |                 | Code échantillon    | Res2011_P8_7_24 |
| Id Sol  | 2389            |                 | Id Sol              | 2389            |
| Masse initiale  | 51.35           | 89.45           | Masse initiale      | 89.45           |
| Masse actuelle  | 49.65           | 87.75           | Masse actuelle      | 87.75           |
| Humidité sol  | 0               | 14              | Humidité sol        | 14              |
| Date réception  | 29-03-2011      |                 | Date réception      | 29-03-2011      |
| Opérateur réception   | Dequiedt Samuel |                 | Opérateur réception | Dequiedt Samuel |
| Etat réception  | tamisé          |                 | Etat réception      | tamisé          |
| Tamissage réception   | 4               |                 | Tamissage réception | 4               |
| Date conservatoire  | 29-03-2011      |                 | Date conservatoire  | 29-03-2011      |
| Etat stockage   | lyophilisé      | frais           | Etat stockage       | frais           |
| Tamissage stockage  | 4               |                 | Tamissage stockage  | 4               |
| Opérateur stockage  | 25              |                 | Opérateur stockage  | 26              |
| Commentaire   |                 |                 | Commentaire         |                 |
| Procédure   | G.P-007         |                 | Procédure           | G.P-007         |
| Loc pièce   | 703             |                 | Loc pièce           | 703             |
| Loc congel  | G-CON-008       |                 | Loc congel          | G-CON-008       |
| Loc tiroir  | 9               |                 | Loc tiroir          | 9               |

## Extraction des données

Dans MicroSol database, les données sont organisées par projets et par sites d'étude. Leur extraction (ou visualisation) à partir de l'interface s'effectue selon l'un de ces deux critères. A l'affichage, elles sont représentées sous la forme d'un tableau que l'utilisateur peut trier à sa convenance en fonction d'une colonne. Par la suite, il a la possibilité d'exporter ces données vers un fichier CSV. Il existe une seconde manière d'accéder aux données : la méthode de recherche. Celle-ci se présente sous deux formes : i) de façon unitaire, par un champ de recherche qui permet de renvoyer toutes les données concernées par la requête soumise et ii) la recherche dite en masse par fichier CSV par laquelle l'utilisateur fournit une liste d'échantillons à MicroSol et récupère en retour un fichier CSV contenant les données associées à ces échantillons.

## Historique

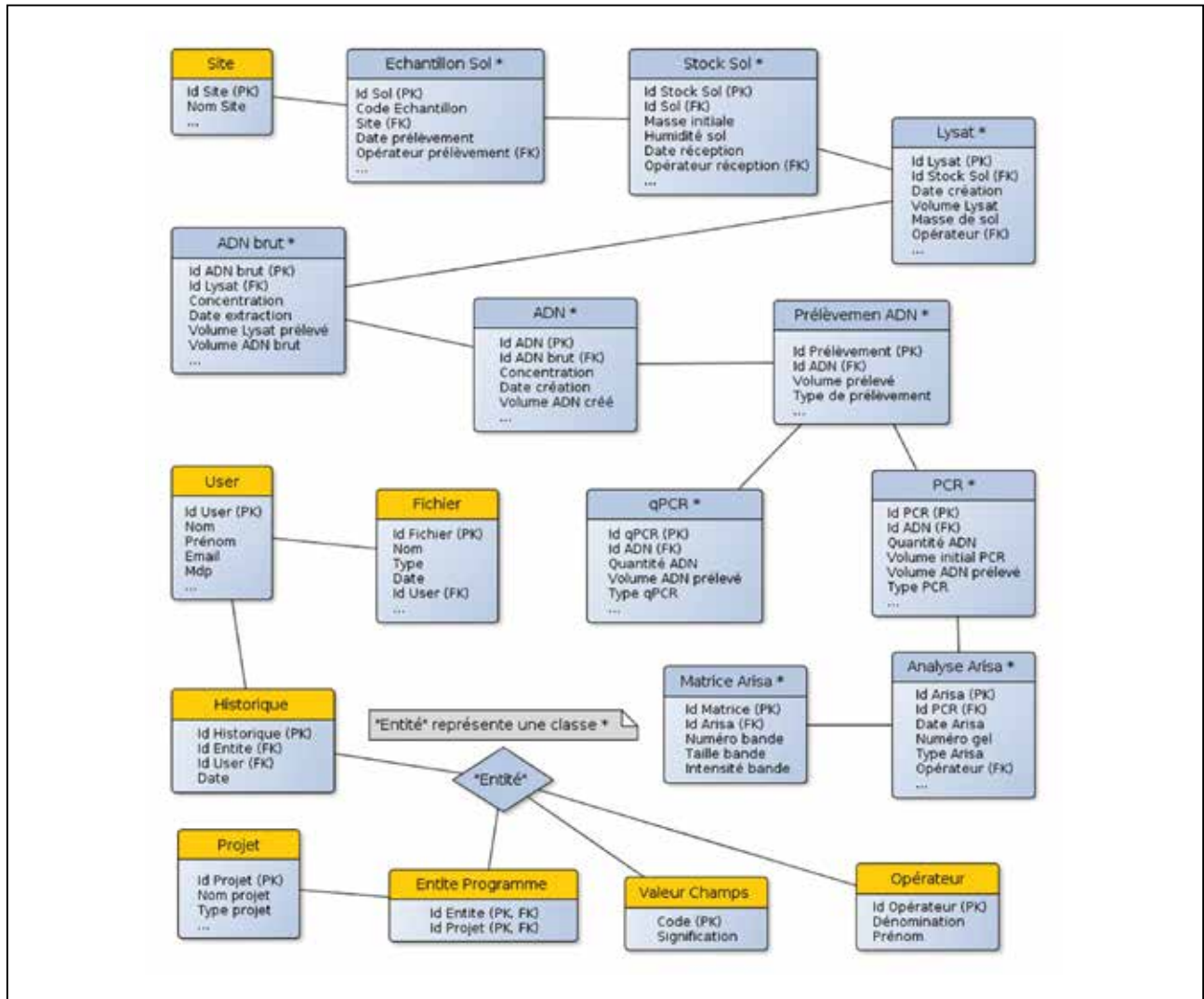
Dans un souci de traçabilité et d'intégrité des données, chaque action d'ajout ou de modification effectuée par un

utilisateur au sein de MicroSol est tracée dans la base. De même, tout fichier importé ou exporté à partir de l'interface y est sauvegardé. Un module « historique » est présent dans l'application et permet aux administrateurs de visualiser l'évolution de ces actions. Grâce à lui, il est possible de savoir précisément pour un ajout la personne qui a inscrit la donnée dans la base et à quel moment, et pour une modification de dire qui a révisé la donnée et quand. A cela s'ajoute un autre outil concernant l'historique de modification : le versioning. Celui-ci consiste à garder l'ensemble des versions d'une donnée et de ses métadonnées. En pratique, dès qu'une donnée subit un changement, sa version pré-modification est automatiquement stockée en base et y est consultable par la suite pour les administrateurs (*figure 3*).

## Traçabilité

Les chaînages entre les étapes du processus métier illustré dans la *figure 2* peuvent être définis comme des relations père-fils entre les entités. Ainsi il est possible d'établir le cycle de vie d'un échantillon sous la forme d'un arbre, composé de



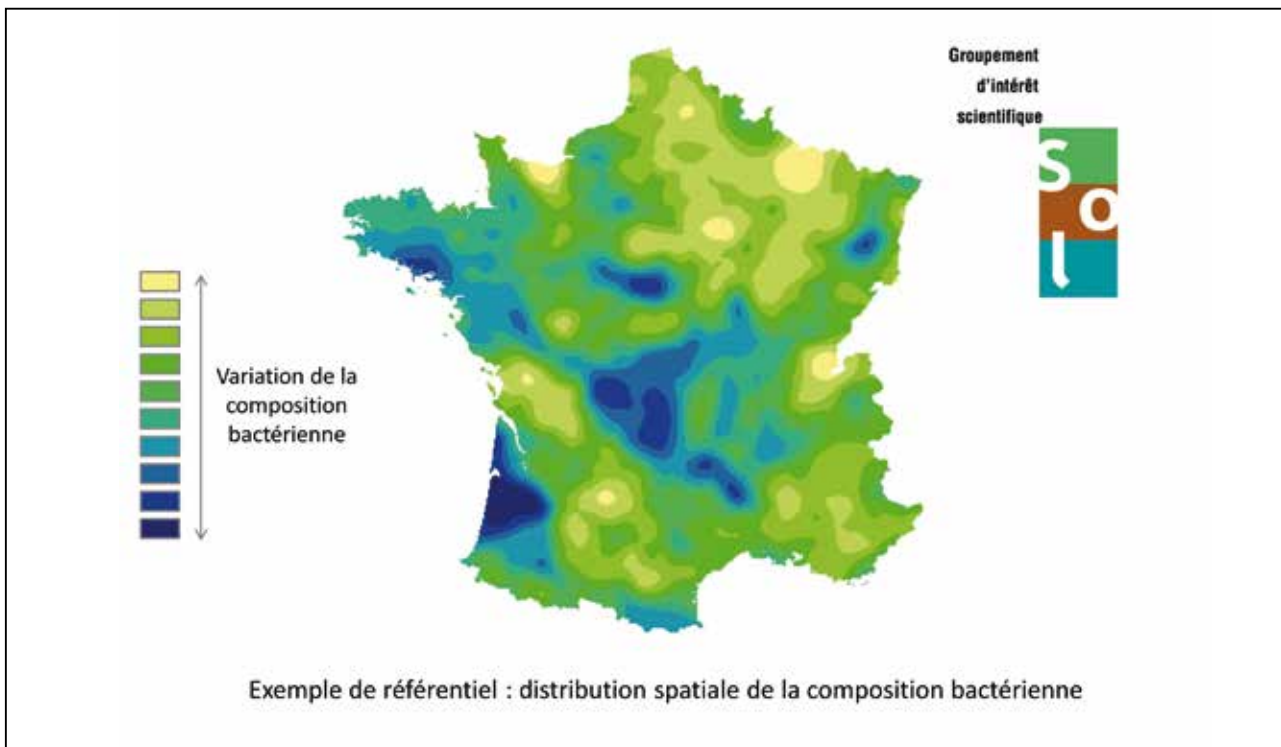
**Figure 4** - Architecture simplifiée de MicroSol database.**Figure 4** - Simplified architecture of MicroSol database.

nœuds père et de nœuds fils. Du fait que pour certaines entités des stocks soient gérés, cet arbre de filiation suit le principe classique de la structure d'un arbre : un nœud père peut avoir plusieurs nœuds fils, un fils n'a qu'un seul père et tous les nœuds ont une racine commune (l'échantillon de sol). Concrètement, dans MicroSol database, une fonctionnalité permet d'afficher la filiation complète de n'importe quelle entité. Quel que soit le niveau où l'on se situe dans le processus, l'application est capable de retrouver la filiation ascendante et descendante d'une donnée. Cette filiation est très importante pour la traçabilité car elle permet de remonter précisément les étapes si jamais un problème a été détecté durant le déroulement du processus.

## Organisationnel

Le système d'information permet d'avoir un contrôle permanent sur les stocks (quantité et localisation) de sols, de lysats et d'ADN gérés par la plateforme GenoSol. Ainsi les utilisateurs sont en mesure d'évaluer précisément et à tout moment les quantités de stocks disponibles et à quel endroit ils se situent dans le conservatoire (congélateur, tiroir, boîte, etc.). De plus, l'application comporte une fonctionnalité qui leur permet de savoir quelles entités n'ont pas encore de fils dans le processus, c'est ce qu'on appelle des « listes à faire ». Ce module permet aux utilisateurs d'établir leur planning d'analyse au laboratoire en leur fournissant un fichier CSV pré-rempli.



**Figure 5** - Exemple de référentiel sur la biomasse moléculaire microbienne.**Figure 5** - Example of standards on microbial molecular biomass.

## MICROSOL CÔTÉ TECHNIQUE

Pour mener à bien le développement du projet MicroSol database, des méthodes dites « méthodes agiles » ont été appliquées. Celles-ci sont beaucoup plus souples que les méthodes traditionnellement employées. Elles consistent à impliquer au maximum les demandeurs pour une plus grande réactivité lors de l'avancement du projet. Cette collaboration régulière minimise les risques d'incompréhension entre les scientifiques et les développeurs. Au final, cette méthode impacte de manière positive le développement de l'application à concevoir, grâce notamment à l'interaction constante entre les acteurs du projet.

Le choix d'une application web s'est imposé dès la genèse du projet MicroSol database. En effet, les utilisateurs souhaitaient disposer d'un outil portable, pouvant être utilisé indépendamment du système d'exploitation et accessible de n'importe où sans nécessité d'installation particulière. C'est pourquoi une architecture de type client-serveur avec client léger a été adoptée. Côté client on trouve l'interface web, utilisable à partir de n'importe quel navigateur internet du marché, et côté serveur la base de données.

Le système de gestion de base de données (SGBD) retenu a été PostgreSQL. Ce choix a été motivé par plusieurs raisons.

Tout d'abord sa grande utilisation dans les projets lourds de l'INRA depuis déjà quelques années, notamment la base de données DONESOL de l'unité Infosol à Orléans, mais aussi pour sa robustesse, ses performances et sa stabilité, éprouvées et largement reconnues par la communauté informatique. Enfin c'est un outil libre, disponible selon les termes d'une licence BSD (Berkeley software distribution licence). Le schéma relationnel de la base de données MicroSol est décrit par la *figure 4*.

L'interface web a quant à elle été développée à l'aide du langage de script libre PHP. Afin d'assurer au maximum la réutilisabilité et la maintenance du code applicatif, il a été décidé d'utiliser le framework PHP Zend Framework pour développer MicroSol database côté client. Sa conception orientée objet procure des outils puissants pour le développement d'applications PHP. Il permet aussi d'utiliser nativement le principe MVC (Modèle Vue Contrôleur), paradigme clarifiant la structure de l'application et par conséquent facilitant grandement son évolution. De plus, dans le but d'améliorer l'expérience utilisateur, de nombreuses fonctionnalités ont été implémentées par le biais de la bibliothèque JavaScript jQuery, ce qui permet d'ajouter de l'interactivité au sein de l'application.

Côté hébergement, la plateforme GenoSol a fait l'acquisition en janvier 2012 d'un serveur dédié et de baies de stockage. Le serveur d'application est un Dell PowerEdge R910 tournant

sous le système d'exploitation CentOS en version 6.2. Les baies sont scindées en deux parties : l'une servant de stockage à proprement parler (d'une capacité de 5,4 To) et l'autre étant un espace de backup afin d'y effectuer les sauvegardes hebdomadaires de MicroSol database. Ce matériel est situé dans la salle serveur du centre INRA de Dijon et a été installé par l'EIC (Equipe Informatique de Centre) qui en assure également la maintenance physique. Quant à l'administration système, elle est effectuée par l'informaticien de la plateforme GenoSol.

## SYSTÈME D'INFORMATION SUR LA MICROBIOLOGIE ENVIRONNEMENTALE

De part les missions de la plateforme GenoSol en terme de standardisation des outils et des méthodes ainsi que de centralisation des échantillons de sol et de mise à disposition des ressources génétiques associées, la base de données MicroSol répond à deux objectifs : i) l'élaboration d'un référentiel d'interprétation et de traduction de la diversité microbienne des sols et ii) la mise à disposition des données associées aux ressources génétiques (résultats de l'analyse moléculaire des communautés microbiennes des sols, etc.). A ce titre, c'est le premier Système d'Information sur la microbiologie environnementale qui confère à la plateforme GenoSol, et aux recherches en écologie microbienne en général, une amélioration de la généricité des conclusions ainsi que des sorties opérationnelles jusque là inaccessibles.

Le référentiel d'interprétation permet *in fine* de définir les gammes de variation (« Normal Operating Range ») de l'abondance et de la diversité microbienne pour chaque type pédo-climatique rencontré sur le territoire national. Cela passe par l'établissement de valeurs basses, moyennes et hautes des composantes de la diversité microbienne (abondance, densité, composition, inventaire, etc.) en regard des caractéristiques environnementales (sol, climat, position géographique, activité anthropique, etc.). Ces gammes de variation sont définies par la comparaison de résultats issus de protocoles standardisés et comparables dont MicroSol assure la traçabilité par le stockage de métadonnées (date de mesure, protocole utilisé, filiation, etc.) associées aux résultats de caractérisation moléculaire des communautés microbiennes. L'extraction de ces jeux de données, couplée à des outils de type SIG, rend possible la composition de cartes de référentiel comme l'illustre la *figure 5*. Le développement de MicroSol dans les standards informatiques des bases de données appliqués à l'INRA (SGBD PostgreSQL) permet de garantir l'interopérabilité et la relation avec les bases de données environnementales existantes (DONESOL, bases de données des SOERes, etc.). A terme, MicroSol possédera un module adapté pour le stockage des données environnementales des partenaires de GenoSol ne possédant pas de solutions adaptées.

Les missions de centralisation et de mise à disposition des ressources génétiques et données associées de la plateforme GenoSol sont garanties par les modalités d'accès définies dans MicroSol. L'accès aux données de MicroSol peut se faire depuis l'interface graphique dédiée ou *via* un accès direct à des vues développées par projet. La sécurité et la confidentialité des données sont assurées par la rédaction d'une convention établie pour chaque projet entre GenoSol et les partenaires. L'accès est sécurisé et restreint aux seules données du projet par identifiant et mot de passe personnel.

## CONCLUSION ET PERSPECTIVES

Le développement et le déploiement de MicroSol database par l'utilisation de méthodes agiles et d'interactions quotidiennes avec les scientifiques de la plateforme permet donc à l'outil de remplir les fonctions essentielles qui avaient été définies en amont du projet ; en particulier le stockage des données, la traçabilité, la sécurité et la gestion organisationnelle (stock, localisation, etc.). D'un point de vue scientifique et opérationnel, l'application constitue aujourd'hui le premier Système d'Information Environnemental, reconnu institutionnellement par une protection intellectuelle délivrée par l'INRA (Inter Deposit Digital Number IDDN.FR.001.370007.000.R.A.2011.000.10300).

Les données contenues dans MicroSol database correspondent à l'aboutissement des missions et développements de la plateforme GenoSol en termes de centralisation des échantillons de sol et de standardisation des méthodes d'analyse moléculaire des communautés microbiennes du sol. MicroSol permet la valorisation de tout ce savoir-faire en augmentant la généricité et la robustesse des conclusions en microbiologie environnementale et notamment l'établissement de référentiels d'interprétation de la diversité microbienne des sols. De plus, en devenant un élément central de la démarche qualité de la plateforme GenoSol, l'application participe grandement à la reconnaissance de celle-ci auprès des partenaires institutionnels, techniques et industriels.

Le positionnement de la plateforme et notamment de son conservatoire et de son référentiel sera renforcé par l'implication récente de la plateforme GenoSol dans le projet ANAEE-Service qui doit mettre en réseau les sites d'observations et expérimentaux des écosystèmes terrestres et aquatiques du CNRS, de l'INRA, de l'IRD et de certaines universités avec des plateformes analytiques en biologie et des plateformes de modélisation.

La réalisation du projet MicroSol database permet à la plateforme GenoSol d'être actuellement reconnue pour son expertise de l'informatique au service de la biologie. Elle participe notamment à la mise en réseau des plateformes analytiques de l'INRA avec la mise en place d'un LIMS

mutualisé mais aussi à d'autres initiatives nationales d'outils de centralisation de données afin d'en accroître la visibilité.

## REMERCIEMENTS

La plateforme GenoSol souhaite tout particulièrement remercier l'unité Infosol d'Orléans pour ses conseils lors de la mise en chantier du projet MicroSol database, spécialement en ce qui concerne les choix technologiques prodigués comme PostgreSQL et comme l'utilisation d'un framework de type MVC.

## BIBLIOGRAPHIE

- Curtis T.P., Sloan W.T., 2005 - Exploring microbial diversity - A vast below, *Science*, 309,1331-1333.
- Gardi C., Montanarella L., Arrouays D., Bispo A., Lemanceau P., Mulder C., Ranjard L., Rombke L., Rutger M., Menta C., 2009 - Soil Biodiversity monitoring in Europe: ongoing activities and challenges. *European Journal of Soil Science*, 60, 807-819.
- Loreau M., 2000 - Biodiversity and ecosystem functioning: recent theoretical advances. *Oikos*, 91: 3-17.
- Maron P.A., Mougél C., Ranjard L., 2011 - Soil microbial diversity: spatial overview, driving factors and functional interest. *CRAS Biology II*, 334:403-411.
- Ranjard L., Dequiedt S., Lelièvre M., Maron P.A., Mougél C., Morin F., Lemanceau P., 2009 - Platform GenoSol: a new tool for conserving and exploring soil microbial diversity. *Environmental Microbiology Reports*, 1: 97-99.
- Torsvik V., Øvreås L., 2002 - Microbial diversity and function in soil: from genes to ecosystems. *Curr Opin Microbiol*,5:240-245.

