Université du Québec
Institut National de la Recherche Scientifique
Institut Armand-Frappier

# L'HYDROGENE MOLECULAIRE EN TANT QUE MODULATEUR DES COMMUNAUTES MICROBIENNES DU SOL

Par

Sarah Piché-Choquette

Thèse présentée pour l'obtention du grade de
Philosophiae doctor (Ph.D.)
en biologie

**Jury d'évaluation**

| | |
|---|---|
| Présidente du jury et Examinatrice interne | Pre Annie Castonguay<br>INRS-Institut Armand-Frappier |
| Examinateur externe | Pr Marc St-Arnaud<br>Institut de recherche en biologie végétale<br>Université de Montréal |
| Examinateur externe | Pr Jean-Philippe Bellenger<br>Université de Sherbrooke |
| Directeur de recherche | Pr Philippe Constant<br>INRS-Institut Armand-Frappier |


Dépôt final : Lundi 14 Janvier 2019

# REMERCIEMENTS

Je remercie d'abord mon directeur de thèse, le professeur Philippe Constant, qui m'a permis d'œuvrer au sein de son laboratoire et d'évoluer d'un point de vue scientifique pendant ma maîtrise et mon doctorat. Merci à tous les membres du laboratoire pour votre collaboration et votre encouragement. Merci Anne, Livie et Julien pour les longues discussions scientifiques, philosophiques et personnelles. Merci Quentin et David pour nos conversations sans queue ni tête. Je tiens aussi à remercier les membres des laboratoires de Richard et Claude et ceux des comités au sein desquels j'ai œuvré.

Je remercie le professeur Josh D. Neufeld et les docteurs Martin Hartmann et Roland Wilhelm pour leurs discussions enrichissantes, l'encouragement et le support moral qu'ils m'ont apportés. Josh, I'll always remember our first chat at IUMS 2014 as a volunteer, or even your audio "pep talk" regarding my involvement in the Armand Frappier Conference as well as your personal welcome in your lab in Waterloo. Martin, in spite of my comprehensive exam, I quickly realized how much of a rare gem you are, both as a person and as a scientist. Our discussions helped me shape my career path, be more confident in my abilities and question myself often to avoid scientific prejudice. I enjoyed your welcome in Zurich, I felt pretty much at home despite the distance. Roli, I'll remember our never-ending discussions under the moonlight while gazing at the silent thunder off in the distance… I never expected to meet a scientist who could understand me so much. I'm glad to know that I am not alone and that people like us can succeed in academia.

Je remercie également ma famille et mes amis, qui m'ont remonté le moral tout au long de ce périple tortueux. Un merci spécial à mes parents, à Sonia, Éric, Esteban, Audrey-Lou, Simon, Stéphanie P, Vincent, Caroline M, Charles, Mathieu, Maxime, Guillaume, Amélie, Philippe I, Camylle, Stéphanie R et Caroline B. Vous êtes irremplaçables.

Je remercie mon conjoint, Jean-Marc, sans qui je n'aurais jamais eu la motivation de terminer cette thèse. Jean-Marc m'a aidé à surmonter les embûches rencontrées pendant mon parcours, dont les présentations, examens, articles, décès dans mon entourage, le syndrome de l'imposteur et les multiples agressions que j'ai subies par des gens de confiance. Je suis heureuse que tu aies toujours été là pour moi. Je t'aime Jean-Marc.

# RÉSUMÉ

L'hydrogène moléculaire ($H_2$) est une source d'énergie ubiquitaire, retrouvée à l'état de trace dans l'atmosphère global, mais en plus fortes concentrations dans de nombreux écosystèmes tels que les terres humides et la rhizosphère des légumineuses fixatrices d'azote. Une grande majorité des bactéries fixatrices d'azote recrutées par les légumineuses utilisées en agriculture sont dépourvues d'hydrogénase, ce qui implique que l'$H_2$ produit par l'enzyme nitrogénase diffuse dans le sol. Compte tenu du fort potentiel énergique de l'$H_2$, celui-ci est rapidement oxydé par les microorganismes oxydant l'$H_2$ (HOM) et ce, dans un rayon de 3 à 4,5 cm de la source. Ces microorganismes ont des capacités physiologiques variées, se retrouvent dans des niches écologiques oxiques et anoxiques et sont distribués parmi plus de 40 embranchements taxonomiques. Bien que des études ont montré que certaines HOM promeuvent la croissance des plantes *in vitro*, l'impact de l'$H_2$ sur les communautés microbiennes du sol est inconnu. Le but de cette thèse est donc d'élucider ce mystère en explorant le potentiel de l'$H_2$ à structurer les processus biogéochimiques du sol. Les hypothèses de recherche sont ainsi que de fortes concentrations d'$H_2$, représentatives d'écosystèmes riches en $H_2$, altéreront directement et indirectement la structure et la fonction des communautés microbiennes du sol et ce, en fonction de la dose d'exposition à l'$H_2$. Un effet direct implique un changement associé à la distribution et l'activité métabolique des HOM, alors que les effets indirects incluent des processus autres que l'oxydation de l'$H_2$. Afin de pouvoir valider ces hypothèses, un système de microcosmes à flux dynamique a été conçu au laboratoire et a permis d'incuber des microcosmes de sols à des rapports stœchiométriques d'$H_2$ précis, s'échelonnant de 0,5 à 10 000 ppmv d'$H_2$. Trois sols prélevés dans une plantation de peupliers, une plantation de mélèzes et une terre agricole ont été utilisés en raison de leurs propriétés biotiques et abiotiques contrastantes. À la suite des incubations, les profils de diversité des communautés microbiennes—bactériennes, fongiques et bactériennes à haute affinité pour l'$H_2$ (HA-HOB)—ainsi que divers processus qui leur sont propres, ont été analysés. D'abord, de nombreux effets directs ont été recensés. Les mesures de potentiel d'oxydation de l'$H_2$ en cours d'incubation ont montré qu'une exposition des HOM à de fortes concentrations d'$H_2$ menait à une activation rapide du métabolisme des HOM à faible affinité, tout en inhibant celui des HA-HOB. Globalement, le $K_m$ et le $V_{max}$ du métabolisme d'oxydation de l'$H_2$ ont augmentés significativement post-incubation. En ce qui a trait à la structure des communautés de HA-HOB, l'abondance relative de centaines de génotypes a été altérée. De manière surprenante, aucun de ces génotypes ne correspondait à ceux actuellement répertoriés dans les bases de données génomiques publiques. Ensuite, plusieurs effets indirects ont aussi été documentés post-incubation. Trois fonctions autres que l'oxydation d'$H_2$ ont été altérées, dont une réduction de l'oxydation du méthane ($CH_4$) et du monoxyde de carbone (CO) et une stimulation de l'utilisation de nombreux substrats carbonés. La réponse globale de ces fonctions fut universelle, bien que l'intensité de l'altération se soit avérée dépendante des conditions biotiques et abiotiques des trois sols étudiés. De plus, l'abondance relative de microorganismes n'oxydant pas l'$H_2$—des Champignons et des assemblages de génomes partiels bactériens exempts de gènes codant pour des hydrogénases—a aussi été affectée et ce, de manière idiosyncratique

entre les trois sols analysés. Les relations de dose-réponse entre la dose d'exposition à l'$H_2$ et la réponse observée furent mineures au niveau de la structure des communautés, mais majeures au niveau de ses fonctions. En effet, bien que l'abondance relative de nombreux phylotypes soit proportionnelle à la concentration d'$H_2$, la contribution de ceux-ci sur l'ensemble de la diversité β fut négligeable. En revanche, les fonctions mesurées, soient la vitesse d'oxydation d'$H_2$ à faible affinité ainsi que celle d'oxydation de l'$H_2$, du $CH_4$ et du CO à haute affinité, étaient proportionnelles à la dose d'exposition. L'altération de processus d'oxydation de gaz traces en fonction de la distance d'une source ponctuelle productrice d'$H_2$ a donc pu être modélisée, permettant ainsi de représenter spatialement ce continuum métabolique. Cette thèse a apporté de nombreuses nouvelles connaissances aux domaines de la biogéochimie et de l'écologie microbienne en démontrant l'interaction de l'$H_2$ avec des processus biogéochimiques variés, en plus de soulever de nombreuses questions. Les diverses interactions entre les cycles biogéochimiques de l'$H_2$ et du carbone impliqueraient entre autres que les HOM, dont l'activité est généralement considérée comme une fonction spécialiste, pourraient en fait être davantage des groupes fonctionnels inhéremment mixotrophes ou même généralistes.

Mots clés: hydrogène, communautés microbiennes, biogéochimie, sol, gaz traces

# ABSTRACT

Molecular hydrogen ($H_2$) is a ubiquitous energy source found in trace amounts in the global atmosphere, but in higher concentrations in ecosystems such as wetlands and the rhizosphere of nitrogen-fixing legumes. Most nitrogen-fixing bacteria recruited by agriculturally-relevant legumes are devoid of hydrogenases, thus $H_2$ generated by nitrogenase enzymes diffuses into soils. Due to its strong energy potential, $H_2$ is quickly oxidized by $H_2$-oxidizing microbes (HOM), that is, within a 3-4.5 cm radius from its source. Those microbes have broad physiological capabilities, are found in both oxic and anoxic ecological niches, and are distributed among more than 40 phyla. While studies have shown that some HOM isolates promote plant growth *in vitro*, the overall impact of $H_2$ on soil microbial communities is unknown. The goal of this thesis is thus to shed light upon the potential of $H_2$ as a driver of soil biogeochemical processes. Specific research hypotheses are that elevated concentrations of $H_2$, representative of $H_2$-rich ecosystems, will directly or indirectly alter the structure and function of soil microbial communities, and that the ensuing responses will be proportional to $H_2$ exposure. Direct effects consist of changes in HOM distribution and metabolic activity, while indirect effects are associated with processes unrelated to $H_2$ metabolism. In order to validate these hypotheses, a dynamic microcosm chambers system was designed in the lab, thus enabling the incubation of soil microcosms to precise $H_2$ stoichiometric ratios ranging from 0.5 to 10,000 ppmv $H_2$. Sites sampled include a poplar monoculture, a larch monoculture and a farmland due to their contrasting biotic and abiotic features. Following the incubations, microbial communities—bacterial, fungal and high-affinity $H_2$-oxidizing bacteria (HA-HOB)—diversity profiles as well as several of their underlying processes were analyzed. Firstly, several direct effects were identified. Potential $H_2$ oxidation rates measured throughout the incubation have shown that HOM exposure to elevated $H_2$ concentrations leads to a swift activation of low-affinity HOM $H_2$ metabolism, all the while inhibiting the $H_2$ metabolism of HA-HOB. Overall, $K_m$ and $V_{max}$ of $H_2$ oxidation metabolism increased significantly post-incubation. In terms of HA-HOB community structure, the relative abundance of hundreds of genotypes was altered. Surprisingly, none of them matched those currently listed in public genomic databases. Secondly, numerous indirect effects were identified post-incubation. Three functions unrelated to $H_2$ oxidation were altered, *i.e.* an abatement in methane ($CH_4$) and carbon monoxide (CO) oxidation and a stimulation of carbon substrates utilization. Global functional responses were universal, yet the intensity of such alterations depended on soil biotic and abiotic conditions. Furthermore, the relative abundance of non-$H_2$-oxidizing microbes—Fungi and partial genome assemblies devoid of hydrogenase-coding genes—was also affected idiosyncratically between the three soils analyzed. Dose-response relationships between $H_2$ exposure and response at the community structure level was minor, however major changes were observed at the functional level. Indeed, despite the fact that the relative abundance of numerous phylotypes was proportional to the $H_2$ concentration, their contribution to overall β-diversity was negligible. However, measured functions, *i.e.* low-affinity $H_2$ oxidation rates and high-affinity $H_2$, $CH_4$ and CO oxidation rates, were proportional to the exposure dose. The alteration of trace gases oxidation processes as a function of distance from an $H_2$ point source was modeled to spatially represent this

metabolic continuum. This thesis has brought significant knowledge to the fields of biogeochemistry and microbial ecology by demonstrating the interaction between $H_2$ and various biogeochemical processes in addition to have raised several questions. For instance, interactions between $H_2$ and carbon biogeochemical cycles would imply that HOM, whose activity is generally assumed to be a specialist function, could in fact be an inherently mixotrophic or generalist functional group.

Keywords: hydrogen, microbial communities, biogeochemistry, soil, trace gases

# AVANT-PROPOS

Cette thèse est écrite sous forme de thèse par articles, incluant cinq articles révisés par les pairs et publiés ainsi qu'une introduction et une discussion globales permettant de lier ces articles. Le premier chapitre (1) constitue une revue de littérature abordant d'une part la diversité des microorganismes produisant ou utilisant l'hydrogène moléculaire ($H_2$), suivi de la proposition du concept que l'$H_2$ est, via les microorganismes qui l'utilisent, un joueur clé du cycle du carbone dans de nombreux écosystèmes. Le second chapitre (2) élabore la mise en contexte, la problématique, les hypothèses et objectifs de recherche de ma thèse. Les chapitres suivants (3-6) consistent en quatre articles publiés dans différents journaux scientifiques. Un récapitulatif des conclusions tirées ainsi que l'intégration de ces articles de manière à justifier leur apport à la littérature scientifique se trouve ensuite au chapitre (7). Quelques projets connexes en cours ainsi que des perspectives par suite de cette thèse sont décrits dans ce même chapitre. Le dernier chapitre (8) fait état de la bibliographie à laquelle cette thèse se réfère. Finalement, des annexes contenant des liens URL permettant d'accéder au matériel supplémentaire de format hors-norme des articles publiés se trouvent à la fin du document. Ces tableaux et figures sont aussi disponibles sur le site web des journaux scientifiques où ont été publiés les articles.

# TABLE DES MATIERES

# LISTE DES TABLEAUX

# LISTE DES FIGURES

# LISTE DES ABRÉVIATIONS

% ..... Percent
$_{(app)}K_m$ ..... Apparent $K_m$
$_{(app)}V_{max}$ ..... Apparent maximum rate achieved by the system
(dw) ..... Dry weight
[ ] ..... Concentration
“ ..... Inch
°C ..... Degree Celsius
µL ..... Microlitre
ACE ..... Abundance-based Coverage Estimator
$aH_2$ ..... Atmospheric hydrogen
AMV ..... Avian myeloblastosis virus
ANOVA ..... Analysis of variance
ASV ..... Amplicon sequence variant
ATP ..... Adenosine triphosphate
BIO ..... Biodiversity composite
BLAST ..... Basic Local Alignment Search Tool
BLASTN ..... Basic Local Alignment Search Tool Nucleotide
BLASTP ..... Basic Local Alignment Search Tool Protein
bp ..... Base pair
BWA ..... Burrows-Wheeler Aligner
C ..... Carbon
C/N ..... Carbon-to-nitrogen ratio
$CH_4$ ..... Methane
cm ..... Centimeter
$cm^3$ ..... Cubic centimeter
CO ..... Carbon monoxide
$CO_2$ ..... Carbon dioxide
COG ..... Clusters of Orthologous Groups of proteins
CPM ..... Count per million
DL ..... Detection limit
DNA ..... Deoxyribonucleic acid
DOE ..... Department of energy
dsDNA ..... Double-stranded DNA
DUK ..... Decontamination Using Kmers
$e^-$ ..... Electron
*e.g.* ..... Exempli gratia (for example)
ECD ..... Electron capture detector
$eH_2$ ..... Elevated hydrogen

F (sample) ... Farmland
FC ..... Fold-change
FID ..... Flame ionization detector
g ..... gram
Gb ..... Gigabase
GC..... Gas chromatograph
GI ..... GenInfo Identifier
h ..... Hour
H (sample)... High mixing ratio (10,000 ppmv)
$H^+$ ..... Proton
$H_2$ ..... Hydrogen
HAH..... High-affinity [NiFe]-hydrogenases
HA-HOB ..... High-affinity hydrogen-oxidizing bacteria
HMM..... Hidden Markov Model
HOB ..... Hydrogen-oxidizing bacteria
HOM..... Hydrogen-oxidizing microbes
*i.e.* ..... Id est (in other words)
ITS..... Internal transcribed spacer
JGI..... Joint Genome Institute
JTT ..... Jones-Taylor-Thornton
KEGG..... Kyoto Encyclopedia of Genes and Genomes
$K_M$..... Michaelis constant
KOG ..... EuKaryotic Orthologous Groups
L (soil) ..... Larch soil
L (treatment) Low mixing ratio (0.5 ppmv)
LG ..... Le and Gascuel
LR..... Likelihood ratio
MBH ..... Membrane-bound hydrogenase
MEGA..... Molecular Evolutionary Genetics Analysis
$H_4$MPT..... Tetrahydromethanopterin
MG-RAST.... Metagenomics Rapid Annotation using Subsystem Technology
min ..... Minute
ml ..... Milliliter
ML ..... Maximum likelihood
mm ..... millimeter
MOB ..... Methane-oxidizing bacteria
MP..... Maximum parsimony
n ..... Sample size
N..... Nitrogen
n.s. ..... Non-significant
$N_2$ ..... Nitrogen

NCBI..... National Center for Biotechnology Information
NJ..... Neighbor joining
nmol ..... Nanomole
nt..... Nucleotide
OD..... Optical density
OTU..... Operational taxonomic unit
P (sample)... Poplar soil
PC ..... Principal component
PCA..... Principal component analysis
PCR..... Polymerase chain reaction
PERMANOVA……..Permutational multivariate analysis of variance
Pfam..... Protein families annotations and multiple sequence alignments
pH..... Potential of hydrogen
pmol ..... picomole
ppmv ..... part per million by volume
QC..... Quality control
qPCR..... Quantitative polymerase chain reaction
qRT-PCR..... Quantitative reverse transcription polymerase chain reaction
$R^2$ ..... Coefficient of determination
RDA..... Redundancy Analysis
RNA..... Ribonucleic acid
rRNA ..... Ribosomal ribonucleic acid
SCCM..... Standard cubic centimeter per minute
SEM ..... Structural equation model
SH ..... Soluble hydrogenase
SIMPROF .... Similarity profile tool
SOIL..... Abiotic parameters composite
ssDNA..... Single-stranded DNA
TCD..... Thermal conductivity detector
Tg..... Teragram
T-RFLP..... Terminal restriction fragment length polymorphism
UPGMA..... Unweighted pair group method with arithmetic mean
URL..... Uniform Resource Locator
whc..... Water holding capacity
y ..... Year
α..... Significance level
β (variable) .. Soft-thresholding power

# 1 MOLECULAR HYDROGEN, A NEGLECTED KEY DRIVER OF SOIL BIOGEOCHEMICAL PROCESSES

**Titre français** : L'hydrogène moléculaire en tant que moteur négligé des cycles biogéochimiques du sol

**Article accepté dans le journal Applied and Environmental Microbiology le 13 janvier 2019**

**Auteurs** : Sarah Piché-Choquette[1]✉ et Philippe Constant[1]

[1]INRS-Institut Armand-Frappier, 531 boulevard des Prairies, Laval (Québec), Canada, H7V 1B7

✉ Auteur correspondant

**Contribution des auteurs**

Sarah Piché-Choquette: Écriture de l'article

Philippe Constant: Écriture de l'article

## 1.1 Résumé en français

Des études ont émis l'hypothèse que l'atmosphère de la Terre primitive était riche en gaz réducteurs comme l'hydrogène ($H_2$). Puisque ce gaz est ubiquitaire dans la biosphère, qu'il peut aisément diffuser à travers les cellules microbiennes et qu'il nécessite une faible énergie d'activation, on suppose que l'$H_2$ pourrait être le premier donneur d'électron ayant donné naissance à la synthèse d'ATP. Même aujourd'hui, les enzymes hydrogénases permettant la production et l'oxydation d'$H_2$ sont retrouvés dans des milliers de génomes originaires des trois domaines du vivant et ce, en provenance d'écosystèmes aquatiques, terrestres ou même associés à un hôte. Bien que l'$H_2$ ait déjà été proposé à titre de source d'énergie universelle pour la croissance et la maintenance, sa contribution en tant que moteur de processus biogéochimiques a reçu peu d'attention. Nous tenons ainsi à remédier à cette lacune en résumant la littérature actuelle sur la classification, la distribution et le rôle physiologique des hydrogénases. La répartition de ces enzymes parmi de nombreux groupes fonctionnels microbiens ainsi que des résultats récents sont par la suite apportés de manière à supporter l'hypothèse que les microorganismes oxydant l'$H_2$ sont des espèces-clés accentuant le recyclage du carbone le long de gradients de concentration d'$O_2$ dans les sols riches en $H_2$. En conclusion, nous suggérons de se concentrer sur la flexibilité métabolique des microorganismes oxydant l'$H_2$ en employant des approches tant au niveau de l'individu que de la communauté microbienne afin de déterminer l'impact de l'$H_2$ sur le recyclage du carbone ainsi que le potentiel de recyclage du carbone des microorganismes oxydant l'$H_2$ et ce, via des techniques dépendantes et indépendantes de la culture en laboratoire afin mieux nous éclairer sur le rôle de l'$H_2$ en tant que moteur des processus biogéochimiques.

## 1.2 Abstract

The atmosphere of the early Earth is hypothesized to have been rich in reducing gases such as hydrogen ($H_2$). $H_2$ has been proposed as the first electron donor leading to ATP synthesis due to its ubiquity throughout the biosphere as well as its ability to easily diffuse through microbial cells and its low activation energy requirement. Even today, hydrogenase enzymes enabling the production and oxidation of $H_2$ are found in

thousands of genomes spanning the three domains of life across aquatic, terrestrial and even host-associated ecosystems. Even though $H_2$ has already been proposed as a universal growth and maintenance energy source, its potential contribution as a driver of biogeochemical cycles has received little attention. Here, we bridge this knowledge gap by providing an overview of the classification, distribution and physiological role of hydrogenases. Distribution of these enzymes in various microbial functional groups and recent experimental evidences are finally integrated to support the hypothesis that $H_2$-oxidizing microbes are keystone species driving C cycling along $O_2$ concentration gradients found in $H_2$-rich soil ecosystems. In conclusion, we suggest focusing on the metabolic flexibility of $H_2$-oxidizing microbes by combining community-level and individual-level approaches aiming to decipher the impact of $H_2$ on C cycling and the C-cycling potential of $H_2$-oxidizing microbes, via both culture-dependent and culture-independent methods to give us more insight into the role of $H_2$ as a driver of biogeochemical processes.

## 1.3 Introduction

Several prebiotic chemistry models have hypothesized that the primary atmosphere of the early Earth was rich in reducing gases such as molecular hydrogen ($H_2$), carbon monoxide (CO) and methane ($CH_4$) (Oparin, 1938; Tian *et al.*, 2005; Trainer, 2013; Zahnle *et al.*, 2010). The ubiquitous availability of $H_2$, its low activation energy requirements and its ability to diffuse through microbial cells led to the hypothesis that $H_2$ was the first electron donor as a mean to generate an ion gradient and consequently, ATP synthesis (Lane *et al.*, 2010; Morita, 1999). In this regard, the iron-sulfur world hypothesis stipulates that NiFe/FeS active centers similar to the active site of hydrogenase enzymes—catalyzing the interconversion of $H_2$ into protons and electrons—could have emerged billion years ago and have led to the first biotic sources and sinks of $H_2$ (Wächtershäuser, 2010). The availability of elevated mixing ratios of $H_2$ in many ecological niches might represent a type of selection pressure maintaining this ancient metabolism within today's microbes, including environmental isolates (*e.g.* newly-cultivated *Verrucomicrobia*) (Carere *et al.*, 2017) and human pathogens (*e.g. Helicobacter pylori*-associated gastric

carcinogenesis) (Wang *et al.*, 2016). Indeed, hydrogenases are well represented within the tree of life, as they are found in thousands of genomes from over 30 phyla ((Greening *et al.*, 2016; Vignais & Billoud, 2007), Fig. 1.1, 1.2 and 1.3). They support autotrophic and mixotrophic lifestyles in various ecosystems encompassing soil, aquatic and animal-associated niches (Morita, 1999).

In soils, microbial growth is often limited by the scarcity of carbon sources and nutrients (Morita, 1999), which translates to bursts of growth in-between intervals of starvation-survival. On the other hand, $H_2$ is ubiquitous, requires a low activation energy and can easily permeate microbial cells (Morita, 1999). The ubiquity of $H_2$ is notably due to its high penetrating power through materials and its continuous abiogenic production in Earth's upper mantle and biogenic production by fermentation, $N_2$ fixation and photoproduction (Morita, 1999). $H_2$ oxidation can be combined with the reduction of a wide variety of oxidants, and therefore it can be performed in most ecosystems. This combination provides a quick energy fix, via $H_2$, to overcome those short-to-long intervals of starvation-survival. Several studies have calculated this minimal energy requirement from thermodynamic and bioenergetic perspectives (Alperin & Hoehler, 2009; Hoehler *et al.*, 2001; Hoehler & Jørgensen, 2013).

When environmental conditions are not suitable for growth, $H_2$ oxidation could provide energy to undergo anabiosis, *i.e.* long-term stationary phase or even dormancy (Keilin, 1959; Morita, 1999). Such lifestyle might combine starvation-survival with another survival mechanism such as anahydrobiosis (*i.e.* desiccation) or cryobiosis (*i.e.* freezing) (Finkel, 2006), not unlike microbes trapped in amber or frozen in permafrost. Metabolically inactive lifestyles drastically reduce energy demands, yet microbial cells still need to prevent amino acids racemization (*i.e.* spontaneous change of isomeric form) and DNA depurination over extended periods of time by constantly repairing cellular damage or re-synthesising DNA (Alperin & Hoehler, 2009; Hoehler & Jørgensen, 2013; Onstott *et al.*, 1998; Poinar *et al.*, 1996), requiring a continuous input of energy. When organic matter

and $H_2$ abound, $H_2$-oxidizing microbes (HOM) can switch from short- or long-term dormancy to a more metabolically-active lifestyle (Morita, 1999).

While recent studies have shown the ubiquity and importance of $H_2$ in natural and engineered ecosystems, this field of study has been fairly overlooked since few research groups have focussed on biotic $H_2$ dynamics at the ecosystem level (Meredith *et al.*, 2017). The aim of this review is to examine and discuss the potential role of $H_2$ as an energy source maintaining microbial viability and activity, and in turn sustaining soil biogeochemical processes. Firstly, a brief overview of hydrogenases enzymes is provided, including their classification and physiological roles, with emphasis on their kinetic properties and $O_2$ tolerance. The following section describes functional groups producing and oxidizing $H_2$ across oxic and anoxic soil habitats. The last section integrates the biochemical properties of hydrogenases to propose succession patterns of HOM along theoretical gradients of $O_2$ and $H_2$ prevailing in soils as well as their potential implications in the carbon cycle. This review concludes with research perspectives aiming to evaluate the contribution of HOM in soil biogeochemical processes.

## 1.4 Hydrogenases classification: an overview

Hydrogenases are metalloenzymes catalyzing the interconversion of $H_2$ to protons and electrons according to the following reaction: $H_2 \rightleftharpoons 2H^+ + 2e^-$ (Constant & Hallenbeck, 2013). These enzymes are used either to provide energy, to reduce cofactors or to disperse reducing equivalents generated through anaerobic processes (Vignais & Billoud, 2007). The conformation of electron bridges between structural subunits as well as the localization of the protein, membrane-bound or soluble, determines their physiological role (Constant & Hallenbeck, 2013). Their classification into [Fe]-hydrogenases, [FeFe]-hydrogenases and [NiFe]-hydrogenases is based on differences between their active sites and amino acid sequences (Armstrong & Albracht, 2005). These classes probably originated from converging rather than sequential evolution, since their tertiary structure, catalytic activity and taxonomic distribution are clearly distinct (Vignais, 2008; Vignais *et al.*, 2001). Horizontal gene transfer is also common in hydrogenases due to their modular

structure (Calteau *et al.*, 2005; Hansen & Perner, 2016; Keeling, 2009; Ludwig *et al.*, 2006). Consequently, several microbes, such as *Cupriavidus necator* (phylum *Proteobacteria*) and *Mycobacterium smegmatis* (phylum *Actinobacteria*), have the genomic capability to express multiple hydrogenases with distinct physiological roles, resulting in an increased metabolic flexibility (Berney *et al.*, 2014; Burgdorf *et al.*, 2005a).

### 1.4.1 *[Fe]-hydrogenases*

[Fe]-hydrogenases are homodimers consisting of two 38 kDa subunits with a single Fe atom in their catalytic center (Constant & Hallenbeck, 2013). In contrast with [FeFe]- and [NiFe]-hydrogenases, [Fe]-hydrogenases do not contain Fe-S clusters, are restricted to hydrogenotrophic methanogenic archaea within the *Euryarchaeota* phylum ((Zirngibl *et al.*, 1992), Fig. 1.1 and 9.1) and are not inactivated by $O_2$ (Fontecilla-Camps *et al.*, 2007; Lyon *et al.*, 2004). In nickel-limiting conditions, [Fe]-hydrogenases work along with $F_{420}$-dependent methylene-$H_4$MPT dehydrogenases to redirect electrons originating from $H_2$ oxidation towards $H_2$- and $CO_2$-dependent methanogenesis (Thauer *et al.*, 1996).

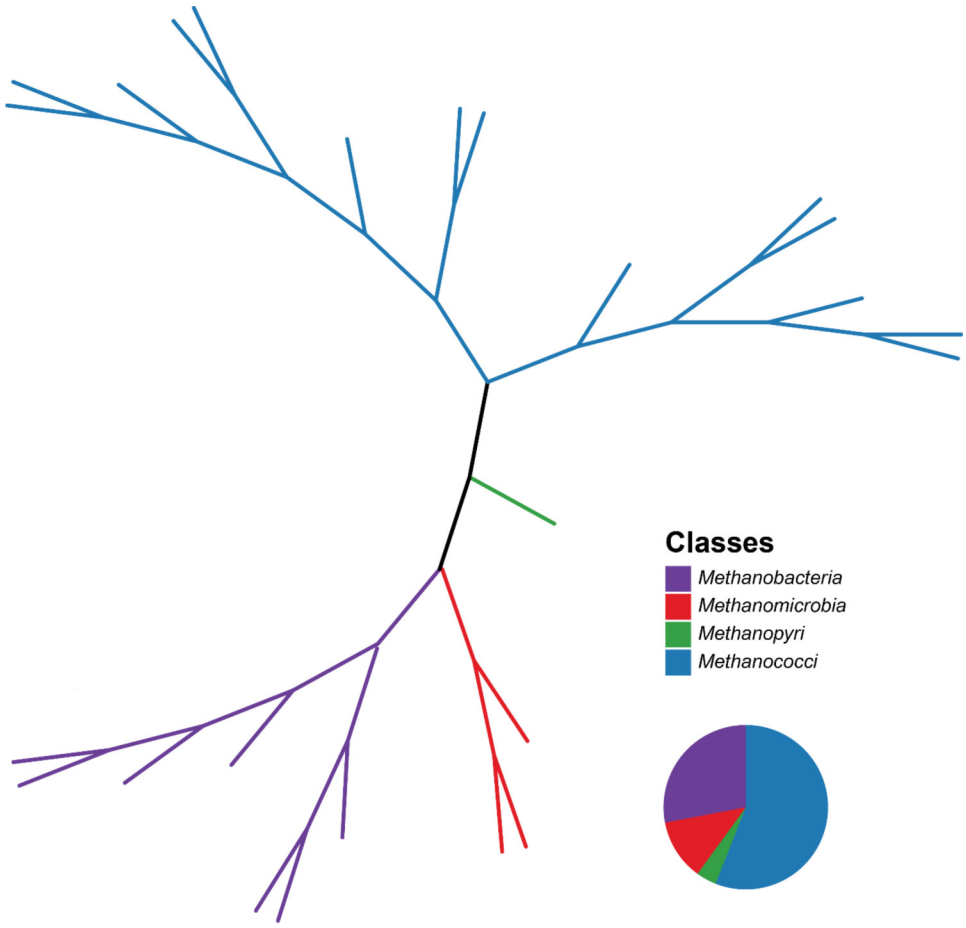



**Figure 1.1** **Consensus tree of 25 [Fe]-hydrogenases sequences, with an alignment length of 389 amino acids.Protein sequences were imported from a study by Greening *et al.* (2016), which included non-redundant putative hydrogenase catalytic subunits from cultured and environmental metagenomes sourced from NCBI Refseq and JGI MDM (Joint Genome Institute, Microbial Dark Matter) databases. More details are found in the article by Greening *et al.* (2016). See Fig. 9.1 for complete taxa names and a more detailed phylogenetic tree. Hydrogenases amino acids sequences were aligned with MUSCLE (Edgar, 2004) and clustered into phylogenetic trees using the following algorithms: Maximum likelihood (Jones-Taylor-Thornton substitution model) using RAxML version 8.2.10 (Stamatakis, 2014); Maximum parsimony using PAUP 4.0 (Swofford, 2003); and Neighbor-joining (Jones-Taylor-Thornton substitution model) using BIONJ (Gascuel, 1997). CIPRES Science Gateway v3.3 (Miller *et al.*, 2011) servers were used for phylogenetic trees construction. Tree branches supported by over 50% of the 1000 bootstrap replications were represented in the final consensus tree. The consensus tree was built using the "ape" package (Paradis *et al.*, 2004) in R (R Development Core Team, 2017). Branch colors represent taxonomic classification at the class level, since all 25 sequences are part of the *Euryarchaeota* phylum, while pie charts show the relative abundance of each class within *Euryarchaeota*.**

### 1.4.2 *[FeFe]-hydrogenases*

[FeFe]-hydrogenases are mono-, di-, tri- or tetrameric enzymes containing an active center with two Fe atoms, called H-cluster, as well as Fe-S clusters (Vignais & Colbeau, 2004). They are strictly anaerobic since their active site is denatured by $O_2$ exposure (Constant & Hallenbeck, 2013). [FeFe]-hydrogenases almost exclusively produce $H_2$ (Burgdorf *et al.*, 2005a; Thauer, 2011), with a few exceptions, such as in *Desulfovibrio vulgaris* (phylum *Proteobacteria*), whose [FeFe]-hydrogenase can oxidize $H_2$ in nickel-limited conditions to circumvent [NiFe]-hydrogenase synthesis (Pohorelic *et al.*, 2002). [FeFe]-hydrogenases are mostly found in anaerobic *Firmicutes* and sulfate reducers (Fig. 1.2 and 9.2) due to their $O_2$-sensitivity (Adams, 1990; Atta & Meyer, 2000; Vignais *et al.*, 2001). They consist of the only class of hydrogenases found in eukaryotes, namely in unicellular green algae *Chlamydomonas reinhardtii* (phylum *Chlorophyta*) (Happe *et al.*, 1994), where they are located within membrane-limited organelles like chloroplasts and hydrogenosomes (Akhmanova *et al.*, 1998).

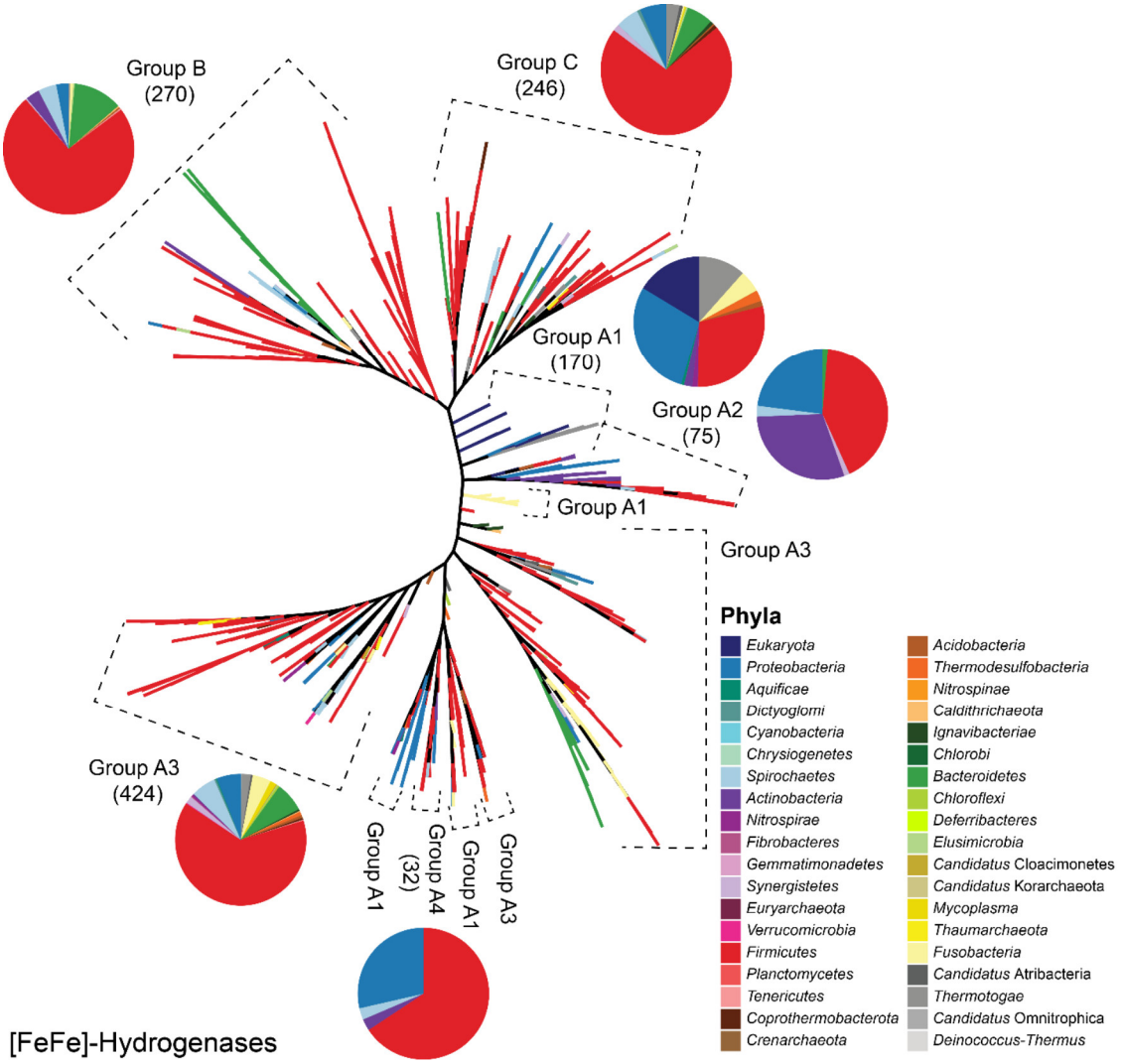



**Figure 1.2** **Consensus tree of 1217 [FeFe]-hydrogenases sequences, with an alignment length of 3525 amino acids, computed as described in Fig. 1.1.See Fig. 9.2 for complete taxa names and a more detailed phylogenetic tree. Branches colors represent taxonomic classification at the phylum level while pie charts show the relative abundance of each phylum within hydrogenases' subgroups. Dashed brackets show which tree branches belong to a specific hydrogenases group. Numbers between parentheses depict the amount of sequences within that group.**

### 1.4.3 *[NiFe]-hydrogenases*

[NiFe]-hydrogenases are the most widespread hydrogenases class (Vignais & Billoud, 2007). Several structural and accessory genes are necessary for their maturation (Lacasse & Zamble, 2016). While [NiFe]-hydrogenases can be multimeric, they contain

at the very least a core heterodimeric component comprising a large α-subunit (*ca.* 60 kDa) and a small β-subunit (*ca.* 30 kDa) (Constant & Hallenbeck, 2013). [NiFe]-hydrogenases represent a heterogenous group of hydrogenases divided into 5 subgroups according to their phylogeny (Fig. 1.3 and 9.3) and physiology (Constant & Hallenbeck, 2013). Among these physiological traits, $O_2$ tolerance and low catalytic affinity are the most common (Constant & Hallenbeck, 2013). Indeed, only a few subgroups are sensitive to $O_2$ and a single subgroup has a high-affinity towards $H_2$, *i.e.* can oxidize atmospheric $H_2$ (0.53 ppmv (Novelli *et al.*, 1999)). This classification scheme supported by concurrence between gene sequences and physiological roles of hydrogenases has been revisited recently using structural gene sequences similarity networks, suggesting a higher number of putative sub-classes within each group (Greening *et al.*, 2016; Søndergaard *et al.*, 2016). Such reclassifications will not be covered here owing to the lack of experimental observations supporting their coherence with genetic, biochemical and physiological features of hydrogenases.

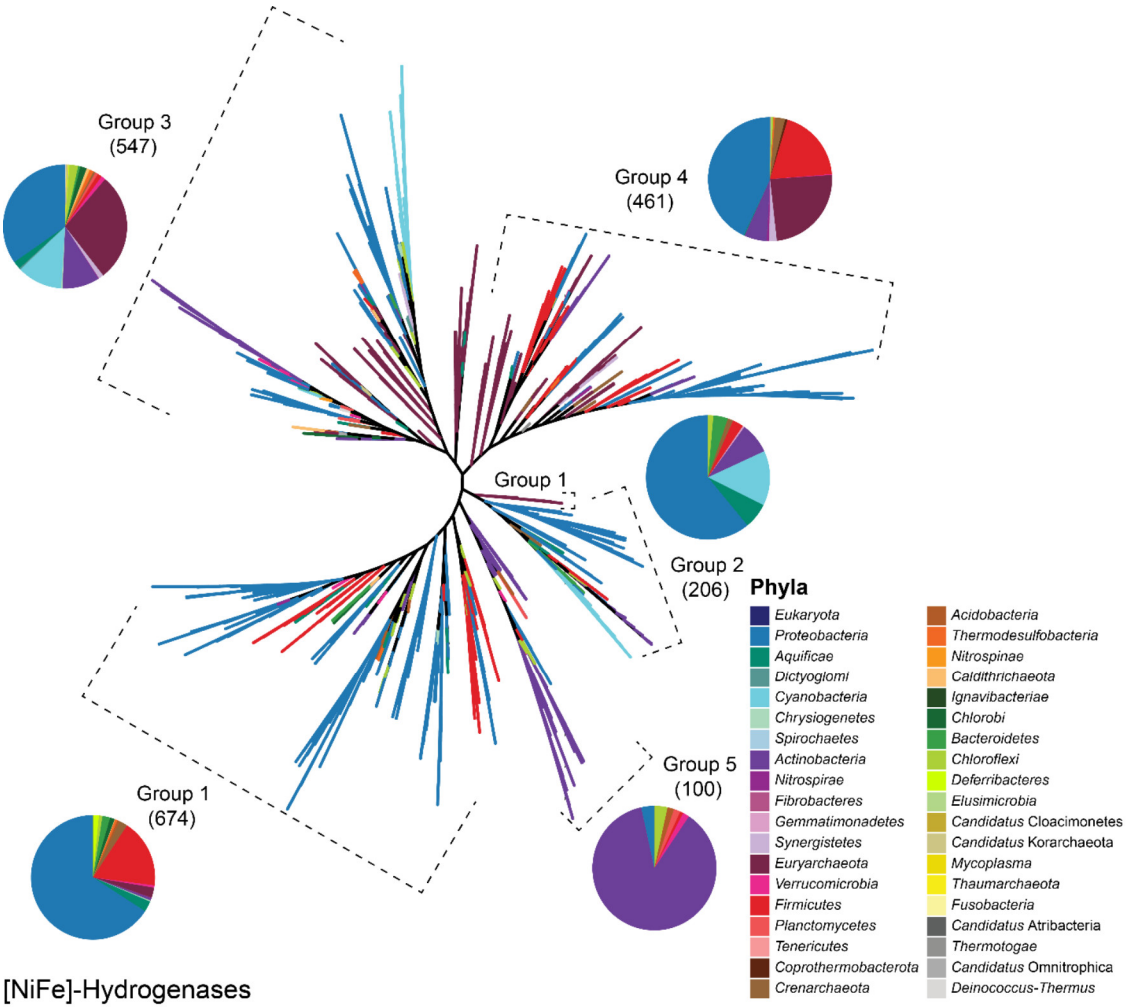



**Figure 1.3** **Consensus tree of 1988 [NiFe]-hydrogenases sequences, with an alignment length of 1850 amino acids, computed as described in Fig. 1.1.See Fig. 9.3 for complete taxa names and a more detailed phylogenetic tree. Branches colors represent taxonomic classification at the phylum level while pie charts show the relative abundance of each phylum within hydrogenases' subgroups. Dashed brackets show which tree branches belong to a specific hydrogenases group. Numbers between parentheses depict the amount of sequences within that group.**

*Group 1 [NiFe]-hydrogenases*

Group 1 [NiFe]-hydrogenases (membrane-bound hydrogenases or MBH) are $H_2$-uptake hydrogenases attached to the cytoplasmic membrane (periplasmic side) (Constant & Hallenbeck, 2013). They are intimately linked to the electron transport chain by supplying the quinone pool with electrons via cytochrome *b*, leading to chemiosmosis and the

subsequent generation of ATP (Pandelia *et al.*, 2012). MBH are used for energy generation by coupling the oxidation of $H_2$ with the reduction of several electron acceptors such as $CO_2$, $SO_4^{-2}$, $NO_3^-$, Fe(III) oxides or $O_2$ (Pandelia *et al.*, 2012). They are mostly found in *Proteobacteria* and *Firmicutes* (62.5% and 17.3%) (Fig. 1.3 and 9.3). Ancestral MBH found in methanogenic archaea and sulfate-reducing bacteria are $O_2$-sensitive (Pandelia *et al.*, 2012). Contrastingly, those associated with Isp complex, in obligate and facultative autotrophs, and with HybA-MBH, in *Gammaproteobacteria*, have a low tolerance towards $O_2$ (Pandelia *et al.*, 2012). $O_2$-tolerant MBH are used to perform the Knallgas reaction, *i.e.* oxidation of $H_2$ using $O_2$ as terminal electron acceptor, or other reactions involving high-potential electron acceptors such as nitrogen oxides or bacteriochlorophylls (Pandelia *et al.*, 2012).

*Group 2 [NiFe]-hydrogenases*

Group 2 [NiFe]-hydrogenases consist of cytoplasmic enzymes divided into 2 distinct functional groups; 2a and 2b (Constant & Hallenbeck, 2013). [NiFe]-hydrogenases encompassing Group 2a are soluble uptake hydrogenases oxidizing $H_2$ endogenously produced by $N_2$ fixation (Tamagnini *et al.*, 2007) or NADH/NADPH oxidation (Berney *et al.*, 2014) and are mainly found in *Cyanobacteria* and other $N_2$-fixing bacteria ((Bothe *et al.*, 2010; Tamagnini *et al.*, 2007), Fig. 1.3 and 9.3). Their role is to channel electrons to the electron transport chain and supply energy to the cell (Vignais & Colbeau, 2004). They are often located inside cyanobacterial heterocysts devoid of oxygenic photosynthesis and are upregulated under anoxic conditions due to their $O_2$ sensitivity (Axelsson & Lindblad, 2002). Contrastingly, Group 2b [NiFe]-hydrogenases are $H_2$-sensing regulatory hydrogenases (RH) acting as two-component signal transduction systems (Constant & Hallenbeck, 2013). The specific shape and size of the hydrophobic gas channel leading exogenously-produced $H_2$ to the active site confers $O_2$-tolerance to the enzyme (Buhrke *et al.*, 2005). When $H_2$ is present, RH induce the transcription of Group 1 and Group 3 uptake hydrogenases operons in Knallgas bacteria (Buhrke *et al.*, 2004; Lenz *et al.*, 2002).

*Group 3 [NiFe]-hydrogenases*

Group 3 [NiFe]-hydrogenases are cytoplasmic enzymes called bidirectional heteromultimeric [NiFe]-hydrogenases or soluble hydrogenases (SH) (Constant & Hallenbeck, 2013). Their role is to provide redox balance within the cell in both bacteria and archaea (Vignais & Billoud, 2007). The structural subunits of those enzymes are in close association with protein modules binding soluble cofactors (Constant & Hallenbeck, 2013). Group 3 [NiFe]-hydrogenases are $O_2$-tolerant due to two extra cyanide ligands bound to the Ni and Fe atoms of their active site (Burgdorf *et al.*, 2005b). These hydrogenases mainly found in *Proteobacteria* (37.4%) and *Euryarchaeota* (27.4%) (Fig. 1.3 and 9.3) are further divided into 4 subgroups (Constant & Hallenbeck, 2013). Group 3a [NiFe]-hydrogenases are heterotrimeric enzymes initially found in methanogenic archaea (Peters *et al.*, 2015), but more recently in some bacterial phyla (Ney *et al.*, 2016), and are used to reduce the $F_{420}$ cofactor. Group 3b [NiFe]-hydrogenases are tetrameric enzymes acting like bifunctional sulfhydrogenases (Constant & Hallenbeck, 2013). They were initially found in hyperthermophilic archaea yet have also been detected in bacterial phyla more recently (Greening *et al.*, 2016). They reduce $S^0$ to $H_2S$ or oxidize NAD(P)H and produce $H_2$ to disperse reducing equivalents generated through fermentation, or alternatively to produce NADPH by using $H_2$ (Ma *et al.*, 1993). Group 3c [NiFe]-hydrogenases are heterodisulfide reductase-associated hydrogenases that catalyze the reduction of compounds such as methyl viologen and CoM-S-S-CoB (Coenzyme M-disulfide bond-Coenzyme B) (Thauer *et al.*, 2010) and are mostly found in methanogens and sulfate reducers. Group 3d [NiFe]-hydrogenases consist of a heteromeric hydrogenase and a NADH-dehydrogenase module that binds to NAD(P)H (Constant & Hallenbeck, 2013). The amount of hydrogenase subunits varies greatly, consisting, for example, of a heterodimeric enzyme in *C. necator* (Lauterbach *et al.*, 2011) or a heteropentameric enzyme in *Cyanobacteria* (Puggioni *et al.*, 2016). These hydrogenases are used to balance the $NAD^+$/NADH pool by reducing $NAD^+$ using $H_2$ (Constant & Hallenbeck, 2013).

*Group 4 [NiFe]-hydrogenases*

Group 4 [NiFe]-hydrogenases or membrane-associated energy-converting hydrogenase are oxygen-sensitive (Bonam *et al.*, 1989), multimeric enzymes comprising two proteins embedded within the cytoplasmic membrane as well as hydrophilic subunits (Constant & Hallenbeck, 2013). Group 4 [NiFe]-hydrogenases are mostly found in methanogenic archaea, chemolithoautotrophic bacteria and in hyperthermophilic archaea (Fig. 1.3 and 9.3). In facultative chemolithoautotrophs, Group 4 hydrogenases disperse reducing equivalents generated through fermentation by coupling the reduction of protons from water with the anaerobic oxidation of C1 compounds like formate or carbon monoxide (Schut *et al.*, 2012). In acetoclastic methanogens, energy-converting hydrogenases (Ech) perform the $H_2$-dependent reduction of ferredoxin, leading to the generation of a proton gradient through the cytoplasmic membrane and the generation of ATP (Welte *et al.*, 2010). In methanogens without a cytochrome, Group 4 [NiFe]-hydrogenases (Eha and Ehb) reduce ferredoxin and catalyze $H_2$ oxidation, which is then coupled with the generation of a sodium ion motive force (Constant & Hallenbeck, 2013). Group 4 [NiFe]-hydrogenases also include energy-converting Mbh-type hyperthermophilic enzymes producing $H_2$ in order to re-establish oxidized ferredoxin levels (Schut *et al.*, 2013).

*Group 5 [NiFe]-hydrogenases*

Group 5 [NiFe]-hydrogenases are $O_2$-tolerant (Schäfer *et al.*, 2013) uptake hydrogenases harboring a high affinity towards $H_2$ enabling them to oxidize atmospheric $H_2$ (Constant *et al.*, 2010). Their activity was first discovered in spores of *Streptomyces* spp. (phylum *Actinobacteria*) (Constant *et al.*, 2008). These enzymes are likely associated to the cytosolic side of the membrane through an electron acceptor, but conflicting results have been reported (Greening *et al.*, 2014a; Schäfer *et al.*, 2013). Genomic database mining demonstrated that Group 5 [NiFe]-hydrogenases are mostly found in *Actinobacteria* (84.0%, Fig. 1.3 and 9.3) (Constant *et al.*, 2011b; Greening *et al.*, 2016). The proposed role of these hydrogenases is to provide a growth advantage (Greening *et al.*, 2014b) or a supplementary energy source within a survival-mixotrophic lifestyle (Liot & Constant, 2016) to sustain microbial seedbanks in oligotrophic environments (Constant *et al.*, 2011b). However, *C. necator* has a weakly expressed Group 5 [NiFe]-hydrogenase that

cannot oxidize atmospheric $H_2$ (Schäfer *et al.*, 2013). Indeed, biochemical properties conferring high affinity for $H_2$ are unknown. While there is a high sequence similarity between the active site of Group 5 [NiFe]-hydrogenases, such hydrogenases display a wide spectrum of affinities, *e.g.* $K_m$ over 1,000 ppmv in *C. necator* in contrast to 11 ppmv in *Streptomyces* sp. PCB7 (Constant *et al.*, 2008). This suggests that either a lateral gene transfer in *C. necator* (Schwartz *et al.*, 2003) or the evolution of unknown biochemical attributes in hydrogenases were found in high-affinity HOM. Alternatively, other factors, such as the currently unknown physiological electron acceptor of Group 5 [NiFe]-hydrogenases (Greening *et al.*, 2014a; Schäfer *et al.*, 2016), could be responsible for their affinity towards $H_2$. Their affinity has not been tested extensively *in vitro*, thus we cannot assume that *C. necator* is an exception to the rule. Also, diverging phylogenetic analyses spawned confusion in hydrogenases terminology, yet Group 5 (Constant *et al.*, 2011b; Piché-Choquette *et al.*, 2017; Schäfer *et al.*, 2013) and Group 1h (Greening *et al.*, 2016) refer to the same enzymes.

## 1.5 Overview of the $H_2$ biogeochemical cycle

Biological sources and sinks of $H_2$ are common in oxic and anoxic ecosystems. In anoxic ecosystems—such as wetlands, freshwaters under the chemocline, marine sediments and animal gastrointestinal tracts—$H_2$ is mainly produced as a reaction intermediate of organic matter degradation by organisms from the three domains of life. The main biological sinks of $H_2$ in anoxic ecosystems are acetogens, methanogens, sulfate-reducing microbes, iron oxide-reducing microbes and nitrate-reducing microbes (Thauer, 2011). In oxic environments—such as upland soils, freshwaters above the chemocline and open oceans—$H_2$ is mainly generated by $N_2$ fixation and is consumed by Knallgas bacteria and high-affinity $H_2$-oxidizing bacteria (Thauer, 2011). Due to its energetic potential, $H_2$ is quickly consumed by microbes within the same microenvironment, indicating that $H_2$ production is likely the limiting step of the $H_2$ biogeochemical cycle (Thauer *et al.*, 1996). Microbial functional groups responsible for $H_2$ production and utilization in anoxic and oxic ecosystems will be presented in the following subsections to showcase their vast diversity and their involvement in the C biogeochemical cycle.

### 1.5.1 $H_2$ production in anoxic ecosystems

In anoxic ecosystems, photo-production (or biophotolysis), dark fermentation and photo-fermentation are responsible for most of the $H_2$ production (Thauer, 2011).

*Photo-production*

Photo-production of $H_2$ in aquatic sediments is performed by unicellular algae such as *Chlorella fusca* (phylum *Chlorophyta*), *Tetradesmus obliquus* (phylum *Chlorophyta*) (Gaffron & Rubin, 1942) (formerly *Scenedesmus obliquus*) and *Chlamydomonas reinhardtii* (Yagi *et al.*, 2016), as well as *Cyanobacteria*. In soils, several purple non-sulfur bacteria perform photo-production, including *Rhodobacter capsulatus* (phylum *Proteobacteria*) (Tsygankov *et al.*, 1998) and *Rhodopseudomonas rutila* (phylum *Proteobacteria*) (Oh *et al.*, 2004). This process occurs under oxygen and $CO_2$ limitation (Benemann, 2000; Ghirardi *et al.*, 2000). Light-excited electrons are transferred across photosystems to reduce ferredoxin, then ferredoxin is re-oxidized in order to produce $H_2$ via an [FeFe]-hydrogenase (Bourke *et al.*, 2017). This process maintains the electron flow within the electron transport chain to resume oxygenic photosynthesis as soon as better conditions arise (Benemann, 2000), such as changing tides or the next photoperiod.

*Photo-fermentation and dark fermentation*

Unicellular algae such as *Fragilariopsis* sp. (phylum *Bacillariophyta*) and *Pyramimonas* sp. (phylum *Chlorophyta*) (Bourke *et al.*, 2017), strict anaerobes like *Clostridium acetobutylicum* (phylum *Firmicutes*) (Zhang *et al.*, 2006) and even facultative anaerobes like *Escherichia coli*, phylum *Proteobacteria*) (Redwood *et al.*, 2008) from both aquatic and terrestrial environments can produce $H_2$ through dark fermentation. Glycolysis followed by the decarboxylation of pyruvate into acetyl-CoA reduces ferredoxin/$NAD^+$ into oxidized ferredoxin/NADH, supplying electrons to a hydrogenase and generating $H_2$ by their re-oxidation (Bourke *et al.*, 2017; Calusinska *et al.*, 2010; Schut & Adams, 2009). Alternatively, the conversion of CoA and pyruvate into formate and acetyl-CoA followed by the cleavage of formate also leads to the generation of $H_2$ and $CO_2$ (Khanna & Das,

2013). Oxidized ferredoxin/NADH are regenerated by discarding reducing equivalents and glycolysis can generate ATP anew. Photo-fermentation is performed by purple non-sulphur bacteria—*e.g. Rhodospirillum rubrum* (phylum *Proteobacteria*) or *Rhodobacter sphaeroides*—that use light energy to convert organic substrates like lactic acid into $H_2$ and $CO_2$ (Khanna & Das, 2013). These processes can be performed sequentially, given that there are dark and light cycles. According to genomic surveys and culture-dependent measurements, in aquatic environments, $H_2$-evolving fermenters include unicellular algae, *Firmicutes*, *Bacteroidetes*, *Spirochetes* and *Proteobacteria* (Bourke *et al.*, 2017; Boyd *et al.*, 2014; Dong *et al.*, 2018; Mei *et al.*, 2016). In anoxic soils, such as paddy fields, molecular surveys targeting [FeFe]-hydrogenases have shown that *Firmicutes* (mostly *Clostridia*), *Proteobacteria* (mostly *Deltaproteobacteria*) and *Chloroflexi* are the most abundant $H_2$ producers (Baba *et al.*, 2016; Baba *et al.*, 2014). Similarly, in a moderately acidic fen slurry enrichment, fermenters were identified as *Firmicutes* (*Clostridia* and *Negativicutes*) and *Deltaproteobacteria* (Schmidt *et al.*, 2010).

*Nitrogen fixation*

Various microbes such as *Clostridium* spp., *Bacillus* spp. (phylum *Firmicutes*) as well as methanogenic archaea and *Cyanobacteria* can fix $N_2$ and consequently produce $H_2$ in anoxic or microoxic environments (Leigh, 2000; Orr *et al.*, 2011). Since current estimates suggest that most of the biological nitrogen fixation is performed in oxic ecosystems (Fowler *et al.*, 2013), especially in surface waters (Jayakumar *et al.*, 2017; Voss *et al.*, 2013) and in the rhizosphere of leguminous and actinorhizal plants (Vitousek *et al.*, 2013), it will be discussed in the next section.

### 1.5.2 $H_2$ production in oxic ecosystems

*Nitrogen fixation*

The main biological process responsible for $H_2$ production in oxic ecosystems is biological $N_2$ fixation (BNF). BNF is performed by the nitrogenase complex according to the following reaction:

$N_2 + 8H^+ + 8e^- \rightarrow 2NH_3 + H_2$ (Burris, 1991; Hunt *et al.*, 1988; Postgate, 1982) (1)

BNF requires a tremendous amount of energy, *i.e.* a minimum of 16 ATP per fixed $N_2$ molecule (2 ATP per electron transferred (Burris, 1991)), thus it is strictly regulated and used only in the absence of bioavailable N (Mylona *et al.*, 1995). Since nitrogenases are irreversibly inactivated by $O_2$ (Berman-Frank *et al.*, 2001; Burris, 1991), microbes can avoid $O_2$ via temporal or spatial separation strategies: separation of $N_2$ fixation and oxygenic photosynthesis (Berman-Frank *et al.*, 2001); avoidance of $O_2$; high respiration rates and the use of $O_2$ transporters to buffer local $O_2$ concentrations (Brewin, 1991; Burris, 1991; Poole & Hill, 1997; Stal & Krumbein, 1987). $H_2$ is an obligate by-product of the BNF reaction, requiring approximately 35% of the electron flux used towards BNF (Hunt & Layzell, 1993) and leading to local accumulations in environments such as legumes rhizosphere (Witty, 1991; Witty & Minchin, 1998). BNF and hydrogen oxidation are often linked due to their complementary outcomes, since $H_2$ oxidation coupled to $O_2$ reduction recovers part of the ATP used towards $N_2$ fixation and leads to a reduction of $O_2$ pressure and $H_2$ buildups, both of which would prevent nitrogen fixation (Bothe *et al.*, 2010; Dixon, 1972). Furthermore, three BNF strategies are used by microbes, including symbiotic $N_2$ fixation (Mus *et al.*, 2016; Mylona *et al.*, 1995), plant-associated $N_2$ fixation (Steenhoudt & Vanderleyden, 2000) and free-living $N_2$ fixation (Postgate, 1982; Tschapek & Giambiagi, 1955). Unlike other nitrogen fixers, most leguminous plants symbionts do not possess the genetic capability to recycle $H_2$ produced through nitrogen fixation (*i.e.* *Hup*$^-$ genotype) (Uratsu *et al.*, 1982). Studies have hypothesized that the additional energy input provided by $H_2$ exerts a fertilization effect promoting growth of the host plant (Dong *et al.*, 2003; Maimaiti *et al.*, 2007).

### 1.5.3 $H_2$ uptake in anoxic ecosystems

In anoxic ecosystems, numerous microbial functional groups use $H_2$ as an electron donor to reduce terminal electron acceptors according to their availability and underlying thermodynamic constraints (Thauer *et al.*, 1996). This section discusses some of them in increasing redox potential order, from acetogenesis to denitrification, although other electron acceptors can be coupled to $H_2$ oxidation, such as $MnO_2$ and $CrO_4$. Many of those microbes are facultative anaerobes, yet these processes are only performed in

anoxic ecosystems since the reduction of $O_2$ would be favored thermodynamically over alternative electron acceptors. These functional groups are not mutually exclusive, *i.e.* numerous bacteria and archaea are able to perform $H_2$ oxidation using more than one electron acceptor.

*Acetogens*

Acetogenesis is performed by several strict anaerobes from the bacterial and archaeal domains of life (He *et al.*, 2016). Acetogens are found in a wide variety of ecosystems: wetlands (Rosencrantz *et al.*, 1999), the gastrointestinal tract of animals (Breznak, 1994; Rey *et al.*, 2010), oceans, subsurface sediments (Liu & Suflita, 1993), hypersaline waters (Ollivier *et al.*, 1994) and oxic soils (Peters & Conrad, 1995). While the latter sounds counterintuitive, several acetogens and methanogens can survive in presence of $O_2$ due to different mechanisms (Karnholz *et al.*, 2002; Küsel *et al.*, 2001). Most acetogens, namely *Clostridium* spp., are comprised within the *Firmicutes* (Ragsdale & Pierce, 2008), yet some are also found among other phyla, including archaeal phylum *Bathyarchaeota* (He *et al.*, 2016). The acetogenesis reaction follows the Wood-Ljungdahl pathway to convert C1 inorganic compounds ($CO_2$) into C2 organic compounds (acetate):

$2\ CO_2 + 4H_2 \rightleftharpoons CH_3COO^- + H^+ + 2\ H_2O$ (2) (Fischer *et al.*, 1932)

Acetogens scavenge $H_2$ generated through the degradation of organic matter. They face fierce competition from methanogens; the latter having a lower threshold for $H_2$ utilization and a higher energy yield (Le Van *et al.*, 1998; Schink, 1997). Acetogens can overcome methanogens in some ecosystems—such as in termites gut (Pester & Brune, 2007)—since the production of acetate consists of an ecological advantage in its animal host. Acetogens are inherently involved in $H_2$ and C cycling processes by coupling the utilization of $CO_2$ and $H_2$. In turn, this leads to the generation of acetate, a bioavailable C source for the surrounding microbial community. As strict or facultative anaerobes, acetogens can also use dissolved organic matter (DOM, also known as labile) through fermentation and produce their own $H_2$ and $CO_2$. Additionally, some of them; *e.g. Clostridium carboxidivorans* and *Clostridium drakei* (Küsel *et al.*, 2000; Liou *et al.*, 2005)), are also involved in CO oxidation to $CO_2$, crystalline cellulose and hemicellulose

degradation, polymeric organic matter (POM, also known as recalcitrant) turnover (Koeck *et al.*, 2014) and humic acids decomposition (Ueno *et al.*, 2016).

*Methanogens*

Methanogenesis is only performed by methanogenic archaea; which are common in wetlands, in the intestinal tract of animals, in extreme environments such as hot springs and hydrothermal vents (Conrad, 2009) and even in oxic soils (Peters & Conrad, 1995). Methanogens consist of the only functional group harbouring [Fe]-hydrogenases, which enables them to use $H_2$ as a reducing agent for the ferredoxin-dependent reduction of $CO_2$ to $CH_4$, according to the simplified reaction:

$CO_2 + 4\ H_2 \rightarrow CH_4 + 2H_2O$ (3) (Stephenson & Stickland, 1933; Thauer, 1998)

Methanogens thrive in ecosystems or layers within those ecosystems where electron acceptors with higher redox potential—such as sulfate and nitrate—are depleted due to the low redox potential of $CO_2$ (Thauer *et al.*, 1996). They scavenge $CO_2$ and $H_2$ derived from the degradation of organic matter, of which they can ferment on their own. Hydrogenotrophic methanogens using electron bifurcation are considered obligate autotrophs (Jablonski *et al.*, 2015); Methylotrophic methanogens using oxidative phosphorylation consume a wider variety of substrates (Singh *et al.*, 2005). Recent studies, however, have shown that hydrogenotrophic methanogens—such as the marine methanogen *Methanococcus maripaludis* (phylum *Euryarchaeota*) (Costa *et al.*, 2013)—might in fact have a greater flexibility than initially thought by using a combination of CO and formate rather than only $CO_2$ and $H_2$. The assumption that methanogens are strict anaerobes has also been challenged with the discovery of "*Candidatus*" Methanothrix paradoxum (phylum *Euryarchaeota*) (non-hydrogenotrophic) (Angle *et al.*, 2017). The involvement of methanogens in the C cycle also includes their ability to reduce humic acids, *e.g.* in *Methanosarcina barkeri* (phylum *Euryarchaeota*) (Bond & Lovley, 2002), which temporarily decreases their methanogenesis potential. Studies even suggest that anaerobic methanotrophic archaea (ANME) could be methanogens as well (Sebastian *et al.*, 2013; Timmers *et al.*, 2017).

*Sulfate-reducing bacteria and archaea*

Sulfate-reducing bacteria (SRB)—such as *Desulfovibrio* spp. and *Desulfotomaculum* spp. (phylum *Firmicutes*)—as well as sulfate-reducing archaea (SRA)—such as *Archaeoglobus* spp. (phylum *Euryarchaeota*)—are anaerobic microbes coupling the oxidation of $H_2$ (or organic compounds) to the reduction of sulfate ($SO_4^{2-}$) into sulfide compounds (Muyzer & Stams, 2008; Thauer & Kunow, 1995).

While they are found in oxic environments such as coastal sediments (Jørgensen, 1977), sulfate-reducing microbes are more common in marine and freshwater sediments, deep subsurface environments and in the gastrointestinal tract of animals (Muyzer & Stams, 2008). They are also found in acid mine drainages and alkaline environments such as soda lakes (Muyzer & Stams, 2008). $H_2$-utilizing sulfate reducers can outcompete acetogens and methanogens within the same ecosystem due to their higher growth kinetics (Muyzer & Stams, 2008). Species of *Desulfitobacterium* (phylum *Firmicutes*) are also able to perform the reduction of humic acids and polyaromatic compounds (Cervantes *et al.*, 2002; Gorny & Schink, 1994). Studies have also shown their potential ability to fix $CO_2$ (Kim *et al.*, 2012; Schauder *et al.*, 1987) and oxidize CO (Parshina *et al.*, 2010). Sulfate-reducing microbes are thought to be the main drivers of carbon cycling in marine sediments (Muyzer & Stams, 2008). $H_2$-oxidizing SRB also act as syntrophs in consortia performing the anaerobic oxidation of $CH_4$ (Boetius *et al.*, 2000).

*Iron oxide reducers*

Iron oxide-reducing bacteria and archaea (Fe(III) reducers) are found in soils, acid mine effluents (Das *et al.*, 1992), aquatic sediments (Lonergan *et al.*, 1996), hot springs (Kashefi *et al.*, 2002) and subsurface ecosystems. Since Fe(III) is insoluble in water, microbes need a direct contact with Fe(III) to reduce it (Gorby *et al.*, 2006). Microbes such as *Shewanella* spp. (phylum *Proteobacteria*), *Geobacter* spp. (phylum *Proteobacteria*) and *Geothrix* spp. (phylum *Acidobacteria*) can produce chelators to solubilize Fe(III) in

water or even produce electrically-conductive bacterial nanowires to reach it (Gorby *et al.*, 2006). They use Fe(III) as an electron acceptor for the oxidation of organic matter or $H_2$ (Lovley, 1993; Lovley *et al.*, 2004; Thauer *et al.*, 1977). Many Fe(III) reducers use other electron acceptors such as $O_2$, $SO_4^{2-}$, $NO_3^-$ or even CO and $CO_2$ (Bond & Lovley, 2002; Lovley *et al.*, 2004). Those Fe(III) reducers include bacteria such as *Geobacter* spp. and *Shewanella* spp. (Lonergan *et al.*, 1996; Lovley *et al.*, 2004), but also methanogens (Bond & Lovley, 2002) and other archaea like *Pyrobaculum islandicum* (phylum *Crenarchaeota*) (Kashefi & Lovley, 2000). Fe(III)-reducing processes are more efficient at scavenging $H_2$ than sulfate-reducing processes (Lovley & Goodwin, 1988). Several Fe(III) reducers can use humic substances as electron acceptors to anaerobically oxidize organic compounds (*e.g.* acetate produced by acetogens) as well as $H_2$ to support their growth (Lovley *et al.*, 1996). Such electron transfers also chelate Fe(III) oxides and make them bioavailable to Fe(III) reducers such as *Geobacter metallireducens*, which also contribute to pollutant degradation (*e.g.* benzene) (Lovley *et al.*, 1996).

*Denitrifying bacteria and archaea*

Denitrifying bacteria and archaea couple the reduction of nitrate ($NO_3^-$) and subsequently reduced nitrogen compounds to the oxidation of $H_2$ (or organic compounds) (Lee & Rittmann, 2002). Denitrification is performed in aquatic and terrestrial ecosystems mostly by heterotrophic microbes such as *Paracoccus denitrificans* (phylum *Proteobacteria*) but is also quite common in autotrophic microbes such as *Thiobacillus denitrificans* (phylum *Proteobacteria*) or within the *Hydrogenophilales* order (phylum *Proteobacteria*) (Ontiveros-Valencia *et al.*, 2013; Seitzinger *et al.*, 2006; Zumft, 1997). It is the most energetically favored anaerobic respiration process and is used by a wide variety of microbes (Greening *et al.*, 2016; Thauer *et al.*, 1996). Since it is performed in numerous sequential reduction reactions, typically $NO_3^- \rightarrow NO_2^- \rightarrow NO \rightarrow N_2O \rightarrow N_2$, many denitrifying microbes do not possess the whole metabolic machinery to perform complete denitrification. Within a single ecosystem, different microbial strains often complement their denitrifying capabilities for more efficient energy retrieval and avoidance of toxic N intermediates (Simon & Klotz, 2013). *Pseudomonas* spp. (phylum *Proteobacteria*) can

perform denitrification as well as POM degradation such as lignin (Tian *et al.*, 2014) and aromatic compounds degradation (Iyer *et al.*, 2017). Like SRB, some denitrifying bacteria are also syntrophs within AOM consortia (Raghoebarsing *et al.*, 2006).

### 1.5.4 $H_2$ uptake in oxic ecosystems

In aerated soils, the oxidation of elevated $H_2$ mixing ratios is performed by Knallgas bacteria, whereas high-affinity $H_2$-oxidizing bacteria (HA-HOB) are able to oxidize much lower $H_2$ mixing ratios, such as atmospheric $H_2$ (*ca.* 0.53 ppmv, (Novelli *et al.*, 1999)).

*Knallgas bacteria*

Knallgas bacteria are aquatic and terrestrial obligate or facultative chemolithoautotrophs that use $O_2$ as a final electron acceptor to oxidize $H_2$ according to the Knallgas reaction:

$H_2 + O_2 \rightleftharpoons H_2O$ (4)

Model Knallgas microbes include *C. necator* (formerly *Ralstonia eutropha*), which possesses 4 [NiFe]-hydrogenases, including Group 1, 2, 3 and 5 (Burgdorf *et al.*, 2005a; Fritsch *et al.*, 2013; Schäfer *et al.*, 2013), yet not all Knallgas bacteria have several hydrogenases. While C*. necator* possesses a Group 5 [NiFe]-hydrogenase, it cannot oxidize atmospheric $H_2$ mixing ratios (Schäfer *et al.*, 2013) and is thus not a HA-HOB, as opposed to *Mycobacterium smegmatis* (Greening *et al.*, 2014a). When $H_2$ concentrations are sufficiently high, *C. necator* fixes $CO_2$ and oxidizes $H_2$ as its sole energy source using its Group 1 [NiFe]-hydrogenase (Burgdorf *et al.*, 2005b). The Group 3 [NiFe]-hydrogenase is rather used to maintain redox balance within the cell by coupling $H_2$ oxidation to $NAD^+$ reduction (Burgdorf *et al.*, 2005a). To avoid wasting energy expressing Group 1 and 3 hydrogenases constitutively when $H_2$ mixing ratios are low, *C. necator* also produces a Group 2 hydrogenase acting as a sensor regulating the transcription of structural and auxiliary genes of the two other hydrogenases (Burgdorf *et al.*, 2005a). Finally, the Group 5 hydrogenase generates ATP at a lower rate than the MBH (Group 1) but is entirely insensitive to $O_2$ (Schäfer *et al.*, 2013). Other Knallgas bacteria have similar $H_2$-uptake mechanisms and encompass various phylogenetic backgrounds (*e.g. Proteobacteria*,

*Firmicutes*, *Cyanobacteria*, *Chloroflexi*, *Acidobacteria*). Furthermore, many $N_2$-fixing microbes, namely actinorhizal plants symbiont *Frankia* spp. (phylum *Actinobacteria*) and phylum *Cyanobacteria*, can oxidize $H_2$ produced through $N_2$ fixation (Rasche & Arp, 1989). Various Knallgas bacteria are also known plant growth-promoting rhizobacteria (PGPR) (Golding *et al.*, 2012; Maimaiti *et al.*, 2007), while many more perform crucial C cycling steps. Several Knallgas bacteria can breakdown polymeric compounds. For instance, *Bacillus pseudofirmus* and *Bacillus thermotolerans* can reduce humic acids when oxidizing Fe(III) oxides (Ma *et al.*, 2012; Yang *et al.*, 2013), while *Burkholderia* spp. (phylum *Proteobacteria*) can breakdown lignin (Tian *et al.*, 2014). Furthermore, many methanotrophs—such as *Methylosinus trichosporium* (phylum *Proteobacteria*) and *Methylacidiphilum fumariolicum* (phylum *Verrucomicrobia*)—can oxidize $H_2$ and $CH_4$ under oxic conditions (Chen & Yoch, 1987; Mohammadi *et al.*, 2017) and are thus also Knallgas bacteria. Various Knallgas bacteria such as *Hydrogenomonas facilis* (Phylum *Proteobacteria*) (Bergmann *et al.*, 1958) and *C. necator* (Bowien & Kusian, 2002) can also fix $CO_2$ or oxidize CO (Weber & King, 2012).

*Atmospheric $H_2$ oxidation*

The oxidation of atmospheric $H_2$ has long been considered as originating from abiontic enzymes due to the failure to isolate microbes responsible for this process at low $H_2$ partial pressure (Conrad *et al.*, 1983a; Guo & Conrad, 2008). Bacteria responsible for the oxidation of $H_2$ at low mixing ratios have only been recently isolated (Constant *et al.*, 2008), hence their small representation in scientific literature. Recent genomic database surveys (Greening *et al.*, 2016) have shown that most microbes possessing Group 5 [NiFe]-hydrogenases are *Actinomycetes* such as *Streptomyces* sp. and *Mycobacterium* sp. Only a handful of high-affinity $H_2$-oxidizing microbes have been isolated and characterized so far, including *Rhodococcus equi* (phylum *Actinobacteria*) (Meredith *et al.*, 2014), various *Streptomyces* sp. (Constant *et al.*, 2008; Liot & Constant, 2016; Meredith *et al.*, 2014) and *Mycobacterium smegmatis* (Greening *et al.*, 2014a). In this regard, the role of Group 5 [NiFe]-hydrogenases within the cell is not understood. HA-HOB are mostly found in soils and they represent a major (70%) sink of atmospheric $H_2$

(Pieterse *et al.*, 2013). While atmospheric $H_2$ provides substantial amounts of energy, it is insufficient for bacterial growth (Constant *et al.*, 2010). So far, HA-HOB have been shown to use $H_2$ as an additional energy source for enhanced growth and survival (Greening *et al.*, 2014b; Meredith *et al.*, 2014) or within a mixotrophic lifestyle called survival mixotrophy (Liot & Constant, 2016). HA-HOB are also involved in C cycling. For instance, *Rhodococcus jostii* and *Rhodococcus equi* can degrade lignin and several other POM compounds (Janusz *et al.*, 2017; Tian *et al.*, 2014). *Streptomyces* species are also known to perform the breakdown of humic acids and even coal, thus contributing to POM cycling (Fodil *et al.*, 2011; Jaouadi *et al.*, 2014; Quigley *et al.*, 1989). Other *Actinobacteria* such as *Mycobacterium smegmatis* can also oxidize atmospheric CO and $H_2$ (King, 2003). Candidate taxa might also have the genetic potential to oxidize atmospheric $H_2$ and perform $CO_2$ fixation (Ji *et al.*, 2017).

## 1.6 Juxtaposition of the $H_2$ and C biogeochemical cycles in soils along theoretical $O_2$ gradients

A wide variety of HOM can hydrolyse DOM, but also take part in the breakdown of POM, including cellulose, hemicellulose, lignin, humic substances and polycyclic aromatic compounds (Martinez *et al.*, 2013) and also catalyze the oxidoreduction of CO, $CO_2$ and $CH_4$ (Fig. 1.4). HOM are able to switch between different modes of nutrition, both in terms of carbon and energy sources (Morita, 1999), which confers them great metabolic versatility when subject to varying C inputs. Many of these C cycling processes also lead to the generation of C sources usable by the broader microbial community, such as acetate. Moreover, in microenvironments rich in $H_2$—*e.g.* legumes rhizosphere, termites' gut, serpentinizing systems—metabolites released by HOM might also prime the decomposition and turnover of organic matter, namely in soils. Recent studies have shown that the exposure of soil to elevated $H_2$ concentrations representative of those found in nature (*i.e.* 10,000 ppmv) increased the amount of C sources used as well as the intensity of their consumption within a defined time period (Khdhiri *et al.*, 2017; Piché-Choquette *et al.*, 2016). However, it is not known if HOM involved in $H_2$ oxidation are responsible for this priming-effect-like response due to their mixotrophic growth strategy, or if positive interactions between HOM and heterotrophic non-HOM led to it. With this in

mind, it could be expected that other C cycling processes performed by HOM, such as decomposition, humification and hydrolysis, could also increase in $H_2$-rich ecosystems (Fig. 1.4).

As for trace gases, a study has shown that atmospheric CO and $H_2$ deposition velocities are positively correlated in both arable and forest soils (Yonemura *et al.*, 1999; Yonemura *et al.*, 2000). While the initial hypothesis was that diffusion was ultimately responsible for this response, a more recent study has shown that $H_2$, CO and $CH_4$ oxidation potential, by microbes, covary as well. Indeed, along an increasing gradient of $H_2$ exposure from 0.5 to 10,000 ppmv $H_2$, low affinity $H_2$ oxidation increased, $CO_2$ production decreased and high affinity $H_2$, CO and $CH_4$ oxidation decreased (Piché-Choquette *et al.*, 2018). $CO_2$ results were not published but were measured at the same time and with the same instrument as the other trace gases (Piché-Choquette *et al.*, 2018). Taken together, those studies imply that ecosystems rich in $H_2$ could be hubs of DOM and POM cycling; HOM oxidizing $H_2$ preferentially if available at elevated mixing ratios instead of using other trace gases.

While potential activity does not guarantee that *in situ* measurements will show the same behavior, examples of $H_2$-driven C cycling exist. Two studies have shown that $CO_2$ dynamics in soil systems continuously fed with air or air supplemented with $H_2$ change from net $CO_2$ production to net $CO_2$ fixation during $H_2$ exposure (Dong & Layzell, 2001; Stein *et al.*, 2005). A recent metagenomic survey has hypothesized similar outcomes in Antarctic desert soils (Ji *et al.*, 2017). Fixed $CO_2$ would increase the input of organic carbon in the system, which can then be distributed to the rest of the microbial community after cell death or secondary metabolites production. Responsible strains in these studies have not been identified, yet several Knallgas bacteria are known to oxidize $H_2$ and fix $CO_2$. While these hypotheses are applied to $H_2$-oxidizing, $CO_2$-fixing microbes, it is expected that $H_2$-oxidizing methanotrophs could also favour certain metabolic pathways in presence of $H_2$ by decreasing $CH_4$ oxidation and thus increasing net C input by methanogens (Fig. 1.4, (Piché-Choquette *et al.*, 2018)).

While HOM are involved in various C cycling processes, their relative contribution to each of them is unknown. Two different types of approach could be used to increase our understanding of $H_2$-C dynamics: community-level and individual-level approaches. As for community-level approaches, a good starting point would be to expose soil (*i.e.* microcosms, mesocosms or even fields) to a baseline mixing ratio of $H_2$ (*e.g.* 0.53 as control, (Novelli *et al.*, 1999)), as well as to lower and higher mixing ratios representative of natural conditions. Soil $H_2$ exposure could be provided by synthetic gas mixtures or through the manipulation of $N_2$-fixing, $H_2$-producing rhizobia ($Hup^+$ and $Hup^-$ phenotypes as isogenic control and treatment strains, respectively) recruited by legume plants. This would enable a direct comparison between $H_2$ availability and C turnover. As for individual-level approaches, three experiments could be performed: culturing novel HOM, single-cell sequencing of novel uncultured HOM or analyzing metagenome-assembled genomes (MAGs) of uncultured HOM. Traditional cultivation would allow for a more direct testing of HOM C-cycling capabilities, while culture-independent methods would give us more insight into their overall versatility. Furthermore, several $H_2$- and C-cycling processes seem to covary in soil, such as atmospheric $CH_4$ and elevated $H_2$ oxidation, yet it is not known if these processes are performed by the same microbes or simply within the same ecological niches. The three proposed individual-level methods could help solving this conundrum, all the while reducing current biases in genomic databases. For example, some taxa have more sequenced representatives and are consequently more represented in databases regardless of their involvement in the cycle (*e.g. Actinomycetales* and high-affinity $H_2$ oxidation). Further work in these directions would shed light on the importance of the overlap between the C and $H_2$ cycle at the species and ecosystem level as well as provide further insights into the overall biogeochemical potential of HOM. In turn, this knowledge could be used to develop techniques that take advantage of the $H_2$ hotspots in soils, *i.e.* for potential increases in C turnover or for agricultural benefits; namely in legumes rhizosphere rich in $H_2$ (Bhowmik *et al.*, 2017). Lastly, this review exemplifies that HOM, as a functional group, is a reductionist concept. That is, HOM should not be reduced to their $H_2$-oxidizing function; they are highly diverse

and many of them perform crucial ecosystem functions that could be more important at the ecosystem level than their individual energy-retrieving potential.

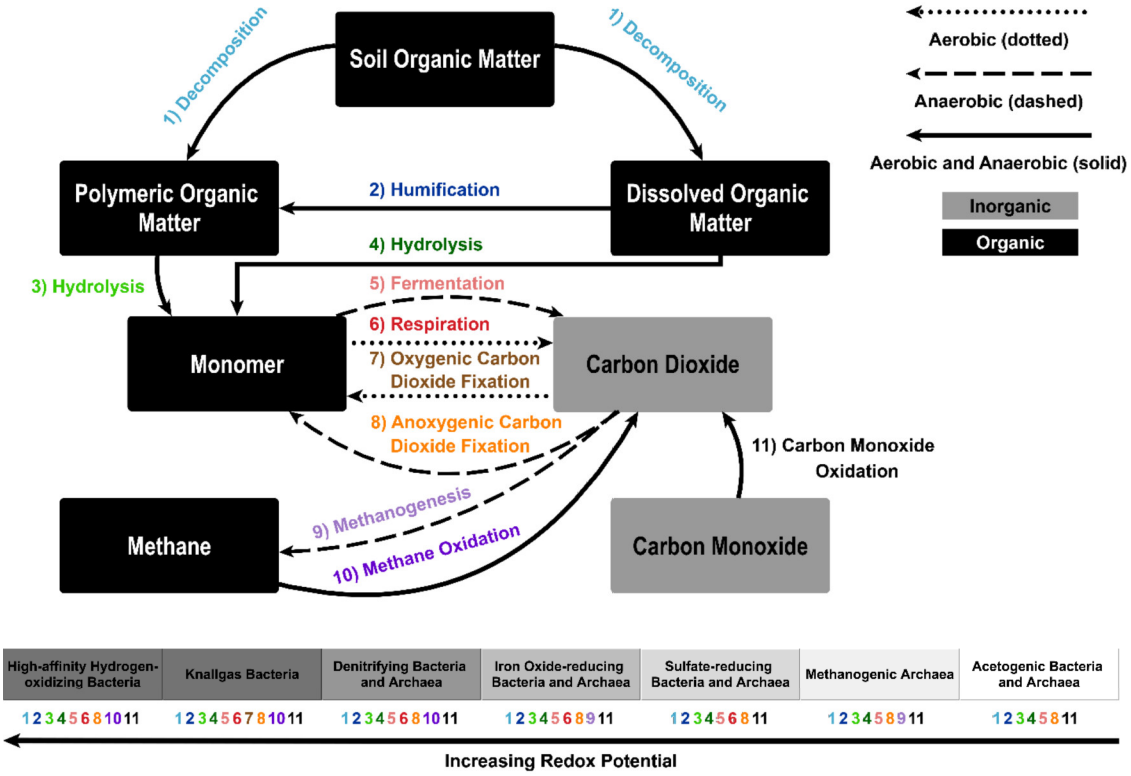



**Figure 1.4** **Juxtaposition of the carbon cycle and main $H_2$-oxidizing functional groups.HOM contribute to all key steps of the C cycle. All numbers below microbial functional groups consist of reactions performed by these groups. While the oxidation of $CH_4$ can be performed in anoxic ecosystems, the anaerobic oxidation of $CH_4$ (AOM) is not performed by HOM, only by their syntrophic bacterial partners (i.e. sulfate- or nitrate-reducing bacteria). $H_2$-utilizing processes occur in a thermodynamically-favored fashion according to available substrates, where $O_2$ is used first, followed by nitrate, iron oxides, sulfate and carbon dioxide, as show in the gradient at the bottom of the figure. The only key C cycling process missing in oxic ecosystems is methanogenesis, yet residual $CH_4$ diffuses from anoxic to oxic layers in various ecosystems. An exception to this is that non-hydrogenotrophic methanogens are also active in oxic layers (Angle *et al.*, 2017), thus providing $CH_4$ directly into oxic ecosystems.**

# 2 MISE EN CONTEXTE DU PROJET DE THESE

## 2.1 Problématique

La précédente revue de littérature a permis de faire l'état des connaissances sur l'écologie des organismes produisant ou oxydant l'$H_2$ et plus concrètement de mettre l'emphase sur le fait que l'$H_2$ représente une source d'énergie pratiquement universelle, c'est-à-dire qu'elle est employée ou produite par une multitude d'organismes taxonomiquement différents et ce, dans des écosystèmes bien distincts. De plus, cette revue a aussi montré l'implication des différents groupes de HOM dans le cycle du carbone. Le contexte plus précis de la présente thèse sera plutôt défini dans ce chapitre.

L'intensification de l'agriculture liée à la révolution verte a permis notamment d'augmenter la production alimentaire mondiale grâce à l'innovation du milieu agricole via l'irrigation, la sélection de cultivars plus productifs et l'ajout d'intrants à la terre (Tilman *et al.*, 2002). Bien que ces avancements aient permis une augmentation considérable de la démographie mondiale, ceux-ci ont aussi apportés un lot d'inconvénients, dont le lessivage d'engrais azotés et phosphorés menant à l'eutrophisation des milieux aquatiques, à une perte de biodiversité, à la pollution de tous les compartiments de la biosphère ainsi que l'érosion et la dégradation de la qualité des sols. Une des pratiques responsables de ces dommages collatéraux consiste en l'ajout d'engrais azotés, contenant de l'azote fixé chimiquement via le procédé d'Haber (Haber, 1905). Ce processus a d'une part permis aux Allemands de produire davantage d'explosifs durant la première guerre mondiale à la suite du blocus de leurs imports, mais d'autre part de pallier le manque d'azote fréquemment observé chez les plantes non légumineuses. En effet, l'azote, un élément essentiel à toute être vivant par sa présence dans les acides nucléiques et acides aminés, est commun dans l'atmosphère sous sa forme inerte $N_2$, mais peu abondant sous forme biodisponible dans la plupart des écosystèmes (Kuypers *et al.*, 2018).

Cependant, la fixation de l'azote se produit aussi biologiquement grâce à l'action de l'enzyme nitrogénase de diverses bactéries aquatiques (*e.g.* Cyanobactéries) ou terrestres (*e.g. Azotobacter*) libres, en association ou en symbiose avec des plantes (Hellriegel & Wilfarth, 1888). La fixation biologique de l'azote chez les plantes légumineuses (*i.e. Fabaceae*) se déroule dans les nodules fixateurs d'azote. L'échange de signaux entre la plante et la bactérie symbiotique (*e.g.* flavonoïdes et bétaïnes chez la plante, facteurs *Nod* chez les bactéries) mène au changement de la bactérie en bactéroïde et à l'altération de la morphologie des racines, ce qui aboutit en la formation de nodules (Mus *et al.*, 2016). Ces organes (*i.e.* nodules) sont microoxiques, ce qui permet ainsi à l'enzyme nitrogénase de fonctionner sans être irréversiblement inactivé par l'$O_2$. Cette symbiose mène donc à l'échange d'azote fixé ($NH_3$) par les bactéries fixatrices d'azote contre du carbone fixé par la plante et des acides aminés. Un aspect intéressant de ce processus est qu'une partie de ses bénéfices est transféré lors de la rotation des cultures, notamment celles pois-blé ou soja-maïs utilisées en Amérique.

La rotation des cultures en elle-même est bénéfique en agriculture, mais plusieurs études estiment que jusqu'à 91% des bénéfices de cette pratique ne proviendraient pas de l'enrichissement des sols en azote (Arcand *et al.*, 2014). Par ailleurs, les bénéfices agricoles associés à la fixation biologique de l'azote ne proviennent pas que de l'azote en lui-même non plus. En effet, l'enzyme nitrogénase catalyse plus exactement la réaction suivante : $N_2 + 8H^+ + 16\ ATP \rightarrow 2NH_3 + H_2 + 16ADP + 16\ P_i$ (Hunt *et al.*, 1988). Celui-ci produit donc obligatoirement de l'$H_2$ à raison d'une molécule d'$H_2$ par molécule d'azote fixé. À l'échelle d'un champ de légumineuses, cela se traduit à environ 5 000 L d'$H_2$ par jour par hectare de soja (Dong & Layzell, 2001). Selon cette même équipe de recherche, il est estimé que jusqu'à 240 000 L d'$H_2$ est produit par saison de culture, par hectare de légumineuses, ce qui correspond approximativement à un investissement de 5% de l'énergie nette issue de la photosynthèse de ces plantes (Dong *et al.*, 2003). Cet investissement n'est cependant pas une perte totale puisque certains nodules fixateurs d'azote, à génotype $Hup^+$ (*i.e.* Hydrogen uptake), possèdent une hydrogénase leur permettant de recycler cet $H_2$ de manière que celui-ci ne s'échappe pas ou presque pas de l'interface nodule-sol.

Contre-intuitivement, la sélection artificielle des cultivars de légumineuses au fil des années en fonction de leur rendement agricole a mené à une forte surreprésentation des nodules de génotype $Hup^-$ (Golding & Dong, 2010; Uratsu *et al.*, 1982), c'est-à-dire laissant l'$H_2$ s'échapper des nodules et causant ainsi une perte d'énergie supplémentaire au système plante-nodule. C'est à ce stade-ci qu'intervient l'autre aspect fertilisant associé à la fixation biologique de l'azote. Des études subséquentes ont montré que cette apparente inefficacité énergétique, voire « mauvaise symbiose » associée aux nodules $Hup^-$, serait en fait bénéfique pour la plante. D'abord, plusieurs études ont montré que l'$H_2$ diffusant des nodules $Hup^-$ est oxydé par les microorganismes du sol jusqu'à 3-4,5 cm de ceux-ci (La Favre & Focht, 1983). Des études en serre et en champs ont montré que l'exposition du sol à des concentrations d'$H_2$ semblables à celles produites par les nodules $Hup^-$ pouvait améliorer la croissance des plantes, soit ce qu'on appelle l'effet fertilisant de l'$H_2$ (Dong *et al.*, 2003) (Figure 2.1). Entre autres, même des plantes non légumineuses comme l'orge ont vu leur poids sec augmenter lorsque cultivées sur un sol prétraité à l'$H_2$. Cette étude a permis de séparer l'effet de l'$H_2$ de celui de la fixation de l'azote.

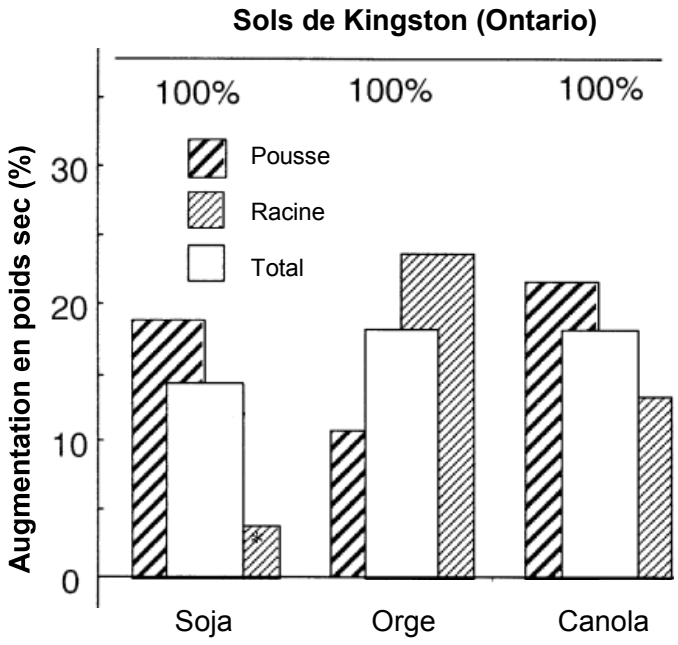



**Figure 2.1** **Augmentation du poids sec de plantes cultivées dans des sols prétraités à de l'air enrichi en $H_2$ comparé à des plantes cultivées dans des sols traités avec de l'air.Toutes les augmentations sont significatives sur un intervalle de confiance de 99% ou 95% (*) (où n = 24-72 plants). Adapté de Dong *et al.* (2003).**

D'autres études ont montré que des isolats bactériens provenant de sols traités à l'$H_2$ et ayant la capacité d'oxyder l'$H_2$ pouvaient améliorer la nodulation et la croissance des plantes par différents mécanismes. Ceux-ci incluent la production d'enzymes clivant des précurseurs de la phytohormone éthylène comme l'ACC (1-aminoocyclopropane-1-carboxylic acid) via l'ACC-désaminase, ou inhibant la synthèse de celle-ci via la rhizobitoxine (Figure 2.2) (Maimaiti *et al.*, 2007), sans toutefois savoir si l'$H_2$ induit ces mécanismes. D'autres études ont plutôt montré que quelques phylotypes bactériens avaient une abondance relative plus élevée ou plus faible dans des microcosmes de sol exposés à des concentrations d'$H_2$ similaires à celles produites par les nodules fixateurs d'azote (Osborne *et al.*, 2010; Stein *et al.*, 2005; Zhang *et al.*, 2009). En somme, ces études ont montré que l'$H_2$ a un effet fertilisant sur les plantes, que les HOM peuvent affecter la croissance des plantes et que les communautés microbiennes elles-mêmes sont affectées par l'$H_2$. Malgré ces constatations fort prometteuses, le manque de consensus entre ces études, probablement causé par la variété de techniques

employées, ainsi que la faible sensibilité de certaines techniques comme le T-RFLP, ne permettent pas d'avoir un portrait global de la situation et donc d'évaluer l'impact écologique de l'$H_2$ dans les sols. Le présent projet de thèse se concentrera donc sur cette problématique.

Inoculé avec *Variovorax paradoxus* JM63 Sans inoculum

**Figure 2.2 Effet de l'isolat *Variovorax paradoxus* JM63 sur l'élongation des racines de plantules de blé d'hiver après deux jours. Adapté de Maimaiti *et al.* (2007).**

## 2.2 But du projet

Considérant le peu de résultats disponibles concernant l'impact écologique de l'$H_2$ sur les communautés microbiennes du sol au moment de débuter ma thèse, le but du projet est plus exploratoire que ciblé. Celui-ci consiste à caractériser la réponse des communautés microbiennes du sol face à une exposition à des sources diffuses d'$H_2$ de concentrations comparables à celles retrouvées dans l'environnement.

## 2.3 Hypothèses

En plus de ce qui a été décrit dans la problématique, les hypothèses qui suivent s'appuient sur des assomptions supplémentaires, soient :

- L'ajout d'énergie dans un système, via la photosynthèse, altère la diversité bactérienne à l'intérieur de ce système (Horner-Devine *et al.*, 2003).

On assume donc qu'en tant qu'analogue aux produits de photosynthèse, l'$H_2$ représente une source d'énergie ajoutée au système qui pourrait potentiellement altérer la diversité microbienne de ce système (*i.e.* le sol). Présumant qu'il y aura un changement de biodiversité, on peut s'attendre conséquemment que, vu les liens diversité-fonctionnement des écosystèmes répertoriés dans plusieurs études (Delgado-Baquerizo *et al.*, 2016; Krause *et al.*, 2014b; Maron *et al.*, 2018; Midgley, 2012), les fonctions de l'écosystème dont la biodiversité est altérée par l'$H_2$ seront altérées à leur tour.

Outre ce lien dépendant directement de la première assomption, on sait que certaines fonctions comme l'oxydation d'$H_2$ pourraient être directement affectées par l'apport en $H_2$. De plus, une autre raison, indépendante de la première, nous pousse à croire que les fonctions de l'écosystème seront affectées par l'$H_2$, soit:

- Les microorganismes oxydant l'$H_2$ (HOM) en milieu oxique (*i.e.* bactéries Knallgas) se trouvent entre autres dans les sols et sont bien représentées chez les rhizobactéries, alors que les bactéries ayant une haute affinité envers l'$H_2$ (HA-HOB) ne représentent qu'approximativement 1% des microorganismes du sol (Khdhiri *et al.*, 2015). Les HOM retrouvées en milieu oxique, tout comme les autres HOM, sont très diverses d'un point de vue taxonomique (Greening *et al.*, 2016; Vignais & Billoud, 2007) et fonctionnel. Elles ne consistent donc pas en un groupe fonctionnel restreint à l'unique fonction d'oxydation d'$H_2$.

De plus, à des fins d'approximation de calculs à l'échelle locale, on assume que la production d'$H_2$ dans les sols, notamment via la fixation de l'azote, est effectuée par des sources ponctuelles (*e.g.* en provenance de chaque nodule fixateur d'azote ou autre source d'$H_2$). On sait que cet $H_2$ est très rapidement oxydé par les HOM situées à proximité (La Favre & Focht, 1983), donc on estime qu'entre le site de production et d'oxydation presque complète de l'$H_2$, c'est-à-dire le point d'équilibre entre l'$H_2$ diffusant

dans le sol via des sources externes ainsi que la production et l'oxydation de l'$H_2$, qu'il se forme un gradient de concentration d'$H_2$. Ces gradients d'$H_2$ pourraient ainsi impliquer des relations de dose-réponse entre la concentration d'$H_2$ et l'impact sur les communautés microbiennes du sol.

Considérant la réflexion décrite ci-haut, plusieurs hypothèses ont été émises par rapport à la problématique initiale :

1. *L'$H_2$ aura des effets directs sur la structure (i.e. composition, richesse spécifique, etc.) et la fonction des communautés microbiennes du sol*
    - L'abondance relative ou la biomasse des HOM sera affectée
    - Le métabolisme de l'$H_2$ sera affecté

2. *L'$H_2$ aura des effets indirects sur la structure et la fonction des communautés microbiennes du sol*
    - L'abondance relative ou la biomasse des microorganismes n'oxydant pas l'$H_2$ seront affectées via des interactions microbe-microbe entre HOM et autres microorganismes
    - Des fonctions microbiennes autres que le métabolisme de l'$H_2$ seront affectées par l'exposition à l'$H_2$

3. *Les gradients de concentration d'$H_2$ comparables à ceux retrouvés dans l'environnement mèneront à un continuum d'activités métaboliques modulées par l'$H_2$ ainsi qu'à une altération de la structure des communautés microbiennes proportionnellement à la concentration d'$H_2$*

## 2.4 **Objectifs**

Similairement, ces hypothèses se traduisent en les objectifs suivants :

1. *Définir les effets directs de l'$H_2$ sur la structure et la fonction des communautés microbiennes du sol*

2. *Définir les effets indirects de l'$H_2$ sur la structure et la fonction des communautés microbiennes du sol*

3. *Définir les relations de dose-réponse entre la concentration d'$H_2$ et les altérations des communautés microbiennes du sol*

Les chapitres suivants permettront donc d'élucider ces différents objectifs en parallèle, c'est-à-dire que chaque chapitre ne correspond pas nécessairement à un objectif, mais plutôt que chacun d'entre eux apporte des pistes de réponses à au moins un d'entre eux.

# 3 $H_2$-SATURATION OF HIGH AFFINITY $H_2$-OXIDIZING BACTERIA ALTERS THE ECOLOGICAL NICHE OF SOIL MICROORGANISMS UNEVENLY AMONG TAXONOMIC GROUPS

**Titre français** : La saturation des bactéries oxydant l'hydrogène avec haute affinité altère la niche écologique des microorganismes du sol de manière inégale parmi les groupes taxonomiques

**Article publié en ligne le 10 mars 2016 dans le journal PeerJ**

*DOI : 10.7717/peerj.1782* (Piché-Choquette *et al.*, 2016)

**Auteurs**

Sarah Piché-Choquette[1], Julien Tremblay[2], Susannah G. Tringe[3] et Philippe Constant[1✉]

[1]INRS-Institut Armand-Frappier, 531 boulevard des Prairies, Laval (Québec), Canada, H7V 1B7

[2]National Research Council Canada, 6100, avenue Royalmount, Montréal (Québec), Canada, H4P 2R2

[3]DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, California 94598, U.S.A.

✉ Auteur correspondant

**Contribution des auteurs**

Sarah Piché-Choquette : Conception et exécution des expériences en laboratoire, analyse de données et écriture de l'article

Julien Tremblay : Contrôle-qualité et analyse des données de séquençage

Susannah G. Tringe : Service de séquençage à haut débit

Philippe Constant : Conception des expériences, analyse de données, écriture de l'article

## 3.1 **Résumé en français**

Les communautés microbiennes du sol sont exposées continuellement à l'hydrogène ($H_2$) diffusant de l'atmosphère au sol. Les nodules fixateurs d'azote ($N_2$) représentent une microniche particulière dans les sols où l'$H_2$ peut atteindre des concentrations jusqu'à 20 000 fois plus élevées que celles de l'atmosphère global (0.53 ppmv). Dans cette étude, nous avons investigué l'impact de l'$H_2$ sur la structure des communautés microbiennes du sol à l'aide de microcosmes à flux dynamique simulant l'exposition du sol à l'$H_2$ provenant de l'atmosphère ou des nodules fixateurs d'azote. Les paramètres de cinétique biphasique gouvernant l'oxydation d'$H_2$ dans les sols ont changés drastiquement à la suite d'une forte exposition à l'$H_2$, correspondant ainsi à une diminution faible mais significative de l'activité des populations microbiennes oxydant l'$H_2$ avec haute affinité conjointement à un enrichissement ou à une activation des microorganismes démontrant une faible affinité pour l'$H_2$. En contraste avec de précédentes études ayant montré une réponse limitée à quelques espèces, l'abondance relative de 958 ribotypes bactériens, provenant de groupes taxonomiques variés plutôt que quelques groupes apparentés, a été influencée par l'exposition à l'$H_2$. De plus, des réseaux de corrélation ont montré des altérations importantes de la covariation entre les ribotypes en réponse à l'exposition d'$H_2$, suggérant que l'$H_2$ affecte les interactions microbe-microbes dans le sol. Pris dans leur ensemble, nos résultats démontrent que les environnements riches en $H_2$ imposent une influence directe sur les bactéries oxydant l'$H_2$ en plus d'un effet indirect sur les autres membres des communautés microbiennes.

## 3.2 **Abstract**

Soil microbial communities are continuously exposed to $H_2$ diffusing into the soil from the atmosphere. $N_2$-fixing nodules represent a peculiar microniche in soil where $H_2$ can reach concentrations up to 20,000 fold higher than in the global atmosphere (0.530 ppmv). In this study, we investigated the impact of $H_2$ exposure on soil bacterial community structure using dynamic microcosm chambers simulating soil $H_2$ exposure from the atmosphere and $N_2$-fixing nodules. Biphasic kinetic parameters governing $H_2$ oxidation activity in soil changed drastically upon elevated $H_2$ exposure, corresponding to a slight

but significant decay of high affinity $H_2$-oxidizing bacteria population, accompanied by an enrichment or activation of microorganisms displaying low-affinity for $H_2$. In contrast to previous studies that unveiled limited response by a few species, the relative abundance of 958 bacterial ribotypes distributed among various taxonomic groups, rather than a few distinct taxa, was influenced by $H_2$ exposure. Furthermore, correlation networks showed important alterations of ribotype covariation in response to $H_2$ exposure, suggesting that $H_2$ affects microbe-microbe interactions in soil. Taken together, our results demonstrate that $H_2$-rich environments exert a direct influence on soil $H_2$-oxidizing bacteria in addition to indirect effects on other members of the bacterial communities.

## 3.3 Introduction

Soil microbial communities are continuously exposed to molecular hydrogen ($H_2$). Trace levels of $H_2$ (0.530 ppmv) diffuse into the soil from the global atmosphere, yet higher concentrations can be found in the rhizosphere of $N_2$-fixing legumes (Constant *et al.*, 2009). Indeed, $H_2$ is an obligate by-product of nitrogenase in $N_2$-fixing free living or symbiotic bacteria, with an $H_2$ molecule produced for every reduced $N_2$ molecule (Hoffman *et al.*, 2009). It has been estimated that 240,000 L of $H_2$ are produced per hectare of legume crop over a growing season (Dong *et al.*, 2003). At the microscale level, there is a steep $H_2$ concentration gradient starting at about 20,000 ppmv at the soil-nodule interface and ending with sub-atmospheric levels a few centimeters away (La Favre & Focht, 1983; Rasche & Arp, 1989; Witty, 1991). Despite the high concentrations of $H_2$ found in legumes rhizosphere, a very small proportion escapes to the atmosphere due to $H_2$-oxidizing bacteria (HOB) thriving in soil. These microorganisms also play a vital role in the global budget of $H_2$, being responsible for about 80% (60 Tg $H_2$ $yr^{-1}$) of the global losses of this trace gas from the atmosphere (Constant *et al.*, 2009; Ehhalt & Rohrer, 2009).

Aerobic soil bacteria scavenging $H_2$ diffusing from the atmosphere and $N_2$-fixing nodules encompass a broad range of taxonomic groups. These microorganisms generally possess one or up to four different types of [NiFe]-hydrogenases catalyzing the

interconversion of $H_2$ into protons and electrons ($H_2 \leftrightarrow 2H^+ + 2e^-$). These hydrogenases are classified into five distinct phylogenetic groups, each displaying particular physiological roles (Constant & Hallenbeck, 2013; Vignais & Billoud, 2007). Hydrogenases encompassing group 1h are generally characterized by a high affinity towards $H_2$ ($_{(app)}K_m$ < 100 ppmv), conferring the ability to oxidize atmospheric $H_2$ (Constant *et al.*, 2010). Genome database mining of *hhyL* gene encoding the large subunit of high affinity hydrogenases are differentially distributed in *Actinobacteria*, with few representatives of *Proteobacteria*, *Acidobacteria* and *Chloroflexi* (Constant *et al.*, 2011a; Constant *et al.*, 2011b). The energy yield associated with the oxidation of atmospheric $H_2$ is insufficient to support chemolithotrophic growth and compelling experimental evidence suggest that $H_2$ supplies maintenance energy requirements and mixotrophic growth in high affinity HOB (Constant *et al.*, 2011a; Constant *et al.*, 2011b; Constant *et al.*, 2010; Greening *et al.*, 2015a; Greening *et al.*, 2014b). The distribution of the four other [NiFe]-hydrogenase groups is broader, including methanogenic archaea and *Cyanobacteria* (Vignais & Billoud, 2007). *Ralstonia eutropha* and *Bradyrhizobium japonicum*, part of the knallgas bacteria functional group, are able to scavenge $H_2$ diffusing from nodules in the presence of $O_2$, yet they are unable to use atmospheric $H_2$ due to the low-affinity ($_{(app)}K_m$ > 1000 ppmv) and high $H_2$ threshold concentration of their [NiFe]-hydrogenases (Conrad *et al.*, 1983a). Knallgas bacteria can use $H_2$ as a sole or supplementary energy source in chemolithotrophic or mixotrophic growth, respectively. The co-occurrence of these two sub-populations of HOB, as defined by substrate affinity, is supported by the biphasic kinetics governing $H_2$ oxidation activity in soil (Häring & Conrad, 1994).

Laboratory incubations simulating $H_2$ fluxes from $N_2$-fixing nodules demonstrated that soil bacterial community composition changed upon $H_2$ exposure (Osborne *et al.*, 2010; Stein *et al.*, 2005; Zhang *et al.*, 2009). However, methods previously used provided a low taxonomic resolution and coverage of bacterial communities responding to $H_2$ exposure. Here, we revisited these experiments using a combination of high-throughput sequencing of the bacterial 16S rRNA gene and an $H_2$ metabolism analysis to compare microbial community structure in dynamic microcosm chambers simulating soil $H_2$ exposure from the atmosphere and from $N_2$-fixing nodules, corresponding to unsaturating and saturating

$H_2$ concentration for high affinity hydrogenases, respectively. Elevated $H_2$ exposure was expected to activate and enrich HOB, consequently leading to indirect impacts on the whole soil bacterial community through competitive and synergistic microbe-microbe interactions.

## 3.4 Material and Methods

### 3.4.1 Soil Sample

Soil sample was collected in an agricultural land located in St. Claude (Québec, Canada) on the south shore of the St. Lawrence River (45.6809°N, -71.9969°W). The field is managed with fallow, potatoes and maize crop rotation. Potato seedlings (approximately 10 cm height) were present on the site during sampling in July 2013. The top layer of the A-horizon (0-10 cm depth) was collected, stored in plastic bags at 4°C and processed within a week. Soil was air-dried for 48 hours in the laboratory and sieved (2 mm mesh size) through a vibratory sieve shaker AS 200® (Retsch GmbH, Haan, Germany) before preparation of soil microcosms. Soil was classified as sandy clay loam according to soil textural class parameters identified with the hydrometer method (Bouyoucos, 1936). Soil pH was determined with 1:2 soil-water suspensions with an Accumet pH-meter (Fisher Scientific, Hampton, NH). Total carbon (3.1 ±0.3%) and total nitrogen (0.3 ±0.0%) content were determined using an elemental combustion system using the protocol described in (Khdhiri *et al.*, 2015).

### 3.4.2 Controlled $H_2$ exposure in dynamic soil microcosm chambers

Dynamic microcosm chambers were designed to expose soil to controlled levels of $H_2$ (Figure 3.7). Microcosm chambers consisted of 0.9 L cell culture flasks (Corning, Tewksbury, MA) equipped with a rubber stopper fitted with two 1/8" outside diameter PTFE (Teflon) tubes: the first supplied gas mixture to the microcosm chamber and the second was vented to the atmosphere. Synthetic gas mixtures supplied the microcosm chambers at a flowrate of 40 ml $min^{-1}$, resulting in a dynamic headspace with a residence time of approximately 22 minutes in the enclosures. Gas mixtures were bubbled in water before entering in microcosms to prevent soil dryness. Two different $H_2$ treatments were

applied in parallel incubations. The first treatment (designated microcosms $eH_2$(a) and $eH_2$(b)), named elevated $H_2$ ($eH_2$), consisted of exposing soil to a dynamic headspace comprising 525 ppmv $H_2$ in synthetic air, simulating $H_2$ concentrations detected around $N_2$-fixing nodules in soil (Hunt & Layzell, 1993). The second was a control treatment (designated microcosms $aH_2$(a) and $aH_2$(b)), named atmospheric $H_2$ ($aH_2$), where soil was exposed to a dynamic headspace comprising 0.54 ppmv $H_2$ in synthetic air, representing $H_2$ concentrations found in the atmosphere. Both treatments were replicated, resulting in four soil microcosms in total. Each microcosm chamber contained 200 $g_{(dw)}$ soil at the beginning of the incubation period. This amount of soil ensured sufficient material for the monitoring of microbiological and physicochemical variables throughout the incubation period, while avoiding diffusion limitation of the $H_2$ soil uptake rate measurements. Diffusion limitation was precluded since preliminary experiments using the same microcosm setup showed proportional $H_2$ uptake activity as a function of the amount of soil in the chamber using 150, 200 and 250 $g_{(dw)}$ soil samples (raw data file). Soil water content was adjusted to 20% water holding capacity before incubation for 10 days at 28°C in the dark. Synthetic gas mixture supply was continuously maintained, with the exception of routine soil subsamples collection and high affinity $H_2$ oxidation rate measurements (see below). A decrease in soil pH was observed over the course of the incubation (from pH 5.9 ±0.2 to 5.1 ±0.1) in all microcosms, without distinction between $aH_2$ and $eH_2$ treatments. Soil moisture was monitored using standard gravimetric method and maintained at 26 ±5% throughout the incubation period. Blank (empty) microcosms incubated prior to the experiment did not show any $H_2$ oxidation or production activity.

### 3.4.3 $H_2$ uptake activity

High affinity $H_2$ oxidation activity was routinely monitored throughout the incubation of the soil microcosms. Briefly, microcosms were disconnected from their respective gas supply, flushed 5 minutes with a synthetic gas mixture (0.54 ppmv $H_2$) and tightly closed with rubber septum caps. The 5-min flush was shown sufficient to avoid residual $H_2$ degassing that would otherwise have led to an underestimation of $H_2$ oxidation rate. A defined volume of $H_2$ gas mixture (525 ± 10 ppmv $H_2$ GST-Welco, Pennsylvania, U.S.A.) was injected to obtain an initial concentration of approximately 3 ppmv in the static headspace.

Decrease of the $H_2$ mixing ratio was monitored as a function of time by analyzing aliquots (10 ml) of the headspace air in a ta3000R gas chromatograph equipped with a reduction gas detector (Ametek Process Instruments®, Delaware, U.S.A.) as previously described (Khdhiri *et al.*, 2015). Considering the low level of $H_2$ added in the headspace, $H_2$ uptake reflected the activity of high affinity HOB, knallgas bacteria showing low affinity for $H_2$ being unable to use these trace amounts of $H_2$ (Conrad *et al.*, 1983a). The biphasic kinetic parameters governing $H_2$ oxidation activity in soil (*i.e.* $_{(app)}K_m$ and $_{(app)}V_{max}$) were measured at the end of the 10-day incubation period after the addition of specified amounts of pure $H_2$ into the headspace of the microcosms, as previously described (Schuler & Conrad, 1990).

### 3.4.4 DNA extraction, qPCR and high-throughput sequencing of PCR-amplified bacterial 16S rRNA gene

Soil subsamples (approximately 10 g per subsample, without replacement) were collected in the four microcosms after 0, 1, 3, 5, 7 and 10 incubation days to investigate the taxonomic structure of microbial communities. The extraction of total genomic DNA was performed using the FastDNA SPIN kit for Soil® (MP Biomedicals, Solon, OH, U.S.A.). DNA was eluted in 100 µL nuclease-free water. Quality of the DNA was then examined on agarose gels and samples were quantified using Quantifluor dsDNA System® (Promega, Fitchburg, WI). The V4 region of bacterial 16S rRNA gene was PCR amplified (Table 3.1) and sequenced on an Illumina MiSeq 2000 instrument as multiplexed libraries to generate paired-end reads (2 x 250bp). Quantification of bacterial 16S rRNA and *hhyL* genes was done by qPCR following (Khdhiri *et al.*, 2015).

### 3.4.5 Quality control of sequencing reads

Our sequencing analysis pipeline using QIIME 1.8.0 (Kuczynski *et al.*, 2011) software was based on the Itagger pipeline (Tremblay *et al.*, 2015). V4 region Illumina sequencing reads were first processed using the *split_library* function implemented in QIIME for samples demultiplexing and forward primer removal. The assembly of paired-end reads was performed using FLASH software version 1.2.10 (Magoc & Salzberg, 2011). The software DUK (Li, 2011), a Kmer-based sequence matching tool, was then used to

remove common sequencing contaminants as well as PhiX sequences added as control. Afterwards, sequences were trimmed to remove low quality ends and a general quality control was also applied. All reads containing at least a single N (ambiguous base), an average quality score of less than 30 or more than 15 bases with a quality score under 20 were discarded. The minimum and maximum read length were respectively set to 200 and 300 bases for the assembled paired-reads. Reverse primers were also removed using the *truncate_reverse_primer* function implemented in QIIME. Raw sequences were deposited in the Sequence Read Archive of the National Center for Biotechnology Information under the Bioproject PRJNA295403.

### 3.4.6 Clustering and taxonomic identification of sequencing reads

Processed 2,554,543 high quality reads were clustered into OTUs (Operational Taxonomic Units) using the function *pick_otus* implemented in QIIME along with the USEARCH (Edgar, 2010) clustering algorithm. Sequences dereplication was performed at 100% sequence identity and clusters were formed by denoising unique sequences at 99% identity. Singletons were removed to avoid diversity bias. Remaining clusters were filtered using UCHIME (Edgar *et al.*, 2011) chimera filter in *de novo* mode, followed by a reference-based filter step against the Gold reference database. These 2 chimera scanning steps allow a better removal of chimeras by removing sequences flagged as chimeras by any of the two methods. Chimera-checked clusters were then clustered into OTUs at 97% identity threshold also using the function *pick_otus* with the USEARCH software. Taxonomic identification of OTUs was performed using the *assign_taxonomy* function along with the naïve Bayesian Ribosomal Database Project (RDP) classifier 2.7 (Wang *et al.*, 2007). Taxonomy was assigned based on the Greengenes taxonomy with the 16S rRNA Greengenes reference database (DeSantis *et al.*, 2006). A rarefaction step was applied to the OTU libraries to standardize all libraries, by random subsampling, to the lowest amount of sequences (74,316 reads) to avoid bias introduced due to unequal sequencing efforts of the samples. This rarefied OTU table comprising the frequency distribution of the OTUs in each sample was used in downstream statistical and co-occurrence network analyses.

### 3.4.7 Statistical analyses

Statistical analyses were performed using the software R version 3.0.2 (R Development Core Team, 2017). Pairwise comparison of $H_2$ uptake activities in soil microcosms incubated under $aH_2$ and $eH_2$ exposure treatments was tested using one-way analysis of variance (ANOVA) and *post hoc* Tukey test. Shapiro-Wilk normality test was applied to confirm normal distribution of data before the ANOVA. The package "nlme" was used to compute nonlinear regression for modeling the first-order decay of *hhyL* gene abundance determined by qPCR. Discrimination of the samples according to their ribotyping profile was performed with the package "vegan". Hellinger transformation of the OTU table (97% identity threshold) was computed due to the presence of zeros in the OTU table (Legendre & Gallagher, 2001). Euclidean distance matrix was used to generate an UPGMA agglomerative clustering of the samples. Identification of statistically different clusters was done by performing 999 permutations of the OTU table dataset separately across the samples and comparing the observed similarity score of each cluster against the expected values under the null hypothesis using the similarity profile tool (SIMPROF) implemented in the package "clustsig" 1.1 (Clarke *et al.*, 2008). Pairwise comparison of relative abundance of OTUs having a higher contribution than average to explain the two dimensions of the PCA space in soil microcosms incubated under $aH_2$ and $eH_2$ exposure treatments was tested using Kruskal-Wallis and *post hoc* Tukey test. Pairwise comparison of relative abundance of all OTUs in soil microcosms incubated under $aH_2$ and $eH_2$ exposure treatments was tested using the likelihood ratio test using the package "edgeR". Covariation among OTUs during the incubation under controlled $H_2$ levels was analyzed by correlation networks using the package "WGCNA" 1.41 (Langfelder & Horvath 2008). A detailed methodology for the computation of correlation networks is provided in supplementary information (Text 3.8.1, 3.8.2 and 3.8.3)

## 3.5 Results and discussion

### 3.5.1 Impact of $H_2$ exposure on the distribution and activity of HOB

Agricultural soil microcosms were incubated under two different $H_2$ level treatments. The first treatment simulated soil exposure to atmospheric level of $H_2$ ($aH_2$; 0.54 ppmv), while the second simulated high affinity hydrogenases substrate saturation ($eH_2$; 525 ppmv). High affinity $H_2$ oxidation rates were not significantly different (ANOVA, $P > 0.05$) between the two pairs of microcosms at the beginning of the incubation (Figure 3.1).

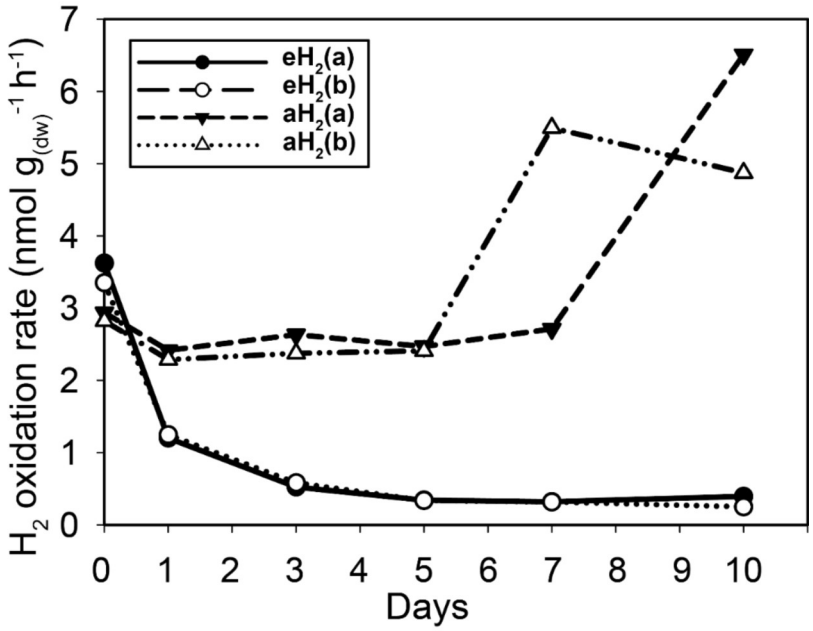



**Figure 3.1** **Time series of the high-affinity $H_2$ oxidation rate measured in soil microcosms exposed to $aH_2$ or $eH_2$ throughout the incubation period.**

The influence of $H_2$ exposure could already be observed after 24 hours. From hour 24 to day 10, a decline of high affinity $H_2$ uptake rate was observed in microcosms exposed to $eH_2$, while an increase of this uptake rate was measured in microcosms exposed to $aH_2$ levels. This trend was maintained over the course of the incubation, with an oxidation rate of 6.2 ± 1.9 and 0.3 ± 0.1 nmol $g_{(dw)}^{-1}$ $h^{-1}$ in $aH_2$ and $eH_2$ treatments at the end of the experiment, respectively. This response of $H_2$ oxidation rate was accompanied by an alteration in the abundance of presumptive high affinity HOB in soil. The abundance of *hhyL* gene determined by qPCR did not change over the course of the incubation under

$aH_2$ (exponential regression, P = 0.06), while a slight but significant exponential decay (exponential regression, P = 0.02) was observed under $eH_2$ (Figure 3.2A). Indeed, the physiological role of high affinity hydrogenases differs within various taxonomic groups of bacteria. In streptomycetes, the enzyme is primarily expressed in spores to support a seed bank under a mixotrophic survival energy mode (Constant *et al.*, 2010; Constant *et al.*, 2008; Liot & Constant, 2016), while *Mycobacterium* express the enzyme in the exponential and stationary phases for mixotrophic growth as well as survival (Berney & Cook, 2010; Greening *et al.*, 2015b; Greening *et al.*, 2014b). Under axenic cultivation conditions, $eH_2$ level would be expected to favour an increase of *Mycobacterium* biomass and persistence of streptomycetes spores. It should be noted that our qPCR data cannot differentiate between active and inactive cells but the decreasing trend of *hhyL* copy number in soil exposed to $eH_2$ suggests that $H_2$ was not sufficient to promote growth and persistence of high affinity HOB.

$H_2$ soil exposure exerted a significant impact on the kinetic parameters governing $H_2$ oxidation activity in soil. At the end of the incubation, microcosms exposed to $aH_2$ level displayed $_{(app)}K_m$ of 40 ± 5 ppmv (Figure 3.3A), while microcosms exposed to $eH_2$ showed lower affinity towards $H_2$ with an $_{(app)}K_m$ of 838 ± 163 ppmv (Figure 3.3B). Similar observations were obtained by Dong and Layzell (Dong & Layzell, 2001), where soil exposed to elevated levels of $H_2$ (600 ppmv) displayed an $_{(app)}K_m$ of 1028 ppmv $H_2$, while soil exposed to low $H_2$ level (0.55 ppmv) was characterized by an $_{(app)}K_m$ of 40 ppmv. The biphasic kinetic parameters characteristics of low and high affinity HOB were also computed using Eadie-Hofstee plots to highlight high- and low-affinity $H_2$ uptake activities in both $H_2$ treatments. The $H_2$ oxidation activity was mainly catalyzed by bacteria demonstrating intermediate and high-affinity in $aH_2$ treatment (Figure 3.3A, insert), while $H_2$-oxidation activity by bacteria displaying low- and high-affinity for $H_2$ were detectable in the $eH_2$ treatment (Figure 3.3B, insert).

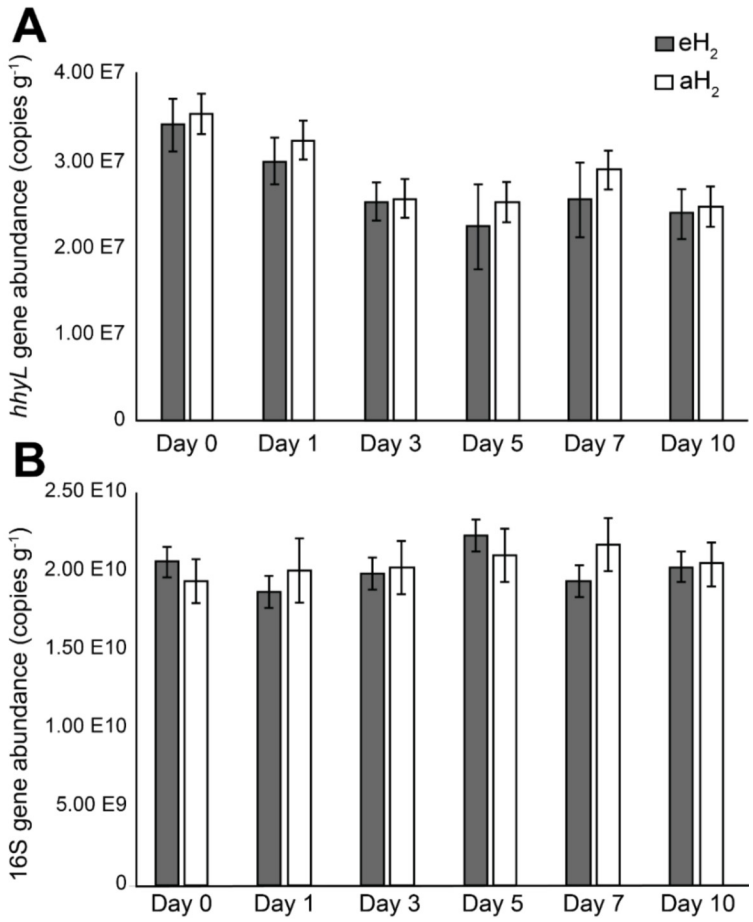



**Figure 3.2** **Time series of (a) hhyL and (b) 16S rRNA gene abundance in soil as determined by qPCR.**

Low $H_2$ exposure supported the metabolism of intermediate- to high-affinity HOB, such as some *Actinobacteria* species of *Streptomyces*, *Rhodococcus* and *Mycobacterium* (Berney *et al.*, 2014; Constant *et al.*, 2011b; Golding *et al.*, 2012; Liot & Constant, 2016; Meredith *et al.*, 2014; Schäfer *et al.*, 2013). Triggering of low-affinity $H_2$ oxidation activity under $eH_2$ treatment might be explained by knallgas bacteria such as *Proteobacteria* species encompassing *Ralstonia*, *Variovorax* and *Bradyrhizobium* using $H_2$ for mixotrophic growth (Rittenberg & Goodman, 1969), as they are known to exhibit a $H_2$ uptake threshold ranging from 1 to 200 ppmv $H_2$ (Conrad, 1996).

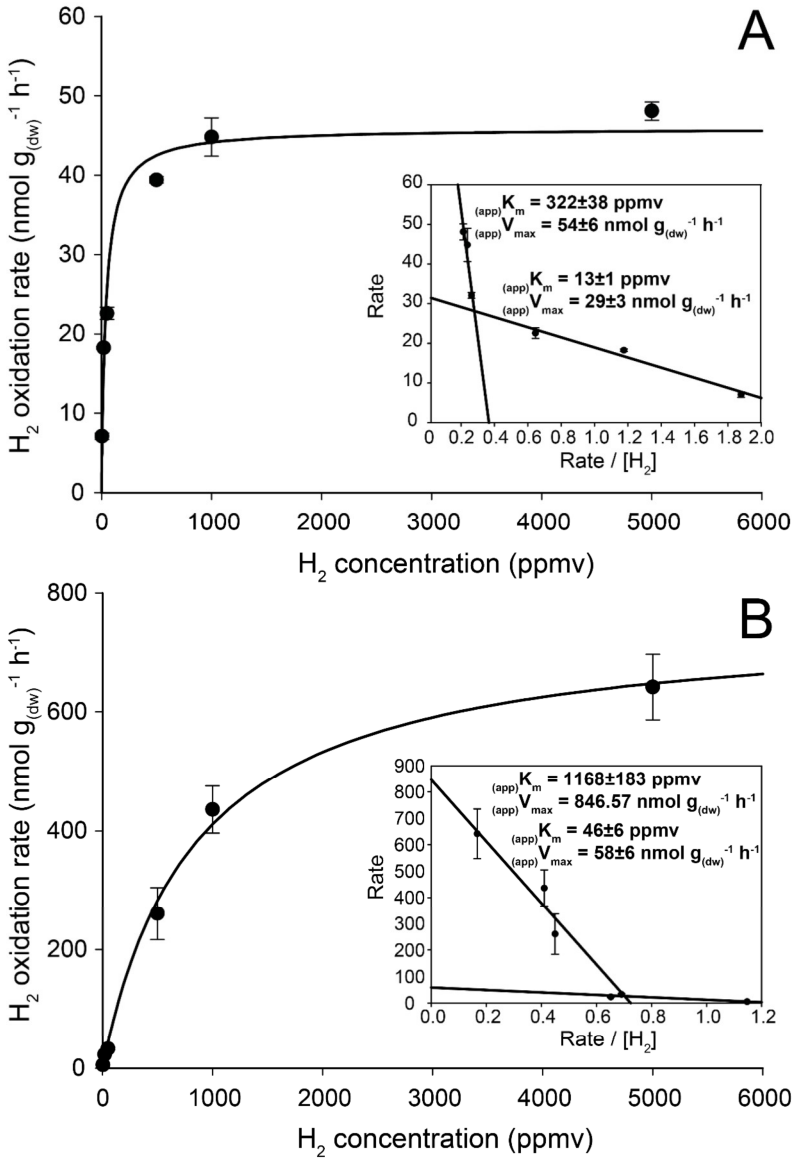



**Figure 3.3** **Kinetic parameters governing $H_2$ oxidation activity in soil microcosms incubated under (a) a$H_2$ and (b) e$H_2$ exposure. Michaelis-Menten graphs are presented with Eadie-Hofstee plots highlighting the biphasic kinetic of the reaction in the inserts.**

### 3.5.2 Impact of $H_2$ exposure on soil bacterial community structure

Soil subsamples were collected in the microcosms during the incubation period to evaluate the temporal variation of soil bacterial community structure. $H_2$ treatment did not influence the abundance of 16S rRNA gene in soil (Figure 3.2B). This is in contrast with

another study that reported more than twofold increase of the amount of bacterial cells upon elevated $H_2$ exposure, but a mistake in the reported $H_2$ concentration units by the authors (*i.e.* the use of a gas mixture of 2,000 ml $H_2$ per L of synthetic air) impairs a sound comparison with the present investigation (Stein *et al.*, 2005). High-throughput sequencing of PCR-amplified 16S rRNA gene unveiled that bacterial communities were dominated by *Proteobacteria* (34%), *Acidobacteria* (20%), *Actinobacteria* (10%), *Verrucomicrobia* (8%) and *Bacteroidetes* (5%) (Figure 3.8).

Species richness was not affected by $H_2$ exposure (ANOVA, $P > 0.05$), with Shannon indices of 9.38 ±0.05 and 9.34 ±0.17 for a$H_2$ and e$H_2$ treatment respectively. An UPGMA hierarchical clustering analysis was computed to compare bacterial community profiles from soil subsamples collected in microcosms exposed to different $H_2$ concentrations. The four microcosms encompassed the same cluster before the incubation, indicating high similarity of their initial ribotyping profile (Figure 3.4A). No coherent impact of $H_2$ exposure on microbial community structure was observed after 1 and 3 incubation days. Ribotyping profiles obtained for sub-samples collected after 5 and 7 incubation days were separated with confidence in different clusters, suggesting a response of microbial communities to $H_2$ exposure. This discrimination was however transient as ribotyping profiles corresponding to the last incubation day could not be discriminated according to their respective $H_2$ treatments (Figure 3.4A). The transient response was likely due to the fact that $H_2$ alone was not sufficient to maintain a sustained change in microbial community structure. Decrease of complex carbon sources and nutrients used along with $H_2$ for mixotrophic growth in soil is a potential explanation for the convergence of the ribotyping profiles after 7 incubations days.

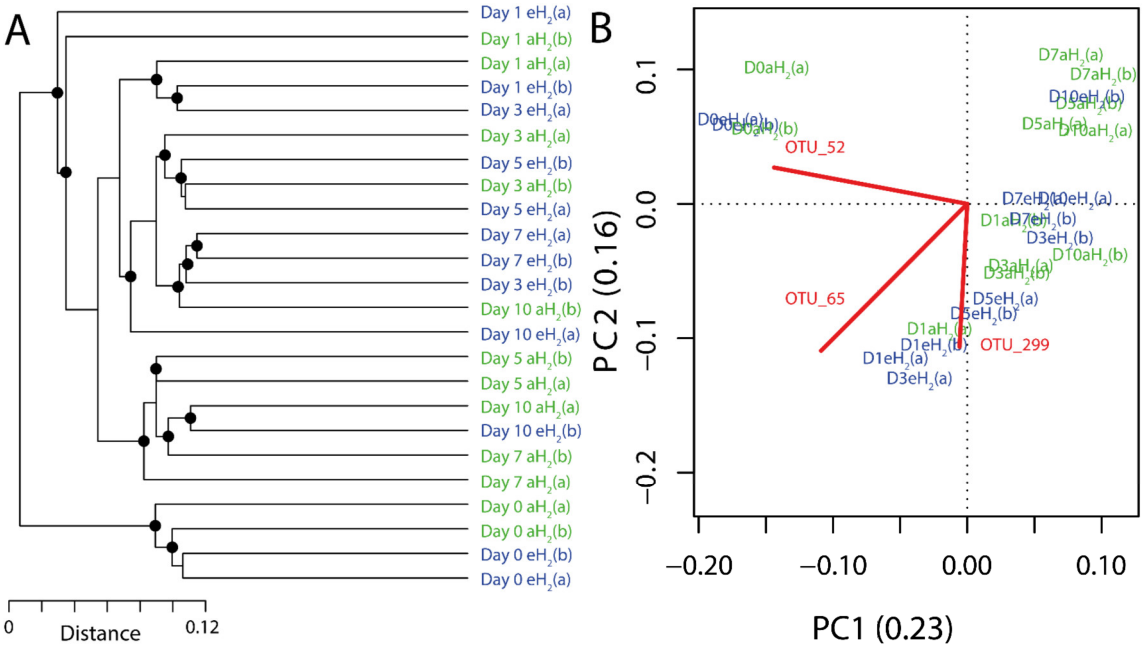



**Figure 3.4** **Influence of $H_2$ exposure on bacterial ribotyping profiles. (a) UPGMA agglomerative clustering of soil samples derived from a matrix of Euclidean distance calculated after Hellinger transformation of OTU (97% identity) absolute abundance in soil microcosms exposed to $eH_2$ and $aH_2$ throughout a 10-day incubation period. The black circles ● represent significant nodes ($P \leq 0.05$). (b) Principal component analysis showing the distribution of soil subsamples in a reduced space defined by the relative abundance of 16S rRNA gene sequences classified at the OTU level (97% identity). Only the OTUs (represented by red lines) having a higher contribution than average to explain the two dimensions of the PCA space are shown to facilitate visualization of the analysis. Soil subsamples collected in microcosms exposed to $eH_2$ and $aH_2$ are shown in blue and green, respectively.**

A PCA was computed to identify OTUs contributing to the clusterization of ribotyping profiles. The ordination space defined by the first two components explained 38.6% of the variation observed (Figure 3.4B). Three OTUs defined an important proportion of the reduced space represented by both axes (Figure 3.4B and 3.5). The relative abundance of OTU 52 (classified as a member of the class *Betaproteobacteria*) contributed to distinguish the four soil subsamples collected before the incubation from the 20 other subsamples collected after 1, 3, 5, 7 and 10 incubation days along the first axis as it decreased after the beginning of the incubation period (Kruskal-Wallis, $P < 0.05$). Two OTUs classified as members of the order *Bacillales* (OTUs 65 and 299) contributed to discriminate both $H_2$ treatments in the ordination space after 5 and 7 incubation days (Kruskal-Wallis, $P < 0.05$). These OTUs were more abundant in microcosms exposed to

$eH_2$ until day 10, where they reached similar relative abundance in $aH_2$ and $eH_2$ treatments. The transient response of these OTUs in $eH_2$ treatment could be a direct consequence of $H_2$ exposure as representatives of the genus *Bacillus* possess putative membrane-bound type 1a and 1d [NiFe]-hydrogenase (Greening *et al.*, 2016).

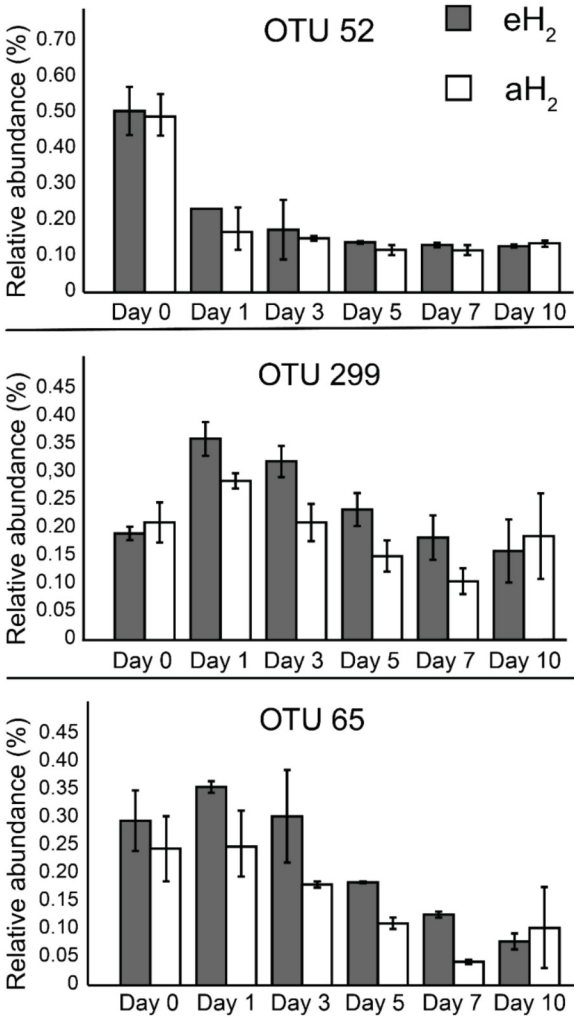



**Figure 3.5** **Time series of the relative abundance of the 3 OTUs having a higher contribution than average to explain the two dimensions of the PCA space in soil microcosms. The average and standard deviation measured in replicated microcosms are represented. The closest taxonomic affiliations of OTUs 52, 299 and 65 are, respectively, the bacterial order MND1 (Betaproteobacteria), genus *Bacillus* (Firmicutes) and the species *Bacillus cereus* (Firmicutes).**

Conflicting results were obtained in two different studies reporting the impact of $H_2$ soil exposure on bacterial community structure based on 16S rRNA terminal restriction fragment profiles analysis. $H_2$ exerted no incidence on ribotyping profile in soil exposed to 250 nmol $H_2$ $cm^{-3}$ $h^{-1}$ (500 ppmv $H_2$ in artificial air, added at 45 ml $min^{-1}$), with a single ribotype related to *Mycobacterium* increased in soils upon elevated $H_2$ exposure (Osborne *et al.*, 2010). On the other hand, an exposure rate of 33 nmol $cm^{-3}$ $h^{-1}$ (79 ppmv $H_2$ at 100 $cm^3$ $min^{-1}$) exerted a significant influence on bacterial community profile, with an increase of T-RFLP peaks belonging to *γ-Proteobacteria* and a decrease of peaks belonging to *Actinobacteria* and *α-Proteobacteria* upon elevated $H_2$ exposure (Zhang *et al.*, 2009). A likelihood ratio analysis, fitting a negative binomial generalized log-linear model to the sequencing data (McCarthy *et al.*, 2012; Robinson *et al.*, 2010), unveiled that distribution of 958 OTUs was influenced by $H_2$ exposure. The relative abundance of OTUs responding to $H_2$ treatment ranged between 0.001 and 1.8%, suggesting an incidence of $H_2$ on members of the rare biosphere and abundant taxa. The influence of $H_2$ was uneven among different taxonomic groups as different representatives of the same taxa (*i.e.* OTUs classified at the phylum and the order taxonomic levels) were found to be advantaged or disfavored in response to e$H_2$ exposure (Table 3.1). Even though the uneven response among taxonomic groups impairs prediction of $H_2$ exposure on metabolic functions in soil (Langille *et al.*, 2013), this observation is sufficient to demonstrate that previous investigations relying on low-resolution community fingerprinting techniques have considerably underestimated the response of soil microbial communities to $H_2$ exposure.

**Table 3.1** **Summary of OTUs showing different relative abundance in $eH_2$ and $aH_2$ treatments (Likelihood ratio test, P < 0.05). Altogether, 406 OTUs were more abundant in $eH_2$ and 552 were more abundant in $aH_2$. The $eH_2$ and $aH_2$ rows indicate the treatment in which the identified phylotypes are more abundant. A single or the two most abundant OTUs are identified for each phylum. A list of the 958 OTUs whose distribution was influenced by $H_2$ treatments is provided in the raw data file accompanying the article.**

| Treatments | Phyla | Most abundant OTU (order level) |
|---|---|---|
| $eH_2$ | *Proteobacteria* (27.6%) | *Myxococcales* |
| | *Planctomycetes* (10.6%) | *Gemmatales*<br>*Phycisphaerales* |
| | *Bacteroidetes* (9.6%) | *Sphingobacteriales* |
| | *Chloroflexi* (9.1%) | *Ktedonobacteria* |
| | *Acidobacteria* (7.1%) | *Acidobacteriales*<br>*Solibacterales* |
| | *Verrucomicrobia* (6.2%) | *Verrucomicrobiales* |
| | *Actinobacteria* (5.9%) | *Actinomycetales*<br>*Solirubrobacterales* |
| | *Elusimicrobia* (3.0%) | *FAC88* |
| | *Gemmatimonadetes* (2.7%) | *Gemmatimonadales* |
| | *Firmicutes* (1.5%) | *Clostridiales*<br>*Bacillales* |
| | *Archaea* (1.0%) | *Methanobacteriales* |
| | *Armatimonadetes* (0.7%) | *CH21* |
| | *Chlorobi* (0.2%) | *SM1B09* |
| | Others (14.8%) | - |
| $aH_2$ | *Proteobacteria* (33.9%) | *Rhodospirillales*<br>*Myxococcales* |
| | *Acidobacteria* (11.4%) | *Solibacterales* |
| | *Planctomycetes* (11.1%) | *Gemmatales* |
| | *Chloroflexi* (6.2%) | *Ktedonobacteria*<br>*A4b* |
| | *Bacteroidetes* (5.1%) | *Sphingobacteriales* |
| | *Actinobacteria* (4.3%) | *Actinomycetales*<br>*Acidimicrobiales* |
| | *Firmicutes* (3.8%) | *Clostridiales* |
| | *Verrucomicrobia* (3.4%) | *Spartobacteriales* |
| | *Gemmatimonadetes* (2.7%) | *Gemmatimonadales* |
| | *Archaea* (1.8%) | *SD-NA (Crenarchaeota)* |
| | *Elusimicrobia* (1.4%) | *Elusimicrobiales*<br>*MVP-88* |
| | *Armatimonadetes* (1.3%) | *Chthonomonadales*<br>*CH21* |
| | *Chlorobi* (0.9%) | *SM1B09* |
| | Others (12.7%) | - |

### 3.5.3 Impact of $H_2$ exposure on the co-occurrence of OTUs

Correlation networks were computed to investigate the impact of $H_2$ exposure on the covariation of OTUs throughout the incubation period and identify OTUs for which the

distribution is influenced by $H_2$. Two separate networks were computed: the first for elevated $H_2$ exposure ($eH_2$ network) and a second for low $H_2$ exposure ($aH_2$ network). Both networks contained 43 modules and module preservation statistics demonstrated that module composition significantly changed across $H_2$ treatments (Text 3.8.2, 3.8.3, Figure 3.6). The significances of measured $H_2$ oxidation rate in explaining network structure were 0.33 and 0.98 (Spearman correlation between modules eigengene and oxidation rates, $P<0.05$) in $aH_2$ and $eH_2$ networks, again pointing out a significant impact of $H_2$ on soil microbial community structure. Indeed, module eigengenes of five modules were significantly correlated with high-affinity $H_2$ oxidation rate time series in $eH_2$ network, while no module was related to the activity in $aH_2$ network (Figure 3.9). Together, these 5 modules represent 1140 OTUs, which is more than a third of the whole $eH_2$ network (3154 OTUs). OTUs belonging to one module were hindered by $eH_2$ exposure, while members of the four other modules were favored by $eH_2$ treatment. The OTUs belonging to these five modules were members of the rare biosphere as well as more abundant ribotypes, with relative abundance ranging between 0.001% and 4.3% encompassing *Proteobacteria*, *Chloroflexi* and *Acidobacteria* (Figure 3.9 and 3.10). Clustering of these OTUs at the class and order levels unveiled that none of the taxonomic groups were restricted to the five modules correlated with $H_2$ oxidation rate, supporting the previous observation that response to $H_2$ exposure and the distribution of hydrogenase genes is uneven within each taxonomic group. The 38 modules in $eH_2$ network whose eigengene showed no significant correlation with $H_2$ oxidation rate are ecologically relevant observations since they represent indirect impacts of $H_2$ exposure on soil microbial communities and highlight a previously overlooked role of $H_2$ in shaping potential microbe-microbe cooperation and competition interactions. Deciphering the impact of these complex interactions on soil biogeochemical processes was beyond the scope of this study but will deserve attention in future investigations.

## 3.6 **Conclusion**

In conclusion, this exploratory study validated our hypothesis that elevated $H_2$ exposure influences the activity of HOB, leading to direct impacts on the HOB but also indirect

impact on the whole soil bacterial community through competitive and synergistic microbe-microbe interactions. Indeed, $H_2$ soil exposure has an impact on the ecological niche of bacteria differentially distributed among taxonomic groups beyond the alteration of HOB reported in early investigations. Our study was limited to a single farmland soil, with two replicated microcosms per treatment. Nonetheless, considering the significant impact that $H_2$ exposure has brought to soil bacterial communities, it is reasonable to expect a similar effect on other soil types. Considering the steep $H_2$ concentration gradient surrounding $N_2$-fixing nodules, from 20,000 ppmv to sub-atmospheric levels along a 4.5 cm radius, the response of microorganisms might vary as a function of the distance from $H_2$ diffusing sources as well as soil types, physicochemical conditions and the structural and functional structure of microbial community (Hartmann *et al.*, 2015). The impact of $H_2$ and HOB on soil microbial communities will definitely deserve more attention in soil microbiology. There is a need to further study whether elevated $H_2$ exposure also exerts noticeable changes in soil ecological functions other than $H_2$ oxidation, such as carbon metabolism, nutrient cycling, trace gas exchanges and xenobiotic degradation. Such a significant impact of a gas molecule on soil was observed in a study investigating the impact of a doubling atmospheric $CO_2$ concentration, simulating climate change, on microbial communities (Zhou *et al.*, 2011). A metagenomics approach on the matter might provide further answers on this rather ubiquitous microniche as well as the succession of hydrogenases along $H_2$ concentration gradients in the environment.

## 3.7 Acknowledgments

This work has been supported by a Natural Sciences and Engineering Research Council of Canada Discovery grant to PC and by the Community Sequencing Program of the Joint Genome Institute (US Department of Energy) to PC, JT and SGT. The work conducted by the US Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the US Department of Energy under Contract No. DE-AC02-05CH11231. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## 3.8 Supplementary material

### 3.8.1 Construction of weighted co-occurrence networks

Covariation among OTUs during the incubation under controlled $H_2$ levels was analyzed by correlation networks using the package "WGCNA" 1.41 (Langfelder & Horvath, 2008). After sequencing, the output OTU table was divided in two distinct datasets: the $eH_2$ exposure treatment (n = 12) and the $aH_2$ exposure treatment (n = 12). Both datasets were analyzed separately, resulting in the computation of two independent weighted co-occurrence networks. A pairwise Spearman correlation matrix (similarity matrix) was calculated between all OTUs. An adjacency matrix was then calculated by raising the correlation coefficient of the similarity matrix to the soft-thresholding power 12 ($\beta = 12$, $eH_2$ treatment) or 14 ($\beta = 14$, $aH_2$ treatment), as detailed in Horvath *et al* (Langfelder & Horvath, 2008; Langfelder *et al.*, 2011). This transformation resulted in a distribution of OTU connectivity conforming to the scale-free topological model - linear regression model fitting of $\log_{10}$-frequency distribution of the connectivity as a function of $\log_{10}$-connectivities resulted in slopes of -1 and $R^2$ above 0.80 (Zhang & Horvath, 2005). The adjacency matrix was used to delineate modules including OTUs showing highly similar distribution profile throughout the incubation, with a minimum size threshold of 10 OTUs per module. Module eigengene was computed for each module and those showing dissimilarities lower than 0.40 were merged together. The consensus modules alignment method (Langfelder *et al.*, 2011) implemented in WGCNA was used to ensure that both replicates of each treatment produced similar networks (*i.e.* modular patterns were conserved across replicates). This technique uses the cross-tabulation preservation statistics described in the next section.

### 3.8.2 Comparison of networks patterns across treatments

Module preservation statistics were used to identify preserved modules among networks. Two methods were used in order to assess modules preservation between $aH_2$ and $eH_2$ networks: cross-tabulation based and composite based preservation statistics. These methods allowed to statistically compare the structure of both networks and assert with confidence whether $H_2$ had an impact on the structure of the soil bacterial community.

The first method, cross-tabulation based statistics, requires module detection of the reference ($aH_2$) and test ($eH_2$) networks as input for the comparison of modules assignments between both networks. The bias of this analysis is that the module separation has already been done in both networks and that the pairwise comparisons are limited. The overlap of nodes between each pair of modules from the $aH_2$ and $eH_2$ networks is then calculated. A contingency table (heatmap) representing these overlaps (data not shown) shows which modules are similar between both networks. Significant module preservation was considered when $P \leq 0.05$. In this case, 4 out of 43 modules were considered similar between both treatments. None of those modules was characterized by eigengene value displaying significant correlation with $H_2$ oxidation rate. The second method used to identify preserved modules among $aH_2$ and $eH_2$ networks was composite preservation statistics. This analysis summarizes 4 density ($Z_{density}$) and 3 connectivity ($Z_{connectivity}$) based preservation statistics into a single variable, $Z_{summary}$ (Langfelder *et al.*, 2011). $Z_{density}$ measures if densely connected nodes in the control network remain as densely connected in the test ($eH_2$) network, while $Z_{connectivity}$ measures if connectivity patterns are similar in both networks. It only requires the module assignment of the reference data (control network) and the adjacency matrix from the test data. This method reduces the bias of cross-tabulation based statistics since it does not need to compare a limited set of modules in the test network. In our analysis, we used the $Z_{summary}$ and medianRank variables. 999 permutations were applied to these calculations. Thresholds defining module similarity between the two networks were: $Z_{summary} > 10$ for highly preserved modules, $Z_{summary}$ between 2 and 10 for weakly to moderately preserved modules and $Z_{summary} < 2$ for modules that are not preserved between networks (Langfelder *et al.*, 2011). In our study, a single module was strongly preserved while 6 were weakly to moderately preserved (Figure 3.6A). The medianRank preservation statistic was also used to complement $Z_{summary}$, as it is less dependent on module size than $Z_{summary}$. Modules with a median rank closer to zero are more preserved than those with higher values. This statistic identified the same preserved modules as the $Z_{summary}$ approach (Figure 3.6A).

### 3.8.3 Conluding remarks regarding the impact of $H_2$ exposure on OTU co-occurrence – Only a few modules are conserved between $aH_2$ and $eH_2$ networks

By complementing cross-tabulation based preservation statistics with composite (comprising density and connectivity-based statistics) based statistics, we have ensured that we could detect modules that have a high proportion of similar nodes, that are as densely connected and that show similar connectivity patterns in both networks. Only 4 modules were considered similar (Fisher's exact test, $\alpha = 0.05$) between both networks when using cross-tabulation based statistics while 1 module was considered highly preserved ($Z_{summary} \geq 10$) and 6 were weakly to moderately preserved ($2 < Z_{summary} < 10$) according to composite based preservation statistics. All modules identified with the first method were identified using the composite based preservation statistics as well. A high proportion of the modules were considered unique (36 out of 43) which suggests that the whole network structure has been altered by $H_2$. On the other hand, cross-tabulation module preservation analysis showed that replicated microcosms shared the same modules, hence they were reproducible replicates.

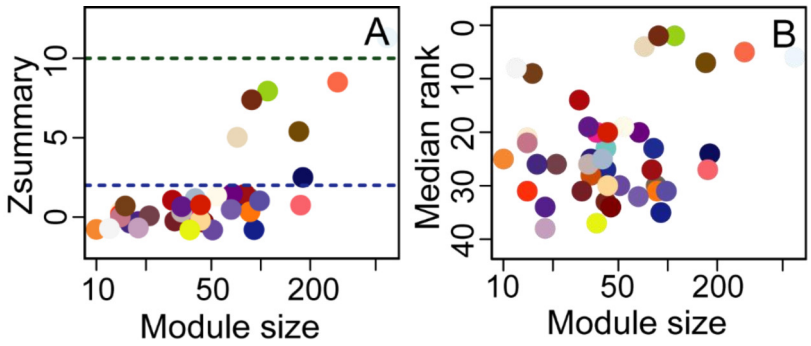



**Figure 3.6** **Composite preservation statistics of modules from the $eH_2$ network (test network) against modules from the $aH_2$ network (reference network). Dots represent the various modules in the reference network. (a) $Z_{summary}$ values of each module as a function of module size (number of nodes). The blue and green dash lines represent minima and maxima of the two thresholds indicating module preservation. $Z_{summary} > 10$ indicates a strong preservation, while $2 \leq Z_{summary} \leq 10$ indicates a weak to moderate module preservation between networks. $Z_{summary} < 2$ indicates unpreserved modules. (b) Median rank values of each module as a function of module size (number of nodes). Lower Median ranks (close to 0) indicate higher module preservation.**

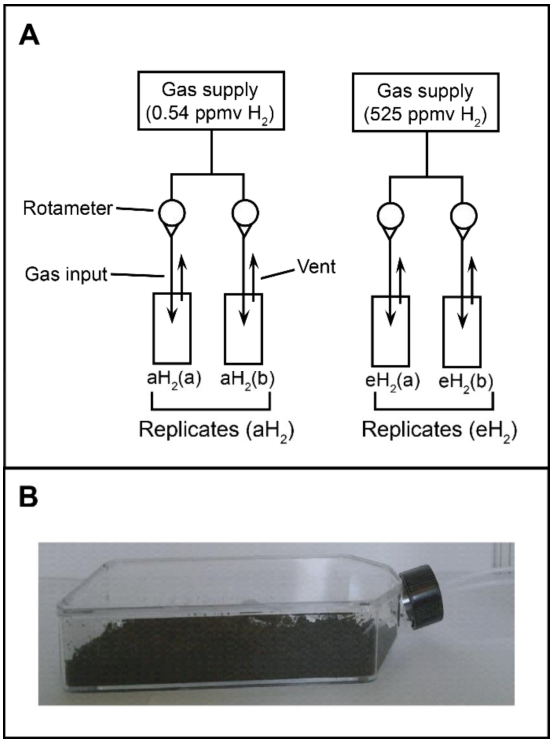



**Figure 3.7** **Schematic representation of the dynamic microcosm chambers. (A) Schematic representation of the dynamic microcosm chambers utilized in this study to expose soil to $aH_2$ or $eH_2$ levels. (B) Photograph of one soil microcosm.**

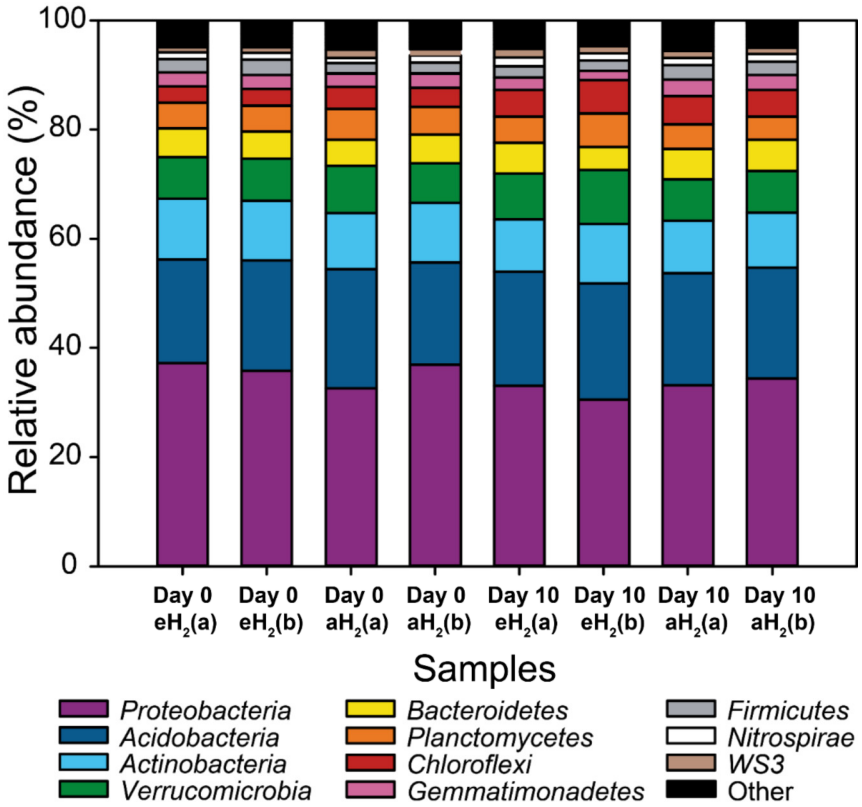



**Figure 3.8** **Taxonomic profiles in soil. Taxonomic profiles of the OTUs clustered at the phylum level for each microcosm at the beginning (day 0) and at the end of the incubation (day 10).**


**Figure 3.9** **Module-Trait relationship heatmap for correlation network computed using OTU covariation profile under $eH_2$ or $aH_2$ exposure. Each line in the heatmaps corresponds to a module. The colors in the heatmap stand for the Spearman correlation coefficient between the module eigengene and high affinity $H_2$-oxidation rate measured in the microcosms. One module out of the 5 has the opposite sign (M1; red color in the heatmap $eH_2$) because it is the only one showing a positive correlation ($P < 0.05$) with high affinity $H_2$-oxidation rate. In contrast, the other modules (M2–M5; green color in the heatmap $eH_2$) displayed a negative correlation ($P < 0.05$) with high affinity $H_2$-oxidation rate.**

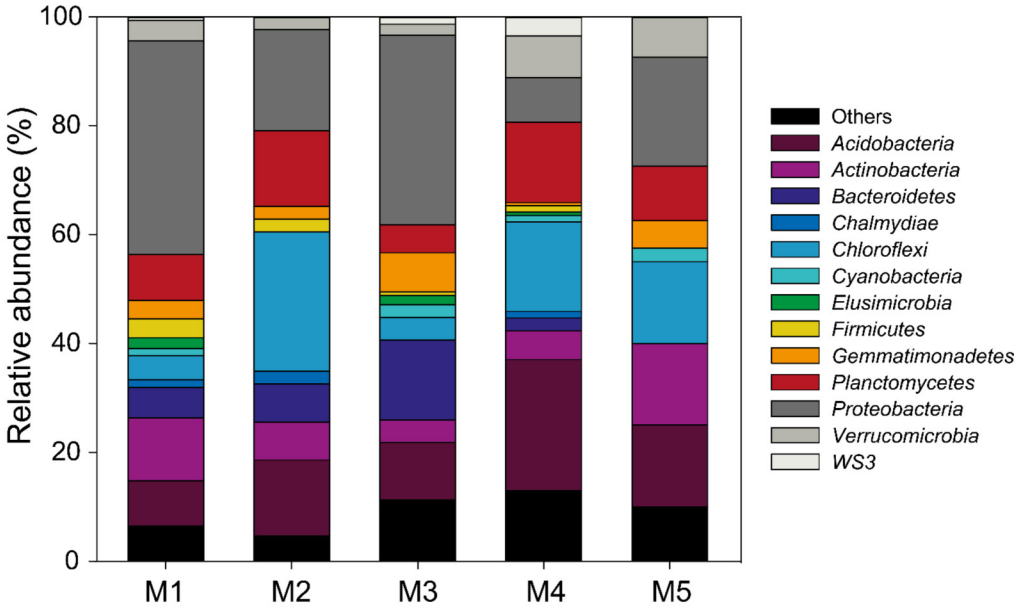



**Figure 3.10** **Taxonomic profile of OTUs found in selected modules of correlation networks. Distribution of the relative abundance of the most abundant phyla in modules whose eigengene is significantly correlated with $H_2$ oxidation rate in the e$H_2$ network.**

**Table 3.2 List of barcodes used to prepare bacterial 16S rRNA gene PCR amplicon libraries[A].**

| Library (internal ID) | Library (sample name) | Barcode |
|---|---|---|
| D1SA | D0-e$H_2$(a) | GCATGGCTCTA |
| D1SB | D0-e$H_2$(b) | AGGCGACCTTA |
| D1SC | D0-a$H_2$(a) | GATAGTGCCAC |
| D1SD | D0- a$H_2$ (b) | GAACACTTCTG |
| D2SA | D1-e$H_2$(a) | TACCGCTAGTA |
| D2SB | D1-e$H_2$(b) | TCTGGAACGCT |
| D2SC | D1- a$H_2$ (a) | GGTATGACTCA |
| D2SD | D1- a$H_2$ (b) | CTTGTAGGACC |
| D3SA | D3-e$H_2$(a) | ACTGTACGCGT |
| D3SB | D3-e$H_2$(b) | TACAGATGGCT |
| D3SC | D3- a$H_2$ (a) | TGCGAACGTAT |
| D3SD | D3- a$H_2$ (b) | TTGAACCAGCT |
| D4SA | D5-e$H_2$(a) | TCAGGCGCTTA |
| D4SB | D5-e$H_2$(b) | ATACTTCGCAG |
| D4SC | D5- a$H_2$ (a) | TGTCGGCTACA |
| D4SD | D5- a$H_2$ (b) | TGGCTCTACAG |
| D5SA | D7-e$H_2$(a) | GCCGCTTAATA |
| D5SB | D7-e$H_2$(b) | GTTACCTCAGA |
| D5SC | D7- a$H_2$ (a) | GTACGAATCCT |
| D5SD | D7- a$H_2$ (b) | TCCTAGCAGTG |
| D6SA | D10-e$H_2$(a) | CACGTGACATG |
| D6SB | D10-e$H_2$(b) | CTTCGGCAGAA |
| D6SC | D10-Air(a) | CATCAGCGTGT |
| D6SD | D10-Air(b) | CCAGTGTATGC |

[A]The V4 region of 16S rRNA gene was amplified using the primers 515F (5'-GTGCCAGCMGCCGCGGTAA-3) and 805R (5'-GGACTACHVGGGTWTCTAAT-3').

# 4 THE TALE OF A NEGLECTED ENERGY SOURCE: ELEVATED HYDROGEN EXPOSURE AFFECTS BOTH MICROBIAL DIVERSITY AND FUNCTION IN SOIL

**Titre français** : L'histoire d'une source d'énergie négligée : Une exposition élevée à l'hydrogène affecte à la fois la diversité et la fonction des microorganismes du sol

**Article publié en ligne le 31 mars 2017 dans le journal Applied and Environmental Microbiology**

*DOI : 10.1128/AEM.00275-17* (Khdhiri *et al.*, 2017)

**Auteurs**

Mondher Khdhiri[1†], Sarah Piché-Choquette[1†], Julien Tremblay[2], Susannah G. Tringe[3] et Philippe Constant[1✉]

[1]INRS-Institut Armand-Frappier, 531 boulevard des Prairies, Laval (Québec), Canada, H7V 1B7

[2]National Research Council Canada, 6100, avenue Royalmount, Montréal (Québec), Canada, H4P 2R2

[3]DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, California 94598, U.S.A.

† Les deux auteurs ont contribué également à l'article

✉ Auteur correspondant

**Contribution des auteurs**

Mondher Khdhiri : Conception des expériences, incubations, analyses en laboratoire (chromatographie en phase gazeuse, mesures de paramètres physicochimiques), analyse des résultats, écriture de l'article

Sarah Piché-Choquette : Conception des expériences, incubations, analyses en laboratoire (chromatographie en phase gazeuse, extraction/purification/quantification de l'ADN, PCR, mesures de paramètres physicochimiques, Ecoplates), analyse des résultats, écriture de l'article

Julien Tremblay : Analyse des données métagénomiques

Susannah G. Tringe : Service de séquençage à haut débit

Philippe Constant : Conception des expériences, analyse des résultats, écriture de l'article

## 4.1 Résumé en français

L'enrichissement des bactéries oxydant l'hydrogène (HOB) via l'$H_2$ généré par les nodules fixateurs d'azote a démontré un effet fertilisant sur de nombreux plants. Le bénéfice des HOB est attribué à leur production de facteurs promoteurs de croissance des plantes, bien que leurs interactions avec les autres membres des communautés microbiennes du sol aient reçu peu d'attention. Nous rapportons que le potentiel énergétique de l'$H_2$, lorsque fourni aux sols, altère la partition de niche écologique des bactéries et champignons, incluant des conséquences à multiples facettes à la fois sur les fonctions microbiennes généralistes et spécialistes. Nous avons utilisé des microcosmes dynamiques afin d'exposer le sol à des ratios de mélange d'$H_2$ typiques de l'atmosphère (0.5 ppmv) diffusant dans les sols, ainsi qu'à des ratios de mélange comparables à ceux retrouvés à l'interface nodule-sol (10 000 ppmv). La forte exposition à l'$H_2$ a causé des effets directs sur deux sous-populations de HOB distinguées par leur affinité envers l'$H_2$, tout en augmentant le potentiel d'utilisation de substrats carbonés au niveau de la communauté et en réduisant l'activité d'oxydation du $CH_4$ dans le sol. Nous avons constaté que l'$H_2$ enclenchait aussi des changements au niveau de l'abondance de microorganismes, ceux-ci étant reproductibles bien qu'inconsistant entre les sols d'un point de vue taxonomique et ceci même pour les HOB. En somme, l'exposition à l'$H_2$ altère la vitesse des procédés d'origine microbienne à une intensité dépendante des paramètres abiotiques et biotiques du sol. Nous soutenons que des investigations supplémentaires des effets directs et indirects de l'$H_2$ sur les communautés microbiennes du sol mèneront à une meilleure compréhension de l'effet fertilisant de l'$H_2$ et des procédés biogéochimiques.

## 4.2 Abstract

The enrichment of $H_2$-oxidizing bacteria (HOB) by $H_2$ generated by nitrogen-fixing nodules has been shown to have a fertilization effect on several different crops. The benefit of HOB is attributed to their production of plant growth promoting factors, yet their interactions with other members of soil microbial communities have received little attention. Here we report that the energy potential of $H_2$, when supplied to soil, alters

ecological niche partitioning of bacteria and fungi, with multifaceted consequences on both generalist and specialist microbial functions. We used dynamic microcosms to expose soil to the typical atmospheric $H_2$ mixing ratio (0.5 ppmv) permeating soils, as well as mixing ratios comparable to those found at the soil-nodule interface (10,000 ppmv). Elevated $H_2$ exposure exerted direct effects on two HOB sub-populations distinguished by their affinity for $H_2$, whilst enhancing community-level carbon substrate utilization potential and lowering $CH_4$ uptake activity in soil. We found that $H_2$ triggered changes in the abundance of microorganisms that were reproducible, yet inconsistent across soils at the taxonomic level and even among HOB. Overall, $H_2$ exposure altered microbial process rates at an intensity that depends upon soil abiotic and biotic features. We argue that further examination of direct and indirect effects of $H_2$ on soil microbial communities will lead to a better understanding of the $H_2$ fertilization effect and soil biogeochemical processes.

## 4.3 Importance

An innovative dynamic microcosm chamber system was used to demonstrate that $H_2$ diffusing in soil triggers changes in the distribution of HOB and non-HOB. Although the response was uneven at the taxonomic level, an unexpected coordinated response of microbial functions was observed, including an abatement of $CH_4$ oxidation activity and a stimulation of carbon turnover. Our work suggests that elevated $H_2$ rewires soil biogeochemical structure through a combination of direct effects on the growth and persistence of HOB and indirect effects on a variety of microbial processes involving HOB and non-HOB.

## 4.4 Introduction

Microorganisms are the metabolic engine of essential biogeochemical processes supplying food and natural resources to mankind, all the while playing a pivotal role in energy balance and plant, animal and human health. Despite the fact that microbial communities often consist of microbial “dark matter” eluding cultivation efforts (Bai *et al.*, 2015), advances in high-throughput sequencing technologies provide new means to

obtain a reasonable portrait of their taxonomic composition and functional potential. Such progress has led to a very promising approach in molecular biogeochemistry: using molecular indicators in combination with environmental variables as parameters for process rate predictive models (Freitag & Prosser, 2009; Lammel *et al.*, 2015; Zak *et al.*, 2006). The inclusion of microbe-microbe interactions to improve these models and harness ecosystem services provided by microbial communities is of paramount importance (Blaser *et al.*, 2016; Daebeler *et al.*, 2014).

As an ubiquitous energy source found in all compartments of the biosphere, molecular hydrogen ($H_2$) has been selected as a case study to infer potential outcomes of stimulating a specialized guild of microbes on other processes in soil. A $H_2$ mixing ratio of 0.53 ppmv is typically found in the global atmosphere (Novelli *et al.*, 1999), which then diffuses into surface soils (Smith-Downey *et al.*, 2008), whereas hotspots can be found in hypersaline cyanobacterial mats, with concentrations between 16,000 and 90,000 ppmv $H_2$ (Nielsen *et al.*, 2015), and inside $N_2$-fixing legume nodules, with concentrations ranging between 9,000 and 27,000 ppmv (Hunt *et al.*, 1988; Rasche & Arp, 1989; Witty, 1991; Witty & Minchin, 1998). Indeed, $H_2$ is produced at a rate of 240,000 L per hectare per growing season in legume fields as an obligate by-product of biological nitrogen fixation (Dong *et al.*, 2003). It is then quickly scavenged by aerobic $H_2$-oxidizing bacteria (HOB) within the first few centimeters surrounding nitrogen-fixing nodules (La Favre & Focht, 1983), after which its remaining concentration is slightly lower than atmospheric mixing ratios. It is assumed that a succession of HOB use this trace gas as an energy source for lithoautotrophic growth or mixotrophic growth and survival according to their affinity for $H_2$ and regulation of hydrogenase gene expression (Liot & Constant, 2016). Under aerobic conditions, $H_2$ oxidation is carried out by [NiFe]-hydrogenases that are unevenly distributed among microbial taxa. So far, *Bradyrhizobium japonicum* (*$\alpha$-Proteobacteria*) and *Cupriavidus necator* (*$\beta$-Proteobacteria*; previously *Ralstonia eutropha*) are the most characterized low affinity soil-dwelling HOB (Schwartz & Friedrich, 2006) while the *Actinobacteria Mycobacterium smegmatis* (Greening *et al.*, 2014b), *Rhodococcus equi* (Meredith *et al.*, 2014) and *Streptomyces* spp. (Constant *et al.*, 2010; Constant *et al.*,

2008) as well as the *Acidobacteria Pyrinomonas methylaliphatogenes* (Greening *et al.*, 2015a) are the most studied high affinity $H_2$ oxidizers. $H_2$ evolved from nitrogen-fixing nodules was shown to exert a fertilization effect in soil (Golding & Dong, 2010), likely caused by an enrichment of HOB displaying plant-growth promoting effects (Maimaiti *et al.*, 2007).

Subsequent studies focussing on HOB have shown that soil $H_2$ exposure shapes the taxonomic fingerprint of microbial communities (Osborne *et al.*, 2010; Zhang *et al.*, 2009). However, neither the conservation of these changes in different soils nor their consequences on microbial community functioning, other than $H_2$ oxidation, was addressed. The current study tested the hypothesis that the response of soil microbial community structure to $H_2$ exposure is idiosyncratic, here defined as an inconsistent response across land-use types, thus under the influence of soil biodiversity and physicochemical properties. On top of that, changes in community structure induced by the activation of HOB exposed to $H_2$ were expected to covary with alterations of soil microbial processes mediated by HOB and non-HOB. To address this prime matter, we have exposed soils embodying three land-use types to different levels of $H_2$ and investigated the impact of the treatment on soil microbial communities using both marker gene PCR-amplicon sequencing and shotgun metagenomics. Soil $H_2$ and $CH_4$ uptake, involving specialized bacterial guilds of microbes, as well as community-level carbon substrate utilization profiles, involving generalist bacterial and fungal guilds, were also monitored. Throughout the article, a “direct effect” of soil $H_2$ exposure is defined as an alteration in $H_2$ oxidation rate of aerobic HOB and “indirect effects” refer to the response of other microbial-mediated processes (*i.e.* $CH_4$ uptake and carbon utilization profile) involving HOB and/or non-HOB.

## 4.5 Results

### 4.5.1 Soil physicochemical properties and microbial metabolism

Soils were incubated in microcosm chambers under a dynamic headspace simulating $H_2$ concentrations found in the atmosphere ($aH_2$; 0.53 ppmv) or at the soil-nodule interface ($eH_2$; 10,000 ppmv). Soils were acidic (pH 4.7-5.2) and encompassed a wide range of carbon (2.4-7.9%) and nitrogen (0.24-0.66%) content (Table 4.1). $H_2$ exposure did not alter measured soil physicochemical properties, with the exception of lower C/N and pH values in larch and poplar soils exposed to $eH_2$, respectively.

**Table 4.1** **Soil physicochemical parameters and process rates measured in soil microcosms. The table shows the mean and standard deviation for triplicate microcosms exposed to $aH_2$ and $eH_2$ treatments. The asterisks (*) represent variables displaying a significant difference ($\alpha < 0.05$) between the two $H_2$ treatments (one-way analysis of variance test for normal distributed variables and Wilcoxon-Mann-Whitney test for non-normal distributed variables).**

| Variables[A] | Soil microcosms | | | | | |
|---|---|---|---|---|---|---|
| | Poplar monoculture | | Farmland | | Larch monoculture | |
| | $eH_2$-P | $aH_2$-P | $eH_2$-A | $aH_2$-A | $eH_2$-L | $aH_2$-L |
| pH | 5.11 (0.02)* | 5.21 (0.02)* | 5.25 (0.02) | 5.20 (0.05) | 4.81 (0.01) | 4.77 (0.01) |
| C (%) | 7.6 (0.4) | 7.9 (0.1) | 2.4 (0.1) | 2.5 (0.3) | 4.3 (0.4) | 4.7 (0.2) |
| N (%) | 0.64 (0.03) | 0.66 (0.01) | 0.24 (0.01) | 0.24 (0.03) | 0.36 (0.02) | 0.37 (0.01) |
| C/N | 11.8 (0.1) | 11.8 (0.1) | 10.1 (0.2) | 10.3 (0.1) | 11.9 (0.2)* | 12.7 (0.31)* |
| $H_2O$ | 0.46 (0.01) | 0.51 (0.01) | 0.19 (0.01)* | 0.25 (0.01)* | 0.34 (0.02) | 0.31 (0.01) |
| High affinity $H_2$-oxidation (pmol $g_{(soil\text{-}dw)}^{-1}h^{-1}$) | 59 (0)* | 1003 (45)* | 22 (2)* | 361 (87)* | 47 (18)* | 1045 (250)* |
| Low affinity $H_2$- oxidation (nmol $g_{(soil\text{-}dw)}^{-1}h^{-1}$) | 264 (32)* | 43 (2)* | 61 (1)* | 19 (3)* | 59 (18)* | 29 (2)* |
| $CH_4$ oxidation (pmol $g_{(soil\text{-}dw)}^{-1}h^{-1}$) | 45 (12)* | 198 (2)* | 29 (7)* | 59 (0)* | 16 (0)* | 45 (9)* |
| $CO_2$ net production (nmol $g_{(soil\text{-}dw)}^{-1}h^{-1}$) | 82 (6)* | 178 (20)* | 9 (1)* | 37 (1)* | 18 (2)* | 58 (3)* |
| Ecoplate (Shannon) | 3.20 (0.01)* | 3.06 (0.01)* | 3.22 (0.04)* | 3.05 (0.02)* | 3.11 (0.01)* | 3.00 (0.02)* |
| Bacteria species richness (Shannon) | 6.39 (0.02) * | 5.85 (0.08)*[B] | 5.83 (0.09)* | 6.26 (0.03)* | 6.02 (0.06) | 6.16 (0.01) |
| Fungi species richness (Shannon) | 4.11 (0.05) | N/A | 4.33 (0.03)[B] | 4.50 (0.05) | 4.23 (0.04) | 3.98 (0.18) |
| Bacteria species evenness (Pielou) | 0.85 (0.00) * | 0.78 (0.01)*[B] | 0.78 (0.01)* | 0.84 (0.00)* | 0.81 (0.01) | 0.82 (0.00) |

| | | | | | | |
|---|---|---|---|---|---|---|
| Fungi species evenness (Pielou) | 0.66 (0.01) | N/A | 0.72 (0.01)[(B)] | 0.73 (0.01) | 0.69 (0.01) | 0.65 (0.03) |

[(A)]The variables 'Water content', 'Ecoplate', 'Bacteria species richness (Shannon)', 'Fungi species richness (Shannon)', 'Bacteria species evenness (Pielou)' and 'Fungi species evenness (Pielou)' followed a normal distribution (Shapiro-Wilk test). Transformation were applied to obtain normal distribution for the other variables. The variable '$CO_2$ net production' was square root-transformed, 'low affinity $H_2$ oxidation' was log-transformed, 'C' was cosine-transformed, and 'C/N' was box-cox-transformed. The variables 'pH', '$CH_4$ oxidation' and 'high affinity $H_2$ oxidation' could not be transformed to obtain normal distribution.

[(B)]The mean and standard deviation were calculated from duplicated results instead of triplicates.

All measured metabolic processes were significantly altered by the $H_2$ treatment in the three soils, albeit with varying extent at the functional and land-use levels (Table 4.1). Direct effect of $H_2$ exposure included a 1.2- to 7-fold higher activity of low-affinity HOB and a 15- to 23-fold lower activity of high-affinity HOB in soils exposed to $eH_2$. Stimulation of low affinity HOB using $CO_2$ and $H_2$ as carbon and energy sources was linked to the lower net $CO_2$ production measured in $eH_2$ treated microcosms when compared to those incubated under $aH_2$. Both carbon utilization profiling and $CH_4$ oxidation rate analyses demonstrated the indirect effect of $H_2$ on soil microbial functions. The highest carbon substrate utilization by bacteria and fungi was observed in soils exposed to $eH_2$ as shown with average well color development (Figure 4.8) and Shannon index (Table 4.1). Furthermore, $CH_4$ oxidation rates in soil microcosms exposed to $eH_2$ were 2 (±0.2), 3 (±0.2) and 4 (±0.3) times lower in farmland, larch and poplar soils when compared with their $aH_2$-exposed counterpart (Table 4.1). As soil-air $CH_4$ exchanges might reflect concomitant production and oxidation processes in aerobic soils (Angel *et al.*, 2011), microcosms were also examined for potential methanogenic activity. No $CH_4$ emission was detected in soil, indicating that $CH_4$ oxidation activity of MOB, rather than $CH_4$ production by methanogenic archaea, was influenced by $H_2$ (Table 4.6).

### 4.5.2 Taxonomic profile of soil microbial communities

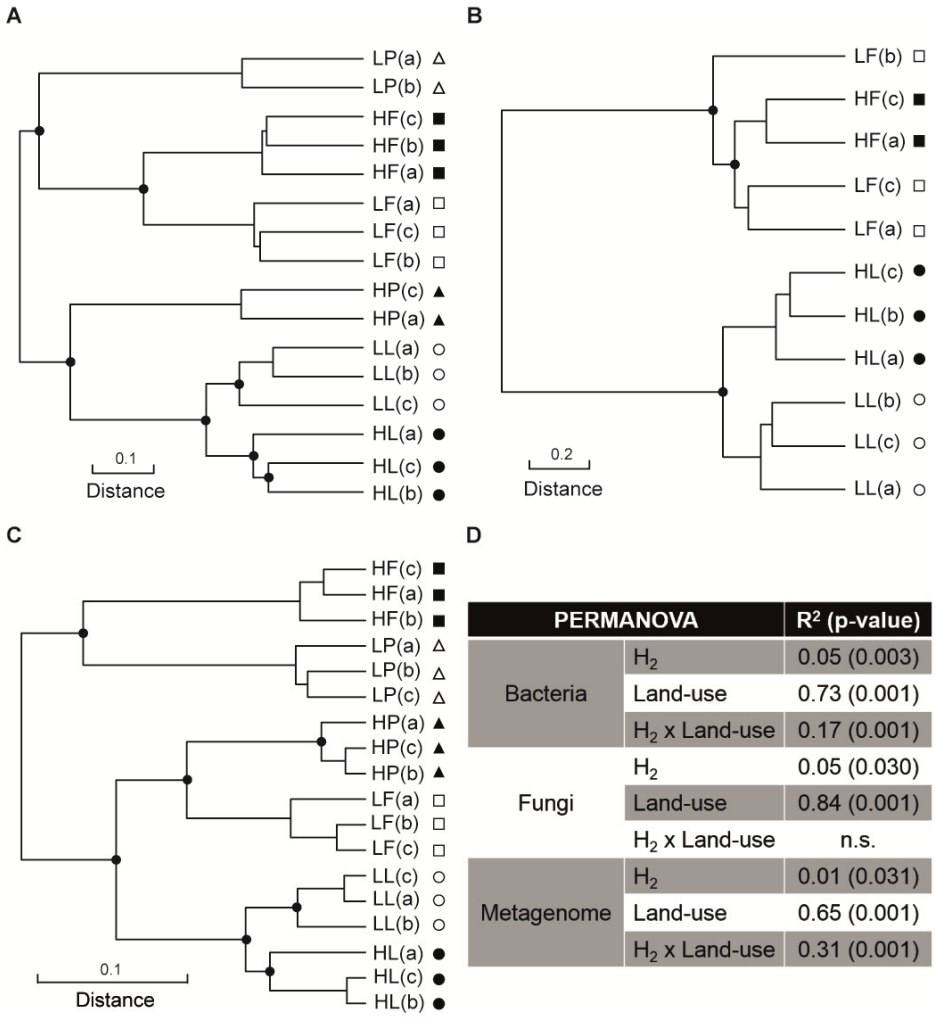
The three soils were dominated by bacteria encompassing the *Proteobacteria*, *Actinobacteria* and *Acidobacteria* phyla and fungi affiliated to the *Ascomycota* phylum (Figure 4.9A). Bacterial and fungal OTUs detected in all three land-use types represented 60% and 18% of the retrieved OTUs (Figure 4.9B). Species richness (Shannon index) and evenness (Pielou index) of bacterial communities were significantly influenced by $H_2$ exposure, although no significant effect was recorded for fungi (Table 4.1). The impact of $H_2$ exposure on bacterial richness was idiosyncratic: $eH_2$ treatments led to higher species richness in poplar soil, whilst farmland soil followed an opposite trend and larch soil did not express any significant response when compared to microcosms exposed to $aH_2$. The idiosyncratic response of soil microbial communities was illustrated with β-diversity analyses of bacterial (Figure 4.1A) and fungal (Figure 4.1B) communities. Land-use type contributed to 74% and 84% of the variation in bacterial and fungal community profiles,

while the contribution of $H_2$ treatment was approximately 5% (Figure 4.1D). Poplar soil was the land-use type showing the strongest response to $H_2$ exposure. Indeed, Euclidean distance between replicated bacterial profiles in microcosms exposed to $aH_2$ and those exposed to $eH_2$ was 11.2, 7.2 and 5.1 for poplar, farmland and larch soil, respectively, with the relative abundance of 857 bacterial and 63 fungal OTUs influenced by $H_2$ treatment (10.1, Data set 1). These responsive OTUs were grouped taxonomically to ascertain whether OTUs within higher level taxa responded similarly to $H_2$ treatments. No reproducible distribution pattern was found at the family level (Table 4.2). While many individual taxa responded to the treatment in one or two soil land-use types, only a few OTUs representative of the rare biosphere (less than 0.01% relative abundance) were characterized by a reproducible response to $H_2$ exposure in the three soils (10.1, Data set 1).

**Table 4.2** **Uneven response of different taxonomic groups (family level) of bacteria and fungi to $H_2$ exposure as a function of land-use type. The numbers expressed in percentage indicate the proportion of OTUs that are more abundant in $aH_2$ or in $eH_2$ treatments for each land-use type. This reduced dataset consists of microbial families containing at least 6 OTUs and of which more than 50% of OTUs responded to the treatment in at least one land-use type. The table shows the mean. Numbers between parentheses depict the amount of OTUs of this specific family in Poplar, Farmland and Larch soils, respectively.**

| | Poplar (%) | | Farmland (%) | | Larch (%) | |
|---|---|---|---|---|---|---|
| | $eH_2$ | $aH_2$ | $eH_2$ | $aH_2$ | $eH_2$ | $aH_2$ |
| **Bacteria** | | | | | | |
| *Nocardioidaceae* (10,9,8) | 0 | 100 | 70 | 0 | 56 | 0 |
| *Xanthomonadaceae* (32,29,33) | 0 | 61 | 69 | 0 | 69 | 0 |
| *Intrasporangiaceae* (7,6,7) | 0 | 43 | 57 | 0 | 100 | 0 |
| *Sphingomonadaceae* (13,11,12) | 0 | 75 | 54 | 0 | 27 | 18 |
| *Hyphomonadaceae* (6,6,7) | 0 | 0 | 50 | 0 | 33 | 0 |
| *Sphingobacteriaceae* (8,8,8) | 0 | 63 | 13 | 13 | 25 | 0 |

| | | | | | | |
|---|---|---|---|---|---|---|
| *Gemmatimonadaceae* (30,26,28) | 11 | 50 | 7 | 13 | 19 | 0 |
| *S47* (6,6,6) | 0 | 50 | 0 | 0 | 0 | 17 |
| *Frankiaceae* (6,6,6) | 0 | 17 | 0 | 0 | 50 | 0 |
| *0319-7L14OR* (18,14,18) | 72 | 0 | 0 | 11 | 0 | 0 |
| *RB25CL* (9,7,9) | 67 | 0 | 33 | 0 | 14 | 0 |
| *Flexibacteraceae* (19,22,20) | 60 | 15 | 16 | 0 | 5 | 18 |
| *OM190OR* (6,7,7) | 57 | 0 | 0 | 17 | 0 | 14 |
| *Bdellovibrionaceae* (13,9,8) | 0 | 0 | 0 | 54 | 0 | 0 |
| *Rhodobacteraceae* (10,8,9) | 0 | 11 | 0 | 10 | 13 | 50 |
| **Fungi** | | | | | | |
| *Cortinariaceae* (6, 14, 0) | N/A | N/A | 0 | 17 | 21 | 64 |
| *Helotiaceae* (8, 22, 0) | N/A | N/A | 25 | 0 | 0 | 50 |




| PERMANOVA | | $R^2$ (p-value) |
|---|---|---|
| Bacteria | $H_2$ | 0.05 (0.003) |
| | Land-use | 0.73 (0.001) |
| | $H_2$ x Land-use | 0.17 (0.001) |
| Fungi | $H_2$ | 0.05 (0.030) |
| | Land-use | 0.84 (0.001) |
| | $H_2$ x Land-use | n.s. |
| Metagenome | $H_2$ | 0.01 (0.031) |
| | Land-use | 0.65 (0.001) |
| | $H_2$ x Land-use | 0.31 (0.001) |

**Figure 4.1** **UPGMA agglomerative clustering of soil microbial communities according to a Euclidean distance matrix calculated with Hellinger-transformed data. UPGMA agglomerative clustering of soil microcosms according to a Euclidean distance matrix calculated with Hellinger-transformed (A) bacteria and (B) fungi ribotyping profiles and (C) a Bray-Curtis distance matrix of genome bins profiles. (D) Permanova displaying the proportion of total β-diversity variance explained by soil land-use type and $H_2$ treatment. Land-use types are distinguished with three different symbols (square; farmland, circle; larch and triangle; poplar). Black symbols indicate soil microcosms exposed to e$H_2$ treatment and white symbols indicate soil microcosms exposed to a$H_2$ treatment. 16S rRNA gene PCR amplicon sequencing failures for samples LP(c) and HP(b) and ITS PCR amplicon sequencing failures for the six poplar samples and HF(b) and LP(c) impaired their inclusion in the analysis.**

### 4.5.3 Metagenomic profile of soil microbial communities

PCR-amplified rRNA marker gene profiles provided evidence that bacterial and fungal response to $H_2$ exposure is idiosyncratic. Nonetheless, uneven distribution of hydrogenases among taxonomic groups impaired the identification of direct and indirect impacts of $H_2$ exposure on soil microbial communities. A metagenomic binning procedure proved efficient to pinpoint microorganisms responding to $H_2$ exposure, search for hydrogenase genes to outline direct and indirect responses from those genomic bins and identify the underlying environmental factors influencing their distribution. Clustering analysis showed that genome bin distribution in soils was mainly driven by land-use type, with a modest influence of $H_2$ treatment (Figure 4.1C and 4.1D). Metagenomic profiles of larch soil were the most resistant to $H_2$ treatment, followed by farmland and poplar soils. The same pattern was observed when the whole metagenomic database was considered in the analysis (Figure 4.10). A PCA was computed to identify the genome bins that contributed the most to distinguish the metagenomic profiles and their relationship with underlying biotic and abiotic environmental factors (Figure 4.2A). The first principal component was correlated with soil nutrients, with C/N having negative loading. The second principal component reflected microbial metabolism, with a negative loading exerted by soil net $CO_2$ production. One genome bin (Bin-1) affiliated to the order *Rhizobiales* and three genome bins affiliated to *Xanthomonadales* bacteria (Bin-2, Bin-3 and Bin-6) exerted more weight than average to define the ordination in the reduced space. The affiliation of these genome bins inferred by gene annotation was supported by a correlation network analysis of the rRNA marker gene amplicon sequencing profile that led to the identification of *Rhizobiales* and *Xanthomonadales* OTU whose elevated relative abundance (up to 7%) and distribution profile corresponded to the four genome bins (Figure 4.11). Network topological properties were analyzed and it has been determined that the Bin-1 representative OTU displayed a very high connectivity (top 10%), Bin-2 had an average connectivity, and Bin-3 and Bin-6 had a lower-than-average connectivity, indicating different contributions in shaping soil microbial community structure.

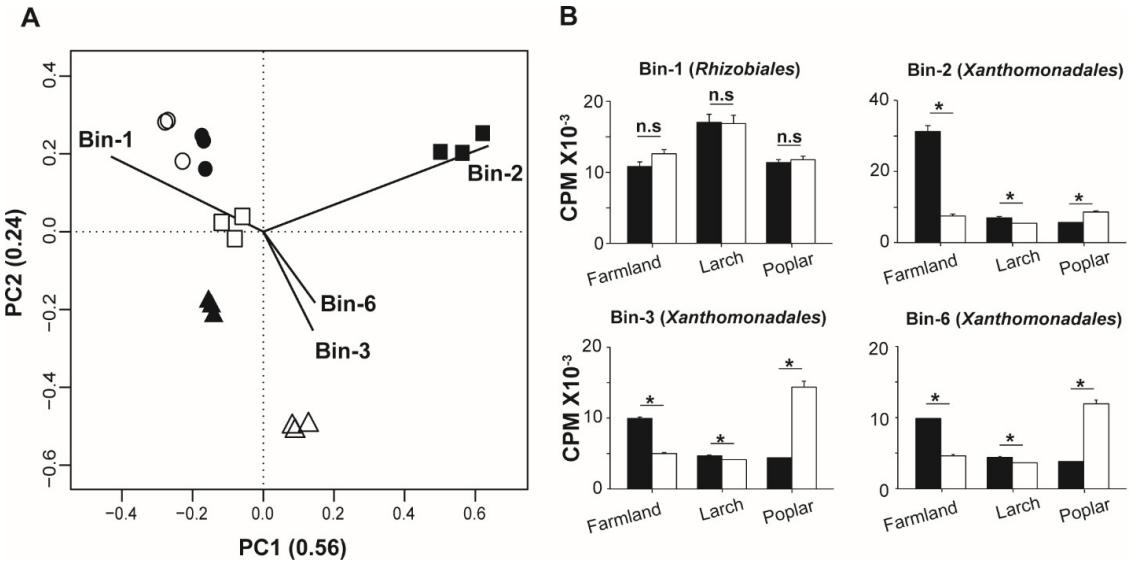



**Figure 4.2** **(A) PCA showing the distribution of soil microcosms in a reduced space defined by the relative abundance of the 93 genome bins. Land-use types are distinguished by three different symbols (square; farmland, circle; larch and triangle; poplar). Black symbols indicate soil microcosms exposed to $eH_2$ treatment and white symbols indicate soil microcosms exposed to $aH_2$ treatment. (B) Relative abundance of the 4 genome bins outlined by the PCA. Asterisks denote genome bins for which the relative abundance was significantly different (Wilcoxon-Mann-Whitney test) between treatments. The term "n.s." represents non-significant relationships.**

In accordance with the rRNA marker gene survey, none of the 93 genome bin distribution profiles displayed a reproducible response to $H_2$ exposure in the three soils (Table 4.7). For instance, *Xanthomonadales* genome bins were significantly more abundant under $eH_2$ conditions (compared to $aH_2$) in larch and farmland soils, while they showed a reverse trend in poplar soil (Figure 4.2B). This idiosyncratic response was attributed to soil factors influencing growth and distribution of bacteria. In fact, multiple regression analyses unveiled that C/N, pH and the interaction between $CO_2$ respiration and $H_2$ treatment were the strongest independent variables, among those measured, to explain the distribution of the 4 genome bins (Table 4.8). Parameters of the multiple regression were consistent to explain variation of OTU candidate representatives of the 4 genome bins (Table 4.8).

Application of a hydrogenase-based Hidden-Markov Models (HMM) on all assembled contigs from this project led to the identification of 122 putative gene fragments encoding

for the structural subunits of [NiFe]-hydrogenases (10.2, Data set 2). Among them, 29 sequenced hydrogenase genes were validated by the examination of the canonical L1 or L2 cysteine motifs binding the metal ions of the catalytic site in [NiFe]-hydrogenases (Greening *et al.*, 2016; Vignais & Billoud, 2007). More than 67% of retrieved sequences belonged to group 3 [NiFe]-hydrogenases, followed by representatives of group 1 (15-23%) and group 2 (10-20%). Among those 29 partial genes, 2 were found in assembled genome bins. These putative HOB genome bins included Bin-1 (*Rhizobiales*), possessing the gene encoding for the large subunit of a group 1 [NiFe]-hydrogenase. The abundance of most potential HOB genome bins did not change according to $H_2$ treatment, indicating that not all HOB take advantage of $H_2$ availability for lithoautotrophic growth (Table 4.9). On the other hand, the higher relative abundance of *Xanthomonadales* genome bins in $eH_2$ reflects the effect of $H_2$ on non-HOB. Indeed, none of these 3 bins (genome sequences completeness estimated at 84, 57 and 40% for genome bins 2, 3 and 6, respectively) displayed genes encoding hydrogenase structural subunits or auxiliary components necessary to harness energy from $H_2$ in their respective genomes based on gene annotation and HMM (10.3, Data set 3).

### 4.5.4 Disentangling the idiosyncratic impact of $H_2$ on microbial communities

A structural equation model (SEM) was developed to test the hypothesis that the idiosyncratic impact of $H_2$ exposure observed on soil microbial community structure and function is explained by soil biotic and abiotic features acting as ecological filters for microbial species and functional groups. The hypothetical model comprised four independent variables selected based on a number of assumptions supported by literature (4.9.1, Method S1) to predict trace gas turnover and carbon utilization profiles (Figure 4.3). Dependent variables included $H_2$ exposure level and soil water content as well as two composite variables representing soil abiotic features (SOIL) and microbial diversity (BIO). Parameterization of the model with experimental data indicated that idiosyncratic response of microbial processes upon $H_2$ exposure is process-specific, owing to the different degree of dependence of each function to physicochemical and biological properties related to their environment. Firstly, loss of high-affinity $H_2$ oxidation activity in soils exposed to $eH_2$ was attenuated by soil biodiversity and higher soil moisture

content (Figure 4.3A). The absence of a significant path between the composite variable SOIL and the high affinity $H_2$-oxidation rate was unique for this functional guild. Secondly, stimulation of low-affinity HOB by $eH_2$ was partly explained by soil physicochemical properties and microbial diversity (Figure 4.3B). Thirdly, the indirect impact of $eH_2$ on $CH_4$ oxidation activity was impaired by elevated species richness and soil moisture content and, to a lesser extent, by soil physicochemical properties (Figure 4.3C). Finally, carbon substrate utilization was mainly driven by $H_2$ exposure and soil abiotic properties and the absence of a significant path from the composite variable BIO was unique to this microbial function (Figure 4.3D). An elevated path coefficient linking $H_2$ treatment to high-affinity $H_2$ oxidation rate (-0.94) and carbon substrate utilization (1.0) indicated that soil abiotic and biotic features exerted less influence than $H_2$ treatment on the inhibition and stimulation of these processes.

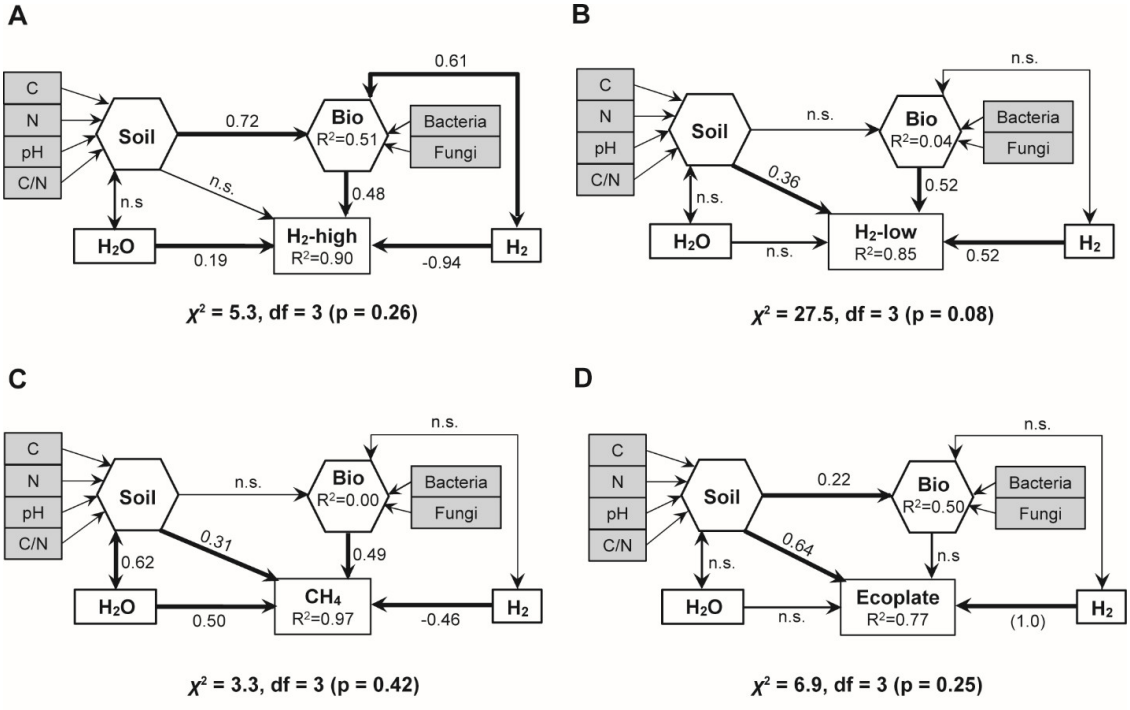



**Figure 4.3** **Structural equation models testing causal assumptions of soil biotic and abiotic variables on microbial processes. (A) Structural equation models to test causal assumptions of soil physicochemical properties, biodiversity, moisture and $H_2$ exposure on (A) high affinity $H_2$-oxidation rate, (B) low-affinity $H_2$-oxidation rate, (C) high-affinity $CH_4$ oxidation rate and (D) carbon utilization profile measured in soil microcosms. Variables shown in hexagons are statistical composites estimated using partial least squares linear regression analysis of observed variables. The $R^2$ values are presented for dependent variables. Path coefficients were standardized**

**and were significant at $p < 0.05$. The term "n.s." represents non-significant relationships.**

## 4.6 Discussion

The structure of microbial communities is driven by a combination of biotic and abiotic environmental features including soil physicochemical properties, climate and vegetation cover. Integration of these microbial community dynamics into an ecological theory framework is necessary to generate predictive models of practical value to mitigate the influence of anthropogenic activities on global biogeochemical cycles (Prosser *et al.*, 2007). For instance, a relationship between energy supply to ecosystem through primary production and taxonomic diversity was proposed to predict ecological patterns of microbial communities in aquatic ecosystems (Horner-Devine *et al.*, 2003). As observed with animals and plants, this relationship is characterized by a variety of patterns, ranging from insignificant trends to humped patterns where microbial diversity exhibits a maximum at intermediate productivity levels (Smith, 2007). Our findings broaden these classical ecological patterns with inorganic energy inputs through $H_2$ addition exerting either positive, negative or insignificant impact on soil microbial diversity (Table 4.1). Hence, we can conclude with confidence that both soil biotic and abiotic properties alter the diversity-energy relationship driven by $H_2$. Furthermore, $H_2$ energy potential influenced the ecological niche of several members of the soil microbial community, which is in contrast with the traditional knowledge that only a few HOB benefit from $H_2$ sources in soil (Osborne *et al.*, 2010; Zhang *et al.*, 2009). This impact on microbial communities is expected to be proportional to $H_2$ concentration, since this actual concentration limits the maximum energy potential accessible to HOB. Indeed, in a previous study analyzing the impact of $H_2$ exposure on soil microbial communities, the relative abundance of 958 taxonomically-diverse OTUs, representing 0.001 to 1.8% of the community, was altered in a farmland soil exposed to 500 ppmv $H_2$, despite having no significant impact on bacterial species richness overall (Piché-Choquette *et al.*, 2016). However, the higher level of $H_2$ used in this study resulted in a more dramatic alteration of microbial community composition, considering that both rare and abundant taxa (up to 7.5% abundance) responded to $H_2$ treatment. Soil biotic and abiotic features have proven to be influential

on the response of individual taxa to $H_2$ exposure as only a few bacterial or fungal OTUs were shown to display the same response in all three soils. This outcome contrasts with observations made in soil amended with nutrients wherein convergent response of taxa was observed in grasslands displaying a broad range of biotic and abiotic properties (Leff *et al.*, 2015). This discrepancy is likely explained by a wider range of microbes benefiting from nitrogen fertilizers when compared to those able to use $H_2$.

The combination of rRNA marker gene amplicon sequencing, metagenomics and process rate measurements in the current investigation led to the first evidence of a combination of direct and indirect impacts of $H_2$ on soil microorganisms. This direct impact consists of an inhibition of high affinity $H_2$ oxidation activity and the enrichment of low affinity HOB in soil exposed to $eH_2$ (Conrad *et al.*, 1983b; La Favre & Focht, 1983). These alterations of HOB led to a shift in kinetic parameters governing soil $H_2$ uptake activity, with an increase of both $_{(app)}V_{max}$ and $_{(app)}K_m$ values (Piché-Choquette *et al.*, 2016). In contrast, potential indirect side effects of the inhibition and stimulation of HOB on the distribution and metabolism of other members of microbial community have received no attention. Here we show that $H_2$ can significantly affect the community as a whole, not only HOB. The first evidence was the response of certain fungal OTUs to $H_2$ exposure, despite recent genome database mining unveiling that no fungus species possesses genes encoding motifs associated with $H_2$-oxidizing enzymatic machinery (Greening *et al.*, 2016; Vignais & Billoud, 2007). A second piece of evidence was the observation that relative abundance of three *Xanthomonadales* genome bins was higher in soil microcosms exposed to $eH_2$, without possessing the genetic potential to oxidize $H_2$. However, considering the incomplete genome coverage of these three candidates (40-84% completeness), their ability to oxidize $H_2$ cannot be completely ruled-out at this stage. A third evidence for an indirect effect of $H_2$ on soil microorganisms was the lower $CH_4$ oxidation rate and higher diversification of carbon substrate utilization potential in soils exposed to $eH_2$. None of these processes are known to be affected by $H_2$, but future studies shedding light upon those underlying mechanisms are crucial considering the importance of these processes with MOB contributing to 15% of the global losses of atmospheric $CH_4$ (Conrad, 2009) and carbon turnover being an indicator for microbial

community physiological state (Garland & Mills, 1991). Although $CH_4$ uptake activity measured in this study involves unknown high affinity MOB and/or conventional MOB under starvation (Cai *et al.*, 2016), their ability to use $H_2$ as an energy source cannot be excluded. In fact, hydrogenases are widespread in MOB (Carere *et al.*, 2017) and the ability to oxidize $H_2$ for energy generation has been reported in certain conventional MOB, such as *Methylococcus capsulatus* (Bath) and *Methylosinus trichosporium* OB3b (Chen & Yoch, 1987; Hanczár *et al.*, 2002). Our data may imply the activation and enrichment of antagonistic bacteria and fungi to MOB under $eH_2$ treatment, but a mixotrophic energy metabolism of MOB with preferential use of $H_2$ under $CH_4$ starvation in $eH_2$ microcosms cannot be completely ruled out at this stage. Similarly, higher community level carbon substrate utilization potential in soils exposed to $eH_2$ could be partly explained by the mixotrophic metabolism of HOB, combining both organic carbon and $H_2$ as energy sources. Regardless of the underlying mechanisms, these observations are in sharp contrast with the uneven response of taxonomic groups and the energy-biodiversity relationships reported before, with a convergent response of microbial functions to $H_2$ exposure.

SEM unveiled that the magnitude of microbial activity gains and losses observed under $eH_2$ exposure fluctuated according to both soil physicochemical properties and microbial diversity. The relative contribution of biodiversity was more important for the metabolic activity of specialized HOB and MOB than the widely distributed metabolism of organic carbon substrates. This is in accordance with previous investigations performed across a land-use gradient where the biodiversity of soil microbial communities represented a poor predictor of $CO_2$ soil respiration compared to the more specialized metabolism of $CH_4$ (Levine *et al.*, 2011). In this work, the contribution of species richness was positive in supporting microbial processes, which is in line with the relationship between biodiversity and ecosystem functioning (Bell *et al.*, 2005). A combination of “selection effect”, by which dominance of certain species with a particular trait influence ecosystem processes, and “complementary effect”, where the level of biodiversity influences the use of available resources, are expected to explain the observed relationship between biodiversity and measured functions (Loreau & Hector, 2001; Peter *et al.*, 2011). In the presence of high

concentrations of $CH_4$, interactions between MOB and heterotrophic bacteria were shown to promote $CH_4$ oxidation activity (Ho *et al.*, 2014; Stock *et al.*, 2013). The SEM provides a framework for future investigations aimed at testing how soil microbial diversity and physicochemical properties mitigate alterations of biogeochemical processes induced by $H_2$. High affinity $H_2$ oxidation rate and community-level carbon substrate utilization were the most sensitive microbial processes to $H_2$ exposure, and their responses are expected to be reproducible in other soil types. In contrast, the fate of high-affinity $CH_4$ and low-affinity $H_2$ oxidation activities in other soil types exposed to $H_2$ is expected to be more variable, relying on soil biotic and abiotic features.

In conclusion, we validated our hypothesis that the taxonomic response of soil microbial community composition to $H_2$ exposure is inconsistent across land-use types. Surprisingly, these changes in community structure occurred along with a common metabolic response mediated by HOB and non-HOB. A few studies have shown the beneficial effect of $H_2$ exposure on plant growth, presumably due to a phytohormonal control (*e.g. via* ACC deaminase) by HOB (Golding & Dong, 2010). Considering both direct and indirect effects on microbial community structure and functions that have been attributed to $H_2$ exposure in this study, the so-called $H_2$ fertilization effect could be attributed to the alteration of biogeochemical cycles through complex microbe-microbe interactions as well as other synergistic factors. However, this work is a proof-of-concept experiment performed under artificial, controlled incubation conditions. The elevated $H_2$ exposure treatment represented extreme conditions in soil as steep $H_2$ concentration gradients occur in the area surrounding nitrogen-fixing nodules. As a result, the impact of $H_2$ on microbial community structure and function observed here might be over-represented when compared to microbial succession along $H_2$ concentration gradients found in natural environments. For instance, other variables such as root exudates, gas diffusion limitation and weather conditions may exert a dominant control on soil microbial communities. Detection of the minimal $H_2$ threshold concentration necessary to induce changes in microbial communities, as well as field investigations, will be important to assess the environmental impacts of $H_2$ in soil. Although preliminary, this work provides the first evidence that $H_2$ exposure disrupts community-level carbon substrate and $CH_4$

utilization profiles predetermined by biotic and abiotic features of soils. This study has focused on aerated upland soils and it remains to be tested whether observed effects of $H_2$ hold true for other soils, including anoxic waterlogged soils exposed to elevated $H_2$ produced during organic matter fermentation and $N_2$-fixation.

## 4.7 Materials and Methods

### 4.7.1 Soil samples

Soil samples were collected in a tree nursery managed by the Quebec Minister of Forests, Wildlife and Parks located in the municipality of St. Claude (Québec, Canada). A 12-year-old larch plantation (45.68114°N, 071.99876°W), a 10-year-old poplar plantation (45.68024°N, 071.99823°W) and a farmland consisting of potato and maize rotation (45.67688°N, 071.99814°W) were visited in June 2014. The litter, fermentation, humus soil horizon (LFH; upper 2-5 cm layer) was removed before collecting the upper layer (first 10 cm of B horizon) in tree plantations. Soil samples were transferred into 37.9-L plastic boxes and stored at 4°C until processing (within 2 months). Soil was first air dried for 60 to 72 hours at room temperature (approximately 22°C), homogenized (2 mm sieve) using a vibratory sieve shaker (AS 200®, Retsch GmbH, Haan, Germany) and then used for physicochemical analyses and microcosm incubations. Soil texture, pH (1:2.5 w/v soil suspension in 0.01 M $CaCl_2$), total carbon, total nitrogen and water contents were determined as described in previous investigations (Khdhiri *et al.*, 2015; Piché-Choquette *et al.*, 2016). All 3 soils were classified as loamy sand according to particle size distribution. Raw values of pH, C, N and water content are provided in Data set 1 (10.1).

### 4.7.2 Soil microcosm incubations

Soil samples were incubated in a dynamic microcosm chamber unit comprising a gas mixing station and a gas distribution network (4.9.2, Method S2). Dilutions made in the gas blending station were controlled by two mass-flow controllers, achieving precise stoichiometric ratios. Gas mixture was distributed into each individual microcosm at 40 ml $min^{-1}$ through seven mass flow controllers. Outlet tubes were vented to the atmosphere. The gas blending station was parameterized to test two $H_2$ exposure treatments: high

concentration (10,000 ppmv $H_2$) and low concentration (0.5 ppmv $H_2$). Our experimental setup was an artificial design used as a proof of concept to explore the potential impact of elevated $H_2$ concentrations found in certain ecosystems comprising $H_2$ hotspots, including $N_2$-fixing nodules-soil interface, hydrothermal vents, hypersaline cyanobacterial mats and hot springs, as opposed to low concentrations (0.5 ppmv $H_2$) found in the global atmosphere and surface soils (Novelli *et al.*, 1999). Both $H_2$ exposure treatments were tested using three independent replicates (3 land-use types × 2 treatments × 3 replicates = 18 microcosms). Microcosms were designated according to $H_2$ treatment (*H* and *L* for elevated and ambient $H_2$ exposure, respectively) and soil type (*F*; farmland, *L*; larch, and *P*; poplar) followed by a letter separating replicates (*a*, *b* or *c*). For instance, the first replicate of farmland soil exposed to 10,000 ppmv $H_2$ was named HF(a). Each microcosm consisted of a 0.9 L polystyrene cell culture flask (Corning, Tewksbury, MA) holding 200 g homogenized soil. Soil water content was adjusted to 30% water holding capacity (whc), using sterile ultrapure water, in order to promote gas diffusion in soils and provide an aerobic ecosystem). Soil incubation in the dynamic microcosm chambers unit lasted for 15 days at 22°C in the dark. Soil water content was monitored on a regular basis using standard gravimetric method. Sterile water was added to keep soil at a steady moisture level.

### 4.7.3 Measurement of gas exchanges

Low affinity $H_2$ oxidation, net $CO_2$ production and high affinity $CH_4$ oxidation rates were measured at the end of the incubation, after soil subsamples were taken for molecular analysis, using a gas chromatographic assay. Microcosms were flushed with a gas mixture comprising 10,000 ppmv $H_2$ and 3 ppmv $CH_4$ in synthetic air (certified standard mixture, Praxair distribution Inc. PA, USA) for 30 minutes. These initial $H_2$ and $CH_4$ concentrations can be used to measure the activity of low affinity HOB and high affinity MOB. Variations in $H_2$, $CH_4$ and $CO_2$ mixing ratios were monitored over an 8-hour time period by analyzing aliquots (10 ml) of the headspace with a gas chromatograph (Agilent technologies GC system SP1 7890-0504, CA, USA) comprising an electron capture detector, a thermal conductivity detector for the analysis of $H_2$ and a flame ionization detector for the detection of $CH_4$ and $CO_2$. Agilent Greenhouse Gas Checkout Sample®

(5 ppmv $CH_4$ and 600 ppmv $CO_2$, balance air, ±2% analytical accuracy) was used for the calibration of $CH_4$ and $CO_2$, while the 10,000 ppmv $H_2$ gas mixture (Praxair distribution, Inc. PA, USA) was used for $H_2$ calibration. The electron capture detector (ECD) used a mixture of 4.95% $CH_4$ in argon (Praxair distribution Inc. PA, USA) as makeup gas. High purity $H_2$ (99.999% $H_2$, Praxair distribution Inc. PA, USA) and hydrocarbon-free synthetic air (zero air, Praxair distribution Inc. PA, USA) were used to ignite the flame and 99.9995% $N_2$ was supplied by a nitrogen generator (Parker Hannifin Corp., QC, Canada) as makeup gas for the flame ionization detector (FID). The makeup gas of the thermal conductivity detector (TCD) was high purity argon (99.999% Argon, Praxair distribution Inc. PA, USA). The GC system consists of two separate channels with 1/8" stainless steel packed columns (HayeSep Q 80/100). The oven temperature was set to 60°C, while temperature of FID, TCD and ECD were maintained at 250°C, 200°C and 350°C, respectively. GC signal acquisition and peak quantification were performed with Agilent OpenLab CDS® software. Each analytical run lasted 8 minutes. High affinity $H_2$ oxidation rate was measured at the end of the incubation period using a gas chromatographic assay, as described by Khdhiri *et al*. (2015). Briefly, microcosms were flushed with synthetic air for 30 minutes before injecting 3 ml of a certified gas mixture containing 500 ppmv $H_2$ (Praxair distribution Inc. PA, USA). As the initial $H_2$ level in the headspace was approximately 3 ppmv, high affinity HOB were responsible for the loss of $H_2$ in the static headspace. Raw values of gaseous exchanges are provided in Data set 1 (10.1).

### 4.7.4 Carbon metabolism

Community-level carbon utilization profiles were calculated using Ecoplates® assay (Biolog, Hayward, CA, U.S.A.) at the end of the incubation period. Soil suspensions were prepared by diluting soil (1:9 w/v) in 0.1 $NaH_2PO_4$ (pH 6) then further diluting it in 0.15M NaCl (1:10 v/v) after mixing with a vortex mixer. 100 µL of the soil suspensions was inoculated in each well of the Ecoplates. Each plate was then incubated for 4 days at 22°C in the dark. After incubation, colour change from clear to purple (reduction of tetrazolium dye) was monitored with an Infinite M1000pro® plate reader (Tecan Group Ltd., Männedorf, Switzerland) at 590 nm, proportionally reflecting microbial use of each substrate. Profiles were standardized by subtracting blank controls from each well

(inoculated wells without substrate). Raw values of Ecoplates are provided in Data set 1 (10.1).

### *4.7.5* Nucleic acid extraction and purification

Soil subsamples were collected in each microcosm at the end of the incubation period for genomic DNA extraction. DNA extraction was performed through a sequence of mechanical lysis (bead beating), phenol-chloroform extraction, ethanol precipitation and purification using PVPP columns (Piché-Choquette *et al.*, 2016). DNA was quantified using Quantifluor dsDNA System® (Promega, Fitchburg, WI, U.S.A.) and Rotorgene 6000® thermocycler (Qiagen, NRW, Germany).

### 4.7.6 Bacterial, archaeal and fungal communities

The V4 region of bacterial and archaeal 16S rRNA gene as well as the ITS2 region of fungal rRNA were PCR-amplified (10.4, Data set 4). Paired-ends sequencing was performed at the Joint Genome Institute (JGI, Walnut Creek, CA, U.S.A.) using Illumina MiSeq system (2x250 bp). Our sequencing analysis pipeline using QIIME 1.8.0 software (Kuczynski *et al.*, 2002) was based on the Itagger pipeline (Piché-Choquette *et al.*, 2016; Tremblay *et al.*, 2015). In total, 8,436,319 bacterial and archaeal as well as 3,391,264 fungal high quality clustered reads remained after quality control. To avoid bias caused by unequal sequencing effort between samples, the amount of sequences of each type of library was subsampled to the library with the least amount of sequences. This standardization resulted in 77,884 bacterial and archaeal sequences (1,925 OTUs) and 53,498 fungal sequences (817 OTUs) per sample (10.1, Data set 1).

### 4.7.7 Metagenomic analysis

Metagenomic libraries were prepared and sequenced on an Illumina HiSeq 2000 system using a 2x150 configuration. Raw sequencing data (256 Gb) were processed through our metagenomic bioinformatics pipelines (4.9.3, Method S3; 10.5, Data set 5) which includes metagenome binning with Metabat v0.26.1 (Kang *et al.*, 2015). Bins obtained from Metabat were further processed/decontaminated by splitting each bin into three sub-bins

based on the assigned taxonomic lineage at the Order level, as each bin typically had a significant amount of contigs associated with the same Order taxon. Genome sequence completeness of sorted bins was estimated to vary between < 0.01 and 89% (10.5, Data set 5D). To increase detection power when searching for hydrogenase genes, we made multiple alignments of all amino acid sequences comprised in an extensive database of [NiFe]-hydrogenase large subunit (Greening *et al.*, 2016) using MUSCLE v.3.8.31 (Edgar, 2004). Hidden Markov Models (named group1, group2, group3 and group4; 10.6, Data set 6) were generated from these alignments with Hmmer v3.1b1 (Eddy, 1998) and gene sequences from metagenome assembly were compared against these training sets with hmmscan (Hmmer v3.1b1). Hits having an e-value ≤ 1e-10, a query length ≥ 100 amino acid residues and an alignment length ≥ 100 were kept, for a total of 122 putative hydrogenase genes. Only sequences containing the canonical L1 or L2 cysteine motifs binding the metal ions of the catalytic site in [NiFe]-hydrogenases (37, 38) were kept. This step was necessary to remove non-specific sequences encoding energy-converting hydrogenase related complexes.

### 4.7.8 Nucleotide sequences accession numbers

Raw sequence reads of rRNA marker genes from PCR amplicons and shotgun metagenomics were deposited in the Sequence Read Archive of the National Center for Biotechnology Information under the Bioprojects PRJNA329645 and PRJNA343121, respectively. Metagenome datasets can be found on the Integrated Microbial Genome and Microbiome Sample (IMG/M) website (https://img.jgi.doe.gov/cgi-bin/m/main.cgi) under the IMG study name “Soil and rhizosphere microbial communities from Centre INRS-Institut Armand-Frappier, Laval, Canada”.

### 4.7.9 Statistical analyses

Statistical analyses were performed using R software v3.1.1 (R Development Core Team, 2014). The impact of $H_2$ treatments on soil physicochemical properties, microbial activities and microbial diversity was tested for each land-use type using one-way analysis of variance for normal distributed variables and a Wilcoxon-Mann-Whitney test for non-normal distributed variables using the package ‘stats’ (R Development Core Team, 2014).

Pairwise comparison of relative abundance of all OTUs in soil microcosms incubated under different $H_2$ exposure treatments was tested using the likelihood ratio test implemented in the package 'edgeR' (Robinson *et al.*, 2010). Unweighted pair group method with arithmetic mean (UPGMA) agglomerative clustering of molecular profiles were computed with the package 'stats' (R Development Core Team, 2014) based on an Euclidean matrix of Hellinger-transformed relative abundance data for ribotyping analysis and a Bray-Curtis distance matrix on non-transformed relative abundance (counts per millions; CPM) for genome bins. Significantly distinct clusters were identified using the Similarity Profile Tool (SIMPROF) implemented in the package 'clustsig' (Clarke *et al.*, 2008) with 999 permutations. Contribution of land-use type and $H_2$ treatments to partition distance matrices was computed with a Permanova using the package 'vegan' (Oksanen *et al.*, 2013). A principal component analysis (PCA) was used to explore soil microcosms partitioning in a reduced space defined by genome bins profile. Equilibrium circle of descriptor with the radius $\sqrt{d/p}$ (where *d* is the number of dimensions of the reduced space: 2 and *p* is the total space: 93) was computed to identify variables significantly contributing to the axes defining the position of soil microcosms. The package 'WGCNA' (Langfelder & Horvath, 2008) was used to analyze covariation among ribotypes and subsequently to provide a potential taxonomic affiliation to genome bins assembled in the metagenomic analysis. All ribotyping libraries were used to generate a single weighted correlation network containing bacteria, archaeal and fungal data. The network comprised 16 samples since HP(b) and LP(c) were removed due to low sequencing effort compared to all other microcosms. Network construction was done as described in Piché-Choquette *et al.* (Piché-Choquette *et al.*, 2016). The soft-thresholding power used was 18 and each pair of modules with less than 0.40 dissimilarity were merged together as a single module. To identify the most probable taxonomic affiliation of genome bins, eigengenes were first computed for all modules. Spearman correlations were computed between module eigengenes and genome bins for taxonomic identification. Modules with the highest module membership for each genome bin were considered to retrieve OTUs with the highest correlation coefficient for each genome bin. The taxonomic affiliation, the relative abundance and the distribution profile of genome bins and OTUs were compared to infer the most probable affiliation for each bin. Multiple regression analyses (stepwise forward

selection of independent variables) were computed with the package 'mfp' (R Development Core Team, 2014) to identify the best predictors for genome bin and OTU relative abundance in soil microcosms. SEM were computed using the package 'Lavaan' (Rosseel, 2012). Missing values of Shannon diversity indices for bacterial and fungal communities in SEM due to PCR amplicons sequencing failures were estimated by linear regression using unpublished RNA-based profiling as independent variables (4.9.1, Method S1).

## 4.8 Acknowledgments

This work has been supported by a Natural Sciences and Engineering Research Council of Canada Discovery grant to PC and by the Community Science Program of the Joint Genome Institute (US Department of Energy) to PC and JT. The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. MK is grateful to the Ministry of Higher Education and Scientific Research of Tunisia for his PhD. scholarship.

## 4.9 Supplementary material

### 4.9.1 Method S1. Description of hypothesized paths constituting the models predicting direct and indirect effects of $H_2$ exposure on measured microbial processes

A structural equation model (SEM) was computed based on the extensive procedure described by Grace (2006) to test the hypothesis that idiosyncratic impact of $H_2$ exposure on soil microbial community structure and function is explained by soil biotic and abiotic features acting as ecological filters for microbial species and functional groups. The overall model was comprised of seven paths involving two composites variables (SOIL and BIO) and two variables ($H_2O$ and $H_2$). Composite variables were computed by linear combination of observed indicators, of which C, N, C\N and pH were combined as a composite variable, named SOIL, and species richness (*i.e.* Shannon index) of bacterial and fungal communities as the composite variable named BIO:

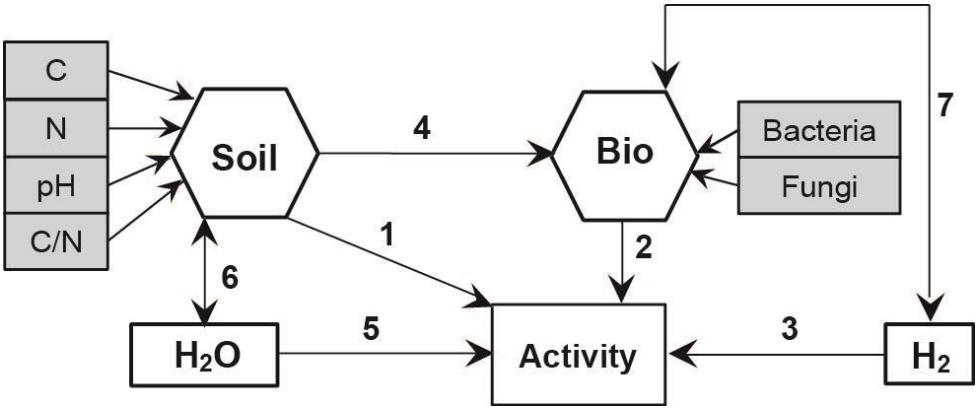



**Figure 4.4 Hypothetical structural equation model**

The hypothetical model was based on a number of assumptions supported by literature to predict trace gas turnover and carbon utilization profiles. These assumptions are described in the following table:

**Table 4.3 Assumptions of the structural equation models**

| Path | Description | Potential mechanism |
|---|---|---|
| 1 | Soil physicochemical effects on microbial activity. | The metabolic activity of microorganisms is influenced by the abiotic conditions of their environment. |
| 2 | Microbial diversity effects on microbial activity. | According to the biodiversity-ecosystem functioning theory, microbial processes are favored by the presence of a high diversity level (herein expressed as species richness). |
| 3 | Direct and indirect effects of $H_2$ exposure on microorganisms catalyzing a specific process. | The results of this study have demonstrated the impact of $H_2$ on microbial metabolism. Direct effects on $H_2$ oxidation rate and indirect effects on $CH_4$ oxidation rate and carbon mineralization potential were observed. The underlying mechanisms are unknown but should involve alteration of microbial community structure (abundance and interactions) induced by HOB favored by $H_2$ exposure. |
| 4 | Impact of soil physicochemical properties on microbial diversity. | Abiotic conditions in soil are known drivers of microbial community structure. |
| 5 | Soil water content effect on microbial activity. | Soil water content influences microbial community structure and function. Under water stress conditions, drought-tolerant microorganisms are favored. Gas diffusion is also a limiting factor for $H_2$ and $CH_4$ oxidation rates measured in soil. Our inability to maintain the exact same water content in soils (*i.e.* lower water content in farmland soil under $eH_2$ treatment when compared to $aH_2$ treatment) justified the addition of this variable in the model (see Table 4.1). |
| 6 | Covariation between soil physicochemical properties and soil water content. | Soil water holding capacity measured in the soil samples representative of the three land-use types defined the amount of water to be added to soil microcosms. |
| 7 | Covariation between microbial diversity and $H_2$ treatment | Different biodiversity levels were expected to occur at different $H_2$ exposure levels. |

Composites variables were estimated using multiple linear regression. The weighting for the relationship between measured and composite variables for the four modeled microbial processes (high affinity $H_2$ oxidation rate, low affinity $H_2$ oxidation rate, $CH_4$ oxidation rate and Ecoplates profile) are shown in the table below. Values reported in the table were used to calculate composite variables estimates that were then included in SEM calculation:

**Table 4.4 Values representing the composite variables**

| **Measured variables** | **Composite variables** | |
|---|---|---|
| | **SOIL** | **BIO** |
| **High affinity $H_2$ oxidation** | | |
| C | -216 | |
| N | -165 | |
| pH | 3797 | |
| C/N | 1171 | |
| Bacteria species richness (Shannon) | | -511 |
| Fungi species richness (Shannon) | | -694 |
| **Low affinity $H_2$ oxidation** | | |
| C | 129 | |
| N | -855 | |
| pH | -423 | |
| C/N | -154 | |
| Bacteria species richness (Shannon) | | 210 |
| Fungi species richness (Shannon) | | -40 |
| **$CH_4$ oxidation** | | |
| C | -10 | |
| N | 81 | |
| pH | 406 | |
| C/N | 78 | |
| Bacteria species richness (Shannon) | | -38 |
| Fungi species richness (Shannon) | | 53 |
| **Ecoplate** | | |
| C | 0.237 | |
| N | -2.295 | |
| pH | -0.381 | |
| C/N | -0.215 | |
| Bacteria species richness (Shannon) | | -0.013 |
| Fungi species richness (Shannon) | | 0.059 |

Due to the lack of two bacterial and four fungal ribotying profiles, missing Shannon indices were derived from RNA-based ribotyping profiles performed on soil samples collected from the microcosms exposed to $eH_2$ and $aH_2$ treatments (RNA-based analyses were obtained from the same incubations reported in the current study but are unpublished as

of yet). Regression analyses between Shannon indices obtained from DNA-based ribotyping profile (independent variable) and Shannon indices obtained from RNA-based ribotyping profile (dependent variable) are shown in the following figures:

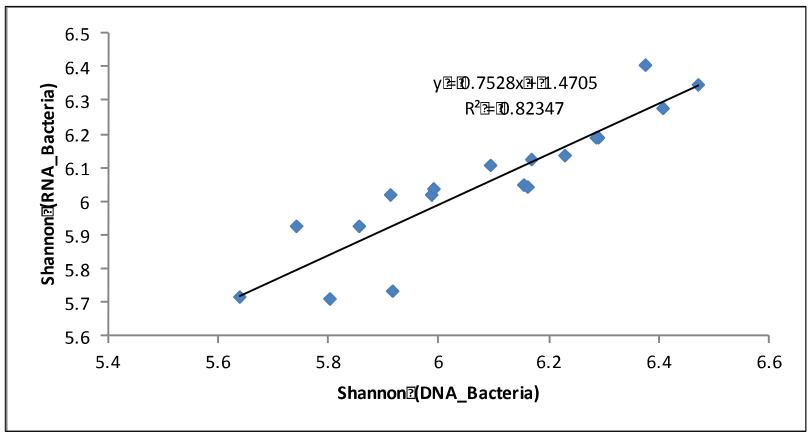



**Figure 4.5** **Regression analysis between bacterial Shannon indices obtained via DNA-based (independent variable) and RNA-based (dependent variable) ribotyping profile**

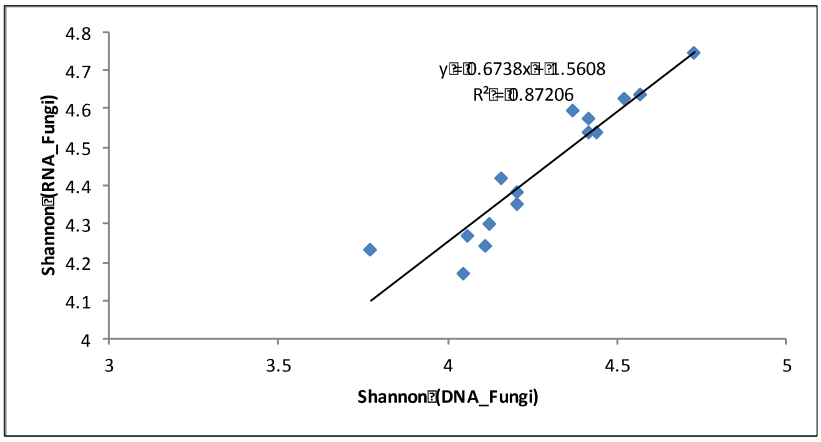



**Figure 4.6** **Regression analysis between fungal Shannon indices obtained via DNA-based (independent variable) and RNA-based (dependent variable) ribotyping profile**

Derived equations were used to estimate the five Shannon index values presented in the following table (ultimately, Shannon index value for microcosm LP(c) (fungi) is still missing since sequencing of both DNA and RNA libraries were unsuccessful):

**Table 4.5 Calculated and estimated Shannon indices of soil samples**

| Sample | Shannon index obtained from RNA-based ribotyping profile | Estimated Shannon index used in SEM |
|---|---|---|
| HP(b) - Bacteria | 6.34 | 6.47 |
| LP(c) - Bacteria | 5.71 | 5.64 |
| HA(b) - Fungi | 4.75 | 4.73 |
| LP(a) - Fungi | 4.64 | 4.57 |
| LP(b) - Fungi | 4.54 | 4.42 |

### 4.9.2 Method S2. Dynamic microcosm chamber unit

The term “dynamic” holds for the continuous air stream circulating in and out of the microcosms, resulting in a user-specified hydraulic residence time and chemical composition of the headspace. The system is comprised of three units: a gas blending station, a gas distribution network and an air sampling station. The gas blending station has been designed to control the dilution of a commercial gas mixture, comprising of 1% $H_2$ in synthetic air (GST-Welco, Pennsylvania, U.S.A), with gas supplied from a Balston zero air generator (Parker Hannifin Corp., QC, Canada). It comprises of two Brooks® Delta II Smart mass flow controllers (0-700 Sccm, accuracy of $\pm$1.0%). The air sampling networks consists of seven Brooks® Delta II Smart mass flow controllers (0-100 Sccm, accuracy of $\pm$1.0%) delivering gas mixture to independent soil microcosms. Microcosms are also vented to the atmosphere through the air sampling station. This station has an option allowing to measure, in real-time, $H_2$ and CO levels when the Trace Analytical ta3000R (Reduction Gas Detector) is connected in series (that option was not used in the present study). Operation of the Dynamic microcosm chamber unit is under the control of a programmable logic controller (Allen-Bradley® PanelView Plus 600) comprising of a digital display of system status and alarms for defaults in flowrate, gas supply and power.

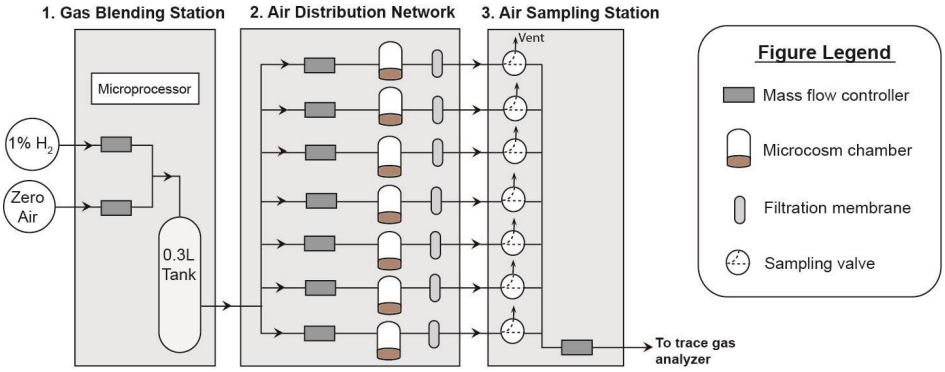



**Figure 4.7 Schematic of the dynamic microcosm chamber unit setup**

### 4.9.3 Method S3. Metagenomic analysis pipeline

Metagenomic libraries were prepared and sequenced on an Illumina HiSeq2000 system on a 2x150 configuration. Sequencing raw data (256 Gb for metagenome) were processed through our metagenomics bioinformatics pipeline. Read count summaries are provided for metagenome sequencing libraries (10.5, Data set 5A). Sequencing adapters were removed from each read (Trimmomatic v0.32) (Bolger *et al.*, 2014) to generate quality controlled (QC) reads. QC-passed reads from each sample were assembled into a large metagenome assembly using Ray software v2.3.1 (Boisvert *et al.*, 2012) with a kmer size of 31 (10.5, Data set 5B). Gene prediction of obtained contigs was performed by calling genes on each assembled contig using MetageneMark v1.0 (Tang *et al.*, 2013). Genes were annotated following the JGI's guidelines (Huntemann *et al.*, 2016) : 1) RPSBLAST (v2.2.29+) (Camacho *et al.*, 2009) against COG database (*e.g.* CDD v3.11); 2) RPSBLAST (v2.2.29+) against KOG database (*e.g.* CDD v3.11). The best hit having at an e-value ≥ 1e-02 was kept for each query; 3) HMMSCAN (v3.1b1) (Eddy, 2011) against PFAM-A (v27.0) database (Finn *et al.*, 2014) best hit having at least an e-value ≥ 1e-02 was kept for each query; 4) TIGRFAM database (v15.0) best hit having at least an e-value ≥ 1e-02 was kept for each queries; 5) BLASTP (v2.2.29+) against KEGG database v71.0, and 6) BLASTN (v2.2.29+) against NCBI`s nucleotide (nt) database (version of May 16th 2013). Contigs (and not genes) sequences were blasted against NCBI's nt database as well for taxonomic assignment. For each of these databases comparisons, the best hit having at least an e-value ≥ 1e-02 was kept for each query.

QC-passed reads were mapped (BWA mem v0.7.10) (unpublished - http://bio-bwa.sourceforge.net) against contigs to assess quality of metagenome assembly and to obtain contigs abundance profiles. Alignment files in bam format were sorted by read coordinates using samtools v1.1 and only properly aligned read pairs were kept for downstream steps. Each bam file (containing properly aligned paired-reads only) was analyzed for coverage of called genes and contigs using bedtools (v2.17.0) (Quinlan & Hall, 2010) using a custom bed file representing gene coordinates of each contig. Only paired-reads both overlapping their contigs or genes were considered for gene counts. Coverage profiles of each sample were merged together to generate an abundance matrix (rows = contig, columns = samples) for which a corresponding CPM (Counts Per Million) abundance matrix (edgeR v3.10.2) (Robinson *et al.*, 2010) was generated as well.

Taxonomy of each contig was assigned using the NCBI taxonomy database (Benson *et al.*, 2013; Sayers *et al.*, 2009) (ftp://ftp.ncbi.nih.gov/pub/taxonomy/taxdump.tar.gz) (as downloaded on June 5th, 2015). Each GIs resulting from BLASTN against nt were used to retrieve full taxonomic lineages (when available) from the NCBI taxonomy database. Taxonomic lineages were integrated to the contig abundance of read counts matrix to generate an OTU table format file (with contigs replacing OTUs as rows). Taxonomic summaries were performed using a combination of in-house Perl and R scripts and Qiime v.1.9.0 (Caporaso *et al.*, 2010). Genome bins abundance tables along with their taxonomic lineages are included in Data set 5C (10.5).

Binning was done using Metabat (v0.26.1) (Kang *et al.*, 2015) using an abundance matrix generated using the jgi_summarize_bam_contig_depths software (Kang *et al.*, 2015) with --minContigLength 1000 --minContigDepth 2 and --minContigIdentity 95 parameters. Genome bins obtained from Metabat were further processed/decontaminated by splitting each bin into three sub-bins based on the assigned taxonomic lineage at the Order level, as each bin typically had a significant amount of contigs associated with the same Order taxon. For instance, in our dataset, the bin labeled “1” had 607 contigs assigned to the *Rhizobiales* Order level, 136 contigs to the *Burkholderiales* and 189 more contigs to an undefined taxonomy value. Consequently, 3 sub-bins were generated and labeled Bin-1

(*Rhizobiales*), Bin-1 (*Burkholderiales*) and Bin-1 (NULL) respectively. All genome bins were characterized using CheckM v1.0.4 (Parks *et al.*, 2015) and detailed in Data set 5D (10.5).

**Table 4.6** **Absence of $CH_4$ production in soil microcosms. The absence of methane production in soil microcosms was tested in poplar and farmland soils (soils where difference in $CH_4$ oxidation rates were most prominent). For that purpose, microcosms were flushed for 15 minutes with a synthetic gas mixture (Zero Air, Praxair distribution Inc. PA, USA) containing non-detectable $CH_4$ (detection limit "DL" of the Agilent GC system below 0.2 ppmv $CH_4$). $CH_4$ level was then monitored for more than 10 hours (compared to 8 hours for $CH_4$ oxidation measurements reported in the article). As no $CH_4$ was detected, we also monitored $CO_2$ as an analytical control of the flame ionization detector (FID) of the gas chromatograph. Average values of three independent replicates are shown with standard deviations.**

| | **$aH_2$ microcosms** | | | | **$eH_2$ microcosms** | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Concentrations (ppmv)** | **$CO_2$** | | **$CH_4$** | | | **$CO_2$** | | **$CH_4$** | |
| **Time (min)** | **Poplar** | **Farmland** | **Poplar** | **Farmland** | **Time (min)** | **Poplar** | **Farmland** | **Poplar** | **Farmland** |
| **0** | 18 (10) | 18 (4) | < DL | < DL | **0** | 4 (1) | 4 (1) | < DL | < DL |
| **66** | 238 (38) | 239 (74) | < DL | < DL | **55** | 49 (4) | 40 (2) | < DL | < DL |
| **132** | 464 (71) | 421 (130) | < DL | < DL | **110** | 96 (5) | 75 (3) | < DL | < DL |
| **802** | 2416 (361) | 1842 (510) | < DL | < DL | **630** | 504 (28) | 398 (21) | < DL | < DL |

**Table 4.7** **List of genome bins for which the distribution was influenced by $H_2$ exposure. Log(FC) = Log-fold-change, number of times this OTU is more/less abundant in $H_2$ treatment. Log(CPM) = Log-fold-change in counts per million. LR = Likelihood ratio or how many times more likely the OTU is more/less abundant in a treatment than the other. P-value = significance level (only p-values < 0.05 were kept). = most abundant in 10,000 ppmv $H_2$ treatment. = most abundant in 0.5 ppmv $H_2$ treatment.**

| FARMLAND | Log(FC) | Log(CPM) | LR | P-value | Response to $H_2$ |
|---|---|---|---|---|---|
| Bin-30 (*Burkholderiales*) | 0.72 | 12 | 5 | 0.03 | 0.71523641 |
| Bin-6 (NULL) | 0.94 | 13 | 8 | 0.004 | 0.93936962 |
| Bin-3 (*Xanthomonadales*) | 1.1 | 16 | 11 | 0.0007 | 1.06909082 |
| Bin-6 (*Xanthomonadales*) | 1.2 | 16 | 13 | 0.0003 | 1.16462972 |
| Bin-2 (*Burkholderiales*) | 2.0 | 12 | 32 | 1.84E-08 | 1.95819705 |
| Bin-2 (*Xanthomonadales*) | 2.1 | 17 | 41 | 1.24E-10 | 2.13231989 |
| | | | | | |
| LARCH | Log(FC) | Log(CPM) | LR | P-value | Response to $H_2$ |
| Bin-21 (NULL) | -3.9 | 11 | 17 | 4.19E-05 | -3.87548775 |
| | | | | | |
| POPLAR | Log(FC) | Log(CPM) | LR | P-value | Response to $H_2$ |
| Bin-3 (*Sphingomonadales*) | -1.7 | 12 | 4 | 0.048 | -1.74271411 |
| Bin-2 (*Xanthomonadales*) | -0.5 | 15 | 5 | 0.03 | -0.52503425 |
| Bin-7 (*Xanthomonadales*) | -0.7 | 16 | 10 | 0.002 | -0.7406181 |
| Bin-6 (*Burkholderiales*) | -1.4 | 13 | 22 | 2.10E-06 | -1.44081785 |
| Bin-3 (NULL) | -1.4 | 14 | 24 | 1.18E-06 | -1.3595836 |
| Bin-6 (*Xanthomonadales*) | -1.6 | 16 | 41 | 1.71E-10 | -1.5638836 |
| Bin-3 (*Xanthomonadales*) | -1.6 | 16 | 44 | 3.11E-11 | .63680157 |

**Table 4.8** **Multiple linear regression equations defining the relative abundance of (A) genome bins outlined in the PCA and the corresponding (B) OTUs (rRNA marker gene amplicon sequencing) according to biotic and abiotic parameters. The interaction between the variables "$H_2$ treatment" and "net $CO_2$ production" was considered due to the ability of HOB to fix $CO_2$, resulting in a significant decrease in net $CO_2$ production rates in e$H_2$ treatments ($CO_2$ alone displayed non-significant explanatory power).**

| **A** | **Equations** | **$R^2$ (p-value)** | **Residual standard error (ε)** |
|---|---|---|---|
| Bin-1 (Rhizobiales) | $= -13\times10^3 (\pm1.7\times10^3)$ pH $+ 77\times10^3 (\pm8.5\times10^3) + \varepsilon$ | 0.78 (0.0001) | 1379 |
| Bin-2 (Xanthomonadales) | $= 2.7\times10^4 (\pm0.65\times10^4)$ pH $- 2.1\times10^{-2} (\pm5.9\times10^{-3})$ $CO_2$ x Treatment $-1.3\times10^5 (\pm3.2\times10^4) + \varepsilon$ | 0.78 (0.0001) | 5091 |
| Bin-3 (Xanthomonadales) | $= -1.6\times10^3 (\pm2.7\times10^2)$ C/N $- 1.1\times10^{-2} (\pm1.1\times10^{-3})$ $CO_2$ x Treatment $+ 1.9\times10^4 (\pm3.1\times10^3) + \varepsilon$ | 0.95 (0.0001) | 977 |
| Bin-6 (Xanthomonadales) | $= 1.6\times10^3 (\pm3.0\times10^2)$ C/N $- 9.7\times10^{-3} (\pm1.3\times10^{-3})$ $CO_2$ x Treatment $+ 1.9\times10^4 (\pm3.4\times10^3) + \varepsilon$ | 0.92 (0.0001) | 1072 |
| **B** | **Equations** | **$R^2$ (p-value)** | **Residual standard error (ε)** |
| Bin-1 (OTU *Bradyrhizobium*) | $= -0.041 (\pm0.003)$ pH $+ 0.245 (0.016) + \varepsilon$ | 0.93 ($2.8 \times 10^{-9}$) | $2.5 \times 10^{-3}$ |
| Bin-1 (OTU *Rhodoplanes*) | $= -0.033 (\pm0.003)$ pH $+ 0.190 (0.013) + \varepsilon$ | 0.92 ($4.6 \times 10^{-9}$) | $2.1 \times 10^{-3}$ |
| Bin-2 (OTU *Luteimonas*) | $= 7.62\times10^{-2} (\pm1.6\times10^{-2})$ pH $- 8.7\times10^{-8} (\pm1.6\times10^{-8})$ $CO_2$ x Treatment $- 3.9\times10^{-1} (\pm7.8\times10^{-2}) + \varepsilon$ | 0.85 ($1.6 \times 10^{-4}$) | $1.2 \times 10^{-2}$ |
| Bin-3 (OTU *Lysobacter*) | $= -4.0\times10^{-3} (\pm1.2\times10^{-3})$ C/N $- 5.3\times10^{-8} (\pm5.7\times10^{-9})$ $CO_2$ x Treatment $+ 2.1\times10^{-2} (\pm1.3\times10^{-2}) + \varepsilon$ | 0.98 ($6.6 \times 10^{-9}$) | $4.1 \times 10^{-3}$ |
| Bin-6 (OTU Xanthomonadaceae) | $= -5.3\times10^{-3} (\pm1.5\times10^{-3})$ C/N $- 3.7\times10^{-8} (\pm7.2\times10^{-9})$ $CO_2$ x Treatment $+ 5.0\times10^{-2} (\pm1.7\times10^{-2}) + \varepsilon$ | 0.89 ($3.9 \times 10^{-5}$) | $5.3 \times 10^{-3}$ |

**Table 4.9** **Genome bins of potential HOB containing hydrogenase gene based on search in metagenomic annotation databases (1 gene) and [NiFe]-hydrogenase HMM (2 genes). The table shows the relative abundance of genome bins (counts per million or CPM, mean and standard deviation) for triplicate microcosms exposed to $aH_2$ and $eH_2$ treatments. The asterisks (*) represent genome bins displaying a significant difference ($\alpha < 0.05$) between the two $H_2$ treatments (Wilcoxon-Mann-Whitney test).**

| Genome bins | Soil microcosms | | | | | |
|---|---|---|---|---|---|---|
| | Poplar monoculture | | Farmland | | Larch monoculture | |
| | $eH_2$-P | $aH_2$-P | $eH_2$-A | $aH_2$-A | $eH_2$-L | $aH_2$-L |
| Bin-1 (Rhizobiales)[A] | 11371 (665) | 11773 (871) | 10824 (1061) | 12593 (1032) | 17021 (1920) | 16856 (1901) * |
| Bin-4 (NULL) | 1102 (28) * | 1273 (110) * | 1205 (92) | 1316 (102) | 906 (35) | 865 (27) |
| Bin-25 (NULL) | 3770 (75) * | 3576 (129) * | 3014 (31) * | 3175 (40) * | 3722 (68) * | 4011 (105) * |

[A]A gene fragment encoding for the small subunit of a group 1 [NiFe]-hydrogenase was found in metagenomic annotation database.

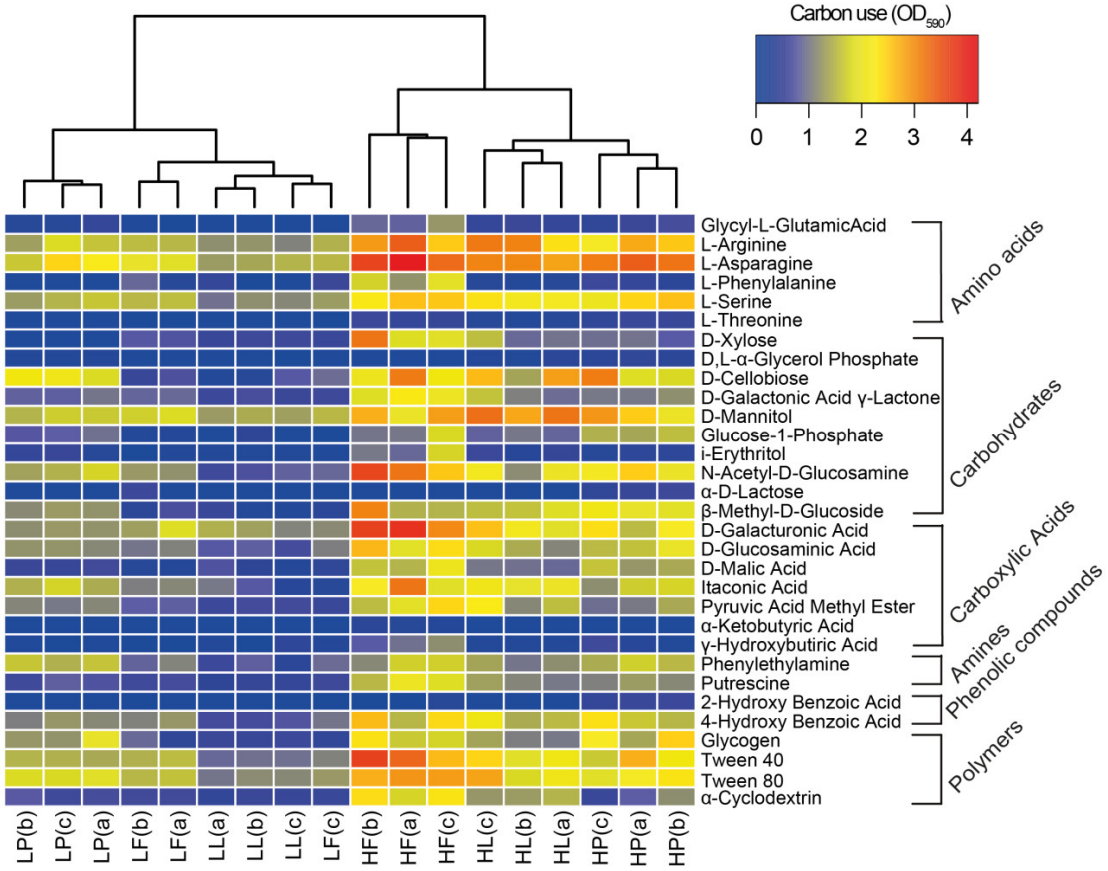



**Figure 4.8** **Community-level carbon substrates utilization profiling. This heatmap reflects the utilization of 31 carbon sources divided into 6 subcategories. The upper part of the graph consists of an UPGMA agglomerative clustering of the profiles, showing a clear dichotomy between both $H_2$ treatments.**

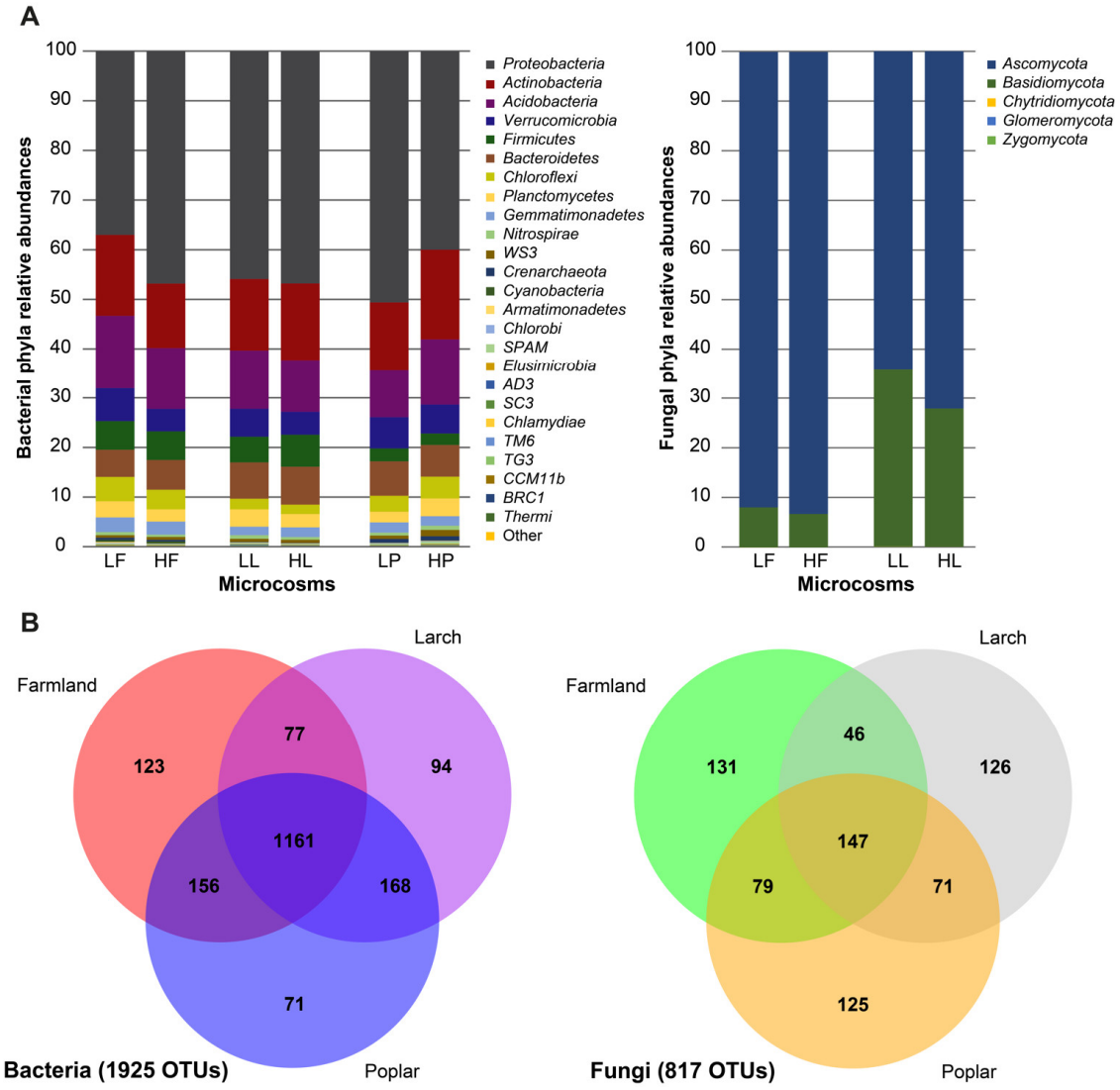



**Figure 4.9** **Overview of the taxonomic composition of bacterial and fungal communities in soils. (A) Relative abundance of bacteria and fungi at the phylum level. (B) Venn diagram showing the distribution of bacteria and fungi OTUs in the three soils.**

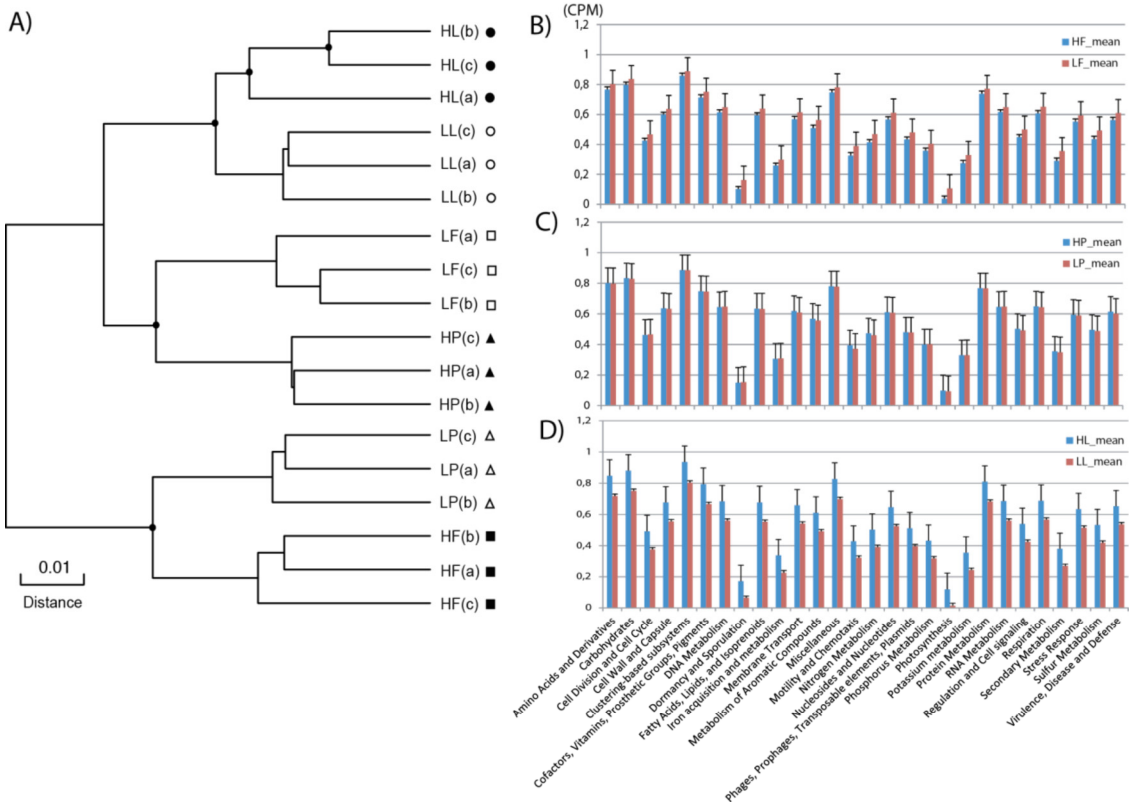



**Figure 4.10** **Metagenomic analysis using MG-RAST pipeline. (A) UPGMA agglomerative clustering of soil microcosms according to a Euclidean distance matrix calculated with a Hellinger distance matrix of annotated gene abundance profiles. Gene annotation (COG database) was performed using unassembled sequenced reads with the pipeline MG-RAST. Land-use types are distinguished by three different symbols (square; farmland, circle; larch and triangle; poplar). Black symbols indicate soil microcosms exposed to $eH_2$ treatment and white symbols indicate soil microcosms exposed to $aH_2$ treatment. The Barplots show distribution of sequences (CPM values) in COG gene categories for (B) farmland, (C) larch and (D) poplar soil exposed to $eH_2$ and $aH_2$.**

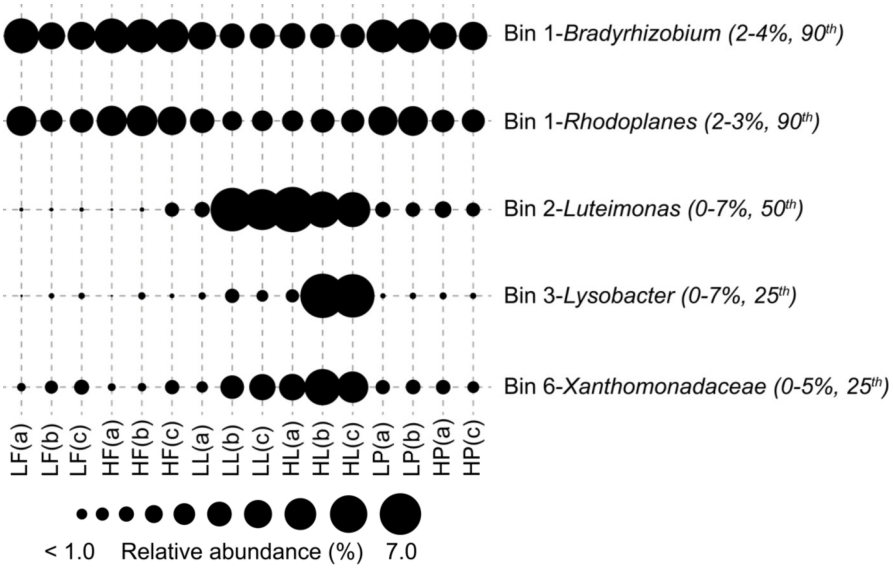



**Figure 4.11** **Bubble-chart representation of OTUs associated with the 4 genome bins of interest. Correspondence between genome bins and OTUs was achieved through correlation network analysis. Dots size is proportional to the relative abundance of OTUs in the 16S rRNA gene amplicon sequencing. Genome bins are presented along with taxonomic affiliation retrieved from the 16S rRNA gene amplicon sequencing analysis. The relative abundance and the rank of the connectivity of each OTU are written in parentheses.**

# 5 SURVEY OF HIGH-AFFINITY $H_2$-OXIDIZING BACTERIA IN SOIL REVEALS THEIR VAST DIVERSITY YET UNDERREPRESENTATION IN GENOMIC DATABASES

**Titre français**: L'étude de bactéries du sol oxydant l'hydrogène avec haute affinité révèle leur vaste diversité ainsi que leur sous-représentation dans les bases de données génomiques

**Article publié en ligne le 17 juin 2017 dans le journal Microbial Ecology**

*DOI: 10.1007/s00248-017-1011-1* (Piché-Choquette *et al.*, 2017)

**Auteurs**

Sarah Piché-Choquette[1], Mondher Khdhiri[1] et Philippe Constant[1]✉

[1]INRS-Institut Armand-Frappier, 531 boulevard des Prairies, Laval (Québec), Canada, H7V 1B7

✉ Auteur correspondant

**Contribution des auteurs**

Sarah Piché-Choquette: Conception des expériences, incubations, analyses en laboratoire, analyse des résultats, écriture de l'article

Mondher Khdhiri: Conception des expériences, incubations

Philippe Constant: Conception des expériences, analyse des résultats, écriture de l'article

## 5.1 Résumé en français

Bien que les bactéries oxydant l'hydrogène avec une haute affinité (HA-HOB) représentent le principal puits d'$H_2$ atmosphérique, l'écologie de ce groupe fonctionnel spécialiste est plutôt inconnue pour cause de leur récente découverte. Le but principal de cette étude est de fournir la première étude exhaustive des HA-HOB dans des sols agricoles, de mélèzes et de peupliers exposés à 0,5 ou 10 000 ppmv $H_2$. À l'aide de dosages par qPCR et qRT-PCR ainsi que du séquençage à haut débit d'amplicons de PCR du gène *hhyL*, codant pour la grande sous-unité des hydrogénases à haute affinité (HAH), nous avons constaté que le ratio d'expression d'*hhyL* expliquait davantage les variations de vitesses d'oxydation d'$H_2$ que la richesse spécifique des HA-HOB. Le carbone, l'azote, le pH et la richesse spécifique bactérienne représentent les plus forts indicateurs de la structure de la communauté des HA-HOB. Notre étude souligne aussi la nécessité de cultiver des HA-HOB compte tenu de leur représentation quasi-nulle dans les bases de données génomiques actuelles.

## 5.2 Abstract

While high-affinity $H_2$-oxidizing bacteria (HA-HOB) serve as the main sink of atmospheric $H_2$, the ecology of this specialist functional group is rather unknown due to its recent discovery. The main purpose of our study is to provide the first extensive survey of HA-HOB in farmland, larch and poplar soils exposed to 0.5 and 10,000 ppmv $H_2$. Using qPCR and qRT-PCR assays along with PCR amplicon high-throughput sequencing of *hhyL* gene encoding for the large subunit of high-affinity [NiFe]-hydrogenases (HAH), we found that *hhyL* gene expression ratio explained better variation in measured $H_2$ oxidation rates than HA-HOB species richness. Carbon, nitrogen, pH and bacterial species richness appeared as the most important drivers of HA-HOB community structure. Our study also highlights the need to cultivate HA-HOB due to the huge gap in current genomic databases.

## 5.3 **Main paper**

State-of-the-art atmospheric chemistry-transport models indicate that nearly 70% of atmospheric $H_2$ losses are caused by microbial $H_2$-oxidizing metabolism in upland soils (Pieterse *et al.*, 2013). Microbes accountable for this crucial ecosystem service remained unknown until *Streptomyces* spp. harboring high-affinity [NiFe]-hydrogenases (HAH) were discovered (Constant *et al.*, 2008; Liot & Constant, 2016). Genome database mining and measurement of $H_2$ uptake in pure cultures demonstrated that other Actinobacteria such as *Mycobacterium* spp. (Greening *et al.*, 2014a) and *Rhodococcus* spp. (Meredith *et al.*, 2014), and Acidobacteria (Greening *et al.*, 2014a; Myers & King, 2016), have the ability to scavenge trace levels of $H_2$. The ecology of this functional group is fairly unknown, yet soil moisture, carbon content and pH were identified as key drivers of high-affinity $H_2$ oxidation in soils (Gödde *et al.*, 2000). While soil moisture limits $H_2$ diffusion (Smith-Downey *et al.*, 2006), the stimulating effect of carbon and pH on $H_2$ uptake isn't fully understood. They both act as ecological niche filters selecting or activating high-affinity $H_2$-oxidizing bacteria (HA-HOB), resulting in higher uptake rates in forests than in grassland and farmland ecosystems (Ehhalt & Rohrer, 2009). Our study aims to provide the first extensive survey of HA-HOB in soil and relate their diversity profile to soil $H_2$ uptake rate and physicochemical properties. We used total genomic DNA and RNA extracts from a previous study in which soil samples encompassing three land-use types were exposed to either ambient $H_2$ ($aH_2$; 0.5 ppmv $H_2$) or elevated $H_2$ ($eH_2$; 10,000 ppmv $H_2$) treatments during 15 days in a dynamic microcosms chamber unit (Khdhiri *et al.*, 2017). Briefly, soil samples were collected in a 12-year-old larch plantation, a 10-year-old poplar plantation and a farmland consisting of potato and maize rotation. The dynamic microcosms chamber unit consisted of a gas mixer and a gas distribution network that continually supplied soil microcosms with synthetic air mixtures, containing precise $H_2$ stoichiometric ratios, and vented them to the atmosphere. Distribution of HA-HOB was examined through qPCR/qRT-PCR and PCR amplicon sequencing of *hhyL* gene, encoding for the large structural subunit of HAH. We hypothesized that bacterial species richness varies in tandem with HA-HOB species richness and that land-use defines HA-HOB abundance and diversity.

**Table 5.1** **HA-HOB species richness and quantification of *hhyL* gene and transcript across $H_2$ treatments and land-uses. Mean and standard deviation (in parentheses) values computed from three independent replicates are presented.**

| Variables | Farmland | | Larch Monoculture | | Poplar Monoculture | |
|---|---|---|---|---|---|---|
| | $aH_2$ | $eH_2$ | $aH_2$ | $eH_2$ | $aH_2$ | $eH_2$ |
| **Number of observed OTUs** | 261 | 273 | 240 | 242 | 270 | 238 |
| **ACE richness estimator** | 266 (7) | 278 (7) | 249 (7) | 245 (7) | 279 (7) | 240 (7) |
| **Chao1 richness estimator** | 267 (5) | 279 (4) | 255 (10) | 248 (5) | 282 (8) | 241 (3) |
| **Shannon richness index** | 3.14 (0.01) | 3.06 (0.27) | 3.20 (0.05) | 3.11 (0.11) | 3.13 (0.04) | 3.09 (0.08) |
| **Pielou eveness index** | 0.62 (0.01) | 0.60 (0.05) | 0.65 (0.02) | 0.63 (0.02) | 0.62 (0.02) | 0.63 (0.01) |
| ***hhyL* gene (copies $g^{-1}_{(dw)}$)** | $1.49 \cdot 10^8$ ($2.43 \cdot 10^7$) | $7.07 \cdot 10^8$ ($1.43 \cdot 10^8$) | $9.30 \cdot 10^8$ ($2.29 \cdot 10^8$) | $5.39 \cdot 10^8$ ($2.56 \cdot 10^8$) | $1.48 \cdot 10^8$ ($1.10 \cdot 10^8$) | $1.09 \cdot 10^9$ ($1.26 \cdot 10^8$) |
| ***hhyL* transcript (copies $g^{-1}_{(dw)}$)** | $2.69 \cdot 10^{11}$ ($8.24 \cdot 10^{10}$)[bc] | $5.10 \cdot 10^8$ ($2.10 \cdot 10^8$)[bc] | $1.16 \cdot 10^{10}$ ($8.56 \cdot 10^8$)[bc] | $3.67 \cdot 10^5$ ($1.35 \cdot 10^5$)[bc] | $6.98 \cdot 10^{11}$ ($2.79 \cdot 10^{11}$)[bc] | $4.09 \cdot 10^9$ ($1.48 \cdot 10^9$)[bc] |
| ***hhyL* gene expression ratio[d]** | $3.45 \cdot 10^2$ ($1.70 \cdot 10^2$)[bc] | 2.59 ($2.38 \cdot 10^{-1}$)[bc] | $1.44 \cdot 10^1$ (1.28)[bc] | $4.71 \cdot 10^{-4}$ ($2.73 \cdot 10^{-4}$)[bc] | $1.66 \cdot 10^4$ ($1.61 \cdot 10^2$)[bc] | 3.42 (2.30)[bc] |
| **High-affinity $H_2$ oxidation rate (pmol $g^{-1}_{(dw)}$ $h^{-1}$)[a]** | 361 (87)[bc] | 22 (2)[bc] | 1045 (250)[bc] | 47 (18)[bc] | 1003 (45)[bc] | 59 (0)[bc] |

[a] Published data (Khdhiri *et al.*, 2017)

[b] Variable showing significant between land-use types (ANOVA, $p<0.05$)

[c] Variable showing significant difference between $H_2$ exposure treatments (ANOVA, $p<0.05$)

[d] The expression ratio of *hhyL* gene was computed by dividing the absolute number of transcripts by the absolute number gene as detected by qRT-PCR and qPCR, respectively

The abundance of *hhyL* gene estimated by qPCR ranged from $10^8$ to $10^9$ gene copies $g^{-1}_{(dw)}$ (Table 5.1), akin to previous soil surveys (Constant *et al.*, 2011b). Single linear regression analysis showed that *hhyL* expression ratios (number of transcripts divided by the number gene detected by qRT-PCR and qPCR) explained variations in high-affinity

$H_2$ oxidation rates ($R^2$=0.45, p=0.04, n=18), which is comparable to the resolution of a previous approach relying on the relative abundance of HA-HOB estimated by *hhyL* and 16S rRNA genes qPCR analysis (Khdhiri *et al.*, 2015). On the other hand, neither *hhyL* gene transcripts ($R^2$=0.21, p >0.05) nor *hhyL* gene counts ($R^2$=0.00, p=0.98) alone reflected measured activities. Sequences of PCR-amplified *hhyL* gene obtained from Illumina MiSeq (2 X 250 bp configuration) sequencing platform were clustered into 664 OTUs using a 99% identity cut-off (5.4, Supplemental Method 1; 11.1, Table S1). This stringent threshold was selected as a trade-off to delineate HA-HOB species, considering both unavoidable technical limitations (*i.e.* sequencing error rate (Glenn, 2011) and potential errors introduced by paired-ends assembly (Magoc & Salzberg, 2011)) and frequent lateral transfers resulting in considerable underestimations of HA-HOB richness (Constant *et al.*, 2011b). An extensive database of [NiFe]-hydrogenase large subunit genes was used for phylogenetic analysis of *hhyL* OTUs (5.4, Supplemental Method 1). Consensus phylogenetic tree of reference sequences displays the expected clusters outlining [NiFe]-hydrogenases encompassing groups 1 to 5 (Figure 5.1A). All OTUs were accurately assigned to group 5 cluster, establishing the specificity of the oligonucleotides targeting *hhyL* gene. However, only nine OTUs were closely associated with reference sequences (11.3, Figure S1). The most abundant OTU genotype from our dataset (OTU_1, 20-40% abundance) has an uncultivated clone retrieved from deciduous forest soil in Germany as its closest neighbor (clone MS12, 100% similarity). Similarly, OTU_858 was classified in a cluster comprising environmental clones LS15 and LS27 (97% similarity) detected in Finnish peatlands, although seven other OTUs were affiliated to clusters comprising known bacterial species. OTU_294 was related to *Rhodococcus sp*. (89% similarity) and OTU_237 encompassed a cluster comprising *Streptosporangium roseum* and the environmental clone MS20 (92% similarity) originating from deciduous forest soil in Germany. OTU_81 and OTU_423 were affiliated to *Streptomyces* spp. (97-99% similarity), while OTU_869, OTU_1305 and OTU_1312 were related to *Conexibacter woesei* (76-81% similarity). Considering this, all HA-HOB harboring *hhyL* genes within this survey remain undiscovered.

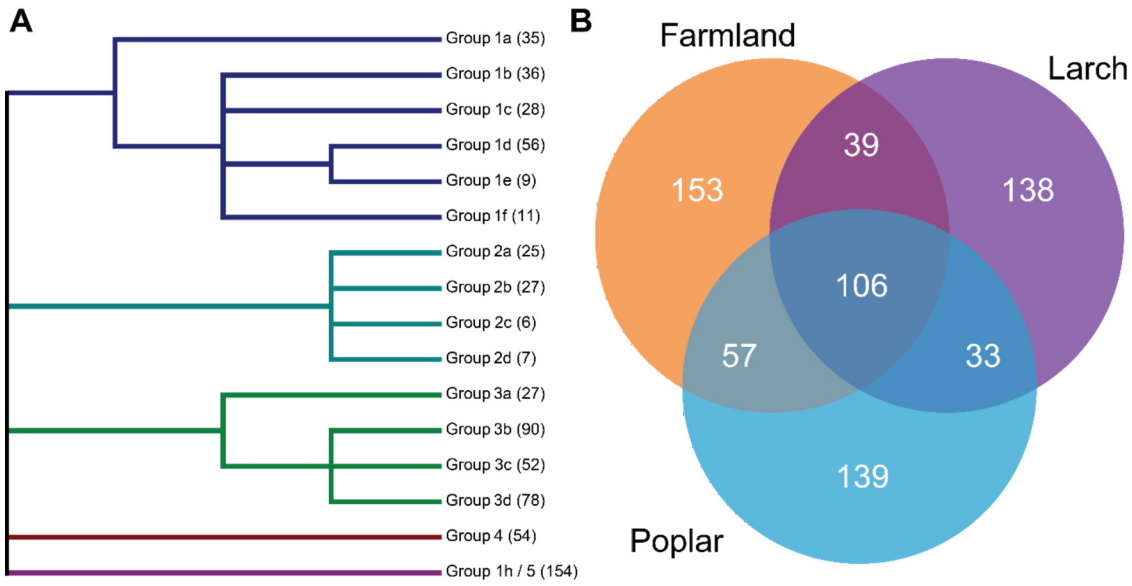



**Figure 5.1** **Classification of *hhyL* OTUs across A) a consensus phylogeny analysis and B) the three land-use types. Consensus phylogenetic tree of translated amino acids sequences from near-complete genes encoding the large subunit of [NiFe]-hydrogenases of group 1 through 5 retrieved from public databases. The consensus tree was generated from the combination neighbor-joining, maximum likelihood and maximum parsimony phylogenetic trees using the "ape" R package (Paradis *et al.*, 2004). Only nodes supported by over 75% of bootstrap replications (out of 500 bootstraps) from all 3 initial trees are shown. Sequences with 100% identity at the amino acid level were removed until only one identical representative remained. Numbers between parentheses represent the number of sequences encompassing each genotype. The Venn diagram represents the number of OTUs present in each combination of land-use types.**

The number of observed OTUs was comparable to species richness estimators (Table 5.1), indicating sufficient sequencing effort to cover HA-HOB diversity. Neither pairwise correlation between 16S rRNA and *hhyL* genes richness (Pearson, P=0.4788) nor evenness (Pearson, P=0.0708) showed significant covariation, suggesting that HA-HOB diversity is independent of bacterial species richness in soil. Furthermore, $H_2$ treatments caused no significant alteration in HA-HOB richness or evenness (Table 5.1). Similarly, the impact of e$H_2$ exposure on *hhyL* OTU distribution profiles was undiscernible at the community level (PERMANOVA, p=0.254). Nevertheless, differential analysis of OTU count data (5.4, Supplemental Method 1) indicated that 27 OTUs were affected by e$H_2$ exposure, mostly with higher relative abundance in e$H_2$ treated soils, with contradicting patterns between land-uses (11.2, Table S2). This agrees with previous observations in

which microbial response to $eH_2$ exposure was idiosyncratic due to the most influential role of abiotic factors such as carbon, nitrogen and pH on microbial communities (Khdhiri *et al.*, 2017; Piché-Choquette *et al.*, 2016). Genotypes showing higher abundance under $eH_2$ could be representatives of HOB possessing both HAH (*i.e.* group 5 hydrogenase) and low-affinity [NiFe]-hydrogenases (*e.g.* group 3 hydrogenase), like *Mycobacterium smegmatis* (Berney *et al.*, 2014). Indeed, multiple hydrogenases provide greater flexibility to switch between survival mixotrophy supported by HAH, at low $H_2$ mixing ratios, and lithoautotrophic growth under $eH_2$ mixing ratios. Bacteria possessing only HAH are expected to be hindered by $eH_2$ exposure (Conrad *et al.*, 1983b). Although the exact biochemical mechanism is unknown, such potential substrate inhibition on HA-HOB metabolism was supported by lower *hhyL* gene expression ratio in $eH_2$-exposed soils (Table 5.1).

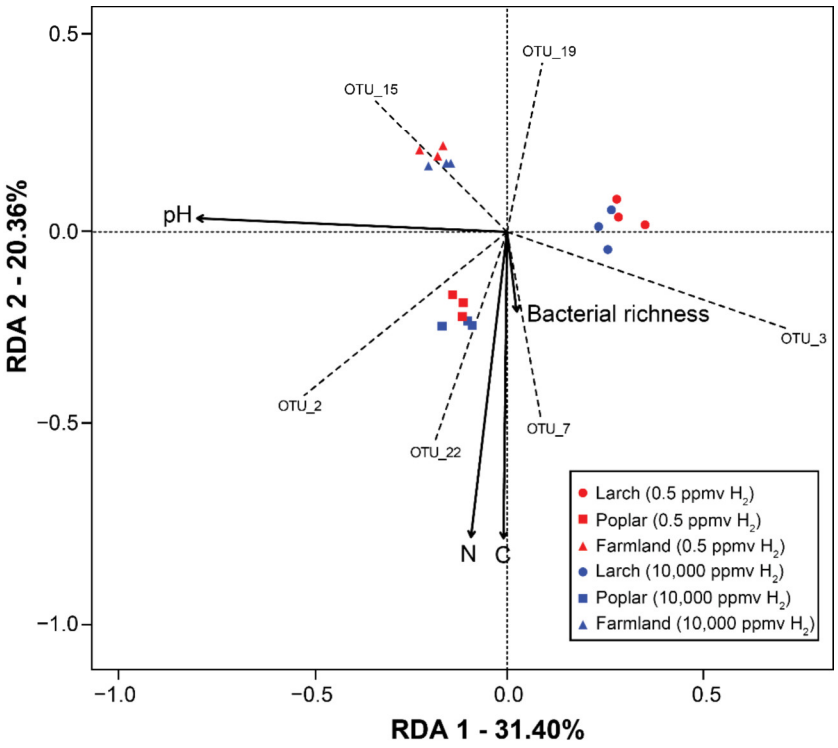



**Figure 5.2** **Parsimonious RDA representing the distribution of soil microcosms in a reduced space according to their *hhyL* OTUs distribution profiles constrained with environmental variables. Land-uses are distinguished by three different symbols (circle; larch, square; poplar, triangle; farmland). Red symbols indicate soil microcosms exposed to $aH_2$ (0.5 ppmv) and blue symbols indicate soil microcosms exposed to $eH_2$ (10,000 ppmv). Only OTUs whose relative abundance exert more**

**weight than the average to separate land use types in terms of *hhyL* genes profiles are depicted.**

Neither HA-HOB species richness nor evenness were distinguishable across the three land-use types (Table 5.1). In contrast to $H_2$ exposure treatments, land-use was a significant driver of HA-HOB community structure, explaining 62% of the variance in *hhyL* gene distribution (PERMANOVA, $p<0.001$). Most OTUs were land-use-specific (Figure 5.1B), yet 106 OTUs common to the three land-uses were most abundant, since they represented up to 88% of *hhyL* genes in any given sample. A parsimonious redundancy analysis (RDA) was then computed to pinpoint which biotic and abiotic parameters measured in soil microcosms (Khdhiri *et al.*, 2017) acted as environmental filters for *hhyL* gene distribution profiles. Nearly 50% variance in HA-HOB community structure was explained by soil carbon and nitrogen contents, pH and bacterial species richness (Figure 5.2). Several OTUs also discriminated land-uses in the reduced space, including OTU_15 which was more abundant in farmland soils. The important contribution of carbon and pH in explaining HA-HOB community structure complies with previous studies investigating environmental drivers of $H_2$ oxidation activity in soil (Gödde *et al.*, 2000; Khdhiri *et al.*, 2015; Schuler & Conrad, 1990).

In short, this study displayed the diversity of soil HA-HOB and the crucial need to cultivate more HA-HOB in order to fill the gap in genomic databases, which are currently not suited to depict soil microbiomes (Choi *et al.*, 2017).

## Accession numbers

Raw sequence reads of *hhyL* genes from PCR amplicon sequencing were deposited in the Sequence Read Archive of the National Center for Biotechnology Information under BioProject PRJNA382186.

## 5.4 Supplemental Method

### 5.4.1 PCR and RT-PCR

Partial *hhyL* gene was PCR amplified using the oligonucleotides NiFe-240f and NiFe-568r (Constant *et al.*, 2011b). RT-PCR was performed using AMV Reverse Transcriptase kit by Promega (Promega, Wisconcin, USA) according to manufacturer instructions.

### 5.4.2 PCR amplicon sequencing

PCR amplicon sequencing of partial *hhyL* gene was performed using Illumina MiSeq and resulted in 4,082,493 high-quality reads. Sequences quality control was then performed using Usearch 9 (Edgar, 2010) software according to (Khdhiri *et al.*, 2017), after which 3,398,942 reads remained. The OTU clustering threshold was set to 99% identity (see main text for justifications). A subsequent singleton removal step brought our sequencing library down to 916,920 reads. Protein-coding sequences were then aligned using UPGMB algorithm within Muscle 3.8.31 software (Edgar, 2004). Frameshift errors detection and OTUs correction using Framebot 1.2 (Wang *et al.*, 2013) were performed using an extensive group 5 [NiFe]-hydrogenase dataset as reference database. That dataset included group 5 [NiFe]-hydrogenases of known organisms available in public databases as well as clonal libraries sequenced in previous studies (Constant *et al.*, 2011b; Greening *et al.*, 2016; Vignais & Billoud, 2007). Erroneous sequences, that either could not be corrected by Framebot or that coded for short proteins (less than 80 amino acids-long), were discarded. 914,796 sequences (50,822 per sample), representing 664 OTUs, remained after frameshift corrections and sequences rarefaction.

### 5.4.3 Phylogenetic analysis

Two [NiFe]-hydrogenases databases were built for phylogenetic analyses: the first one included near full-length (at least 1383bp) [NiFe]-hydrogenases genes encompassing group 1 to 5, and the second one included short-length [NiFe]-hydrogenases genes (approximately 340bp) as well as sequenced OTUs. The near full-length database was used in order to assess phylogenetic similarity between [NiFe]-hydrogenases groups. The short-length database included an approximately 340 bp-long region located between two

double cysteine conserved motifs, and was used to infer taxonomic affiliation of sequenced OTUs as well as to detect potential non-specific amplification (none were detected). Nucleic acids sequences from both databases were aligned separately using MUSCLE 3.8.31 (Edgar, 2004) software according to predicted protein-coding sequences. Due to numerous misclassifications among hydrogenases included in NCBI databases, conserved metal-binding motifs of all [NiFe]-hydrogenases were filtered carefully according to conserved domains (Greening *et al.*, 2016) to avoid including genes similar to [NiFe]-hydrogenases in our databases. Evolutionary relationships were inferred using phylogenetic trees built under MEGA7 (Kumar *et al.*, 2016). A consensus [NiFe]-hydrogenases phylogenetic tree was then computed by combining NJ (Neighbor-joining, JTT model (Jones *et al.*, 1992)), MP (Maximum parsimony) and ML (Maximum likelihood, LG model (Le & Gascuel, 2008)) phylogenetic trees to add robustness to the analysis. All phylogenetic trees relied upon amino acids substitution models considering uniformly-distributed rates across residue sites. Gaps and missing data treatment was performed according to a pairwise deletion scheme for NJ and MP trees and a partial deletion scheme for the ML tree. 461 (near full-length) and 113 (short-length) residues were considered in the final datasets and 500 bootstrap replications were performed for trees computation. Consensus trees were computed using “ape” and “phytools” packages in R (Revell, 2012).

### 5.4.4 Statistical analyses

Various statistical analyses were performed to assess α- and β-diversity of group 5 [NiFe]-hydrogenases across land-uses and $H_2$ treatments. Shannon diversity indices, ACE estimators, Chao1 estimators, PERMANOVA (1000 permutations) and parsimonious RDA (Redundancy Analysis, 1000 permutations, Hellinger-transformed data) were computed using “vegan” package in R (Oksanen *et al.*, 2013). EdgeR and Deseq2 R packages were used to detect differentially abundant OTUs across land-uses and $H_2$ treatments (Love *et al.*, 2014; Robinson *et al.*, 2010). Only OTUs outlined ($p < 0.05$) by both algorithms were kept for further analysis.

# 6 DOSE-RESPONSE RELATIONSHIPS BETWEEN ENVIRONMENTALLY-RELEVANT $H_2$ CONCENTRATIONS AND THE BIOLOGICAL SINKS OF $H_2$, $CH_4$ AND CO IN SOIL

**Titre français** : Relations de dose-réponse entre des concentrations d'$H_2$ comparables à celles retrouvées dans l'environnement et les puits biologiques d'$H_2$, de $CH_4$ et de CO du sol

**Article publié en ligne le 19 mai 2018 dans le journal Soil Biology and Biochemistry**

*DOI : 10.1016/j.soilbio.2018.05.008* (Piché-Choquette *et al.*, 2018)

**Auteurs**

Sarah Piché-Choquette[1], Mondher Khdhiri[1] et Philippe Constant[1]✉

[1]INRS-Institut Armand-Frappier, 531 boulevard des Prairies, Laval (Québec), Canada, H7V 1B7

✉ Auteur correspondant

**Contribution des auteurs**

Sarah Piché-Choquette: Conception des expériences, expériences en laboratoire, analyse des résultats, écriture de l'article

Mondher Khdhiri : Expériences en laboratoire

Philippe Constant: Conception des expériences, analyse des résultats, écriture de l'article

## 6.1 **Résumé en français**

Des accumulations locales d'$H_2$ sont retrouvées dans les sols, notamment dans les agroécosystèmes de culture de légumineuses, où l'$H_2$ est un sous-produit obligatoire de la fixation de l'azote. De récentes études ont montré que les flux diffusifs d'$H_2$ agissent à titre d'intrants énergétiques additionnels modulant la structure et la fonction des communautés microbiennes du sol. Le but de cette étude est donc de définir les relations de dose-réponse entre la dose d'exposition à l'$H_2$ et la dynamique des communautés microbiennes du sol. La structure des communautés et l'activité d'oxydation de gaz traces ($H_2$, $CH_4$ et CO) ont été investiguées à la suite de l'incubation de sols à des ratios de mélange d'$H_2$ comparables à ceux retrouvés dans l'environnement. Malgré l'absence d'une altération de la diversité microbienne, des relations de dose-réponse coordonnées entre les cinétiques d'oxydation de gaz traces et l'exposition à l'$H_2$ ont été observées. Les vitesses d'oxydation d'$H_2$ mesurées ont été implémentées dans un cadre théorique modélisant la diminution d'$H_2$ en fonction de la distance d'une source ponctuelle émettant de l'$H_2$. Les profils de concentrations d'$H_2$ théoriques et les relations de dose-réponse entre la concentration d'$H_2$ et la vitesse d'oxydation de gaz traces ont été intégrés afin de prédire l'impact de l'$H_2$ sur le fonctionnement des communautés microbiennes. Bien que la majorité de l'$H_2$ soit oxydé à moins de 1 cm des sources ponctuelles d'$H_2$, l'oxydation des gaz traces serait, selon les prévisions du modèle, altérée jusqu'à un rayon de 10 cm. Les capacités d'oxydation du $CH_4$ et du CO à haute affinité ont chutées de jusqu'à 78% et 84% le long du gradient de concentration d'$H_2$, respectivement. Les distances théoriques des sources ponctuelles émettant de l'$H_2$ requises pour réactiver 50% de la vitesse d'oxydation du $CH_4$ étaient de 0.3 cm dans le sol agricole et de 0.2 cm dans le sol peuplier. Une plus longue distance était requise pour réactiver 50% de la vitesse d'oxydation du CO, soit 1.5 et 2.7 cm dans les sols agricole et peuplier, respectivement. La perte du potentiel d'oxydation du $CH_4$ observée sous une forte exposition à l'$H_2$ était corrélée avec un gain d'activité d'oxydation d'$H_2$ à faible affinité, alors qu'une inhibition de la vitesse d'oxydation d'$H_2$ à haute affinité par son substrat a été plutôt corrélée à une tendance décroissante de l'activité d'oxydation de CO. En plus de mettre en lumière des interactions potentielles entre les procédés biogéochimiques de l'$H_2$, du $CH_4$ et du CO

dans le sol, ces nouvelles découvertes démontrent que l'$H_2$ supporte la flexibilité énergétique et métabolique chez des microorganismes fournissant une variété de services écosystémiques.

## 6.2 **Abstract**

Local $H_2$ accumulations can be found in soil, especially within legume crop agroecosystems, where $H_2$ is an obligate by-product of nitrogen fixation. Recent investigations show that diffusive fluxes of $H_2$ act as additional energy inputs shaping microbial community structure and function in soil. The goal of this study is thus to define dose-response relationships between $H_2$ exposure and soil microbial community dynamics. Community structure and trace gases (*i.e.* $H_2$, $CH_4$ and CO) oxidation activities were investigated following soil incubation to environmentally-relevant $H_2$ mixing ratios. Despite no evidence of an alteration of microbial diversity, coordinated dose-response relationships between trace gases oxidation rates and $H_2$ exposure were recorded. Measured $H_2$ oxidation rates were implemented into a theoretical framework modeling $H_2$ decay as a function of distance from $H_2$-emitting point sources. Theoretical $H_2$ concentration profiles and dose-response relationships between $H_2$ concentration and trace gases oxidation rates were integrated to predict the impact of $H_2$ on microbial community functioning. While most $H_2$ is oxidized within 1 cm of $H_2$ point sources, trace gases oxidation is predicted to be altered within a 10 cm radius. High-affinity $CH_4$ and CO oxidation capacities dropped by up to 78% and 84% along $H_2$ concentration gradients, respectively. Theoretical distances from $H_2$-emitting point sources required to reactivate 50% of maximal $CH_4$ oxidation rate were 0.3 cm in farmland soil and 0.2 cm in poplar soil. A longer distance was required to reactivate 50% of CO oxidation rate, *i.e.* 1.5 and 2.7 cm in farmland and poplar soils, respectively. Loss of $CH_4$ oxidation potential observed under elevated $H_2$ exposure was correlated with a gain of low-affinity $H_2$ oxidation activity, while a substrate inhibition of high affinity $H_2$ oxidation rate was paralleled with a decreasing trend of CO oxidation activity. In addition to shedding light on potential interactions between $H_2$, $CH_4$ and CO biogeochemical processes in soil, these novel

findings provide evidence that $H_2$ supports metabolic and energetic flexibility in microorganisms supplying a variety of ecosystem services.

## 6.3 **Introduction**

Biological $N_2$-fixation (BNF) is performed either in $N_2$-fixing nodules (Mus *et al.*, 2016) or by free-living $N_2$-fixing bacteria (Orr *et al.*, 2011) according to the following enzymatic reaction: $N_2 + 8H^+ + 8e^- + 16ATP \rightarrow 2NH_3 + H_2 + 16ADP + 16P_i$ (Hoffman *et al.*, 2009). While providing bioavailable N to surrounding organisms, BNF also produces substantial amounts of $H_2$ as an obligate by-product. The high density of $N_2$-fixing bacteria in nodules leads to a local accumulation of $H_2$, with $H_2$ mixing ratios ranging from 9 000 to 27 000 ppmv in nodules (Hunt *et al.*, 1988; Witty, 1991; Witty & Minchin, 1998), of which diffusion losses can account for up to 240 000 L $H_2$ per hectare of legumes per growing season (Dong *et al.*, 2003). Metabolic capabilities of rhizobacterial symbionts recruited by plants represent the main determinants of the intensity of these $H_2$ emissions in soil. Indeed, plant roots can establish symbiosis with rhizobacteria displaying either a *Hup*$^+$ or *Hup*$^-$ genotype, according to the presence or absence of an uptake [NiFe]-hydrogenase, respectively. *Hup*$^+$ rhizobacteria have an energetic advantage since their [NiFe]-hydrogenase enables them to recycle the energy potential of $H_2$ that would otherwise be lost during $N_2$ fixation. In *Hup*$^-$ legumes rhizosphere, $H_2$ is thus released into surrounding soil, resulting in a net loss of energy at the plant-nodule level, which roughly translates to 5-6% of net photosynthesis energy output (Dong & Layzell, 2001). At first glance, the existence of this genotype looks like a flawed symbiosis, yet analyses of $N_2$-fixing nodules *in situ* have shown that less energy-efficient nodules, *i.e. Hup*$^-$ nodules, were more widespread than their *Hup*$^+$ counterpart (Dong *et al.*, 2003; Uratsu *et al.*, 1982). This is especially true in annual crops without well-established root systems (Annan *et al.*, 2012). A potential argument supporting preferential recruitment of less energy-efficient plant symbionts is the fact that $H_2$ released into surrounding soils, analogously to *Hup*$^-$ nodules, has been shown to increase plant biomass by 15-48% in both legumes and non-legumes (Dong *et al.*, 2003). Although the exact mechanisms behind $H_2$ fertilization remain elusive, enrichment of Knallgas bacteria, *i.e.* aerobic $H_2$-oxidizing bacteria (HOB), including plant

growth-promoting rhizobacteria (PGPR) showing either 1-aminocyclopropane-1-carboxylate (ACC) deaminase or rhizobitoxine activity promoting nodulation and root elongation, was proposed (Maimaiti *et al.*, 2007). *In situ*, $H_2$ is quickly used-up by HOB, within a 3-4.5 cm radius from the soil-nodule interface (La Favre & Focht, 1983). The magnitude of this biological sink is supported by the negligible amount of $H_2$ escaping from legume fields to the atmosphere (Conrad & Seiler, 1979). While $N_2$-fixing nodules and their underlying plant-microbe interactions have been extensively studied over the years (Dixon & Kahn, 2004; Mus *et al.*, 2016), the spatial patterns of microbial communities and processes along their resulting $H_2$ gradients has received little attention.

As a potent energy source for soil microbial communities, diffusive $H_2$ fluxes released by *Hup*$^-$ nodules are indubitably expected to stimulate $H_2$-oxidizing microbes such as Knallgas bacteria, with $K_m$ *ca.* 1 000 ppmv, and high-affinity HOB, with $K_m$ *ca.* 100 ppmv (Dong & Layzell, 2001; Häring & Conrad, 1994; Piché-Choquette *et al.*, 2016; Schuler & Conrad, 1990). Knallgas bacteria can quickly oxidize high loads of $H_2$, enabling them to use it as an energy source for facultative chemolithoautotrophic growth. However, the low-affinity of their [NiFe]-hydrogenase enzymes prevents them from using $H_2$ below a specific threshold concentration, namely at atmospheric mixing ratios (Conrad *et al.*, 1983a). Therein, high-affinity HOB, possessing unusual high-affinity [NiFe]-hydrogenases, take over the $H_2$-oxidizing process for their own maintenance energy (Constant *et al.*, 2010; Greening *et al.*, 2014b; Liot & Constant, 2016) at low $H_2$ concentrations. The stimulation of these contrasting HOB sub-populations along $H_2$ concentration gradients is expected to result in a differential alteration of surrounding soil microbial communities, notably considering that HOB are broadly distributed in both bacterial and archeal domains of life (Greening *et al.*, 2016; Piché-Choquette *et al.*, 2017; Vignais & Billoud, 2007) and thus, can be responsible for a wide variety of ecological functions apparently unrelated to $H_2$-oxidation.

Previous investigations have shown that elevated $H_2$ mixing ratios reflecting those found in natural ecosystems could indeed result in a shift in soil microbial communities' structure

(Khdhiri *et al.*, 2017; Osborne *et al.*, 2010; Piché-Choquette *et al.*, 2016; Stein *et al.*, 2005; Zhang *et al.*, 2009) and more importantly, in a coordinated feedback of communities' functions (Khdhiri *et al.*, 2017). Soil exposure to 10 000 ppmv $H_2$ led to, amongst other responses, a 2- to 4-fold decrease in $CH_4$ oxidation potential and a diversification in carbon utilization potential (Khdhiri *et al.*, 2017). Considering steep $H_2$ concentration gradients found in the environment, the present study aims to define dose-response relationships between $H_2$ exposure and the alteration of soil microbial community structure and function. For this purpose, soil microbial community structure and trace gases oxidation rates were investigated in soil exposed to various $H_2$ mixing ratios encompassing gradients found in natural environments. These observations were implemented into a theoretical framework modeling the decay of $H_2$ mixing ratios as a function of distance from $H_2$-emitting point sources, therein built according to Fick's second law of diffusion while also taking microbiological constraints into account. Dose-response relationships between $H_2$ mixing ratios and microbial processes were finally used to model $H_2$, carbon monoxide (CO) and methane ($CH_4$) oxidation rates as a function of distance from $H_2$-emitting point sources. Experimental observations and theoretical frameworks provide new experimental evidence into the potential impact of $H_2$ on soil biogeochemical processes.

## 6.4 Material and Methods

### 6.4.1 Soil sampling and microcosms incubation

Soil samples were collected from a farmland and a poplar plantation. Site location, sampling protocol, soil physicochemical properties and the dynamic microcosm chambers unit are all described in (Khdhiri *et al.*, 2017). Briefly, soils were air-dried and sieved (2 mm mesh size) prior to their incubation in a dynamic microcosms chamber unit comprising a programmable gas mixer enabling exposure of up to seven independent soil microcosms to a user-defined $H_2$ mixing ratio. Incubations were performed under a systematic experimental design in which six experimental units, consisting of three independent replicates of farmland and poplar soil microcosms, were exposed to one of the five selected $H_2$ exposure treatments (0.5, 50, 500, 5 000, and 10 000 ppmv $H_2$). Both

$H_2$ exposure treatment and position of each microcosm inside the dynamic microcosm chamber unit were randomly selected between each set of incubation. Each incubation lasted for 15 days before the destructive sampling of microcosms for microbial communities and trace gases oxidation activity measurements. Soil subsamples dedicated to microbial community structure analyses were stored at -80°C until DNA extraction was performed, while other analyses were processed immediately.

### 6.4.2 Nucleic acids extraction and purification

Soil DNA and RNA were co-extracted using a protocol based on (Mettel *et al.*, 2010), with a few exceptions to the original protocol detailed as follows. 35µL 20% SDS (sodium dodecyl sulfate) was added to each 0.5g soil sample alongside the TPM buffer. Microbial cells were lysed by bead-beating using a FastPrep®-24 homogenizer (MP Biomedicals, Solon, USA) at 6.5 m $s^{-1}$ for 45 s. All centrifugation steps before nucleic acids precipitation lasted for 5 minutes at 4°C rather than 1 minute. The pH 4.5 phenol-isoamyl alcohol used in the first extraction step was supplemented with 1 g $L^{-1}$ 8-hydroxyquinoline as anti-oxidant. For the precipitation step, 0.1 volume 3.0 M sodium acetate (pH 5.2) was first added to the supernatant, followed by 2 volumes 95% ethanol. Samples were then slightly shaken and were left to precipitate for 12h at -20°C. Samples were then centrifuged at 20,000g for 60 minutes at 4°C, pellets were washed twice with 70% ethanol and were resuspended as stated in the original protocol. RNA was preserved for other studies, while DNA was treated with 5 µL 100 µg $ml^{-1}$ RNaseA and left at room temperature for 15 minutes. DNA was then purified using acid-washed polyvinylpolypyrrolidone columns according to (Holben *et al.*, 1988). DNA samples were assessed for integrity on a 1.0% agarose gel and quantified using Quantifluor dsDNA System® (Promega, Fitchburg, WI, U.S.A.) and a Rotorgene 6000® thermocycler (Qiagen, NRW, Germany). Nucleic acids were preserved at -80°C until sent for sequencing.

### 6.4.3 PCR amplicon sequencing and sequences analysis pipeline

PCR amplification of the V6-V8 regions of bacterial 16S rRNA gene was performed with B969F-CS1 (5'-ACGCGHNRAACCTTACC-3') and BA1406R-CS2 (5'-ACGGGCRGTGWGTRCAA-3') (Comeau *et al.*, 2011) primers, while the ITS2 region of

fungal rRNA was PCR-amplified with ITS3_KYO2 (5'-GATGAAGAACGYAGYRAA-3') and ITS4_KYO3 (5'-CTBTTVCCKCTTCACTCG-3') (Toju *et al.*, 2012) primers. Both sequencing library preparation and Illumina Miseq reactions (2x250 paired-ends configuration) were performed at McGill University Génome Québec Innovation Centre (Montreal, Quebec, Canada). Raw sequencing reads were processed using Usearch10 (Edgar, 2010). In short, paired-ends were first merged to a total length of 400-500 bases, for bacteria, or 200-450 bases, for fungi, to account for the variability in sequence length among sequenced regions. Maximum overlap mismatches accepted for downstream processing consisted of 5 bases or 10%, whichever came first. Primers were trimmed from merged reads, while reads without their respective primers were discarded from the dataset. Reads with over 1 expected error (Edgar & Flyvbjerg, 2015), *i.e.* when the sum of nucleotides errors predicted on the basis of Phred quality scores adds up to over 1 error per read, were discarded during reads quality control. Reads were then dereplicated and singletons were discarded. Before ASV (amplicon sequence variant) assignation, reads were denoised using Unoise3 (Edgar, 2016b). Following denoising, only reads with a frequency over 8 were considered as representative sequences for ASV assignation. ASV clustering was chosen over OTU clustering (less than 100% sequence identity) due to its higher reproducibility and resolution across studies (Callahan *et al.*, 2017). Taxonomic affiliation of ASV was predicted by k-mer similarity of ASV representative sequences to the RDP (Ribosomal Database Project) v.16 training set of 16S rRNA gene for bacteria and RDP Warcup training set v.2. for fungi (Cole *et al.*, 2014), altogether considering taxa agreed upon by over 85% of bootstrap replications using the SINTAX algorithm (Edgar, 2016a). Reads assigned to chloroplasts were removed from the dataset, alongside ASV representing less than 0.005% of read counts. All libraries were subsampled to match the library with the least amount of sequences, *i.e*. 51,105 bacterial sequences and 38,448 fungal sequences, ultimately corresponding to 3,185 bacterial and 786 fungal ASV. Rarefaction curves (data not shown) and diversity indices and indicators were computed using "vegan" (Oksanen *et al.*, 2013) package in R (R Development Core Team, 2017) to ensure sufficient sequencing efforts.

### 6.4.4 Trace gases exchange measurements

At the end of the incubation period, 50 g soil (wet weight) subsamples were transferred into 500 ml Wheaton glass bottles for trace gases oxidation rate measurements using two gas chromatographic assays (Khdhiri *et al.*, 2015; Khdhiri *et al.*, 2017; Lalonde & Constant, 2016). The static headspace was first flushed with synthetic air (certified standard mixture; Praxair Distribution Inc., PA, USA) in order to remove $H_2$ mixing ratios associated with each treatment. High-affinity $H_2$ and CO oxidation rates in soil subsamples were measured after the injection of $H_2$ (524 ±10 ppmv, Praxair Distribution Inc., PA, USA) and CO (508 ±10 ppmv, Praxair Distribution Inc., PA, USA) gas mixtures to obtain an approximate initial mixing ratio of 3 ppmv in the static headspace. Low-affinity $H_2$ oxidation rates and high-affinity $CH_4$ oxidation rates were rather measured after flushing the microcosms' static headspace with a synthetic gas mixture comprising 10 000 ppmv $H_2$ (± 2%) and 3 ppmv $CH_4$ (± 2%) (Praxair Distribution Inc., PA, USA) instead of previously used synthetic air. First-order $H_2$, CO and $CH_4$ oxidation rates were calculated by integrating gas mole fraction time series, using at least 5 time points.

### 6.4.5 $H_2$ fluxes and concentrations theoretical framework

Based on the steep $H_2$ concentration gradients found in nature, one would expect the impact of $H_2$ to be restricted to a distance of a few centimeters away from a point source. Testing that assumption in field or pot experiments is a challenging task due to the fine spatial sampling resolution required to measure spatial patterns of $H_2$ concentrations and process rates in soil. As an alternative, we elaborated a theoretical framework simulating $H_2$ concentration profiles as a function of distance from $H_2$-emitting point sources found in the surroundings of *Hup*$^-$ $N_2$-fixing nodules. The model, based upon Karstens *et al.* (2015), assumes a diffusive $H_2$ flux evolving from a single point source within a homogeneous soil texture and air-filled porosity environment and which cannot oxidize $H_2$ at the source-soil interface. It is also assumed that only that point source produces $H_2$. The framework itself was built upon the equation of continuity:

$$\frac{\partial[H_2]}{\partial t} = \frac{\partial}{\partial d}\left(D_s(d)\frac{\partial[H_2]}{\partial d}\right) + Q + S \quad (1)$$

Equation (1) describes the net concentration change of $H_2$ ( $\frac{\partial[H_2]}{\partial t}$ ), expressed in nmol $cm^{-3}$ $s^{-1}$, determined by the divergence of the $H_2$ diffusive flux $\left(D_s(d)\frac{\partial[H_2]}{\partial d}\right)$ and constrained by diffusion limitation in soils, herein represented by $H_2$ diffusivity coefficient $D_s(d)$ ($cm^2$ $s^{-1}$) set to 0.1 $cm^2$ $s^{-1}$ according to Smith-Downey (2008), as well as source Q (nmol $cm^{-3}$ $s^{-1}$) and sink S (nmol $cm^{-3}$ $s^{-1}$) terms. Simulations were also performed using a $D_s(d)$ of 0.01 $cm^2$ $s^{-1}$ to show sensitivity of $H_2$ concentration profiles to enhanced diffusion constraints in soil. Terms *t* and *d* represent time (s) and distance (cm), respectively. Q is set to 0 nmol $cm^{-3}$ $s^{-1}$ owing to the absence of $H_2$ production in soil, while S can be transposed to "microbe-mediated $H_2$-oxidizing activity" defined as high-affinity and low-affinity biological $H_2$ sinks:

$$\frac{\partial[H_2](d,t)}{\partial t} = \frac{\partial}{\partial d}\left(D_s(d)\frac{\partial[H_2](d,t)}{\partial d}\right) + \lambda_{HA}([H_2],d,t) + \lambda_{LA}([H_2],d,t) \quad (2)$$

where $\lambda_{HA}([H_2],d,t)$ is the high-affinity $H_2$ uptake rate (nmol $cm^{-3}$ $s^{-1}$) and $\lambda_{LA}([H_2],d,t)$ is the low affinity $H_2$ uptake rate (nmol $cm^{-3}$ $s^{-1}$). Both variables were measured (section 2.4) and shown to be dependent on residual $H_2$ mixing ratios [$H_2$] ($H_2$ concentration, nmol $cm^{-3}$), therefore they are also dependent on both *t* and *d*. The model assumes steady state conditions, *i.e.* $\frac{\partial[H_2]}{\partial t} = 0$, thus we obtain equation (3):

$$\left(D_s\frac{\partial^2[H_2](d)}{\partial d^2}\right) + \lambda_{HA}([H_2],d) + \lambda_{LA}([H_2],d) = 0 \quad (3)$$

At a distance of $d$ = 0, *i.e.* at the source-soil interface, we have a fixed concentration $[H_2]_0$ and then $H_2$ uptake occurs from $d \rightarrow 0$ to $d \rightarrow \infty$. If we solve the homogeneous differential equation in (3), we can calculate residual $H_2$ mixing ratios as a function of distance from a $H_2$-emitting point source:

$$[H_2](d) = [H_2]_0 \, e^{\frac{-d}{\bar{d}}} \qquad (4)$$

Where $[H_2]_0$ is the initial $H_2$ mixing ratio (*i.e.* set to 10 000 ppmv in the first boundary condition, see below) evolved by a nodule and $\bar{d}$ (cm) is the relaxation length, *i.e.* the distance at which the residual concentration is equal to $\frac{[H_2]_0}{e}$.

Then, the second derivative of (4) is equivalent to:

$$\frac{\partial^2 [H_2](d)}{\partial d^2} = \frac{1}{\bar{d}^2} [H_2]_0 e^{\frac{-d}{\bar{d}}} \qquad (5)$$

If we combine (4) and (5) into (3), we obtain:

$$\left( D_s \frac{1}{\bar{d}^2} [H_2]_0 e^{\frac{-d}{\bar{d}}} \right) + \lambda_{HA} [H_2]_0 \, e^{\frac{-d}{\bar{d}}} + \lambda_{LA} [H_2]_0 \, e^{\frac{-d}{\bar{d}}} = 0 \qquad (6)$$

We can further isolate $\bar{d}$ from (6) and obtain:

$$\bar{d} = \sqrt{\frac{D_s}{\lambda_{LA} + \lambda_{HA}}} \qquad (7)$$

By combining the relaxation length $\overline{d}$ from (7) into (4), $H_2$ mixing ratios can be calculated for a given distance using:

$$[H_2](d) = [H_2]_0 \, e^{\frac{-d}{\sqrt{\frac{D_s}{\lambda_{LA}+\lambda_{HA}}}}} \qquad (8)$$

Finally, equation (8) was used for the simulation of theoretical $H_2$ concentration gradients following its production by a point source, represented by setting a fixed boundary condition (A) at *d* = 0, $[H_2]_0 = 10{,}000\,ppmv$. The model also assumed that at an infinite distance from the point source, $H_2$ mixing ratio tends towards 0.53 ppmv, *i.e.* resulting from the diffusion of atmospheric $H_2$ into the soil column (*ca.* 0.53 ppmv) (Novelli *et al.*, 1999), although in some ecosystems without $H_2$ points sources, $H_2$ mixing ratios can be lower than 0.53 ppmv (Smith-Downey *et al.*, 2008). Therefore, our boundary conditions were:

(A) $[H_2]_{(d\to 0)} = 10{,}000\,ppmv$

(B) $[H_2]_{(d\to \infty)} = 0.53\,ppmv$

Furthermore, model runs were computed by 0.1 cm increments until $H_2$ mixing ratios decreased below 0.53 ppmv, as per boundary condition (B). It should be noted that low- and high-affinity HOB community-level $H_2$ oxidation rates (*i.e.* $\lambda_{LA}$ and $\lambda_{HA}$, respectively) were adjusted according to residual $H_2$ mixing ratios before each simulation step. This adjustment was performed according to computed Michaelis-Menten models depicting measured $H_2$ oxidation rate as a function of $H_2$ concentration (see results section and Table 6.1).

### 6.4.6 Distribution of trace gases oxidation rates altered by $H_2$ on a spatial scale

Trace gases exchange measurements led to the determination of oxidation rates as a function of $H_2$ mixing ratios for 5 distinct mixing ratios, namely 0.5, 50, 500, 5 000, and 10 000 ppmv $H_2$. These first-order oxidation rates were implemented into best fitting regression models in the software R (R Development Core Team, 2017) using the "glm2" package (Marschner, 2011) in order to approximate a continuous dose-response relationship from discrete datasets. These dose-response relationships between residual $H_2$ mixing ratios and trace gases ($H_2$, CO and $CH_4$) oxidation rates were fitted to theoretical $H_2$ concentration gradients derived from the $H_2$ fluxes theoretical framework, leading to projections of traces gases oxidation rates as a function of distance from a $H_2$-emitting point source.

### 6.4.7 Statistical analyses

Statistical analyses were performed using the software R (R Development Core Team, 2017). Samples were first compared in terms of β-diversity to determine if $H_2$ gradients led to contrasting trends in microbial community structure, using Principal Component Analysis (PCA). PCA were computed using Euclidean distance matrices derived from Hellinger-transformed data and using R packages "vegan" (Oksanen *et al.*, 2013) and "gclus" (Hurley, 2012). In order to test if $H_2$ mixing ratios could explain the distribution of samples in an unconstrained reduced space and thus β-diversity, a smoothness model was fitted to predict $H_2$ exposure using PC1 and PC2 as predictor variables. A response surface was then calculated by a general additive model (gam) and plotted using packages "vegan" (Oksanen *et al.*, 2013) and "mgcv" (Wood, 2011). Response surfaces were computed according to maximum likelihood smoothness model selection. Model assumptions regarding residuals were also met for all four models. Afterwards, a likelihood ratio test, computed with the package "edgeR" (Robinson *et al.*, 2010), was used to detect bacterial and fungal ASV that were differentially abundant between extreme $H_2$ exposures, *i.e.* 0.5 ppmv and 10 000 ppmv $H_2$. Biological replicates were treated like different samples of a same $H_2$ exposure treatment for the likelihood ratio test. Spearman rank correlation tests were performed to define if and which ASV's relative

abundance was correlated with $H_2$ exposure, using the package “stats” (R Development Core Team, 2017). Rank correlations were calculated between the five $H_2$ mixing ratios and the relative abundance of ASVs from all $H_2$ exposure treatments. Spearman rank correlation tests were also performed to explore covariation between trace gases oxidation rates.

## 6.5 Results

Findings reported in this article are the result of work derived from two complementary phases. Firstly, series of laboratory scale experiments were conducted to define dose-response relationships between $H_2$ mixing ratio and soil microbial community structure and trace gases oxidation activities. Secondly, experimental measurements were integrated into theoretical models aimed at elaborating potential scenarios regarding the spatial influence of $H_2$ point sources on soil biogeochemical processes. A first model simulated theoretical spatial distribution of $H_2$ mixing ratios from its production site (10 000 ppmv $H_2$) to its near-complete depletion (0.5 ppmv $H_2$) a few cm away. Dose-response relationships between $H_2$ mixing ratio and trace gases oxidation rates, measured in laboratory scale experiments, were superimposed with theoretical spatial distribution of $H_2$ mixing ratios in a second model in order to predict spatial patterns of process rates as a function of distance from the simulated $H_2$ point source.

### 6.5.1 Changes in microbial community structure over the $H_2$ concentration gradient

Potential relationship between $H_2$ mixing ratio and soil microbial diversity was first explored using a principal component analysis (PCA) of ASV profiles, including a linear integration and response surface of $H_2$ mixing ratios. However, the distribution of soil microcosms in the reduced space defined by ASV relative abundances profiles showed no clear clustering within replicates and between $H_2$ treatments and thus unveiled a lack of coherent patterns of microbial community structure across $H_2$ concentration gradients (Figure 6.1). In farmland soil, neither linear integration nor maximum likelihood response surfaces of $H_2$ concentrations showed significant covariation with bacterial (Figure 6.1A)

and fungal (Figure 6.1C) community profiles ($P > 0.05$). In contrast, variation of bacterial community structure was partly explained by maximum likelihood response surfaces of $H_2$ concentrations in poplar soil (Figure 6.1B, $R^2 = 0.91$, $P = 6.86 \cdot 10^{-6}$), while no significant trend was observed under a linear integration of $H_2$ concentration gradient (Figure 6.1B, $R^2 = 0.27$, $P = 0.15$). Variation in fungal communities in poplar soil was not explained by the response surface ($R^2 = 0.43$, $P = 0.079$) nor linear ($R^2 = 0.088$, $P = 0.58$) integration of $H_2$ concentrations (Figure 6.1D). These observations imply that $H_2$ gradients have no obvious impact on microbial community structure at the community level. Furthermore, $H_2$ exposure was not a significant driver of bacterial and fungal species richness and evenness indices either (12.1, Table S1). However, at the ASV level, a significant response of microbial community structure to $H_2$ exposure was detected (ASV tables available in 12.1, Table S1) when comparing between 0.5 and 10 000 ppmv $H_2$ treatments. Indeed, EdgeR likelihood ratio test showed that 516 bacterial and 77 fungal ASV significantly differed (p-value < 0.05) in relative abundance between 0.5 and 10 000 ppmv $H_2$ treatments (12.2, Table S2) across all tested land uses. In poplar soils, around half of the most abundant bacterial ASV in samples exposed to 0.5 ppmv $H_2$ were either Actinobacteria (24.2%, while only 7.0% in 10 000 ppmv $H_2$) or Alphaproteobacteria (23.5%, similarly with 18.3% in 10 000 ppmv $H_2$), while Gammaproteobacteria (18.4%) were dominant in 10 000 ppmv $H_2$ (6.2% in 0.5 ppmv $H_2$). In farmland soils, the most abundant bacterial ASV in both extreme $H_2$ exposures were Alphaproteobacteria (36.6% and 23.7% in 0.5 and 10 000 ppmv $H_2$, respectively). In both soils, fungal ASV responded more evenly to $H_2$ exposure. While many ASV were significantly more abundant in one of the two extremes of $H_2$ exposure treatments, we also wanted to test if the abundance of these ASV was proportional to $H_2$ mixing ratios. Among the ASV displaying a significant differential abundance between treatments, 183 bacterial and 22 fungal ASV were also significantly correlated with $H_2$ mixing ratios (12.3, Table S3, Spearman rank correlation, p-value < 0.05) when the mixing ratios of the five $H_2$ exposure treatments were related with their corresponding ASV relative abundance. In farmland soil, bacterial ASV affiliated to Alphaproteobacteria were the most positively (22.5%) and negatively (34.5%) correlated with $H_2$ exposure. In poplar soils, the most positively correlated ASV were Alphaproteobacteria (20.0%) and Betaproteobacteria (20.0%) and the most negatively

correlated ASV were affiliated to Alphaproteobacteria (25.8%) and Actinobacteria (19.4%). Only 2 of those bacterial ASV (ASV3789, Planctomycetes and ASV1505, Actinobacteria, taxonomy available in 12.1, Table S1) responded likewise in both soil land use types. Fungal ASV did not show any particular trend. Taken together, these observations indicate that the impact of $H_2$ exposure on bacteria and fungi occurred only at the species-level, with no evidence of significant alteration in biodiversity at the community level.

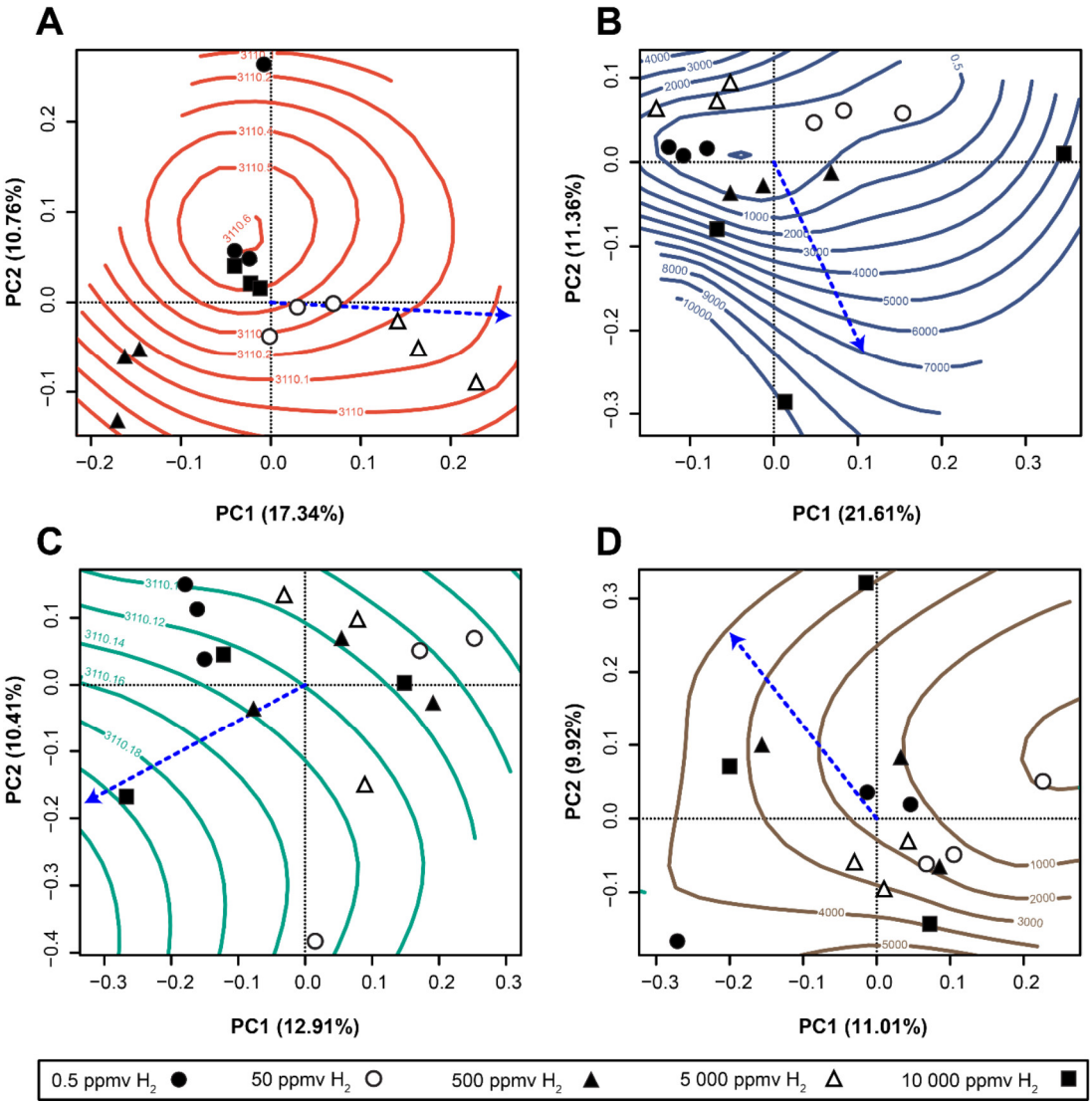


**Figure 6.1** **PCA depicting the distribution of soil microcosms in a reduced space defined by the relative abundances profiles of A) Bacterial ASV in farmland soil, B) Bacterial ASV in poplar soil, C) Fungal ASV in farmland soil and D) Fungal ASV in poplar soil. ASV profiles are overlayed by a response surface of $H_2$ mixing ratio values predicted by a Maximum Likelihood smoothness model. The blue arrow represents the predicted linear variation of $H_2$ mixing ratios across the ordination plot.**

### 6.5.2 Dose-response relationships between $H_2$ mixing ratio and trace gases oxidation rates

Trace gases oxidation rates were first measured for each $H_2$ exposure (0.5, 50, 500, 5 000 and 10 000 ppmv $H_2$, 12.4, Table S4) after incubation of soil microcosms for 15 days. Changes in high-affinity and low-affinity $H_2$ oxidation rates as a function of $H_2$ exposure were comparable to those detailed in previous studies (Khdhiri *et al.*, 2017; Piché-Choquette *et al.*, 2016). High-affinity $H_2$ oxidation rates decreased (Figure 6.2A) while low-affinity $H_2$ oxidation rates increased (Figure 6.2B) along the $H_2$ concentration gradient. Both $H_2$ oxidation activities followed typical Michaelis-Menten kinetics and were integrated under logarithmic regression models (Table 6.1). The abatement of $CH_4$ oxidation rates as a function of $H_2$ exposure (Figure 6.2C) followed a linear trend (Table 6.1). A loss of 78% and 67% of $CH_4$ oxidation activity was measured along the $H_2$ concentration gradient in farmland and poplar soils, respectively. $CH_4$ oxidation rates were negatively correlated with low-affinity $H_2$ oxidation rates in farmland (Spearman; $\rho = -0.53$, p-value = 0.04) and poplar (Spearman; $\rho = -0.85$, $P = 6 \cdot 10^{-5}$) soils. The activity of high-affinity CO-oxidizing bacteria was more affected by elevated $H_2$ exposure than methanotrophy, with 4.4- and 6.4-fold lower activity in soil exposed to 10 000 ppmv when compared with 0.5 ppmv treatments for farmland and poplar soils, respectively (Figure 6.2D). The loss of CO oxidation activity along $H_2$ concentration gradient was exponential (Table 6.1) and correlated with the abatement of high-affinity $H_2$ oxidation rate in farmland (Spearman, $\rho = 0.67$, $P = 0.02$) and poplar (Spearman, $\rho = 0.97$, $P = 2.3 \cdot 10^{-9}$) soils.

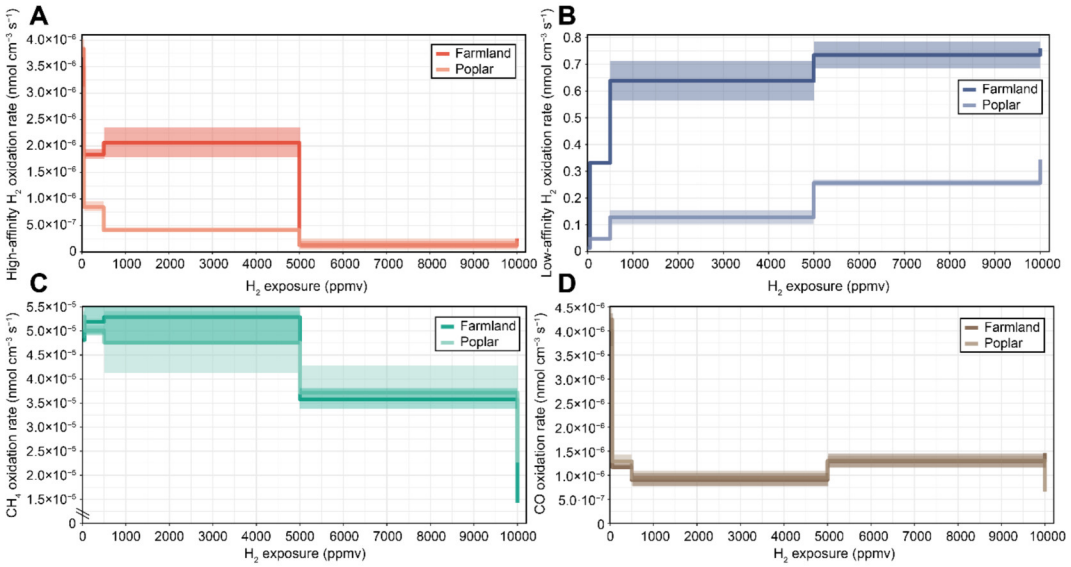



**Figure 6.2** **Stair-step plot of measured process rates as a function of $H_2$ exposure. Darker colors represent farmland soil while lighter colors represent poplar soil. Processes correspond to A) High-affinity $H_2$ oxidation, B) Low-affinity $H_2$ oxidation, C) High-affinity $CH_4$ oxidation and D) High-affinity CO oxidation. Shaded areas represent the standard deviation of the plotted mean of three biological replicates. Raw data of trace gases oxidation rates measured in soil exposed to 0.5, 50, 500, 5 000 and 10 000 ppmv $H_2$ are shown in Table 6.2. The measured oxidation rates (y-axis) reflect the discrete values of soil exposure to the five aforementioned $H_2$ mixing ratios. Fitted relationships calculated from these data can be found in Table 1.**

**Table 6.1** **Regression models depicting trace gases oxidation processes as a function of $H_2$ exposure in both land-uses. $H_2$ mixing ratio is in ppmv.**

| Biogeochemical processes[A] | Farmland | p-value | Poplar | p-value |
|---|---|---|---|---|
| **High-affinity $H_2$ oxidation ($\lambda_{HA}$)** | uptake rate = $-7.61\times10^{-7} \times \log([H_2]) + 3.36\times10^{-6}$ | $2.70\times10^{-7}$ | uptake rate = $-8.32\times10^{-7} \times \log([H_2]) + 3.05\times10^{-6}$ | $6.33\times10^{-7}$ |
| **Low-affinity $H_2$ oxidation ($\lambda_{LA}$)** | uptake rate = $0.18 \times \log([H_2]) + 0.07$ | $1.62\times10^{-10}$ | uptake rate = $0.07 \times \log([H_2]) - 0.001$ | $3.63\times10^{-6}$ |
| **High-affinity $CH_4$ oxidation** | uptake rate = $-3.25\times10^{-9} \times [H_2] + 4.88\times10^{-5}$ | $1.50\times10^{-3}$ | uptake rate = $-2.80\times10^{-9} \times [H_2] + 5.10\times10^{-5}$ | $2.52\times10^{-5}$ |
| **High-affinity CO oxidation** | uptake rate = $-5.91\times10^{-7} \times \log([H_2]) + 3.07\times10^{-6}$ | $1.36\times10^{-5}$ | uptake rate = $-7.50\times10^{-7}\cdot\log([H_2]) + 3.40\times10^{-6}$ | $4.48\times10^{-7}$ |

[A]Uptake rates are expressed as nmol $cm^{-3}$ $s^{-1}$

### 6.5.3 Projection of $H_2$ concentration gradients in soils exposed to a $H_2$ point source

Now that a dose-response relationship between $H_2$ mixing ratio and trace gases oxidation rates is demonstrated (Table 6.1), the next step is to evaluate the potential range of $H_2$ interference with biogeochemical processes in soil. Dose-response relationships between $H_2$ mixing ratios and $H_2$ oxidation activities (Table 6.1) were integrated into a theoretical framework to predict distribution of $H_2$ concentrations in soil as a function of distance from $H_2$-emitting point sources in both farmland and poplar soils. $H_2$ mixing ratios sharply decreased over a short distance from the simulated point sources due to the activity of low-affinity HOB, with losses of 91% and 77% $H_2$ within the first cm in farmland and poplar soils, respectively (Figure 6.3). Residual $H_2$ mixing ratios above atmospheric conditions, *i.e.* 0.53 ppmv, are predicted at a distance up to 6 and 13.5 cm away from the point source in farmland and poplar soils, respectively (Figure 6.3, Insert). Simulating more important diffusion constraints with $D_s$ value set to 0.01 $cm^2\ s^{-1}$ instead of 0.1 $cm^2\ s^{-1}$, as mentioned in the method section, decreased these distances to 2.0 and 4.2 cm in farmland and poplar soils, respectively. Theoretical $H_2$ concentration gradients are exploited in the next section to compute spatial patterns of trace gases oxidation rates in soil.

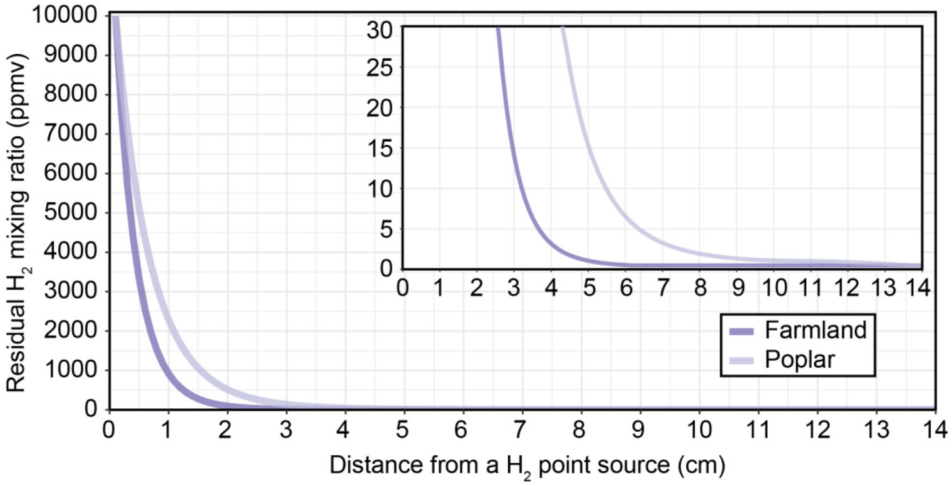



**Figure 6.3** **Modeled $H_2$ mixing ratios as a function of distance from a $H_2$-emiting point source. The main plot depicts the whole $H_2$ mixing ratios gradient from 1-14 cm distance from a $H_2$-emitting point source, while the insert represents a zoom on lower $H_2$ mixing ratios.**

### 6.5.4 Trace gases oxidation rates as a function of distance from $H_2$-emitting point sources

Dose-response relationships between $H_2$ mixing ratio and trace gases oxidation activities, fitted using regression models shown in Table 6.1 and projection of $H_2$ concentration gradient in soil exposed to $H_2$, derived in the previous section, were integrated in order to predict the spatial distribution of $H_2$, CO and $CH_4$ oxidation activities in soils. Theoretical distance from $H_2$-emitting point-source required to reactivate 50% of maximal high-affinity $H_2$-oxidation rate were 2.0 and 3.6 cm in farmland and polar soil, respectively (Figure 6.4A). The theoretical distance over which a minimal $H_2$ concentration threshold is sufficient to support 50% of maximal low-affinity $H_2$-oxidation rates measured at 10 000 ppmv was shorter in farmland than in polar soil, at 2.2 and 3.3 cm from the $H_2$-emitting point source (Figure 6.4B). According to the model, a complete recovery of the $CH_4$-oxidation activity potential occurs in the first 2 cm for farmland soil and first 4 cm in poplar soil (Figure 6.4C). Recovery of CO-oxidation activity is expected to occur over a longer distance than the activity of methanotrophic bacteria, with minimal distances of 6 and 10 cm for farmland and poplar soils, respectively (Figure 6.4D).

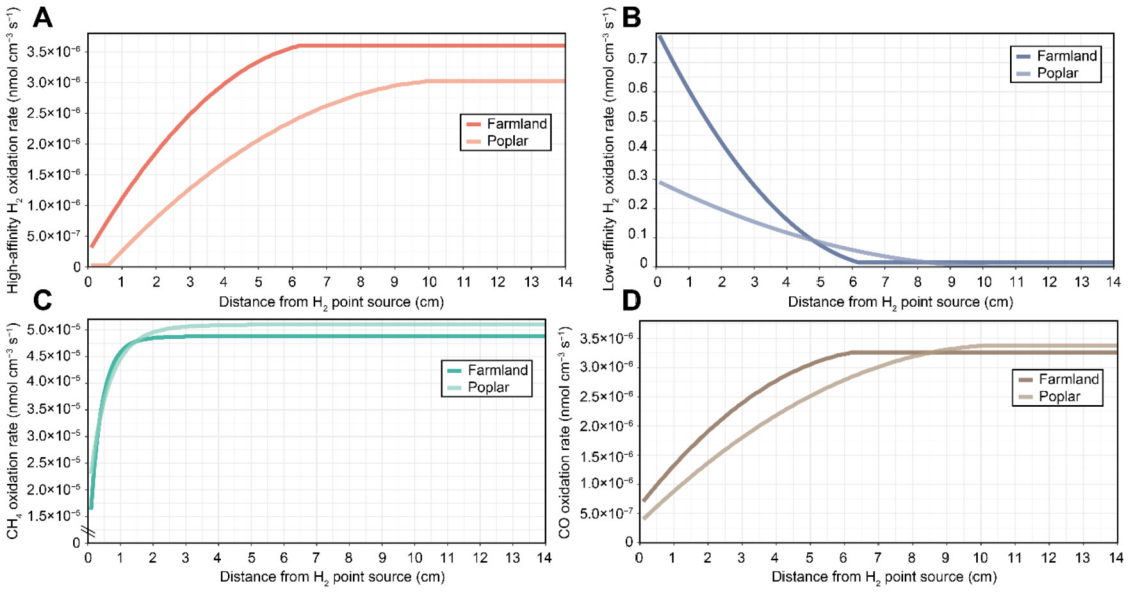



**Figure 6.4** **Simulated $H_2$, CO and $CH_4$ oxidation rates as a function of distance from a $H_2$-emiting point source. Darker colors represent farmland soil while lighter colors represent poplar soil. Processes correspond to A) High-affinity $H_2$ oxidation, B) Low-affinity $H_2$ oxidation, C) High-affinity $CH_4$ oxidation and D) High-affinity CO oxidation.**

## 6.6 **Discussion**

This study explores the dose-response relationship between $H_2$ exposure and microbial community structure and function. Previous investigations have shown that the energy potential of elevated $H_2$ concentrations found at the soil-nodule interface could alter microbial community composition and species richness (Khdhiri *et al.*, 2017; Piché-Choquette *et al.*, 2017). Here, the energy potential of $H_2$ was shown to exert a filtering effect on bacterial and fungal species distribution. While the relative abundance of a few ASV was strongly correlated with $H_2$ mixing ratios, these changes in community structure resulted in little to no impact on β-diversity. Nevertheless, this idiosyncratic shift in community structure was accompanied by a contrastingly coordinated dose-response relationship in specialist microbial functions.

Alteration in HOB oxidation dynamics was expected to occur in soils exposed to different $H_2$ concentrations due to the simultaneous inhibition of high-affinity HOB and activation of Knallgas bacteria $H_2$ metabolism under elevated $H_2$ levels (Dong & Layzell, 2001; Häring & Conrad, 1994; Piché-Choquette *et al.*, 2016; Schuler & Conrad, 1990). Theoretical frameworks from this study predict that most of the $H_2$ evolved by $H_2$ point sources is depleted within a 2 cm radius. Residual $H_2$ mixing ratios are expected to remain above atmospheric conditions (*ca.* 0.53 ppmv $H_2$) up to 6-13 cm from $H_2$ point sources in farmland and poplar monoculture soils, respectively. These distances are of the same magnitude as the 3-4.5 cm range previously measured in the vicinity of pigeon peas $N_2$-fixing nodules, where depletion of $H_2$ was correlated with the density of Knallgas bacteria (La Favre & Focht, 1983). Although spatial patterns of HOB communities along $H_2$ concentration gradients found in nature have received little attention, numerous pieces of evidence are suggesting ecological consequences of $H_2$ diffusive fluxes in soil. Microorganisms responsible for the observed $H_2$-oxidizing activity are broadly distributed taxonomically, forasmuch that genes encoding [NiFe]-hydrogenases encompass 36 bacterial and 6 archaeal phyla (Greening *et al.*, 2016; Vignais & Billoud, 2007). Such a widespread taxonomic distribution of those enzymes implies versatile physiological capabilities in HOB, including CO-, $CH_4$- and nitrite-oxidizing bacteria. From an ecological

perspective, this is in line with the hypothesis that $H_2$ is a universal energy source supplying survival energy requirements for the seed bank of numerous microbial functions (Morita, 1999). Huge knowledge gaps within the ecophysiology of HOB still remain, thus impairing the elaboration of a predictive framework forecasting how natural sources of $H_2$ in soil drive the juxtaposition of HOB and the biogeochemical processes they provide. Theoretical frameworks reported in this study are a first step towards that direction, highlighting potential $H_2$ oxidation activity in high affinity $CH_4$- and CO-oxidizing bacteria in addition to suggest an impact of $H_2$ sources on soil-to-air $CH_4$ and CO net exchange rates.

Our results highlight that the proximity of $H_2$ point sources could sharply reduce soil atmospheric $CH_4$ oxidation rates by 67% and 78% in farmland and poplar soils, respectively. This is consistent with previous investigations showing that those same soils, when exposed to 10 000 ppmv $H_2$, had a 2-4-fold decrease in atmospheric $CH_4$ oxidation rates along with an absence of methanogenesis activity (Khdhiri *et al.*, 2017). The frameworks predict that this decrease in $CH_4$ oxidation capacity slowly reverts to its original state within 2-4 cm from $H_2$-emitting point sources, implying that only high $H_2$ mixing ratios inhibit soil $CH_4$ oxidation activity. This distance corresponds to residual $H_2$ mixing ratios ranging between 38 and 50 ppmv (Figure 6.3), thus close to the minimum threshold concentration at which Knallgas bacteria can oxidize $H_2$ (Conrad *et al.*, 1983a; Constant *et al.*, 2008). The strong negative correlation between low-affinity $H_2$ oxidation and high-affinity $CH_4$ oxidation suggests that these processes might be performed by the same microbes depending on the availability of both substrates. Indeed, there is an increasing body of evidence supporting the notion that methanotrophic bacteria have a more versatile metabolic potential than initially thought, including mixotrophic lifestyles relying on alternative carbon and energy sources. According to recent genome database mining, more than 25 aerobic $CH_4$-oxidizing bacteria harbor gene encoding [NiFe]-hydrogenase (Greening *et al.*, 2016; Vignais & Billoud, 2007). For instance, soluble low-affinity [NiFe]-hydrogenases are expected to provide the necessary reducing power to maintain $CH_4$ monooxygenase enzyme activity, *e.g.* via balancing the $NAD^+$/NADH pool, for $CH_4$ oxidation in *Methylococcus capsulatus* (Bath) (Hanczár *et al.*, 2002) and

*Methylosinus trichosporium* OB3b (Chen & Yoch, 1987; Shah *et al.*, 1995). Furthermore, a membrane-bound [NiFe]-hydrogenase was shown to supply supplementary energy under $CH_4$ starvation in *Methylosinus trichosporium* (Chen & Yoch, 1987). More recently, an oxygen-sensitive membrane-bound [NiFe]-hydrogenase was shown to provide the ability to switch between growth strategies either based on $CH_4$ as a unique substrate or on the concurrent use of $H_2$ and $CO_2$ as energy and carbon sources, in the Verrucomicrobia *Methylacidiphilum fumariolicum* SolV (Mohammadi *et al.*, 2017). Microorganisms responsible for the $CH_4$ oxidation activity measured in this study comprise unknown high-affinity methanotrophic bacteria, including those possessing methane monooxygenase enzymes belonging to USCα (Pratscher *et al.*, 2018; Ricke *et al.*, 2005) and USCγ (Holmes *et al.*, 1999; Knief *et al.*, 2003) genotypes and/or conventional methanotrophic bacteria under starvation (Cai *et al.*, 2016). Consequently, two main mechanisms could explain $CH_4$ oxidation activity losses in soil microcosms exposed to elevated $H_2$ concentrations. Indeed, high-affinity methanotrophic bacteria, or conventional methanotrophs under starvation, could possess (i) oxygen-tolerant membrane-bound [NiFe]-hydrogenases, thus conferring a flexible Knallgas-methanotrophic lifestyle that uses $H_2$ preferentially when both gases are present due to its lower activation energy requirements, or (ii) high-affinity [NiFe]-hydrogenases for which saturation at elevated $H_2$ concentrations causes the accumulation of reducing equivalents and ultimately compromise overall cell viability.

In addition to a decrease in methanotrophic activity, the proximity of $H_2$ point sources also drastically reduces atmospheric CO oxidation rates, including a 4.4-6.4-fold reduction in high-affinity CO oxidation rate at the soil-point source interface. Strikingly, we found strong covariation between high-affinity $H_2$ and CO oxidation rates throughout the $H_2$ gradient. This is in accordance with field observations of both trace gases oxidizing activities in arable field, forest, andisol and volcanic deposit ecosystems that suggested either the existence of microorganisms possessing the necessary enzymatic machinery for both atmospheric $H_2$ and CO uptake or that both functional groups shared the same ecological niche (King *et al.*, 2008; Yonemura *et al.*, 1999; Yonemura *et al.*, 2000). Detection of both genes encoding [NiFe]-hydrogenase and CO-dehydrogenase in

Proteobacteria (Pohlmann *et al.*, 2006), Actinobacteria (King, 2003) and Chloroflexi (King & King, 2014) tends to support the hypothesis that atmospheric $H_2$ and CO oxidation is performed simultaneously by the same bacteria as part of a mixotrophic lifestyle. Whether this metabolic versatility towards trace gases is widespread in environmentally-relevant HOB remains elusive as none of the 664 and 136 genotypes of HOB and CO-oxidizing bacteria detected in soil corresponded to reference sequences from public genomic databases (Lalonde & Constant, 2016; Piché-Choquette *et al.*, 2017). Furthermore, metagenomic surveys performed in desert and Antarctic soils led to the assembly of partial genome sequences of *Pseudonocardia* (Lynch *et al.*, 2014) and Actinobacteria, AD3 and WPS-2 (Ji *et al.*, 2017) genotypes, respectively, encoding for both $H_2$ and CO metabolism. It is expected that the energy brought by high-affinity $H_2$- and CO-oxidation helps in fulfilling energetic maintenance needs to support dormant cell survival in oligotrophic environments until better conditions arise (Conrad, 1999; Constant *et al.*, 2011b; Constant *et al.*, 2010; Quiza *et al.*, 2014) or support mixotrophic growth in conjunction with organic compounds (Greening *et al.*, 2014b).

In conclusion, $H_2$ concentration gradients found across various environments such as earthworms tunnel networks (Wüst *et al.*, 2009), cyanobacterial mats (Nielsen *et al.*, 2015) or oxic-anoxic interfaces could have an impact on soil biogeochemical processes. Close examination of microbial spatial patterns in the rhizosphere and bulk soil of legume plants represents a promising approach to investigate the environmental impact of $H_2$ and challenge theoretical models depicting successions and interactions between $H_2$, CO and $CH_4$ oxidation activities observed in the soil microcosms utilized in this study. Indeed, our framework predicted that soil CO- and $CH_4$-oxidation activities could be affected beyond the rhizosphere (*i.e.* 1 mm from plant roots). While the impact of $N_2$-fixing legume crops on the soil atmospheric CO sink remains unexplored, a few investigations reported either a significant transient decrease or no impact on atmospheric $CH_4$ uptake activity (Guardia *et al.*, 2016a; Guardia *et al.*, 2016b; Pandey & Agrawal, 2015; Sanz-Cobena *et al.*, 2014). Incidence of environmental factors such as gas diffusion in soil, nutrients, root exudates, $N_2$-fixing nodule density and depth of the $H_2$-emitting point sources within the soil profile will also need to be examined for their potential interference on the biogeochemical

interactions and spatial patterns predicted in this study. Variation of $N_2$-fixation activity during plant developmental stages (Jensen, 1987) and in response to soil temperature (Cabeza *et al.*, 2015) will need to be considered in the elaboration of these future investigations. Finally, isolation efforts towards microorganisms consuming atmospheric trace gases must be pursued in order to shed light on the metabolic flexibility of HOB and explore the importance of $H_2$ in supplying a microbial seedbank sustaining ecosystem services.

## 6.7 Accession numbers

Raw sequence reads of 16S rRNA and ITS genes from PCR amplicon sequencing were deposited in the Sequence Read Archive of the National Center for Biotechnology Information under BioProject PRJNA423817.

## 6.8 Acknowledgments

This work was supported by a Natural Sciences and Engineering Research Council of Canada Discovery grant to P.C. The authors acknowledge use of McGill University and Génome Québec Innovation Centre for the preparation of 16S rRNA and ITS gene libraries and sequencing services. S.P.-C. is grateful to the Fondation Universitaire Armand-Frappier for her Ph.D. scholarship. The authors are very grateful to Prof. Ingeborg Levin from Institut für Umweltphysik, Germany, for her valuable advice regarding the hydrogen fluxes theoretical model and underlying equations. The authors also thank Prof. Richard Villemur from INRS-Institut Armand-Frappier for his input regarding the paper.

## 6.9 Supplementary material

Tables S1 to S4 are found in the Appendix.

# 7 DISCUSSION

La présente section met en perspective l'intégration des chapitres 3 à 6 aux objectifs énoncés au chapitre 2 ainsi que de nouveaux résultats, interprétations ou perspectives associés à ceux-ci. Les 3 objectifs se succèdent, suivis d'une section conclusion et perspectives résumant l'apport et l'importance de cette thèse parmi la littérature scientifique. À titre de rappel, les objectifs de recherche énoncés au chapitre 2 étaient les suivants :

1. *Définir les effets directs de l'$H_2$ sur la structure et la fonction des communautés microbiennes du sol*

2. *Définir les effets indirects de l'$H_2$ sur la structure et la fonction des communautés microbiennes du sol*

3. *Définir les relations de dose-réponse entre la concentration d'$H_2$ et les altérations des communautés microbiennes du sol*

## 7.1 Effets directs de l'$H_2$ sur la structure et la fonction des communautés microbiennes du sol

Le chapitre 3 a entre autres revisité les premières études (Osborne *et al.*, 2010; Stein *et al.*, 2005; Zhang *et al.*, 2009) exposant des sols à de fortes concentrations d'$H_2$ (*ca.* 500 ppmv) afin d'en évaluer l'impact sur le métabolisme de l'$H_2$ et la diversité des communautés microbiennes. Les résultats présentés dans ce chapitre ont montré qu'en effet, l'exposition de sol à l'$H_2$, pendant 10 jours, altérait la cinétique d'oxydation de l'$H_2$, globalement en augmentant le $K_{m(app)}$ et le $V_{max}$ du système. En observant la variation de la vitesse d'oxydation d'$H_2$ au cours d'une série temporelle, on en déduit que la réponse des microorganismes à un enrichissement en $H_2$ s'opère dès les premières 24h, la vitesse d'oxydation d'$H_2$ à haute affinité ayant déjà chuté drastiquement par rapport aux

microcosmes témoins exposés à des concentrations d'$H_2$ ambiantes. Enfin, l'existence de ce type d'effet « direct » de l'$H_2$ sur les communautés microbiennes du sol était attendue, sachant entre autres que les cinétiques d'oxydation d'$H_2$ de sols exposés à l'$H_2$ avaient été réalisées par le passé (Dong & Layzell, 2001) et avaient démontré un changement de $K_m$ et $V_{max}$ suite à l'exposition du sol à l'$H_2$. La plupart des enzymes ne ciblant qu'un seul substrat suivent une cinétique enzymatique de type Michaelis-Menten, *i.e.* qu'elles démontrent une augmentation de vitesse d'utilisation du substrat lors de l'augmentation de la concentration du substrat, ce qui est le cas chez les hydrogénases en général. Chez certaines hydrogénases à haute affinité, lorsque la concentration en substrat dépasse un certain seuil de concentration critique d'$H_2$ après l'atteinte de la $V_{max}$, plutôt que l'activité atteigne un plateau peu importe l'augmentation en substrat, on observe plutôt une diminution de l'activité jusqu'à ce que celle-ci tende vers 0. Cela implique que la cinétique de Michaelis-Menten inclus non seulement un $K_m$ mais aussi un $K_i$ (constante d'inhibition). Cette tendance a été observée chez d'autres microorganismes démontrant une haute affinité pour leur substrat tel que *Stappia* sp. KB902 approvisionnée avec plus de 500 ppmv de CO (Weber & King, 2007). Cette même tendance pourrait avoir lieu chez les HA-HOB mais il faudrait revisiter les cinétiques d'oxydation de l'$H_2$ antérieures, car celles-ci ne comprenaient pas de concentrations d'$H_2$ suffisamment élevées pour atteindre une inhibition de l'activité d'oxydation d'$H_2$ (Constant *et al.*, 2008). Advenant que le $K_i$ d'une HA-HOB soit beaucoup plus faible que les concentrations retrouvées dans un écosystème donné (*e.g.* $K_i$ de 500 ppmv à proximité de nodules fournissant 20 000 ppmv $H_2$), cela impliquerait que l'$H_2$ ne serait plus viable à titre de source d'énergie supplémentaire pour la croissance ou la survie puisqu'il ne permettrait plus d'obtenir de l'énergie. Cela serait d'autant plus nuisible si l'hydrogénase du groupe est exprimée constitutivement. Chez des microorganismes comme *Streptomyces avermitilis*, dont la viabilité en milieu oligotrophe dépend de l'$H_2$ par mixotrophie de survie (Liot & Constant, 2016), un surplus d'$H_2$ inactiverait l'activité d'oxydation d'$H_2$, ce qui se traduirait par une perte de viabilité de la cellule au même titre que celle observée chez les souches mutantes ayant une délétion de l'hydrogénase à haute affinité. On dénoterait donc un désavantage significatif des HA-HOB face aux bactéries Knallgas dans un milieu riche en $H_2$, voire un désavantage face aux autres

stratèges k. En contrepartie, chez des HA-HOB comme *Mycobacterium smegmatis* (Greening *et al.*, 2014b), dont les multiples hydrogénases apportent un avantage de croissance, on peut estimer que leur compétitivité dans un écosystème riche en $H_2$ serait augmentée. Leur avantage évolutif réside donc en leur capacité à s'adapter aux milieux pauvres ou riches en $H_2$, alors que les bactéries Knallgas dominent dans les milieux riches en $H_2$ comme la rhizosphère des légumineuses. Certaines espèces intermédiaires comme la Knallgas thermophile *Methylacidiphilum fumariolicum* (Mohammadi *et al.*, 2017) possède à la fois des hydrogénases à haute et faible affinité, ce qui leur permet de s'adapter à une niche écologique instable et donc d'alterner entre un métabolisme autotrophe, mixotrophe et hétérotrophe. À noter que des gènes codant pour une hydrogénase à haute affinité ont été détectés chez ce microorganisme, mais les propriétés catalytiques de celle-ci n'ont pas été mesurées. Il est donc possible que cette hydrogénase du groupe 5 soit aussi à faible affinité comme celle de *Cupriavidus necator*, donc la précédente affirmation demeure hypothétique.

Une innovation de cette étude par rapport aux précédentes est que l'observation d'effets directs de l'$H_2$ a été effectuée sous forme de série temporelle, qu'on se réfère à l'oxydation de l'$H_2$ ou à la structure des communautés microbiennes. De plus, la structure de la communauté bactérienne a été analysée via des données de séquençage à haut débit d'amplicons du gène codant pour l'ARNr 16S. Contrairement à ses prédécesseurs, cette étude a permis d'avoir un portrait plus global de la structure de la communauté bactérienne, c'est-à-dire comprenant à la fois la biosphère majoritaire et rare, ainsi qu'une résolution taxonomique plus fine. En effet, plutôt que d'observer une variation significative de l'abondance relative de quelques groupes taxonomiques larges (*e.g.* augmentation des γ-Proteobacteria et une diminution des α-Proteobacteria et des Actinobacteria dans l'étude de Zhang *et al.* (2009) par exemple), nos analyses ont plutôt montré une altération de l'abondance relative de centaines d'unités taxonomiques opérationnelles (OTU). Plus exactement, 958 OTU ont été affectés par une exposition à l'$H_2$, la plupart d'entre eux étant issus de la biosphère rare, ce qui peut expliquer certaines nuances obtenues entre cette étude et les précédentes considérant la différence de sensibilité entre les deux techniques. La meilleure résolution taxonomique de notre étude a aussi permis de

constater, pour une première fois, que parmi un même groupe taxonomique, les différents OTU le constituant avaient des réponses parfois complètement divergentes. Chez les microorganismes, les hauts rangs taxonomiques apportent peu d'informations cohérentes en termes de rôles écologiques et physiologiques (Philippot *et al.*, 2010). Ainsi, la réponse divergente d'espèces au sein d'un haut rang taxonomique peut masquer la réponse observée à un plus bas rang, d'où l'utilité d'employer des unités taxonomiques plus étroites comme les OTU ou, depuis plus récemment, les ASV ou ESV (Amplicon Sequence Variant ou Exact Sequence Variant selon les textes) (Callahan *et al.*, 2017). L'altération des réseaux de corrélations associés aux profils d'abondance relative des OTU bactériens a aussi soulevé ce même point. Plusieurs des OTU affectés par l'$H_2$ semblent avoir une affiliation taxonomique pouvant correspondre à des HOM (*e.g.* ceux du genre *Bacillus*), mais nombre d'entre eux ne semblent pas correspondre à ceux-ci (*e.g.* ceux de l'ordre Phycisphaerales). Ce point laisse présager un effet indirect de l'$H_2$, c'est-à-dire que l'$H_2$ pourrait affecter l'ensemble du microbiote du sol plutôt que seulement les HOM, via des interactions microbe-microbe synergiques ou antagonistes par exemple. Ce volet « effet indirect » est davantage exploité au chapitre 4, où le choix des analyses effectuées a davantage permis de discriminer les HOM des non-HOM.

Le chapitre 5 a aussi permis d'étudier les effets directs de l'$H_2$, celui-ci concernant strictement les HOM possédant une hydrogénase du groupe 5, c'est-à-dire les HA-HOB. Ce chapitre a documenté l'altération du profil d'abondance relative des HA-HOB et de leur richesse spécifique face à un influx d'$H_2$, en plus de démontrer la vaste diversité des HA-HOB dans les sols et de mesurer le potentiel explicatif de marqueurs du processus d'oxydation d'$H_2$ à haute affinité. Le dispositif expérimental employé fut le même qu'au chapitre 4. Sachant que l'étendue de la diversité des HA-HOB était inconnue, ce chapitre a tenté de remédier à la situation via une analyse phylogénétique du gène codant pour la grande sous-unité des hydrogénases à haute affinité, *hhyL*. Les séquences du gène partiel *hhyL* des 664 OTU *hhyL* (*i.e.* ici, les OTU correspondent à des séquences du gène *hhyL* et non de l'ARNr 16S) provenant des microcosmes de sol ont été comparées, en utilisant un arbre phylogénétique consensus impliquant trois algorithmes différents, avec celles des séquences du gène *hhyL* déjà présentes dans les bases de données

génomiques. Or, il n'y avait aucune correspondance entre la base de données expérimentales et la base de données génomiques. Cet aspect est tout de même surprenant considérant que ce groupe fonctionnel de microorganismes représentent 70-80% du puits de l'$H_2$ atmosphérique (Constant *et al.*, 2009; Ehhalt & Rohrer, 2009; Pieterse *et al.*, 2013) et pourtant, l'identité de ce groupe reste largement inconnue. Cela s'explique par le fait que les hydrogénases présentes dans les bases de données publiques sont associées à des microorganismes dont le génome a été séquencé pour d'autres motifs que la caractérisation des HA-HOB ou de leur diversité, c'est-à-dire que le génome de la plupart des HA-HOB n'a probablement pas encore été séquencé. Par exemple, plusieurs Streptomycètes ont probablement été séquencés pour leur potentiel à produire des antibiotiques d'intérêt industriel et non pour leur potentiel en tant que HA-HOB.

Un autre aspect important associé à cette analyse phylogénétique est qu'elle contredit une reclassification phylogénétique qui a été effectuée récemment (Greening *et al.*, 2016). L'analyse mentionnée dans cet article impliquait que les hydrogénases à haute affinité, *i.e.* groupe 5 dans cette thèse ou 1h dans l'article cité, seraient inclues parmi les hydrogénases du groupe 1 (MBH) plutôt que de représenter un groupe distinct d'hydrogénases tel que déterminé par le passé par (Constant *et al.*, 2011b). Deux principales raisons pourraient expliquer notre incapacité à reproduire l'analyse phylogénétique de Greening *et al.* (2016). D'abord, l'alignement des gènes structuraux associée à notre analyse ne prenait en compte que la portion du gène amplifiée par PCR alors que celle de l'autre équipe de recherche reposait sur le gène complet. Ensuite, l'analyse phylogénétique effectuée par l'équipe de Greening a été effectuée sur des données pré-compartimentées par groupe hypothétique (donc groupes 1 et 5 analysés ensemble sans les groupes 2 à 4) et que seul l'algorithme « neighbour-joining » a été employé. L'analyse phylogénétique du chapitre 5, ainsi que celle réalisée avec le gène complet (Constant *et al.*, 2011b), ont été effectuées selon les règles de l'art (*i.e.* consensus de trois algorithmes avec un nombre de « bootstraps » élevé) et remettent donc en question la phylogénie des hydrogénases du groupe 5/1h. En effet, l'analyse phylogénétique du chapitre 5, incluant tous les groupes d'hydrogénases dans le même

alignement, a montré que les hydrogénases à haute affinité forment une branche séparée des autres groupes et donc qu'elles consistent en un groupe distinct (groupe 5) plutôt qu'un sous-groupe du groupe 1 (groupe 1h). Une reclassification complète des 5 groupes d'hydrogénases basée sur le nombre et la similarité des séquences associées aux différentes sous-unités, gènes structuraux et gènes accessoires (*e.g.* arbre phylogénétique concaténé) pourrait être effectuée afin d'obtenir une classification fiable.

D'autre part, le chapitre 5 a montré que, comme pour les communautés microbiennes des chapitres 3 et 4, l'altération des profils d'abondance relative des différents génotypes (*i.e.* OTU) de HA-HOB est idiosyncratique, bien que les changements observés soient strictement au niveau de l'individu (*i.e.* OTU *hhyL*). Ceci met de l'avant un défi considérable associé au domaine de la biogéochimie, où les tendances actuelles visent à utiliser la distribution des génotypes présents dans un milieu donné à titre d'outil de prédiction des processus environnementaux, tels que ceux liés au cycle du carbone et de l'azote, plutôt que des proxys actuellement employés comme les paramètres physicochimiques des sols. Bien que cette approche soit en apparence valide, la corrélation entre la structure de la communauté globale, ou celle d'un groupe fonctionnel spécifique, et le processus ciblé, est rarement élevée (Bailey *et al.*, 2018; Graham *et al.*, 2016; Graham *et al.*, 2014; Krause *et al.*, 2014a; Rocca *et al.*, 2015). Une approche de caractérisation des traits écologiques (Krause *et al.*, 2014b) associés à certains groupes de HOM permettrait d'anticiper la réponse de ce groupe fonctionnel face aux changements environnementaux et donc d'effectuer une gestion plus éclairée au niveau des pratiques de protection de l'environnement. Les travaux de Ho *et al.* (2016) ont par exemple permis de classer les bactéries méthanotrophes (MOB) en fonction de leur niveau de tolérance à des perturbations environnementales telles que leur résilience à des périodes successives de dessication/humidification. Cette classification consiste en un cadre théorique utilisé en tant que référence afin d'évaluer la susceptibilité de cette guilde fonctionnelle à des modes de gestion environnementale (Kim *et al.*, 2014; Reumer *et al.*, 2018). De manière homologue, pour les HOM, plusieurs isolats environnementaux sont actuellement caractérisés en termes de cinétique d'oxydation d'$H_2$ (*i.e.* utilité du groupe fonctionnel en question) en fonction de la concentration d'$H_2$, de manière à déterminer leur $K_{m(app)}$ pour l'$H_2$, ainsi que leur cinétique de croissance en fonction de leur

mode de vie, *i.e.* autotrophe, mixotrophe, hétérotrophe. Le mode de vie préférentiel ainsi que l'affinité envers l'$H_2$ consistent en des traits écologiques permettant de classer les HOM en sous-groupes. Advenant que d'autres traits soient mesurés et qu'on puisse déterminer comment les sous-groupes répondent à un stress anticipé (*e.g.* hausse de température via les changements climatiques) ou induit (*e.g.* coupes à blanc, ajout d'intrants azotés), il serait possible de prédire l'impact des changements environnementaux sur la fonction d'intérêt (oxydation d'$H_2$). Par exemple, lors d'une étude longitudinale ou d'un changement environnemental par gradient (*e.g.* un stress ou un stimulus), on pourrait ainsi effectuer le criblage de la guilde fonctionnelle via séquençage à haut débit et par la suite estimer comment la capacité d'oxydation d'$H_2$ du milieu sera affectée en fonction de la structure de la guilde fonctionnelle et ce, en se basant sur la base de référence de traits écologiques construite préalablement.

Mis à part la répartition des différents individus d'une guilde fonctionnelle, dont la contribution relative au processus a été préalablement évaluée, la quantification du nombre d'individus de cette guilde via qPCR/RT-qPCR peut aussi s'avérer utile lors de l'élaboration de modèles biogéochimiques. Cette notion fut supportée au chapitre 5, où les variables de richesse spécifique des HA-HOB ainsi que du nombre de copies et de transcrits du gène *hhyL* furent testés à titre de prédicteurs de l'activité d'oxydation d'$H_2$ à haute affinité (mesurée au chapitre 4). Les résultats ont montré que seule la quantification du ratio du nombre de transcrits sur le nombre de gènes *hhyL* permettait d'expliquer substantiellement l'activité mesurée. Cela est peu étonnant considérant une méta-analyse montrant que l'abondance absolue de gènes associés à diverses fonctions généralistes et spécialistes n'était pas corrélée à ladite fonction (Rocca *et al.*, 2015). La combinaison de ce marqueur au contenu en C des sols étudiés, ce dernier étant un bon indicateur de l'activité des HA-HOB (Khdhiri *et al.*, 2015), pourrait d'ailleurs permettre d'estimer plus adéquatement les puits d'$H_2$ atmosphérique locaux.

## 7.2 **Effets indirects de l'$H_2$ sur la structure et la fonction des communautés microbiennes du sol**

Un des aspects davantage novateurs de cette thèse est la démonstration de l'existence d'effets indirects de l'$H_2$ sur la structure et la fonction des communautés microbiennes du sol. Ayant comme rationnel que l'ajout d'une source d'énergie spécifique (*i.e.* $H_2$) devrait stimuler le groupe fonctionnel ciblé (*i.e.* HOM), on suppose que ce groupe fonctionnel affectera les autres microorganismes du milieu, via des interactions microbe-microbe agonistes ou antagonistes. Cette prédiction s'est appuyé sur 3 constats; un gradient d'intrant énergétique (*i.e.* production primaire) affecte différentiellement la diversité de génotypes au sein d'embranchements bactériens (Horner-Devine *et al.*, 2003); l'$H_2$ est une source d'énergie ubiquitaire, qui diffuse facilement et qui nécessite une faible énergie d'activation (Morita, 1999); les HOM sont nombreuses en termes de nombres de génotypes et ont une très grande diversité taxonomique (Greening *et al.*, 2016; Piché-Choquette *et al.*, 2017; Vignais & Billoud, 2007), ce qui implique potentiellement une grande diversité fonctionnelle.

D'abord, au chapitre 4, les microcosmes de sol ont été exposés à des concentrations d'$H_2$ de 10 000 ppmv pendant 15 jours de manière à refléter davantage les « hotspots » d'$H_2$ environnementaux. Trois sols issus d'utilisation des terres différentes et ayant des paramètres physicochimiques contrastant ont d'ailleurs été employés afin de déterminer si les réponses observées étaient universelles ou non. On constate d'abord, via des données de séquençage à haut débit ciblant les gènes codants pour l'ARNr 16S (bactérien) ou l'ITS2 (fongique), que la structure des communautés microbiennes est largement altérée par l'exposition à l'$H_2$ et ce, au même titre qu'au chapitre 3. On remarque une altération globale au niveau de la communauté, par exemple via l'altération de la richesse spécifique, ainsi qu'au niveau de l'individu. Les éléments susmentionnés n'apportent cependant pas une preuve de l'existence d'effets indirects. En effet, le premier effet indirect observé fut l'altération significative de l'abondance relative des Champignons à la suite de leur exposition à l'$H_2$, considérant qu'aucun Champignon connu ne possède d'hydrogénase. Ensuite, le séquençage d'amplicons de PCR a été

complémenté par une étude métagénomique qui a permis de distinguer les HOM des non-HOM chez les microorganismes les plus abondantes. Les différents contigs des métagénomes ont d'abord été reconstitués sous forme de génome partiels (*i.e.* genome bins ou metagenome-assembled genomes) issus de quelques microorganismes apparentés. Par la suite, des modèles de Markov caché, associés aux motifs conservés des différents groupes d'hydrogénases, ont été utilisés afin de cibler les hydrogénases présentes parmi l'ensemble des contigs, inclus ou non parmi les génomes partiels. Cela a donc permis de constater que l'abondance relative de plusieurs microorganismes ne possédant pas le potentiel génétique d'oxyder l'$H_2$ a tout de même été altérée par l'exposition à l'$H_2$. Ces 2 aspects ont donc confirmé l'existence de l'effet indirect de l'$H_2$ sur la structure des communautés microbiennes du sol. De manière surprenante, l'altération de l'abondance relative des OTU et génomes partiels des sols exposés à 10 000 ppmv $H_2$, par rapport à l'exposition à 0.5 ppmv $H_2$, divergeait entre les sols, *idem* pour leur richesse spécifique. On en déduit que le lien entre la diversité microbienne et l'ajout d'énergie, ici via l'$H_2$, est variable selon les sols et donc que la structure de la communauté microbienne du sol n'est pas affectée de manière universelle entre les utilisations des terres employées. L'effet indirect de l'$H_2$ sur les fonctions des communautés microbiennes du sol était tout autre chez les fonctions généralistes et spécialistes étudiées.

La fonction généraliste employée dans ce cadre-ci consistait en l'utilisation de 31 substrats carbonés via des plaques Ecoplate®, dont l'activité peut provenir d'une vaste gamme de microorganismes, incluant notamment les Champignons. En contrepartie, la fonction spécialiste étudiée consistait en l'oxydation du $CH_4$ atmosphérique, qui est en principe limitée aux bactéries méthanotrophes. L'exposition des sols à l'$H_2$ a ainsi diversifié l'utilisation des sources de carbone et a diminué d'environ 50-75% l'activité d'oxydation du $CH_4$ et ce, pour les trois sols. Cela implique que l'$H_2$ a des effets directs et indirects sur les fonctions des communautés microbiennes du sol et que ces fonctions répondent d'ailleurs de manière universelle à l'ajout de substrat, contrairement à la structure de cette même communauté. Bien que l'intensité de ces réponses fonctionnelles fût variable d'un sol à l'autre, celles-ci ont pu être expliquées par différents

paramètres biotiques (richesse spécifique bactérienne et fongique) et abiotiques ($H_2$, paramètres physicochimiques du sol, etc.) via des modèles d'équations structurales. Comme les réponses absolues de ces processus face à l'exposition à l'$H_2$ fut universelle chez les trois sols mesurés, cela laisse présager que la modélisation de ces processus en termes de dose-réponse entre la concentration d'exposition à l'$H_2$ et l'altération du processus sera possible. On en déduit aussi que les interactions microbe-microbe entre HOM et non-HOM permettent de réorganiser la dynamique des populations microbiennes du sol de manière à moduler certains processus issus du cycle du carbone (*i.e.* substrats carbonés et $CH_4$).

Par ailleurs, la structure des communautés microbiennes semble plus résistante à l'$H_2$ que les fonctions du microbiote, effets directs et indirects confondus. Cette affirmation provient du fait que la distance entre la structure des communautés (diversité β) est plus faible entre traitements chez un même sol qu'entre les sols, alors que chez les fonctions du microbiote, on constate plutôt l'inverse. Une explication possible est que l'apport en $H_2$, bien qu'utile d'un point de vue énergétique, ne soit pas suffisant pour augmenter la biomasse des HOM en conditions limitantes de carbone organique et inorganique. De nombreuses HOM peuvent fixer le $CO_2$, alors que d'autres bénéficient des plus fortes concentrations de carbone organique retrouvées dans la rhizosphère, ce à quoi elles n'avaient pas accès dans cette étude. De plus, il est possible que la biomasse présente, donc séquencée ou amplifiée par qPCR, ne soit pas forcément active et donc que la proportion du microbiote affectée (*i.e.* vivant et/ou actif) sur le microbiote total soit faible. Cela implique que le bruit de fond causé par l'ADN relique ou les microorganismes dormants (*i.e.* seed banks) limiterait ou inhiberait complètement les tendances qu'on aurait pu observer.

Une autre hypothèse concernant cette « résistance » de la structure des communautés au sein d'une même utilisation des terres vient probablement davantage du fait que le système de microcosmes est un environnement fermé. Cela implique que l'ajout de diversité provenant de l'extérieur du système est impossible (*e.g.* dépôts de litière,

transport de microorganismes par aérosols), ce qui réduit donc les potentiels variations via l'invasion d'espèces allochtones. On peut donc avoir une diminution mais pas une augmentation du nombre d'espèces *a priori*. Dans cette optique, il serait possible de tester cette assomption, en microcosmes de manière à contrôler les autres paramètres, en conservant cet environnement fermé et en ajoutant à certains moments des « inocula » de microorganismes afin de voir si l'exposition à l'$H_2$ a un impact sur la permissivité de la colonisation de microorganismes allochtones. On pourrait imaginer une étude effectuée sur le terrain ayant un but similaire utilisant comme traitement des infections nodulaires isogènes *Hup*$^+$ et *Hup*$^-$ (traitements à faible et haute concentrations d'$H_2$, respectivement) dont les inocula microbiens seraient des intrants organiques dont le contenu microbien pourrait s'établir, ou non, parmi le microbiote présent dans les sols.

## 7.3 **Relations de dose-réponse entre la concentration d'$H_2$ et les altérations des communautés microbiennes du sol**

Ce dernier objectif de recherche se limite strictement au chapitre 6 et concerne la relation de dose-réponse entre la dose d'exposition du sol à l'$H_2$ et la dynamique des communautés microbiennes du sol dans le but de modéliser, d'un point de vue spatial, la réponse des communautés microbiennes en fonction de leur distance d'une source ponctuelle productrice d'$H_2$. En ce qui a trait à l'altération de la structure des communautés microbiennes du sol en réponse à l'$H_2$, de nombreux ASV avaient une abondance relative corrélée significativement à la dose d'exposition d'$H_2$. Cependant, la structure globale des communautés microbiennes ainsi que leur richesse spécifique ne suivaient pas le gradient de concentration d'$H_2$, cela laissant croire que l'altération de la richesse spécifique observée au chapitre 4 ne serait pas reproductible ou que les variations annuelles de ces sols (*i.e.* facteurs météorologiques, infiltration de litière différentielle, déposition d'espèces par le vent ou les précipitations, etc.) pourrait interférer avec l'effet de l'$H_2$ sur la biodiversité. Cette analyse a été testée avec des OTU plutôt que des ASV et la conclusion reste la même.

Ces résultats contrastent grandement avec l'altération des fonctions d'oxydation de gaz traces ($H_2$, $CH_4$ et CO) en réponse à l'exposition à différents ratios de mélange d'$H_2$. Le cadre théorique a d'abord été calculé de manière à estimer la distribution spatiale des ratios de mélange d'$H_2$ à proximité de sources ponctuelles d'$H_2$. À chaque bond de 0.1 cm de la source d'$H_2$ théorique, les vitesses d'oxydation d'$H_2$ (haute et faible affinités) ont été ajustées en fonction du ratio de mélange résiduel d'$H_2$ via les équations logarithmiques représentant la dose-réponse entre la concentration d'$H_2$ et la vitesse d'oxydation d'$H_2$. Ce cadre théorique prédit d'abord que l'oxydation quasi-complète de l'$H_2$ est réalisée en 13,5 cm ou moins de la source ponctuelle d'$H_2$, ce qui est comparable (*i.e.* même échelle de grandeur) à ce qui a été rapporté par le passé dans une étude menée en serre (La Favre & Focht, 1983). À la suite de l'obtention de cette distribution spatiale de l'$H_2$ (concentration résiduelle en fonction de la distance de la source), les équations associées à la dose-réponse entre la concentration d'$H_2$ et la vitesse d'oxydation de gaz traces ($H_2$, $CH_4$ et CO) ont été implémentées dans le cadre théorique de manière à projeter l'altération des fonctions (*i.e.* vitesses d'oxydation) en fonction de la distance d'une source ponctuelle d'$H_2$. La projection de ces processus sur le gradient d'$H_2$ prédit implique que l'oxydation de gaz traces serait affectée jusqu'à pratiquement 10 cm de la source d'$H_2$, avec parfois une diminution de jusqu'à 90% de l'activité à proximité de la source, l'oxydation du CO étant davantage affectée que celle du $CH_4$. Deux corrélations significatives ont été obtenues en comparant l'altération de ces processus face à l'exposition à l'$H_2$, soient une corrélation positive entre l'oxydation d'$H_2$ et de CO à haute affinité ainsi qu'une corrélation négative entre l'oxydation d'$H_2$ à faible affinité et du $CH_4$ à haute affinité. Ceci laisse présager que ces combinaisons de processus pourraient être interreliés, soient par la juxtaposition de niches écologiques entre deux groupes fonctionnels ou via l'utilisation des deux substrats par un même organisme. Deux mécanismes sont possibles par rapport à cette constatation, les deux n'étant pas mutuellement exclusifs et pouvant donc avoir lieu simultanément.

Premièrement, il est possible qu'un même microorganisme puisse réaliser ces 2 processus. Bien que non détecté dans les sols employés au cours de cette thèse, on sait que des microorganismes tels que *Mycobacterium smegmatis* (King, 2003) possèdent

une hydrogénase du groupe 5 et une monoxyde de carbone déshydrogénase, ceux-ci peuvent utiliser l'$H_2$ et le CO à des concentrations comparables à celles retrouvées dans l'atmosphère. De plus, certaines bactéries méthanotrophes conventionnelles sont aussi des bactéries Knallgas utilisant l'$H_2$ aérobiquement afin de promouvoir la fixation du $CO_2$ (Carere *et al.*, 2017; Chen & Yoch, 1987; Mohammadi *et al.*, 2017). En revanche, les méthanotrophes en question utilisent des concentrations de $CH_4$ bien plus élevées que celles employées au cours de cette thèse (*ca.* concentrations atmosphériques). Une étude a montré que des bactéries méthanotrophes conventionnelles (*i.e.* faible affinité) pouvaient utiliser le $CH_4$ atmosphérique (*i.e.* haute affinité) sous certaines conditions, ce qui implique que l'hypothèse précédente n'est pas systématiquement écartée. Une autre hypothèse consiste en l'existence de méthanotrophes pouvant utiliser le $CH_4$ atmosphérique grâce à des enzymes spécialisés à haute affinité, dont le groupe USCα (Pratscher *et al.*, 2018). Outre le fait que la réelle identité de ces méthanotrophes oxydant le $CH_4$ atmosphérique soit toujours sujet de débat, on peut tout de même présumer que les 2 groupes de bactéries mixotrophes susmentionnées pourraient être responsables des activités modélisées.

Deuxièmement, il est possible que les microorganismes responsables des processus corrélées soient simplement actifs dans une même niche écologique, c'est-à-dire dans les mêmes conditions environnementales et le même écosystème. Plusieurs microorganismes isolés de sols ont démontré la capacité d'effectuer l'un ou l'autre de ces processus (*e.g. Burkholderia xenovorans* (Weber & King, 2012) et *Streptomyces* sp PCB7 (Constant *et al.*, 2008) pour l'oxydation de CO et d'$H_2$ atmosphérique, respectivement, ou *Cupriavidus necator* pour l'oxydation d'$H_2$ en fortes concentrations (Klüber *et al.*, 1995)), ce qui implique que certains d'entre eux partagent déjà le même écosystème, il s'avère donc possible qu'ils occupent également la même niche écologique et donc que leur activité soit régie par sensiblement les mêmes paramètres physicochimiques et agents stressants. Davantage de microorganismes prodiguant ces activités devront cependant être caractérisés afin de pouvoir répondre à cette assomption. Puisque peu de bactéries Knallgas ou oxydant l'$H_2$ avec haute affinité ont été isolées et caractérisées jusqu'à présent, il devient essentiel d'isoler davantage de ces

souches et de caractériser leur activité enzymatique (*e.g.* $K_m$ et $V_{max}$) d'oxydation d'$H_2$ ainsi que leur mode de vie (*i.e.* autotrophes, mixotrophes, etc.). Une première vague d'isolement de ce type de microorganismes, sur 12 milieux de cultures susceptibles de permettre l'isolement de bactéries oligo- ou autotrophes, a été entamée. Les isolats proviennent de deux expériences, 1) des sols exposés à 10 000 ppmv d'$H_2$ pendant un mois et 2) des extraits de sols exposés à 5 à 50% d'$H_2$, 5 à 50% de $CH_4$, 5% d'$O_2$ et 10% $CO_2$ transférés dans du milieu frais à toutes les semaines pendant 5 semaines. Cela a mené à l'isolement de 165 souches, dont au moins 80 (certaines peuvent être des duplicas) peuvent oxyder soit l'$H_2$ ou le CO atmosphérique, ou l'$H_2$ ou le $CH_4$ en forte concentration.

Sur un autre ordre d'idées, si on retourne aux distances des sources productrices d'$H_2$ où une altération des processus d'oxydation d'$H_2$, de CO et de $CH_4$ est prédite, on constate que ces processus seraient potentiellement affectés au-delà de la rhizosphère. Si l'impact de ces sources d'$H_2$ sur les processus modélisés est le même dans les écosystèmes qu'ils émulent, cela impliquerait que la production d'$H_2$ dans la rhizosphère aurait un impact significatif sur les puits d'$H_2$, de CO et de $CH_4$ atmosphériques. Ainsi, il est possible que l'exposition à l'$H_2$ de bactéries carboxydovores, méthanotrophes ou HA-HOB réduisent leur capacité à oxyder ces gaz. Ainsi, les émissions actuelles seraient susceptibles d'augmenter advenant une amplification des sources d'$H_2$ dans les sols, via une plus grande démocratisation de la culture de légumineuses par exemple, sous réserve que l'ajout des légumineuses dans l'écosystème ne tamponne pas l'effet modélisé. Bien entendu, le dispositif expérimental employé au cours de cette thèse comprend plusieurs limites outre l'absence de plantes dans les microcosmes, entre autres car il assume que la production d'$H_2$ est stable dans le temps, c'est-à-dire au cours d'une journée, mais aussi à travers une saison de culture, ce qui n'est pas le cas *in situ*. Deux principales directions de recherche permettraient d'évaluer l'impact de ces sources d'$H_2$ sur le bilan des échanges sol-air du $CH_4$ et du CO. Premièrement, un dispositif expérimental comportant des plantes légumineuses en pots inoculés avec *Rhizobium leguminosarum* à génotype $Hup^+$ ou $Hup^-$ permettrait de mettre en lumière les interférences que pourraient causer le carbone labile et récalcitrant via les

rhizodépositions ainsi que les cycles diurnes-nocturnes de photosynthèse et de fixation de l'azote sur les impacts de l'$H_2$ observés dans cette thèse. La seconde approche viserait un suivi temporel des flux nets d'$H_2$, de $CH_4$ et de CO provenant de champs de légumineuses *Hup*$^+$ et *Hup*$^-$ et/ou de légumineuses (présumablement *Hup*$^-$) et non légumineuses. En effet, peu d'études en champs comparant les flux de gaz à effets de serre entre plantes légumineuses et non-légumineuses ont été effectuées. Parmi celles-ci, l'application différentielle de fertilisant entre les espèces est un facteur confondant important, sachant que l'ajout de fertilisants azotés augmente la production de $N_2O$ et de $CH_4$ selon plusieurs études (Guardia *et al.*, 2016a; Guardia *et al.*, 2016b; Sanz-Cobena *et al.*, 2014).

## 7.4 Conclusion et perspectives

En somme, les recherches présentées dans cette thèse ont permis de mettre en lumière la vaste diversité taxonomique et fonctionnelle des microorganismes oxydant ou produisant de l'$H_2$ ainsi que la manière dont l'$H_2$ moléculaire module les communautés microbiennes du sol. Plus concrètement, ces recherches ont permis de confirmer l'existence d'effets directs et indirects de l'$H_2$ sur la structure et la fonction des communautés microbiennes de trois sols ayant une physicochimie contrastante, de manière à inférer l'universalité des réponses observées à l'échelle locale, c'est-à-dire sur une même zone géographique et par conséquent le même historique de climat depuis leur conversion. Ainsi, l'exposition des sols à des concentrations d'$H_2$ représentatives de celles retrouvées dans l'environnement affecte de manière idiosyncratique (*i.e.* différemment d'un sol à l'autre et d'une année à l'autre) la structure des communautés microbiennes, tant au niveau de la composition que de la richesse et de l'équitabilité des espèces bactériennes et fongiques. En revanche, à l'échelle locale, la réponse fonctionnelle de ces communautés microbiennes est bel et bien universelle chez les fonctions mesurées. Il semble donc y avoir une redondance fonctionnelle marquée en ce qui a trait aux fonctions régies par l'$H_2$. Des marqueurs moléculaires et génomiques ont aussi été mis à l'épreuve afin de tester leur capacité de prédiction de l'activité des HA-HOB, la conclusion étant que le ratio d'expression du gène *hhyL* est la seule mesure

prédisant significativement cette activité. Des variables biotiques et abiotiques permettent d'ailleurs d'expliquer la variation d'intensité des différentes fonctions généralistes et spécialistes étudiées. Un autre pas dans cette voie a été franchi via la modélisation de l'impact des gradients d'$H_2$ sur la distribution spatiale des processus d'oxydation de gaz traces ($H_2$, $CH_4$ et CO), extrapolant davantage ce à quoi on pourrait s'attendre dans l'environnement. De plus, bien qu'une étude ait montré la vaste diversité des HOM présents dans les bases de données génomiques publiques, la présente thèse a aussi montré que les HA-HOB sont, elles aussi, diverses, mais mal classées et actuellement quasi-absentes des bases de données génomiques. Cette thèse a donc apporté des notions substantielles aux domaines de la biogéochimie et de l'écologie microbienne des sols. Bien que ces nouvelles connaissances permettent de répondre aux questions soulevées initialement, elles apportent aussi un lot de nouvelles questions et hypothèses qui devront être explorées dans le futur.

# 8 REFERENCES


Adams MW (1990) The structure and mechanism of iron-hydrogenases. *Biochimica et biophysica acta* 1020(2):115-145.

Akhmanova A, Voncken F, van Alen T, van Hoek A, Boxma B, Vogels G, Veenhuis M & Hackstein JH (1998) A hydrogenosome with a genome. *Nature* 396(6711):527-528.

Alperin MJ & Hoehler TM (2009) Anaerobic methane oxidation by archaea/sulfate-reducing bacteria aggregates: 2. Isotopic constraints. *American Journal of Science* 309(10):958-984.

Angel R, Matthies D & Conrad R (2011) Activation of methanogenesis in arid biological soil crusts despite the presence of oxygen. *PLoS ONE* 6(5):e20453.

Angle JC, Morin TH, Solden LM, Narrowe AB, Smith GJ, Borton MA, Rey-Sanchez C, Daly RA, Mirfenderesgi G, Hoyt DW, Riley WJ, Miller CS, Bohrer G & Wrighton KC (2017) Methanogenesis in oxygenated soils is a substantial fraction of wetland methane emissions. *Nature Communications* 8(1):1567.

Annan H, Golding A-L, Zhao Y & Dong Z (2012) Choice of hydrogen uptake (Hup) status in legume-rhizobia symbioses. *Ecology and Evolution* 2(9):2285-2290.

Arcand MM, Lemke R, Farrell RE & Knight JD (2014) Nitrogen supply from belowground residues of lentil and wheat to a subsequent wheat crop. *Biology and Fertility of Soils* 50(3):507-515.

Armstrong FA & Albracht SPJ (2005) [NiFe]-hydrogenases: spectroscopic and electrochemical definition of reactions and intermediates. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 363(1829):937-954.

Atta M & Meyer J (2000) Characterization of the gene encoding the [Fe]-hydrogenase from *Megasphaera elsdenii*. *Biochimica et biophysica acta* 1476(2):368-371.

Axelsson R & Lindblad P (2002) Transcriptional regulation of *Nostoc* hydrogenases: effects of oxygen, hydrogen, and nickel. *Applied and Environmental Microbiology* 68(1):444-447.

Baba R, Asakawa S & Watanabe T (2016) $H_2$-producing bacterial community during rice straw decomposition in paddy field soil: estimation by an analysis of [FeFe]-hydrogenase gene transcripts. *Microbes and Environments* 31(3):226-233.

Baba R, Kimura M, Asakawa S & Watanabe T (2014) Analysis of [FeFe]-hydrogenase genes for the elucidation of a hydrogen-producing bacterial community in paddy field soil. *FEMS Microbiology Letters* 350(2):249-256.

Bai Y, Müller DB, Srinivas G, Garrido-Oter R, Potthoff E, Rott M, Dombrowski N, Münch PC, Spaepen S, Remus-Emsermann M, Hüttel B, McHardy AC, Vorholt JA & Schulze-Lefert P (2015) Functional overlap of the *Arabidopsis* leaf and root microbiota. *Nature* 528(7582):364-369.

Bailey VL, Bond-Lamberty B, DeAngelis K, Grandy AS, Hawkes CV, Heckman K, Lajtha K, Phillips RP, Sulman BN, Todd-Brown KEO & Wallenstein MD (2018) Soil carbon cycling proxies: Understanding their critical role in predicting climate change feedbacks. *Global Change Biology* 24(3):895-905.

Bell T, Newman JA, Silverman BW, Turner SL & Lilley AK (2005) The contribution of species richness and composition to bacterial services. *Nature* 436(7054):1157-1160.

Benemann JR (2000) Hydrogen production by microalgae. *Journal of Applied Phycology* 12(3):291-300.

Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J & Sayers EW (2013) GenBank. *Nucleic Acids Research* 41:D36-D42.

Bergmann FH, Towne JC & Burris RH (1958) Assimilation of carbon dioxide by hydrogen bacteria. *Journal of Biological Chemistry* 230(1):13-24.

Berman-Frank I, Lundgren P, Chen Y-B, Küpper H, Kolber Z, Bergman B & Falkowski P (2001) Segregation of nitrogen fixation and oxygenic photosynthesis in the marine cyanobacterium *Trichodesmium*. *Science* 294(5546):1534-1537.

Berney M & Cook GM (2010) Unique flexibility in energy metabolism allows mycobacteria to combat starvation and hypoxia. *PLoS One* 5(1):e8614.

Berney M, Greening C, Hards K, Collins D & Cook GM (2014) Three different [NiFe]-hydrogenases confer metabolic flexibility in the obligate aerobe *Mycobacterium smegmatis*. *Environmental Microbiology* 16(1):318-330.

Bhowmik A, Cloutier M, Ball E & Bruns MA (2017) Underexplored microbial metabolisms for enhanced nutrient recycling in agricultural soils. *AIMS Microbiology* 3(4):826-845.

Blaser MJ, Cardon ZG, Cho MK, Dangl JL, Donohue TJ, Green JL, Knight R, Maxon ME, Northen TR, Pollard KS & Brodie EL (2016) Toward a predictive understanding of Earth's microbiomes to address 21st century challenges. *mBio* 7(3).

Boetius A, Ravenschlag K, Schubert CJ, Rickert D, Widdel F, Gieseke A, Amann R, Jørgensen BB, Witte U & Pfannkuche O (2000) A marine microbial consortium apparently mediating anaerobic oxidation of methane. *Nature* 407:623.

Boisvert S, Raymond F, Godzaridis E, Laviolette F & Corbeil J (2012) Ray Meta: scalable de novo metagenome assembly and profiling. *Genome Biology* 13(12):R122.

Bolger AM, Lohse M & Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114-2120.

Bonam D, Lehman L, Roberts GP & Ludden PW (1989) Regulation of carbon monoxide dehydrogenase and hydrogenase in *Rhodospirillum rubrum*: effects of CO and oxygen on synthesis and activity. *Journal of Bacteriology* 171(6):3102-3107.

Bond DR & Lovley DR (2002) Reduction of Fe(III) oxide by methanogens in the presence and absence of extracellular quinones. *Environmental Microbiology* 4(2):115-124.

Bothe H, Schmitz O, Yates MG & Newton WE (2010) Nitrogen fixation and hydrogen metabolism in Cyanobacteria. *Microbiology and Molecular Biology Reviews* 74(4):529-551.

Bourke MF, Marriott PJ, Glud RN, Hasler-Sheetal H, Kamalanathan M, Beardall J, Greening C & Cook PLM (2017) Metabolism in anoxic permeable sediments is dominated by eukaryotic dark fermentation. *Nature Geoscience* 10(1):30-35.

Bouyoucos G-J (1936) Directions for making mechanical analyses of soils by the hydrometer method. *Soil Science* 42(3):225-230.

Bowien B & Kusian B (2002) Genetics and control of $CO_2$ assimilation in the chemoautotroph *Ralstonia eutropha*. *Archives of Microbiology* 178(2):85-93.

Boyd ES, Hamilton TL, Swanson KD, Howells AE, Baxter BK, Meuser JE, Posewitz MC & Peters JW (2014) [FeFe]-hydrogenase abundance and diversity along a vertical redox gradient in Great Salt Lake, USA. *International Journal of Molecular Sciences* 15(12):21947-21966.

Brewin NJ (1991) Development of the legume root nodule. *Annual Review of Cell Biology* 7(1):191-226.

Breznak JA (1994) Acetogenesis from carbon dioxide in termite guts. *Acetogenesis*, Drake HL (Édit.) Springer US, Boston, MA10.1007/978-1-4615-1777-1_11. p 303-330.

Buhrke T, Lenz O, Krauss N & Friedrich B (2005) Oxygen tolerance of the $H_2$-sensing [NiFe] hydrogenase from *Ralstonia eutropha* H16 is based on limited access of oxygen to the active site. *Journal of Biological Chemistry* 280(25):23791-23796.

Buhrke T, Lenz O, Porthun A & Friedrich B (2004) The $H_2$-sensing complex of *Ralstonia eutropha*: interaction between a regulatory [NiFe] hydrogenase and a histidine protein kinase. *Molecular Microbiology* 51(6):1677-1689.

Burgdorf T, Lenz O, Buhrke T, van der Linden E, Jones AK, Albracht SP & Friedrich B (2005a) [NiFe]-hydrogenases of *Ralstonia eutropha* H16: modular enzymes for oxygen-tolerant biological hydrogen oxidation. *Journal of Molecular Microbiology and Biotechnology* 10(2-4):181-196.

Burgdorf T, Loscher S, Liebisch P, Van der Linden E, Galander M, Lendzian F, Meyer-Klaucke W, Albracht SP, Friedrich B, Dau H & Haumann M (2005b) Structural and oxidation-state changes at its nonstandard Ni-Fe site during activation of the NAD-reducing hydrogenase from *Ralstonia eutropha* detected by X-ray absorption, EPR, and FTIR spectroscopy. *Journal of the American Chemical Society* 127(2):576-592.

Burris RH (1991) Nitrogenases. *Journal of Biological Chemistry* 266(15):9339-9342.

Cabeza RA, Liese R, Fischinger SA, Sulieman S, Avenhaus U, Lingner A, Hein H, Koester B, Baumgarten V, Dittert K & Schulze J (2015) Long-term non-invasive and continuous measurements of legume nodule activity. *The Plant Journal* 81(4):637-648.

Cai Y, Zheng Y, Bodelier PLE, Conrad R & Jia Z (2016) Conventional methanotrophs are responsible for atmospheric methane oxidation in paddy soils. *Nature Communications* 7:11728.

Callahan BJ, McMurdie PJ & Holmes SP (2017) Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME Journal* 11(12):2639-2643.

Calteau A, Gouy M & Perriere G (2005) Horizontal transfer of two operons coding for hydrogenases between bacteria and archaea. *Journal of Molecular Evolution* 60(5):557-565.

Calusinska M, Happe T, Joris B & Wilmotte A (2010) The surprising diversity of clostridial hydrogenases: a comparative genomic perspective. *Microbiology* 156(6):1575-1588.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K & Madden TL (2009) BLAST+: architecture and applications. *BMC Bioinformatics* 10:421-421.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J & Knight R (2010) QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* 7(5):335-336.

Carere CR, Hards K, Houghton KM, Power JF, McDonald B, Collet C, Gapes DJ, Sparling R, Boyd ES, Cook GM, Greening C & Stott MB (2017) Mixotrophy drives niche expansion of verrucomicrobial methanotrophs. *ISME Journal* 11(11):2599-2610.

Cervantes FJ, de Bok FA, Duong-Dac T, Stams AJ, Lettinga G & Field JA (2002) Reduction of humic substances by halorespiring, sulphate-reducing and methanogenic microorganisms. *Environmental Microbiology* 4(1):51-57.

Chen Y & Yoch D (1987) Regulation of two nickel-requiring (inducible and constitutive) hydrogenases and their coupling to nitrogenase in *Methylosinus trichosporium* OB3b. *Journal of Bacteriology* 169(10):4778-4783.

Choi J, Yang F, Stepanauskas R, Cardenas E, Garoutte A, Williams R, Flater J, Tiedje JM, Hofmockel KS, Gelder B & Howe A (2017) Strategies to improve reference databases for soil microbiomes. *ISME Journal* 11(4):829-834.

Clarke KR, Somerfield PJ & Gorley RN (2008) Testing of null hypotheses in exploratory community analyses: similarity profiles and biota-environment linkage. *Journal of Experimental Marine Biology and Ecology* 366(1–2):56-69.

Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, Brown CT, Porras-Alfaro A, Kuske CR & Tiedje JM (2014) Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Research* 42(1):633-642.

Comeau AM, Li WKW, Tremblay J-É, Carmack EC & Lovejoy C (2011) Arctic ocean microbial community structure before and after the 2007 record sea ice minimum. *PLoS ONE* 6(11):e27492.

Conrad R (1996) Soil microorganisms as controllers of atmospheric trace gases ($H_2$, CO, $CH_4$, OCS, $N_2O$, and NO). *Microbiological Reviews* 60(4):609-640.

Conrad R (1999) Soil microorganisms oxidizing atmospheric trace gases ($CH_4$, CO, $H_2$, NO). *Indian Journal of Microbiology* 39:193-203.

Conrad R (2009) The global methane cycle: recent advances in understanding the microbial processes involved. *Environmental Microbiology Reports* 1(5):285-292.

Conrad R, Aragno M & Seiler W (1983a) The inability of hydrogen bacteria to utilize atmospheric hydrogen is due to threshold and affinity for hydrogen. *FEMS Microbiology Letters* 18(3):207-210.

Conrad R & Seiler W (1979) The role of hydrogen bacteria during the decomposition of hydrogen by soil. *FEMS Microbiology Letters* 6(3):143-145.

Conrad R, Weber M & Seiler W (1983b) Kinetics and electron transport of soil hydrogenases catalyzing the oxidation of atmospheric hydrogen. *Soil Biology and Biochemistry* 15(2):167-173.

Constant P, Chowdhury SP, Hesse L & Conrad R (2011a) Co-localization of atmospheric $H_2$ oxidation activity and high affinity $H_2$-oxidizing bacteria in non-axenic soil and sterile soil amended with *Streptomyces* sp. PCB7. *Soil Biology and Biochemistry* 43(9):1888-1893.

Constant P, Chowdhury SP, Hesse L, Pratscher J & Conrad R (2011b) Genome data mining and soil survey for the novel group 5 [NiFe]-hydrogenase to explore the diversity and ecological importance of presumptive high-affinity $H_2$-oxidizing Bacteria. *Applied and Environmental Microbiology* 77(17):6027-6035.

Constant P, Chowdhury SP, Pratscher J & Conrad R (2010) Streptomycetes contributing to atmospheric molecular hydrogen soil uptake are widespread and encode a putative high-affinity [NiFe]-hydrogenase. *Environmental Microbiology* 12(3):821-829.

Constant P & Hallenbeck PC (2013) Chapter 5 - Hydrogenase. *Biohydrogen,* Elsevier, Amsterdam10.1016/B978-0-444-59555-3.00005-2. p 75-102.

Constant P, Poissant L & Villemur R (2008) Isolation of *Streptomyces* sp. PCB7, the first microorganism demonstrating high-affinity uptake of tropospheric $H_2$. *ISME Journal* 2(10):1066-1076.

Constant P, Poissant L & Villemur R (2009) Tropospheric $H_2$ budget and the response of its soil uptake under the changing environment. *Science of the Total Environment* 407(6):1809-1823.

Costa KC, Lie TJ, Jacobs MA & Leigh JA (2013) $H_2$-Independent growth of the hydrogenotrophic methanogen *Methanococcus maripaludis*. *mBio* 4(2).

Daebeler A, Bodelier PL, Yan Z, Hefting MM, Jia Z & Laanbroek HJ (2014) Interactions between Thaumarchaea, Nitrospira and methanotrophs modulate autotrophic nitrification in volcanic grassland soil. *ISME Journal* 8(12):2397-2410.

Das A, Mishra AK & Roy P (1992) Anaerobic growth on elemental sulfur using dissimilar iron reduction by autotrophic *Thiobacillus ferrooxidans*. *FEMS Microbiology Letters* 97(1):167-172.

Delgado-Baquerizo M, Maestre FT, Reich PB, Jeffries TC, Gaitan JJ, Encinar D, Berdugo M, Campbell CD & Singh BK (2016) Microbial diversity drives multifunctionality in terrestrial ecosystems. *Nature Communications* 7:10541.

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P & Andersen GL (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology* 72(7):5069-5072.

Dixon R & Kahn D (2004) Genetic regulation of biological nitrogen fixation. *Nature Reviews Microbiology* 2:621.

Dixon ROD (1972) Hydrogenase in legume root nodule bacteroids: Occurrence and properties. *Archiv für Mikrobiologie* 85(3):193-201.

Dong X, Greening C, Brüls T, Conrad R, Guo K, Blaskowski S, Kaschani F, Kaiser M, Laban NA & Meckenstock RU (2018) Fermentative Spirochaetes mediate necromass recycling in anoxic hydrocarbon-contaminated habitats. *ISME Journal* 12(8):2039-2050.

Dong Z & Layzell DB (2001) $H_2$ oxidation, $O_2$ uptake and $CO_2$ fixation in hydrogen treated soils. *Plant and Soil* 229(1):1-12.

Dong Z, Wu L, Kettlewell B, Caldwell CD & Layzell DB (2003) Hydrogen fertilization of soils – is this a benefit of legumes in rotation? *Plant, Cell & Environment* 26(11):1875-1879.

Eddy SR (1998) Profile hidden Markov models. *Bioinformatics* 14(9):755-763.

Eddy SR (2011) Accelerated profile HMM searches. *PLoS Computational Biology* 7(10):e1002195.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32(5):1792-1797.

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26(19):2460-2461.

Edgar RC (2016a) SINTAX: a simple non-Bayesian taxonomy classifier for 16S and ITS sequences. *bioRxiv* 10.1101/074161.

Edgar RC (2016b) UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. *bioRxiv* 10.1101/081257.

Edgar RC & Flyvbjerg H (2015) Error filtering, pair assembly and error correction for next-generation sequencing reads. *Bioinformatics* 31(21):3476-3482.

Edgar RC, Haas BJ, Clemente JC, Quince C & Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27(16):2194-2200.

Ehhalt DH & Rohrer F (2009) The tropospheric cycle of $H_2$: a critical review. *Tellus B* 61(3):500-535.

Finkel SE (2006) Long-term survival during stationary phase: evolution and the GASP phenotype. *Nature Reviews Microbiology* 4:113.

Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J & Punta M (2014) Pfam: the protein families database. *Nucleic Acids Research* 42:D222-D230.

Fischer F, Lieske R & Winzer K (1932) Biologische gasreaktionen. II. Gber die bildung von essigs ure bei der biologischen umsetzung von kohlenoxyd und kohlens ure mit wasserstoff zu methan. *Biochemische Zeitschrift* 245:2-12.

Fodil D, Badis A, Jaouadi B, Zaraî N, Ferradji FZ & Boutoumi H (2011) Purification and characterization of two extracellular peroxidases from *Streptomyces* sp. strain AM2, a decolorizing actinomycetes responsible for the biodegradation of natural humic acids. *International Biodeterioration & Biodegradation* 65(3):470-478.

Fontecilla-Camps JC, Volbeda A, Cavazza C & Nicolet Y (2007) Structure/function relationships of [NiFe]- and [FeFe]-hydrogenases. *Chemical Reviews* 107(10):4273-4303.

Fowler D, Coyle M, Skiba U, Sutton MA, Cape JN, Reis S, Sheppard LJ, Jenkins A, Grizzetti B, Galloway JN, Vitousek P, Leach A, Bouwman AF, Butterbach-Bahl K, Dentener F, Stevenson D, Amann M & Voss M (2013) The global nitrogen cycle in the twenty-first century. *Philos. Trans. R. Soc. B-Biol. Sci.* 368(1621):13.

Freitag TE & Prosser JI (2009) Correlation of methane production and functional gene transcriptional activity in a peat soil. *Applied and Environmental Microbiology* 75(21):6679-6687.

Fritsch J, Lenz O & Friedrich B (2013) Structure, function and biosynthesis of $O_2$-tolerant hydrogenases. *Nature Reviews Microbiology* 11(2):106-114.

Gaffron H & Rubin J (1942) Fermentative and photochemical production of hydrogen in algae. *The Journal of General Physiology* 26(2):219-240.

Garland JL & Mills AL (1991) Classification and characterization of heterotrophic microbial communities on the basis of patterns of community-level sole-carbon-source utilization. *Applied and Environmental Microbiology* 57(8):2351-2359.

Gascuel O (1997) BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Molecular Biology and Evolution* 14(7):685-695.

Ghirardi ML, Zhang L, Lee JW, Flynn T, Seibert M, Greenbaum E & Melis A (2000) Microalgae: a green source of renewable $H_2$. *Trends in Biotechnology* 18(12):506-511.

Glenn TC (2011) Field guide to next-generation DNA sequencers. *Molecular Ecology Resources* 11(5):759-769.

Gödde M, Meuser K & Conrad R (2000) Hydrogen consumption and carbon monoxide production in soils with different properties. *Biology and Fertility of Soils* 32(2):129-134.

Golding A-L & Dong Z (2010) Hydrogen production by nitrogenase as a potential crop rotation benefit. *Environmental Chemistry Letters* 8(2):101-121.

Golding AL, Zou Y, Yang X, Flynn B & Dong Z (2012) Plant growth promoting $H_2$-oxidizing bacteria as seed inoculants for cereal crops. *Agricultural Sciences* 3(4):510-516.

Gorby YA, Yanina S, McLean JS, Rosso KM, Moyles D, Dohnalkova A, Beveridge TJ, Chang IS, Kim BH, Kim KS, Culley DE, Reed SB, Romine MF, Saffarini DA, Hill EA, Shi L, Elias DA, Kennedy DW, Pinchuk G, Watanabe K, Ishii Si, Logan B, Nealson KH & Fredrickson JK (2006) Electrically conductive bacterial nanowires produced by *Shewanella oneidensis* strain MR-1 and other microorganisms. *Proceedings of the National Academy of Sciences* 103(30):11358-11363.

Gorny N & Schink B (1994) Anaerobic degradation of catechol by *Desulfobacterium* sp. strain Cat2 proceeds via carboxylation to protocatechuate. *Applied and Environmental Microbiology* 60(9):3396-3400.

Grace JB (2006) *Structural equation modeling and natural systems.* Cambridge University Press, Cambridge

Graham EB, Knelman JE, Schindlbacher A, Siciliano S, Breulmann M, Yannarell A, Beman JM, Abell G, Philippot L, Prosser J, Foulquier A, Yuste JC, Glanville HC, Jones DL, Angel R, Salminen J, Newton RJ, Bürgmann H, Ingram LJ, Hamer U, Siljanen HMP, Peltoniemi K, Potthast K, Bañeras L, Hartmann M, Banerjee S, Yu R-Q, Nogaro G, Richter A, Koranda M, Castle SC, Goberna M, Song B, Chatterjee A, Nunes OC, Lopes AR, Cao Y, Kaisermann A, Hallin S, Strickland MS, Garcia-Pausas J, Barba J, Kang H, Isobe K, Papaspyrou S, Pastorelli R, Lagomarsino A, Lindström ES, Basiliko N & Nemergut DR (2016) Microbes as engines of ecosystem function: when does community structure enhance predictions of ecosystem processes? *Frontiers in Microbiology* 7(214).

Graham EB, Wieder WR, Leff JW, Weintraub SR, Townsend AR, Cleveland CC, Philippot L & Nemergut DR (2014) Do we need to understand microbial communities to predict ecosystem function? A comparison of statistical models of nitrogen cycling processes. *Soil Biology and Biochemistry* 68:279-282.

Greening C, Berney M, Hards K, Cook GM & Conrad R (2014a) A soil actinobacterium scavenges atmospheric $H_2$ using two membrane-associated, oxygen-dependent NiFe-hydrogenases. *Proceedings of the National Academy of Sciences* 111(11):4257-4261.

Greening C, Biswas A, Carere CR, Jackson CJ, Taylor MC, Stott MB, Cook GM & Morales SE (2016) Genomic and metagenomic surveys of hydrogenase distribution indicate $H_2$ is a widely utilised energy source for microbial growth and survival. *ISME Journal* 10(3):761-777.

Greening C, Carere CR, Rushton-Green R, Harold LK, Hards K, Taylor MC, Morales SE, Stott MB & Cook GM (2015a) Persistence of the dominant soil phylum Acidobacteria by trace gas scavenging. *Proceedings of the National Academy of Sciences* 112(33):10497-10502.

Greening C, Constant P, Hards K, Morales SE, Oakeshott JG, Russell RJ, Taylor MC, Berney M, Conrad R & Cook GM (2015b) Atmospheric hydrogen scavenging: from enzymes to ecosystems. *Applied and Environmental Microbiology* 81(4):1190-1199.

Greening C, Villas-Boas SG, Robson JR, Berney M & Cook GM (2014b) The growth and survival of *Mycobacterium smegmatis* is enhanced by co-metabolism of atmospheric $H_2$. *PLoS One* 9(7):e103034.

Guardia G, Abalos D, García-Marco S, Quemada M, Alonso-Ayuso M, Cárdenas LM, Dixon ER & Vallejo A (2016a) Effect of cover crops on greenhouse gas emissions in an irrigated field under integrated soil fertility management. *Biogeosciences* 13(18):5245-5257.

Guardia G, Tellez-Rio A, Garcia-Marco S, Martin-Lammerding D, Tenorio JL, Ibanez MA & Vallejo A (2016b) Effect of tillage and crop (cereal versus legume) on greenhouse gas emissions and Global Warming Potential in a non-irrigated mediterranean field. *Agriculture, Ecosystems & Environment* 221:187-197.

Guo R & Conrad R (2008) Extraction and characterization of soil hydrogenases oxidizing atmospheric hydrogen. *Soil Biology and Biochemistry* 40(5):1149-1154.

Haber F (1905) *Thermodynamik technischer Gasreaktionen: Sieben Vorlesungen.* R. Oldenbourg, München und Berlin, Deutschland. 321 p

Hanczár T, Csáki R, Bodrossy L, Murrell CJ & Kovács KL (2002) Detection and localization of two hydrogenases in *Methylococcus capsulatus* (Bath) and their potential role in methane metabolism. *Archives of Microbiology* 177(2):167-172.

Hansen M & Perner M (2016) Hydrogenase gene distribution and $H_2$ consumption ability within the *Thiomicrospira* lineage. *Frontiers in Microbiology* 7(99).

Happe T, Mosler B & Naber JD (1994) Induction, localization and metal content of hydrogenase in the green alga *Chlamydomonas reinhardtii*. *European Journal of Biochemistry* 222(3):769-774.

Häring V & Conrad R (1994) Demonstration of two different $H_2$-oxidizing activities in soil using an $H_2$ consumption and a tritium exchange assay. *Biology and Fertility of Soils* 17(2):125-128.

Hartmann M, Frey B, Mayer J, Mader P & Widmer F (2015) Distinct soil microbial diversity under long-term organic and conventional farming. *ISME Journal* 9(5):1177-1194.

He Y, Li M, Perumal V, Feng X, Fang J, Xie J, Sievert SM & Wang F (2016) Genomic and enzymatic evidence for acetogenesis among multiple lineages of the archaeal phylum Bathyarchaeota widespread in marine sediments. *Nature Microbiology* 1:16035.

Hellriegel H & Wilfarth H (1888) *Untersuchungen über die Stickstoffnahrung der Gramineen und Leguminosen.* Buchdruckerei der "Post" Kayssler, Berlin

Ho A, de Roy K, Thas O, De Neve J, Hoefman S, Vandamme P, Heylen K & Boon N (2014) The more, the merrier: heterotroph richness stimulates methanotrophic activity. *ISME Journal* 8(9):1945-1948.

Ho A, van den Brink E, Reim A, Krause SMB & Bodelier PLE (2016) Recurrence and frequency of disturbance have cumulative effect on methanotrophic activity, abundance, and community structure. *Frontiers in Microbiology* 6(1493).

Hoehler TM, Alperin MJ, Albert DB & Martens CS (2001) Apparent minimum free energy requirements for methanogenic Archaea and sulfate-reducing bacteria in an anoxic marine sediment. *FEMS Microbiology Ecology* 38(1):33-41.

Hoehler TM & Jørgensen BB (2013) Microbial life under extreme energy limitation. *Nature Reviews Microbiology* 11:83.

Hoffman BM, Dean DR & Seefeldt LC (2009) Climbing nitrogenase: toward a mechanism of enzymatic nitrogen fixation. *Accounts of Chemical Research* 42(5):609-619.

Holben WE, Jansson JK, Chelm BK & Tiedje JM (1988) DNA probe method for the detection of specific microorganisms in the soil bacterial community. *Applied and Environmental Microbiology* 54(3):703-711.

Holmes AJ, Roslev P, McDonald IR, Iversen N, Henriksen K & Murrell JC (1999) Characterization of methanotrophic bacterial populations in soils showing atmospheric methane uptake. *Applied and Environmental Microbiology* 65(8):3312-3318.

Horner-Devine CM, Leibold MA, Smith VH & Bohannan BJ (2003) Bacterial diversity patterns along a gradient of primary productivity. *Ecology Letters* 6(7):613-622.

Hunt S, Gaito S & Layzell D (1988) Model of gas exchange and diffusion in legume nodules. *Planta* 173(1):128-141.

Hunt S & Layzell DB (1993) Gas exchange of legume nodules and the regulation of nitrogenase activity. *Annual Review of Plant Physiology and Plant Molecular Biology* 44(1):483-511.

Huntemann M, Ivanova NN, Mavromatis K, Tripp HJ, Paez-Espino D, Tennessen K, Palaniappan K, Szeto E, Pillay M, Chen I-MA, Pati A, Nielsen T, Markowitz VM & Kyrpides NC (2016) The standard operating procedure of the DOE-JGI Metagenome Annotation Pipeline (MAP v.4). *Standards in Genomic Sciences* 11(1):17.

Hurley C (2012) gclus: Clustering Graphics. *R package version 1.3.1* (R package version 1.3.1).

Iyer R, Iken B & Damania A (2017) Genome of *Pseudomonas nitroreducens* DF05 from dioxin contaminated sediment downstream of the San Jacinto River waste pits reveals a broad array of aromatic degradation gene determinants. *Genomics Data* 14:40-43.

Jablonski S, Rodowicz P & Lukaszewicz M (2015) Methanogenic archaea database containing physiological and biochemical characteristics. *International Journal of Systematic and Evolutionary Microbiology* 65(Pt 4):1360-1368.

Janusz G, Pawlik A, Sulej J, Świderska-Burek U, Jarosz-Wilkołazka A & Paszczyński A (2017) Lignin degradation: microorganisms, enzymes involved, genomes analysis and evolution. *FEMS Microbiology Reviews* 41(6):941-962.

Jaouadi B, Rekik H, Badis A, Jaouadi NZ, Belhoul M, Hmidi M, Kourdali S, Fodil D & Bejar S (2014) Production, purification, and characterization of a highly thermostable and humic acid biodegrading peroxidase from a decolorizing *Streptomyces albidoflavus* strain TN644 isolated from a Tunisian off-shore oil field. *International Biodeterioration & Biodegradation* 90:36-44.

Jayakumar A, Chang BX, Widner B, Bernhardt P, Mulholland MR & Ward BB (2017) Biological nitrogen fixation in the oxygen-minimum region of the eastern tropical North Pacific ocean. *ISME Journal* 11:2356.

Jensen ES (1987) Seasonal patterns of growth and nitrogen fixation in field-grown pea. *Plant and Soil* 101(1):29-37.

Ji M, Greening C, Vanwonterghem I, Carere CR, Bay SK, Steen JA, Montgomery K, Lines T, Beardall J, Van Dorst J, Snape I, Stott MB, Hugenholtz P & Ferrari BC (2017) Atmospheric trace gases support primary production in Antarctic desert surface soil. *Nature* 552(7685):400-403.

Jones DT, Taylor WR & Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *Computer Applications in the Biosciences* 8(3):275-282.

Jørgensen BB (1977) The sulfur cycle of a coastal marine sediment (Limfjorden, Denmark). *Limnology and Oceanography* 22(5):814-832.

Kang DD, Froula J, Egan R & Wang Z (2015) MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* 3:e1165.

Karnholz A, Küsel K, Gössner A, Schramm A & Drake HL (2002) Tolerance and metabolic response of acetogenic bacteria toward oxygen. *Applied and environmental microbiology* 68(2):1005-1009.

Karstens U, Schwingshackl C, Schmithüsen D & Levin I (2015) A process-based $^{222}$radon flux map for Europe and its comparison to long-term observations. *Atmospheric Chemistry and Physics* 15(22):12845-12865.

Kashefi K & Lovley DR (2000) Reduction of Fe(III), Mn(IV), and toxic metals at 100°C by *Pyrobaculum islandicum*. *Applied and Environmental Microbiology* 66(3):1050-1056.

Kashefi K, Tor JM, Holmes DE, Gaw Van Praagh CV, Reysenbach AL & Lovley DR (2002) *Geoglobus ahangari* gen. nov., sp. nov., a novel hyperthermophilic archaeon capable of oxidizing organic acids and growing autotrophically on hydrogen with Fe(III) serving as the sole electron acceptor. *International Journal of Systematic and Evolutionary Microbiology* 52(Pt 3):719-728.

Keeling PJ (2009) Functional and ecological impacts of horizontal gene transfer in eukaryotes. *Current Opinion in Genetics & Development* 19(6):613-619.

Keilin D (1959) The problem of anabiosis or latent life: history and current concept. *Proceedings of the Royal Society of London* 150(939):149-191.

Khanna N & Das D (2013) Biohydrogen production by dark fermentation. *Wiley Interdisciplinary Reviews: Energy and Environment* 2(4):401-421.

Khdhiri M, Hesse L, Popa ME, Quiza L, Lalonde I, Meredith LK, Röckmann T & Constant P (2015) Soil carbon content and relative abundance of high affinity $H_2$-oxidizing bacteria predict atmospheric $H_2$ soil uptake activity better than soil microbial community composition. *Soil Biology and Biochemistry* 85:1-9.

Khdhiri M, Piché-Choquette S, Tremblay J, Tringe SG & Constant P (2017) The tale of a neglected energy source: elevated hydrogen exposure affects both microbial diversity and function in soil. *Applied and Environmental Microbiology* 83(11).

Kim SH, Harzman C, Davis JK, Hutcheson R, Broderick JB, Marsh TL & Tiedje JM (2012) Genome sequence of *Desulfitobacterium hafniense* DCB-2, a Gram-positive anaerobe capable of dehalogenation and metal reduction. *BMC Microbiology* 12:21.

Kim SY, Pramanik P, Bodelier PLE & Kim PJ (2014) Cattle manure enhances methanogens diversity and methane emissions compared to swine manure under rice paddy. *PLoS ONE* 9(12):e113593.

King CE & King GM (2014) Description of *Thermogemmatispora carboxidivorans* sp. nov., a carbon monoxide-oxidizing member of the class Ktedonobacteria isolated from a geothermally heated biofilm, and analysis of carbon monoxide oxidation by members of the class Ktedonobacteria. *International Journal of Systematic and Evolutionary Microbiology* 64(4):1244-1251.

King GM (2003) Uptake of carbon monoxide and hydrogen at environmentally relevant concentrations by mycobacteria. *Applied and Environmental Microbiology* 69(12):7266-7272.

King GM, Weber CF, Nanba K, Sato Y & Ohta H (2008) Atmospheric CO and hydrogen uptake and CO oxidizer phylogeny for Miyake-jima, Japan volcanic deposits. *Microbes and Environments* 23(4):299-305.

Klüber HD, Lechner S & Conrad R (1995) Characterization of populations of aerobic hydrogen-oxidizing soil bacteria. *FEMS Microbiology Ecology* 16(2):167-175.

Knief C, Lipski A & Dunfield PF (2003) Diversity and activity of methanotrophic bacteria in different upland soils. *Applied and Environmental Microbiology* 69(11):6703-6714.

Koeck DE, Pechtl A, Zverlov VV & Schwarz WH (2014) Genomics of cellulolytic bacteria. *Current Opinion in Biotechnology* 29:171-183.

Krause S, Bodegom PM, Cornwell WK & Bodelier PLE (2014a) Weak phylogenetic signal in physiological traits of methane-oxidizing bacteria. *Journal of Evolutionary Biology* 27(6):1240-1247.

Krause S, Le Roux X, Niklaus PA, Van Bodegom PM, Lennon JT, Bertilsson S, Grossart H-P, Philippot L & Bodelier PLE (2014b) Trait-based approaches for understanding microbial biodiversity and ecosystem functioning. *Frontiers in Microbiology* 5(251).

Kuczynski J, Stombaugh J, Walters WA, González A, Caporaso JG & Knight R (2002) Using QIIME to analyze 16S rRNA gene sequences from microbial communities. *Current Protocols in Bioinformatics,* John Wiley & Sons, Inc., 10.1002/0471250953.bi1007s36.

Kuczynski J, Stombaugh J, Walters WA, González A, Caporaso JG & Knight R (2011) Using QIIME to analyze 16S rRNA gene sequences from microbial communities. *Current Protocols in Bioinformatics* Chapter 10:Unit10.17-Unit10.17.

Kumar S, Stecher G & Tamura K (2016) MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* 33(7):1870-1874.

Küsel K, Dorsch T, Acker G, Stackebrandt E & Drake HL (2000) *Clostridium scatologenes* strain SL1 isolated as an acetogenic bacterium from acidic sediments. *International Journal of Systematic and Evolutionary Microbiology* 50 Pt 2:537-546.

Küsel K, Karnholz A, Trinkwalter T, Devereux R, Acker G & Drake HL (2001) Physiological ecology of Clostridium glycolicum RD-1, an aerotolerant acetogen isolated from sea grass roots. *Applied and Environmental Microbiology* 67(10):4734-4741.

Kuypers MMM, Marchant HK & Kartal B (2018) The microbial nitrogen-cycling network. *Nature Reviews Microbiology* 16:263.

La Favre JS & Focht DD (1983) Conservation in Soil of $H_2$ Liberated from $N_2$ Fixation by $Hup^-$ Nodules. *Applied and Environmental Microbiology* 46(2):304-311.

Lacasse MJ & Zamble DB (2016) [NiFe]-hydrogenase maturation. *Biochemistry* 55(12):1689-1701.

Lalonde I & Constant P (2016) Identification of unknown carboxydovore bacteria dominant in deciduous forest soil via succession of bacterial communities, *coxL* genotypes, and carbon monoxide oxidation activity in soil microcosms. *Applied and Environmental Microbiology* 82(4):1324-1333.

Lammel DR, Feigl B, Cerri CC & Nüsslein K (2015) Specific microbial gene abundances and soil parameters contribute to C, N, and greenhouse gas process rates after land use change in Southern Amazonian Soils. *Frontiers in Microbiology* 6:1057.

Lane N, Allen JF & Martin W (2010) How did LUCA make a living? Chemiosmosis in the origin of life. *Bioessays* 32(4):271-280.

Langfelder P & Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9(1):559.

Langfelder P, Luo R, Oldham MC & Horvath S (2011) Is my network module preserved and reproducible? *PLoS Computational Biology* 7(1):e1001057.

Langille MGI, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Clemente JC, Burkepile DE, Vega Thurber RL, Knight R, Beiko RG & Huttenhower C (2013) Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature Biotechnology* 31(9):814-821.

Lauterbach L, Liu J, Horch M, Hummel P, Schwarze A, Haumann M, Vincent KA, Lenz O & Zebger I (2011) The hydrogenase subcomplex of the $NAD^+$-reducing [NiFe]-hydrogenase from *Ralstonia eutropha* – Insights into catalysis and redox interconversions. *European Journal of Inorganic Chemistry* 2011(7):1067-1079.

Le SQ & Gascuel O (2008) An improved general amino acid replacement matrix. *Molecular Biology and Evolution* 25(7):1307-1320.

Le Van TD, Robinson JA, Ralph J, Greening RC, Smolenski WJ, Leedle JAZ & Schaefer DM (1998) Assessment of reductive acetogenesis with indigenous ruminal bacterium populations and Acetitomaculum ruminis. *Applied and Environmental Microbiology* 64(9):3429-3436.

Lee K-C & Rittmann BE (2002) Applying a novel autohydrogenotrophic hollow-fiber membrane biofilm reactor for denitrification of drinking water. *Water Research* 36(8):2040-2052.

Leff JW, Jones SE, Prober SM, Barberán A, Borer ET, Firn JL, Harpole WS, Hobbie SE, Hofmockel KS, Knops JMH, McCulley RL, La Pierre K, Risch AC, Seabloom EW, Schütz M, Steenbock C, Stevens CJ & Fierer N (2015) Consistent responses of soil microbial communities to elevated nutrient inputs in grasslands across the globe. *Proceedings of the National Academy of Sciences* 112(35):10967-10972.

Legendre P & Gallagher E (2001) Ecologically meaningful transformations for ordination of species data. *Oecologia* 129(2):271-280.

Leigh JA (2000) Nitrogen fixation in methanogens: the archaeal perspective. *Current Issues in Molecular Biology* 2(4):125-131.

Lenz O, Bernhard M, Buhrke T, Schwartz E & Friedrich B (2002) The hydrogen-sensing apparatus in *Ralstonia eutropha*. *Journal of Molecular Microbiology and Biotechnology* 4(3):255-262.

Levine UY, Teal TK, Robertson GP & Schmidt TM (2011) Agriculture's impact on microbial diversity and associated fluxes of carbon dioxide and methane. *ISME Journal* 5(10):1683-1691.

Li M (2011) DUK-A Fast and efficient Kmer based sequence matching tool. *Lawrence Berkeley National Laboratory*.

Liot Q & Constant P (2016) Breathing air to save energy - new insights into the ecophysiological role of high-affinity [NiFe]-hydrogenase in *Streptomyces avermitilis*. *MicrobiologyOpen* 5(1):47-59.

Liou JS-C, Balkwill DL, Drake GR & Tanner RS (2005) *Clostridium carboxidivorans* sp. nov., a solvent-producing clostridium isolated from an agricultural settling lagoon, and reclassification of the acetogen *Clostridium scatologenes* strain SL1 as *Clostridium drakei* sp. nov. *International Journal of Systematic and Evolutionary Microbiology* 55(5):2085-2091.

Liu S & Suflita JM (1993) $H_2$-$CO_2$-dependent anaerobic O-demethylation activity in subsurface sediments and by an isolated bacterium. *Applied and Environmental Microbiology* 59(5):1325-1331.

Lonergan DJ, Jenter HL, Coates JD, Phillips EJ, Schmidt TM & Lovley DR (1996) Phylogenetic analysis of dissimilatory Fe(III)-reducing bacteria. *Journal of Bacteriology* 178(8):2402-2408.

Loreau M & Hector A (2001) Partitioning selection and complementarity in biodiversity experiments. *Nature* 412(6842):72-76.

Love MI, Huber W & Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15(12):550.

Lovley DR (1993) Dissimilatory metal reduction. *Annual Review of Microbiology* 47:263-290.

Lovley DR, Coates JD, Blunt-Harris EL, Phillips EJP & Woodward JC (1996) Humic substances as electron acceptors for microbial respiration. *Nature* 382:445.

Lovley DR & Goodwin S (1988) Hydrogen concentrations as an indicator of the predominant terminal electron-accepting reactions in aquatic sediments. *Geochimica et Cosmochimica Acta* 52(12):2993-3003.

Lovley DR, Holmes DE & Nevin KP (2004) Dissimilatory Fe(III) and Mn(IV) reduction. *Advances in Microbial Physiology* 49:219-286.

Ludwig M, Schulz-Friedrich R & Appel J (2006) Occurrence of hydrogenases in cyanobacteria and anoxygenic photosynthetic bacteria: implications for the phylogenetic origin of cyanobacterial and algal hydrogenases. *Journal of Molecular Evolution* 63(6):758-768.

Lynch RC, Darcy JL, Kane NC, Nemergut DR & Schmidt SK (2014) Metagenomic evidence for metabolism of trace atmospheric gases by high-elevation desert Actinobacteria. *Frontiers in Microbiology* 5:698.

Lyon EJ, Shima S, Buurman G, Chowdhuri S, Batschauer A, Steinbach K & Thauer RK (2004) UV-A/blue-light inactivation of the 'metal-free' hydrogenase (Hmd) from methanogenic archaea. *European Journal of Biochemistry* 271(1):195-204.

Ma C, Zhuang L, Zhou SG, Yang GQ, Yuan Y & Xu RX (2012) Alkaline extracellular reduction: isolation and characterization of an alkaliphilic and halotolerant bacterium, *Bacillus pseudofirmus* MC02. *Journal of Applied Microbiology* 112(5):883-891.

Ma K, Schicho RN, Kelly RM & Adams MW (1993) Hydrogenase of the hyperthermophile *Pyrococcus furiosus* is an elemental sulfur reductase or sulfhydrogenase: evidence for a sulfur-reducing hydrogenase ancestor. *Proceedings of the National Academy of Sciences of the United States of America* 90(11):5341-5344.

Magoc T & Salzberg SL (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27(21):2957-2963.

Maimaiti J, Zhang Y, Yang J, Cen Y-P, Layzell DB, Peoples M & Dong Z (2007) Isolation and characterization of hydrogen-oxidizing bacteria induced following exposure of soil to hydrogen gas and their impact on plant growth. *Environmental Microbiology* 9(2):435-444.

Maron P-A, Sarr A, Kaisermann A, Lévêque J, Mathieu O, Guigue J, Karimi B, Bernard L, Dequiedt S, Terrat S, Chabbi A & Ranjard L (2018) High microbial diversity promotes soil ecosystem functioning. *Applied and Environmental Microbiology* 10.1128/AEM.02738-17 %.

Marschner IC (2011) glm2: Fitting generalized linear models with convergence problems. *The R Journal* 3(2):12-15.

Martinez CM, Alvarez LH, Celis LB & Cervantes FJ (2013) Humus-reducing microorganisms and their valuable contribution in environmental processes. *Applied Microbiology and Biotechnology* 97(24):10293-10308.

McCarthy DJ, Chen Y & Smyth GK (2012) Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research* 40(10):4288-4297.

Mei N, Postec A, Monnin C, Pelletier B, Payri CE, Ménez B, Frouin E, Ollivier B, Erauso G & Quéméneur M (2016) Metagenomic and PCR-Based diversity surveys of [FeFe]-hydrogenases combined with isolation of alkaliphilic hydrogen-producing bacteria from the serpentinite-hosted Prony Hydrothermal Field, New Caledonia. *Frontiers in Microbiology* 7(1301).

Meredith LK, Commane R, Keenan TF, Klosterman ST, Munger JW, Templer PH, Tang JW, Wofsy SC & Prinn RG (2017) Ecosystem fluxes of hydrogen in a mid-latitude forest driven by soil microorganisms and plants. *Global Change Biology* 23(2):906-919.

Meredith LK, Rao D, Bosak T, Klepac-Ceraj V, Tada KR, Hansel CM, Ono S & Prinn RG (2014) Consumption of atmospheric hydrogen during the life cycle of soil-dwelling actinobacteria. *Environmental Microbiology Reports* 6(3):226-238.

Mettel C, Kim Y, Shrestha PM & Liesack W (2010) Extraction of mRNA from soil. *Applied and Environmental Microbiology* 76(17):5995-6000.

Midgley GF (2012) Biodiversity and Ecosystem Function. *Science* 335(6065):174-175.

Miller MA, Pfeiffer W & Schwartz T (2011) The CIPRES science gateway: a community resource for phylogenetic analyses. in *Proceedings of the 2011 TeraGrid Conference: Extreme Digital Discovery* (ACM, Salt Lake City, Utah), p 1-8.

Mohammadi S, Pol A, van Alen TA, Jetten MSM & Op den Camp HJM (2017) *Methylacidiphilum fumariolicum* SolV, a thermoacidophilic 'Knallgas' methanotroph with both an oxygen-sensitive and -insensitive hydrogenase. *ISME Journal* 11(4):945-958.

Morita RY (1999) Is $H_2$ the universal energy source for long-term survival? *Microbial Ecology* 38(4):307-320.

Mus F, Crook MB, Garcia K, Garcia Costas A, Geddes BA, Kouri ED, Paramasivan P, Ryu M-H, Oldroyd GED, Poole PS, Udvardi MK, Voigt CA, Ané J-M & Peters JW (2016) Symbiotic nitrogen fixation and the challenges to its extension to non-legumes. *Applied and Environmental Microbiology* 82(13):3698-3710.

Muyzer G & Stams AJM (2008) The ecology and biotechnology of sulphate-reducing bacteria. *Nature Reviews Microbiology* 6:441.

Myers MR & King GM (2016) Isolation and characterization of *Acidobacterium ailaaui* sp nov., a novel member of Acidobacteria subdivision 1, from a geothermally heated Hawaiian microbial mat. *International Journal of Systematic and Evolutionary Microbiology* 66:5328-5335.

Mylona P, Pawlowski K & Bisseling T (1995) Symbiotic nitrogen fixation. *The Plant Cell* 7(7):869-885.

Ney B, Ahmed FH, Carere CR, Biswas A, Warden AC, Morales SE, Pandey G, Watt SJ, Oakeshott JG, Taylor MC, Stott MB, Jackson CJ & Greening C (2016) The methanogenic redox cofactor $F_{420}$ is widely synthesized by aerobic soil bacteria. *ISME Journal* 11:125.

Nielsen M, Revsbech NP & Kühl M (2015) Microsensor measurements of hydrogen gas dynamics in cyanobacterial microbial mats. *Frontiers in Microbiology* 6(726):726.

Novelli PC, Lang PM, Masarie KA, Hurst DF, Myers R & Elkins JW (1999) Molecular hydrogen in the troposphere: Global distribution and budget. *Journal of Geophysical Research: Atmospheres* 104(D23):30427-30444.

Oh Y-K, Seol E-H, Kim M-S & Park S (2004) Photoproduction of hydrogen from acetate by a chemoheterotrophic bacterium *Rhodopseudomonas palustris* P4. *International Journal of Hydrogen Energy* 29(11):1115-1121.

Oksanen J, Blanchet F, Kindt R, Legendre P, Minchin P, O'Hara R, Simpson G, Solymos P, Henry M, Stevens H & Wagner H (2013) *vegan: community ecology package. R package version 2.4-2.*

Ollivier B, Caumette P, Garcia JL & Mah RA (1994) Anaerobic bacteria from hypersaline environments. *Microbiological Reviews* 58(1):27-38.

Onstott TC, Phelps TJ, Colwell FS, Ringelberg D, White DC, Boone DR, McKinley JP, Stevens TO, Long PE, Balkwill DL, Griffin WT & Kieft T (1998) Observations pertaining to the origin and ecology of microorganisms recovered from the deep subsurface of Taylorsville Basin, Virginia. *Geomicrobiology Journal* 15(4):353-385.

Ontiveros-Valencia A, Ilhan ZE, Kang D-W, Rittmann B & Krajmalnik-Brown R (2013) Phylogenetic analysis of nitrate- and sulfate-reducing bacteria in a hydrogen-fed biofilm. *FEMS Microbiology Ecology* 85(1):158-167.

Oparin AI (1938) *The origin of life.* Dover Publications,

Orr CH, James A, Leifert C, Cooper JM & Cummings SP (2011) Diversity and activity of free-living nitrogen-fixing bacteria and total bacteria in organic and conventionally managed soils. *Applied and Environmental Microbiology* 77(3):911-919.

Osborne CA, Peoples MB & Janssen PH (2010) Detection of a reproducible, single-member shift in soil bacterial communities exposed to low levels of hydrogen. *Applied and Environmental Microbiology* 76(5):1471-1479.

Pandelia M-E, Lubitz W & Nitschke W (2012) Evolution and diversification of Group 1 [NiFe] hydrogenases. Is there a phylogenetic marker for $O_2$-tolerance? *Biochimica et Biophysica Acta* 1817(9):1565-1575.

Pandey D & Agrawal M (2015) Greenhouse gas fluxes from sugarcane and pigeon pea cultivated soils. *Agricultural Research* 4(3):245-253.

Paradis E, Claude J & Strimmer K (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20(2):289-290.

Parks DH, Imelfort M, Skennerton CT, Hugenholtz P & Tyson GW (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research* 25(7):1043-1055.

Parshina SN, Sipma J, Henstra AM & Stams AJM (2010) Carbon monoxide as an electron donor for the biological reduction of sulphate. *International Journal of Microbiology* 2010:9.

Pester M & Brune A (2007) Hydrogen is the central free intermediate during lignocellulose degradation by termite gut symbionts. *ISME Journal* 1(6):551-565.

Peter H, Beier S, Bertilsson S, Lindstrom ES, Langenheder S & Tranvik LJ (2011) Function-specific response to depletion of microbial diversity. *ISME Journal* 5(2):351-361.

Peters JW, Schut GJ, Boyd ES, Mulder DW, Shepard EM, Broderick JB, King PW & Adams MWW (2015) [FeFe]- and [NiFe]-hydrogenase diversity, mechanism, and maturation. *Biochimica et Biophysica Acta* 1853(6):1350-1369.

Peters V & Conrad R (1995) Methanogenic and other strictly anaerobic bacteria in desert soil and other oxic soils. *Applied and Environmental Microbiology* 61(4):1673-1676.

Philippot L, Andersson SG, Battin TJ, Prosser JI, Schimel JP, Whitman WB & Hallin S (2010) The ecological coherence of high bacterial taxonomic ranks. *Nature Reviews Microbiology* 8(7):523-529.

Piché-Choquette S, Khdhiri M & Constant P (2017) Survey of high-affinity $H_2$-oxidizing bacteria in soil reveals their vast diversity yet underrepresentation in genomic databases. *Microbial Ecology* 74(4):771-775.

Piché-Choquette S, Khdhiri M & Constant P (2018) Dose-response relationships between environmentally-relevant $H_2$ concentrations and the biological sinks of $H_2$, $CH_4$ and CO in soil. *Soil Biology and Biochemistry* 123:190-199.

Piché-Choquette S, Tremblay J, Tringe SG & Constant P (2016) $H_2$-saturation of high affinity $H_2$-oxidizing bacteria alters the ecological niche of soil microorganisms unevenly among taxonomic groups. *PeerJ* 4:e1782.

Pieterse G, Krol MC, Batenburg AM, M. Brenninkmeijer CA, Popa ME, O'Doherty S, Grant A, Steele LP, Krummel PB, Langenfelds RL, Wang HJ, Vermeulen AT, Schmidt M, Yver C, Jordan A, Engel A, Fisher RE, Lowry D, Nisbet EG, Reimann S, Vollmer MK, Steinbacher M, Hammer S, Forster G, Sturges WT & Röckmann T (2013)

Reassessing the variability in atmospheric $H_2$ using the two-way nested TM5 model. *Journal of Geophysical Research: Atmospheres* 118(9):3764-3780.

Pohlmann A, Fricke WF, Reinecke F, Kusian B, Liesegang H, Cramm R, Eitinger T, Ewering C, Pötter M, Schwartz E, Strittmatter A, Voß I, Gottschalk G, Steinbüchel A, Friedrich B & Bowien B (2006) Genome sequence of the bioplastic-producing "Knallgas" bacterium *Ralstonia eutropha* H16. *Nature Biotechnology* 24:1257.

Pohorelic BK, Voordouw JK, Lojou E, Dolla A, Harder J & Voordouw G (2002) Effects of deletion of genes encoding Fe-only hydrogenase of *Desulfovibrio vulgaris* Hildenborough on hydrogen and lactate metabolism. *Journal of Bacteriology* 184(3):679-686.

Poinar HN, Hoss M, Bada JL & Paabo S (1996) Amino acid racemization and the preservation of ancient DNA. *Science* 272(5263):864-866.

Poole RK & Hill S (1997) Respiratory protection of nitrogenase activity in *Azotobacter vinelandii*—roles of the terminal oxidases. *Bioscience Reports* 17(3):303-317.

Postgate JR (1982) Biology nitrogen fixation: fundamentals. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 296(1082):375-385.

Pratscher J, Vollmers J, Wiegand S, Dumont MG & Kaster A-K (2018) Unravelling the identity, metabolic potential and global biogeography of the atmospheric methane-oxidizing upland soil cluster α. *Environmental Microbiology* 10.1111/1462-2920.14036.

Prosser JI, Bohannan BJM, Curtis TP, Ellis RJ, Firestone MK, Freckleton RP, Green JL, Green LE, Killham K, Lennon JJ, Osborn AM, Solan M, van der Gast CJ & Young JPW (2007) The role of ecological theory in microbial ecology. *Nature Reviews Microbiology* 5(5):384-392.

Puggioni V, Tempel S & Latifi A (2016) Distribution of hydrogenases in cyanobacteria: A phylum-wide genomic survey. *Frontiers in Genetic* 7(223).

Quigley DR, Ward B, Crawford DL, Hatcher HJ & Dugan PR (1989) Evidence that microbially produced alkaline materials are involved in coal biosolubilization. *Applied Biochemistry and Biotechnology* 20(1):753-763.

Quinlan AR & Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841-842.

Quiza L, Lalonde I, Guertin C & Constant P (2014) Land-use influences the distribution and activity of high affinity CO-oxidizing bacteria associated to type I-*coxL* genotype in soil. *Frontiers in Microbiology* 5:271.

R Development Core Team (2014) *The R Stats Package - version 3.1.3.*

R Development Core Team (2017) R: A language and environment for statistical computing. *R Foundation for Statistical Computing*.

Raghoebarsing AA, Pol A, van de Pas-Schoonen KT, Smolders AJP, Ettwig KF, Rijpstra WIC, Schouten S, Damsté JSS, Op den Camp HJM, Jetten MSM & Strous M

(2006) A microbial consortium couples anaerobic methane oxidation to denitrification. *Nature* 440:918.

Ragsdale SW & Pierce E (2008) Acetogenesis and the Wood–Ljungdahl pathway of $CO_2$ fixation. *Biochimica et Biophysica Acta* 1784(12):1873-1898.

Rasche ME & Arp DJ (1989) Hydrogen inhibition of nitrogen reduction by nitrogenase in isolated soybean nodule bacteroids. *Plant Physiology* 91(2):663-668.

Redwood MD, Paterson-Beedle M & Macaskie LE (2008) Integrating dark and light bio-hydrogen production strategies: towards the hydrogen economy. *Reviews in Environmental Science and Biotechnology* 8(2):149.

Reumer M, Harnisz M, Lee HJ, Reim A, Grunert O, Putkinen A, Fritze H, Bodelier PLE & Ho A (2018) Impact of peat mining and restoration on methane turnover potential and methane-cycling microorganisms in a northern bog. *Applied and Environmental Microbiology* 84(3).

Revell LJ (2012) phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 3(2):217-223.

Rey FE, Faith JJ, Bain J, Muehlbauer MJ, Stevens RD, Newgard CB & Gordon JI (2010) Dissecting the in vivo metabolic potential of two human gut acetogens. *Journal of Biological Chemistry* 285(29):22082-22090.

Ricke P, Kube M, Nakagawa S, Erkel C, Reinhardt R & Liesack W (2005) First genome data from uncultured upland soil cluster alpha methanotrophs provide further evidence for a close phylogenetic relationship to *Methylocapsa acidiphila* B2 and for high-affinity methanotrophy involving particulate methane monooxygenase. *Applied and Environmental Microbiology* 71(11):7472-7482.

Rittenberg SC & Goodman NS (1969) Mixotrophic growth of *Hydrogenomonas eutropha*. *Journal of Bacteriology* 98(2):617-622.

Robinson MD, McCarthy DJ & Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26(1):139-140.

Rocca JD, Hall EK, Lennon JT, Evans SE, Waldrop MP, Cotner JB, Nemergut DR, Graham EB & Wallenstein MD (2015) Relationships between protein-encoding gene abundance and corresponding process are commonly assumed yet rarely observed. *ISME Journal* 9(8):1693-1699.

Rosencrantz D, Rainey FA & Janssen PH (1999) Culturable populations of Sporomusa spp. and Desulfovibrio spp. in the anoxic bulk soil of flooded rice microcosms. *Applied and Environmental Microbiology* 65(8):3526-3533.

Rosseel Y (2012) lavaan: An R package for structural equation modeling. *Journal of Statistical Software* 48(2):1-36.

Sanz-Cobena A, García-Marco S, Quemada M, Gabriel JL, Almendros P & Vallejo A (2014) Do cover crops enhance $N_2O$, $CO_2$ or $CH_4$ emissions from soil in Mediterranean arable systems? *Science of The Total Environment* 466-467:164-174.

Sayers EW, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Edgar R, Federhen S, Feolo M, Geer LY, Helmberg W, Kapustin Y, Landsman D, Lipman DJ, Madden TL, Maglott DR, Miller V, Mizrachi I, Ostell J, Pruitt KD, Schuler GD, Sequeira E, Sherry ST, Shumway M, Sirotkin K, Souvorov A, Starchenko G, Tatusova TA, Wagner L, Yaschenko E & Ye J (2009) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research* 37(Database issue):D5-15.

Schäfer C, Bommer M, Hennig Sandra E, Jeoung J-H, Dobbek H & Lenz O (2016) Structure of an Actinobacterial-Type [NiFe]-Hydrogenase Reveals Insight into $O_2$-Tolerant $H_2$ Oxidation. *Structure* 24(2):285-292.

Schäfer C, Friedrich B & Lenza O (2013) Novel, oxygen-insensitive group 5 [NiFe]-hydrogenase in *Ralstonia eutropha*. *Applied and Environmental Microbiology* 79(17):5137-5145.

Schauder R, Widdel F & Fuchs G (1987) Carbon assimilation pathways in sulfate-reducing bacteria II. Enzymes of a reductive citric acid cycle in the autotrophic *Desulfobacter hydrogenophilus*. *Archives of Microbiology* 148(3):218-225.

Schink B (1997) Energetics of syntrophic cooperation in methanogenic degradation. *Microbiology and Molecular Biology Reviews* 61(2):262-280.

Schmidt O, Drake HL & Horn MA (2010) Hitherto unknown [FeFe]-hydrogenase gene diversity in anaerobes and anoxic enrichments from a moderately acidic fen. *Applied and Environmental Microbiology* 76(6):2027-2031.

Schuler S & Conrad R (1990) Soils contain two different activities for oxidation of hydrogen. *FEMS Microbiology Letters* 73(1):77-83.

Schut G, Nixon W, Lipscomb G, Scott R & Adams M (2012) Mutational analyses of the enzymes involved in the metabolism of hydrogen by the hyperthermophilic archaeon *Pyrococcus furiosus*. *Frontiers in Microbiology* 3(163).

Schut GJ & Adams MW (2009) The iron-hydrogenase of *Thermotoga maritima* utilizes ferredoxin and NADH synergistically: a new perspective on anaerobic hydrogen production. *Journal of Bacteriology* 191(13):4451-4457.

Schut GJ, Boyd ES, Peters JW & Adams MW (2013) The modular respiratory complexes involved in hydrogen and sulfur metabolism by heterotrophic hyperthermophilic archaea and their evolutionary implications. *FEMS Microbiology Reviews* 37(2):182-203.

Schwartz E & Friedrich B (2006) The $H_2$-Metabolizing Prokaryotes. *The Prokaryotes*, Dworkin M, Falkow S, Rosenberg E, Schleifer K-H & Stackebrandt E (Édit.) Springer New York, 10.1007/0-387-30742-7_17. p 496-563.

Schwartz E, Henne A, Cramm R, Eitinger T, Friedrich B & Gottschalk G (2003) Complete nucleotide sequence of pHG1: a *Ralstonia eutropha* H16 megaplasmid encoding key enzymes of $H_2$-based lithoautotrophy and anaerobiosis. *Journal of Molecular Biology* 332(2):369-383.

Sebastian B, Martin B, Walter M, Michael S, Martin K & Richard S (2013) Methanogenic capabilities of ANME-archaea deduced from $^{13}$C-labelling approaches. *Environmental Microbiology* 15(8):2384-2393.

Seitzinger S, Harrison JA, Böhlke JK, Bouwman AF, Lowrance R, Peterson B, Tobias C & Drecht GV (2006) Denitrification across landscapes and waterscapes: a synthesis. *Ecological Applications* 16(6):2064-2090.

Shah NN, Hanna ML, Jackson KJ & Taylor RT (1995) Batch cultivation of *Methylosinus trichosporium* OB3B: IV. Production of hydrogen-driven soluble or particulate methane monooxygenase activity. *Biotechnology and Bioengineering* 45(3):229-238.

Simon J & Klotz MG (2013) Diversity and evolution of bioenergetic systems involved in microbial nitrogen compound transformations. *Biochimica et Biophysica Acta* 1827(2):114-135.

Singh N, Kendall MM, Liu Y & Boone DR (2005) Isolation and characterization of methylotrophic methanogens from anoxic marine sediments in Skan Bay, Alaska: description of *Methanococcoides alaskense* sp. nov., and emended description of *Methanosarcina baltica*. *International Journal of Systematc and Evolutionary Microbiology* 55(Pt 6):2531-2538.

Smith-Downey NV, Randerson JT & Eiler JM (2006) Temperature and moisture dependence of soil $H_2$ uptake measured in the laboratory. *Geophysical Research Letters* 33(14):L14813.

Smith-Downey NV, Randerson JT & Eiler JM (2008) Molecular hydrogen uptake by soils in forest, desert, and marsh ecosystems in California. *Journal of Geophysical Research: Biogeosciences* 113(G3).

Smith VH (2007) Microbial diversity–productivity relationships in aquatic ecosystems. *FEMS Microbiology Ecology* 62(2):181-186.

Søndergaard D, Pedersen CNS & Greening C (2016) HydDB: A web tool for hydrogenase classification and analysis. *Scientific Reports* 6:34212.

Stal LJ & Krumbein WE (1987) Temporal separation of nitrogen fixation and photosynthesis in the filamentous, non-heterocystous cyanobacterium *Oscillatoria* sp. *Archives of Microbiology* 149(1):76-80.

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312-1313.

Steenhoudt O & Vanderleyden J (2000) *Azospirillum*, a free-living nitrogen-fixing bacterium closely associated with grasses: genetic, biochemical and ecological aspects. *FEMS Microbiology Reviews* 24(4):487-506.

Stein S, Selesi D, Schilling R, Pattis I, Schmid M & Hartmann A (2005) Microbial activity and bacterial composition of $H_2$-treated soils with net $CO_2$ fixation. *Soil Biology and Biochemistry* 37(10):1938-1945.

Stephenson M & Stickland LH (1933) Hydrogenase: The bacterial formation of methane by the reduction of one-carbon compounds by molecular hydrogen. *The Biochemical Journal* 27(5):1517-1527.

Stock M, Hoefman S, Kerckhof F-M, Boon N, De Vos P, De Baets B, Heylen K & Waegeman W (2013) Exploration and prediction of interactions between methanotrophs and heterotrophs. *Research in Microbiology* 164(10):1045-1054.

Swofford DL (2003) Phylogenetic Analysis Using Parsimony. (Sinauer Associates, Sunderland, Massachusetts).

Tamagnini P, Leitão E, Oliveira P, Ferreira D, Pinto F, Harris DJ, Heidorn T & Lindblad P (2007) Cyanobacterial hydrogenases: diversity, regulation and applications. *FEMS Microbiology Reviews* 31(6):692-720.

Tang S, Antonov I & Borodovsky M (2013) MetaGeneTack: ab initio detection of frameshifts in metagenomic sequences. *Bioinformatics* 29(1):114-116.

Thauer RK (1998) Biochemistry of methanogenesis: a tribute to Marjory Stephenson:1998 Marjory Stephenson Prize Lecture. *Microbiology* 144(9):2377-2406.

Thauer RK (2011) Hydrogenases and the Global $H_2$ Cycle. *European Journal of Inorganic Chemistry* 2011(7):919-921.

Thauer RK, Jungermann K & Decker K (1977) Energy conservation in chemotrophic anaerobic bacteria. *Bacteriological Reviews* 41(1):100-180.

Thauer RK, Kaster AK, Goenrich M, Schick M, Hiromoto T & Shima S (2010) Hydrogenases from methanogenic archaea, nickel, a novel cofactor, and $H_2$ storage. *Annual Reviews in Biochemistry* 79:507-536.

Thauer RK, Klein AR & Hartmann GC (1996) Reactions with molecular hydrogen in microorganisms: evidence for a purely organic hydrogenation catalyst. *Chemical Reviews* 96(7):3031-3042.

Thauer RK & Kunow J (1995) Sulfate-Reducing Archaea. *Sulfate-Reducing Bacteria*, Barton LL (Édit.) Springer US, Boston, MA10.1007/978-1-4899-1582-5_2. p 33-48.

Tian F, Toon OB, Pavlov AA & De Sterck H (2005) A hydrogen-rich early earth atmosphere. *Science* 308(5724):1014-1017.

Tian J-H, Pourcher A-M, Bouchez T, Gelhaye E & Peu P (2014) Occurrence of lignin degradation genotypes and phenotypes among prokaryotes. *Applied Microbiology and Biotechnology* 98(23):9527-9544.

Tilman D, Cassman KG, Matson PA, Naylor R & Polasky S (2002) Agricultural sustainability and intensive production practices. *Nature* 418:671-677.

Timmers PHA, Welte CU, Koehorst JJ, Plugge CM, Jetten MSM & Stams AJM (2017) Reverse methanogenesis and respiration in methanotrophic archaea. *Archaea* 2017:22.

Toju H, Tanabe AS, Yamamoto S & Sato H (2012) High-coverage ITS primers for the DNA-based identification of Ascomycetes and Basidiomycetes in environmental samples. *PLoS ONE* 7(7):e40863.

Trainer MG (2013) Atmospheric prebiotic chemistry and organic Hazes. *Current Organic Chemistry* 17:1710-1723.

Tremblay J, Singh K, Fern A, Kirton ES, He S, Woyke T, Lee J, Chen F, Dangl JL & Tringe SG (2015) Primer and platform effects on 16S rRNA tag sequencing. *Frontiers in Microbiology* 6:771.

Tschapek M & Giambiagi N (1955) Nitrogen fixation of *Azotobacter* in soil —Its inhibition by oxygen. *Archiv für Mikrobiologie* 21(4):376-390.

Tsygankov AA, Fedorov AS, Laurinavichene TV, Gogotov IN, Rao KK & Hall DO (1998) Actual and potential rates of hydrogen photoproduction by continuous culture of the purple non-sulphur bacterium *Rhodobacter capsulatus*. *Applied Microbiology and Biotechnology* 49(1):102-107.

Ueno A, Shimizu S, Tamamura S, Okuyama H, Naganuma T & Kaneko K (2016) Anaerobic decomposition of humic substances by *Clostridium* from the deep subsurface. *Scientific Reports* 6:18990.

Uratsu SL, Keyser HH, Weber DF & Lim ST (1982) Hydrogen uptake (Hup) activity of *Rhizobium Japonicum* from major United-States soybean production areas. *Crop Science* 22(3):600-602.

Vignais PM (2008) Hydrogenases and $H^+$-reduction in primary energy conservation. *Results and problems in cell differentiation* 45:223-252.

Vignais PM & Billoud B (2007) Occurrence, classification, and biological function of hydrogenases: an overview. *Chemical Reviews* 107(10):4206-4272.

Vignais PM, Billoud B & Meyer J (2001) Classification and phylogeny of hydrogenases. *FEMS Microbiology Reviews* 25(4):455-501.

Vignais PM & Colbeau A (2004) Molecular biology of microbial hydrogenases. *Current Issues in Molecular Biology* 6(2):159-188.

Vitousek PM, Menge DNL, Reed SC & Cleveland CC (2013) Biological nitrogen fixation: rates, patterns and ecological controls in terrestrial ecosystems. *Philosophical Transactions of the Royal Society B: Biological Sciences* 368(1621):20130119.

Voss M, Bange HW, Dippner JW, Middelburg JJ, Montoya JP & Ward B (2013) The marine nitrogen cycle: recent discoveries, uncertainties and the potential relevance of climate change. *Philosophical Transactions of the Royal Society B: Biological Sciences* 368(1621).

Wächtershäuser G (2010) Chemoautotrophic origin of life: the iron–sulfur world hypothesis. *Geomicrobiology: Molecular and Environmental Perspective*, Barton LL, Mandl M & Loy A (Édit.) Springer Netherlands, Dordrecht10.1007/978-90-481-9204-5_1. p 1-35.

Wang G, Romero-Gallo J, Benoit SL, Piazuelo MB, Dominguez RL, Morgan DR, Peek RM & Maier RJ (2016) Hydrogen metabolism in *Helicobacter pylori* plays a role in gastric carcinogenesis through facilitating CagA translocation. *mBio* 7(4).

Wang Q, Garrity GM, Tiedje JM & Cole JR (2007) Naïve bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology* 73(16):5261-5267.

Wang Q, Quensen JF, Fish JA, Kwon Lee T, Sun Y, Tiedje JM & Cole JR (2013) Ecological patterns of *nifH* genes in four terrestrial climatic zones explored with targeted metagenomics using FrameBot, a new informatics tool. *mBio* 4(5):e00592-00513.

Weber CF & King GM (2007) Physiological, ecological, and phylogenetic characterization of *Stappia*, a marine CO-oxidizing bacterial genus. *Applied and Environmental Microbiology* 73(4):1266-1276.

Weber CF & King GM (2012) The phylogenetic distribution and ecological role of carbon monoxide oxidation in the genus *Burkholderia*. *FEMS Microbiology Ecology* 79(1):167-175.

Welte C, Krätzer C & Deppenmeier U (2010) Involvement of Ech hydrogenase in energy conservation of *Methanosarcina mazei*. *The FEBS Journal* 277(16):3396-3403.

Witty JF (1991) Microelectrode measurements of hydrogen concentrations and gradients in legume nodules. *Journal of Experimental Botany* 42(6):765-771.

Witty JF & Minchin FR (1998) Hydrogen measurements provide direct evidence for a variable physical barrier to gas diffusion in legume nodules. *Journal of Experimental Botany* 49(323):1015-1020.

Wood SN (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73(1):3-36.

Wüst PK, Horn MA & Drake HL (2009) *In situ* hydrogen and nitrous oxide as indicators of concomitant fermentation and denitrification in the alimentary canal of the earthworm *Lumbricus terrestris*. *Applied and Environmental Microbiology* 75(7):1852-1859.

Yagi T, Yamashita K, Okada N, Isono T, Momose D, Mineki S & Tokunaga E (2016) Hydrogen photoproduction in green algae *Chlamydomonas reinhardtii* sustainable over 2 weeks with the original cell culture without supply of fresh cells nor exchange of the whole culture medium. *Journal of Plant Research* 129(4):771-779.

Yang G, Zhou X, Zhou S, Yang D, Wang Y & Wang D (2013) *Bacillusthermotolerans* sp. nov., a thermophilic bacterium capable of reducing humus. *International Journal of Systematic and Evolutionary Microbiology* 63(10):3672-3678.

Yonemura S, Kawashima S & Tsuruta H (1999) Continuous measurements of CO and $H_2$ deposition velocities onto an andisol: uptake control by soil moisture. *Tellus B* 51(3):688-700.

Yonemura S, Kawashima S & Tsuruta H (2000) Carbon monoxide, hydrogen, and methane uptake by soils in a temperate arable field and a forest. *Journal of Geophysical Research: Atmospheres* 105(D11):14347-14362.

Zahnle K, Schaefer L & Fegley B (2010) Earth's earliest atmospheres. *Cold Spring Harbor Perspectives in Biology* 2(10):a004895.

Zak DR, Blackwood CB & Waldrop MP (2006) A molecular dawn for biogeochemistry. *Trends in Ecology and Evolution* 21(6):288-295.

Zhang B & Horvath S (2005) A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology* 4:17.

Zhang H, Bruns MA & Logan BE (2006) Biological hydrogen production by *Clostridium acetobutylicum* in an unsaturated flow reactor. *Water Research* 40(4):728-734.

Zhang Y, He X & Dong ZM (2009) Effect of hydrogen on soil bacterial community structure in two soils as determined by terminal restriction fragment length polymorphism. *Plant and Soil* 320(1-2):295-305.

Zhou J, Deng Y, Luo F, He Z & Yang Y (2011) Phylogenetic molecular ecological network of soil microbial communities in response to elevated $CO_2$. *mBio* 2(4):e00122-00111.

Zirngibl C, Van Dongen W, Schworer B, Von Bunau R, Richter M, Klein A & Thauer RK (1992) $H_2$-forming methylenetetrahydromethanopterin dehydrogenase, a novel type of hydrogenase without iron-sulfur clusters in methanogenic archaea. *European Journal of Biochemistry* 208(2):511-520.

Zumft WG (1997) Cell biology and molecular basis of denitrification. *Microbiology and Molecular Biology Reviews* 61(4):533-616.

# 9 ANNEXE 1 – MATÉRIEL SUPPLÉMENTAIRE DE FORMAT HORS-NORMES ASSOCIÉ À L'ARTICLE "MOLECULAR HYDROGEN, A NEGLECTED KEY DRIVER OF SOIL BIOGEOCHEMICAL PROCESSES "

Cet article contient de nombreuses bases de données (appelées Data set dans l'article) fournies en format tableur (.xlsx) pouvant difficilement être insérées à la suite de l'article lui-même. Elles sont donc inclues ici via des liens URL « Google Drive » de manière à les rendre plus accessibles. Cependant, veuillez noter que ces fichiers se retrouvent aussi sur le site du journal Applied and Environmental Microbiology avec l'article en question sous la section réservée au matériel supplémentaire.

## 9.1 **Figure S1**

La figure S1 représente l'arbre phylogénétique consensus des [Fe]-hydrogénases. La figure se trouve au lien suivant:
https://drive.google.com/open?id=1uac-7nr9Q-a1XxPcEzCbyIQvKCxfEqaU

## 9.2 **Figure S2**

La figure S1 représente l'arbre phylogénétique consensus des [FeFe]-hydrogénases. La figure se trouve au lien suivant:
https://drive.google.com/open?id=1NL8jn60FfAXVd16nTRny8OQXaoCBFj4m

## 9.3 **Figure S3**

La figure S1 représente l'arbre phylogénétique consensus des [NiFe]-hydrogénases. La figure se trouve au lien suivant:
https://drive.google.com/open?id=1Vx1ZCm0eBFCkFSKw-ereXVnFZMEQNqOi

# 10 ANNEXE 2 – MATÉRIEL SUPPLÉMENTAIRE DE FORMAT HORS-NORMES ASSOCIÉ À L'ARTICLE "THE TALE OF A NEGLECTED ENERGY SOURCE: ELEVATED HYDROGEN EXPOSURE AFFECTS BOTH MICROBIAL DIVERSITY AND FUNCTION IN SOIL"

Cet article contient de nombreuses bases de données (appelées Data set dans l'article) fournies en format tableur (.xlsx) pouvant difficilement être insérées à la suite de l'article lui-même. Elles sont donc inclues ici via des liens URL « Google Drive » de manière à les rendre plus accessibles. Cependant, veuillez noter que ces fichiers se retrouvent aussi sur le site du journal Applied and Environmental Microbiology avec l'article en question sous la section réservée au matériel supplémentaire.

## 10.1 **Data set S1**

Le fichier Data set S1 se trouve à l'adresse suivante :

https://drive.google.com/open?id=1yIXHehI3VHHBS9K3f2tC5KkSRIt5aUks. Le tableau suivant montre le contenu de chaque onglet du fichier tableur Data set S1.

| Onglets | Titre | Description du contenu |
|---|---|---|
| **Onglet 1:** | Explanation of EdgeR output | Description du contenu des onglets 2-11 |
| **Onglet 2:** | Differential Abundance - Bacteria Farmland | OTU bactériens différentiellement abondants entre les traitements $aH_2$ et $eH_2$ pour le sol agricole |
| **Onglet 3:** | Differential Abundance - Bacteria Larch | OTU bactériens différentiellement abondants entre les traitements $aH_2$ et $eH_2$ pour le sol mélèzes |
| **Onglet 4:** | Differential Abundance - Bacteria Poplar | OTU bactériens différentiellement abondants entre les traitements $aH_2$ et $eH_2$ pour le sol peuplier |
| **Onglet 5:** | Differential Abundance - Fungi Farmland | OTU fongiques différentiellement abondants entre les traitements $aH_2$ et $eH_2$ pour le sol agricole |
| **Onglet 6:** | Differential Abundance - Fungi Larch | OTU fongiques différentiellement abondants entre les traitements $aH_2$ et $eH_2$ pour le sol mélèzes |
| **Onglet 7:** | Differential Abundance - Genome Bins | Bins génomiques différentiellement abondants dans les 3 utilisations des terres pour un même traitement à l'$H_2$ |
| **Onglet 8:** | Differential Abundance - 16S rRNA OTUs with coherent response in the 3 land-use types | OTU bactériens différentiellement abondants dans les 3 utilisations des terres pour un même traitement à l'$H_2$ |
| **Onglet 9:** | OTU - Table - Bacteria | Nombre de séquence des OTU bactériens suite au contrôle-qualité |
| **Onglet 10:** | OTU - Table - Fungi | Nombre de séquence des OTU fongiques suite au contrôle-qualité |
| **Onglet 11:** | Raw data (Ecoplates) | Données brutes des plaques Ecoplates |
| **Onglet 12:** | Raw data (Physicochemical, trace gas exchanges, diversity indexes) | Données brutes associées aux données physicochimiques, richesse spécifique, activité d'oxydation de gaz trace, etc. |

## 10.2 **Data set S2**

Abondance relative (CPM) des gènes codant pour la grande sous-unité des hydrogénases [NiFe] détectées par modèles de Markov cachés:
https://drive.google.com/open?id=1T18HkOORgKKfshcgpkIyyh47voIgIr0u

## 10.3 **Data set S3**

Gènes annotés des bins génomiques d'intérêt :
https://drive.google.com/open?id=1IJcupKnApK9qHkru5mAiLTtuR676k2PB

## 10.4 **Data set S4**

Le fichier Data set S4 contient les différentes amorces employées lors du séquençage d'amplicons codant pour la région V4 de l'ARNr 16S (Bacteria) ou pour l'ITS2 (Fungi) :

https://drive.google.com/open?id=171bXPFI0IOMjVNO-iuV3ymY-9RxOUer3

## 10.5 **Data set S5**

Contrôle-qualité des métagénomes et données concernant les bins génomiques d'intérêts parmi les différents échantillons :
https://drive.google.com/open?id=1U0jYHoLK3dSC8jmJea35dFP_Sx0OpAdf

## 10.6 **Data set S6**

Ce fichier contient les modèles de Markov cachés utilisés afin d'identifier les hydrogénases présentes dans les échantillons:

https://drive.google.com/open?id=1AGcyqerHWNMWfgZmqEDilz3s8e8REO3J

# 11 ANNEXE 3 – MATÉRIEL SUPPLÉMENTAIRE DE FORMAT HORS-NORMES ASSOCIÉ À L'ARTICLE "SURVEY OF HIGH-AFFINITY $H_2$-OXIDIZING BACTERIA IN SOIL REVEALS THEIR VAST DIVERSITY YET UNDERREPRESENTATION IN GENOMIC DATABASES"

Les tableaux et l'arbre phylogénétique sont inclus ici sous format tableur (.xlsx) ou vectoriel (eps), respectivement, puisque leur format ne se prête pas à celui de la thèse, c'est pourquoi ils sont inclus ici via des liens URL « Google Drive ». Veuillez noter que ces fichiers se retrouvent aussi sur le site du journal Microbial Ecology avec l'article en question dans la section réservée au matériel supplémentaire.

## 11.1 **Table S1**

Le tableau S1 consiste en un tableau d'abondance des différents OTU *hhyL*. Il se trouve au lien suivant:
https://drive.google.com/open?id=1DcObV-PyJAx2wuRutEzfOxR9RcbjWjXr

## 11.2 **Table S2**

Le tableau S2 contient les OTU *hhyL* différentiellement abondant selon le traitement à l'$H_2$. Il se trouve au lien suivant:
https://drive.google.com/open?id=1CstCx_7vaDQd2TPxSE_g7nLlDZtNGSfM

## 11.3 **Figure S1**

La Figure S1 (arbre phylogénétique) se trouve au lien suivant:
https://drive.google.com/open?id=1pojDpZ7FEagvgqtEDR2V1tAaGglhpEQh

# 12 ANNEXE 4 – MATÉRIEL SUPPLÉMENTAIRE OU BASES DE DONNÉES ASSOCIÉS À L'ARTICLE "DOSE-RESPONSE RELATIONSHIPS BETWEEN ENVIRONMENTALLY-RELEVANT $H_2$ CONCENTRATIONS AND THE BIOLOGICAL SINKS OF $H_2$, $CH_4$ AND CO IN SOIL"

Cet article contient de nombreux fichiers de grand format comme les tableaux d'abondances des ASV. Ceux-ci sont originalement fournis en format tableur (.xlsx) pouvant difficilement être insérés dans la thèse. Ils sont donc inclus ici via des liens URL « Google Drive » de manière à les rendre plus accessibles. Cependant, ils se trouvent aussi sur le site web du journal Soil Biology and Biochemistry dans la section de l'article réservée au matériel supplémentaire.

## 12.1 **Table S1**

Le tableau S1 consiste en un ensemble d'indices de diversité, de tableaux d'abondance relative des différents ASV, de la taxonomie associée à ces ASV ainsi que les séquences représentatives de ces ASV. Le tableau se trouve au lien suivant: https://drive.google.com/open?id=1A5z4pRO5M2v8I8ku2d7L2Kls6c4tCs-U

## 12.2 **Table S2**

Le tableau S2 consiste en un ensemble de tableaux des résultats obtenus suite à l'utilisation du logiciel « edgeR » afin d'identifier les ASV ayant une abondance différentielle entre les traitements et entre les différentes utilisations des terres. Le tableau se trouve au lien suivant: https://drive.google.com/open?id=1pyjFtzmX4abnbuiZPORiH7il8M4GgPCd

## 12.3 **Table S3**

Le tableau S3 consiste en un ensemble de tableaux montrant la force des corrélations entre l'abondance relative de certains ASV et le ratio de mélange d'$H_2$ et ce, en fonction de l'utilisation des terres. Le tableau se trouve au lien suivant: https://drive.google.com/open?id=1vi6qBLPd2tMTGi_7xwGIDG34g7Ywidu8

## 12.4 **Table S4**

Le tableau S4 contient les données brutes d'oxydation d'$H_2$, de $CH_4$ et de CO pour tous les échantillons testés. Le tableau se trouve au lien suivant :https://drive.google.com/open?id=1-sLlo3ed3ha2gSRoTDVjxyg8Ht2aPxON