Title: Meta-omics survey of [NiFe]-hydrogenase genes fails to capture drastic variations in $H_2$-oxidation activity measured in soil

Authors: Mondher Khdhiri[a], Sarah Piché-Choquette[a], Julien Tremblay[b], Susannah G. Tringe[c] and Philippe Constant[a*]

[a]Institut National de la Recherche Scientifique-Institut Armand-Frappier, 531 boulevard des Prairies, Laval (Québec), Canada, H7V 1B7.

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

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

*Corresponding author: Philippe Constant, INRS-Institut Armand-Frappier, 531 boulevard des Prairies, Laval (Québec), Canada, H7V 1B7, Phone: (450) 687-5010 ext. 4117, Email address: Philippe.Constant@iaf.inrs.ca

**Abstract**

~~In a debatable framework regarding the use of incomplete portraits of~~ ~~m~~Microbial community structure and functions obtained from metagenomic data are often used to infer biogeochemical processes. H~~, h~~ere we used the biological sink of $H_2$ as a case study to test the hypothesis that [NiFe]-hydrogenase gene distribution and expression profiles explain variations in $H_2$ oxidation rate measured in soils. Shotgun meta-omics (metagenomic and metatranscriptomic) analyses of soil samples exposed to elevated or low $H_2$ concentration led to the identification of 45 genes encoding the large subunit of [NiFe]-hydrogenases belonging to 8 distinct phyla. Our results indicate that despite significant sequencing effort, retrieved hydrogenase sequences are not ~~relevant~~ in themselves adequate surrogates of $H_2$ oxidation activity in soil. ~~Furthermore~~In fact, land-use exerted a ~~more important~~greater ~~contribution~~ influence than $H_2$ exposure on both hydrogenase gene distribution and expression, though expression of certain genes responded to $H_2$. We argue that approaches relying on PCR/RT-PCR amplicon~~s~~ sequencing or quantification combined with physicochemical parameters are currently the best option to infer the activity of $H_2$-oxidizing bacteria, as well as other specialist functional guilds with similar population size in soil.

Keywords: Metagenomic, Biogeochemistry, Microbial Ecology, $H_2$ oxidation activities

Microbial community profiles are increasingly used to refine biogeochemical process models relying on abiotic factors (Todd-Brown et al., 2012; Wieder et al., 2013; Pérez-Valera et al., 2015; Powell et al., 2015). Nevertheless, the appeal of this approach remains controversial. For instance, modeling diverse C and N process rates using PCR amplicon sequencing of functional or 16S rRNA genes did not improve every single model previously based on edaphic factors (Graham et al., 2014; Graham et al., 2016). Thus, in this study we investigate the application of meta-omics (metagenomic and metatranscriptomic) approaches in molecular biogeochemistry using molecular hydrogen ($H_2$) as a case study. Aerated upland soils are the most important $H_2$ sinks. Indeed, high-affinity $H_2$-oxidizing bacteria (HOB) are responsible for 70% of global losses of tropospheric $H_2$ (Pieterse et al., 2013) while low-affinity HOB scavenge $H_2$ produced by nodules on plant roots, preventing its diffusion to the atmosphere (Schuler and Conrad, 1991). According to their physiological role and cellular localization, [NiFe]-hydrogenases are classified into 5 distinct functional groups. Briefly, group 1 includes membrane-bound hydrogenases supplying electrons to the respiratory chain; group 2 are either $H_2$ sensors regulating the transcription of hydrogenases operons or recyclers of $H_2$ by-products of nitrogen fixation; group 3 are used to regenerate reducing equivalents in the cells (*e.g.* NAD, ferredoxin-420); group 4 hydrogenases are membrane-bound $H_2$-evolving enzymes operating under anaerobic conditions (Vignais and Billoud, 2007; Greening et al., 2016) and lastly, group 5 encompasses high-affinity hydrogenases used for mixotrophic growth and survival (Constant et al., 2011).

Here, we test the hypothesis that~~, albeit incomplete,~~ retrieved hydrogenase gene profiles from meta-omics analyses, although incomplete, represent a valuable surrogate of high- and low-affinity HOB activity in soil. We used data from previous experiments in which soil samples collected from a larch plantation, a poplar plantation and a farmland were exposed to low (0.5 ppmv) or elevated (10,000 ppmv) $H_2$ concentrations for 15 days (Khdhiri et al., 2017). Low- and high-affinity $H_2$ oxidation rates, soil physicochemical properties and metagenomic profiles measured in the previous report were supplemented with a metatranscriptomic analysis. Quality control-passed metatranscriptomic reads were mapped, using the Burrows-Wheeler aligner BWA v0.7.10 (Li and Durbin, 2010), against contigs from the metagenome assembly after alignment sorting with samtools v1.1 (Li et al., 2009) and gene abundance computation with bedtools v.2.17.0 (Quinlan and Hall, 2010). In this paper, 45 genes encoding the large subunit of [NiFe]-hydrogenases retrieved from the metagenomic analysis based on a trained set of Hidden Markov Models (Khdhiri et al., 2017) and complementary phylogenetic analysis (Figure S1) were used as a template to retrieve corresponding hydrogenase gene transcripts from metatranscriptomic profiles (Table S1A). Comparative analyses were computed using either gene or transcript counts from meta-omics analysis or gene expression ratios (*i.e.* transcript count divided by gene count). This case study was well-suited to challenge our hypothesis, as $H_2$ exposures caused drastic variations in $H_2$ oxidation rates measured in the three soils, including a 94-96% loss of high-affinity oxidation activity and a 103-514% gain of low-affinity activity in soil upon elevated $H_2$ exposure (Khdhiri et al., 2017).

Hydrogenase genes encompassed 8 distinct phyla, all within the bacterial kingdom (Figure S2). Unclassified sequences at the phylum level presented a minor proportion, not exceeding 9% of identified genes. In accordance with recent metagenomic database mining (Greening et al., 2016), Acidobacteria represented only a small proportion of HOBs (less than 2%), while Proteobacteria and Actinobacteria were dominant in all land-uses (40% and 25%, respectively). Hydrogenase gene expression ratios were maintained at an even level in Proteobacteria, while an under- and over-expression was observed in Acidobacteria and Actinobacteria, respectively (Figure S3). The impact of $H_2$ exposure on hydrogenase transcription was analyzed using sequencing reads clustered at the phylum, functional group and gene levels. At the taxonomic level, no phylum showed a significant response to $H_2$ exposure (Table S1B). At the functional group level, a specific response was expected due to the physiological role of group 2B [NiFe]-hydrogenases in regulating the expression of hydrogenase gene operons according to $H_2$ availability (Lenz and Friedrich, 1998). Moreover, a significant decrease of group 5 [NiFe]-hydrogenase gene expression was observed in those same land-uses using qPCR and qRT-PCR techniques (Piche-Choquette et al., 2017). Nevertheless, there was no consistent response to $H_2$ treatment when examining the five groups separately (Table S1C). At the gene level, the expression of 3 hydrogenase genes showed a significant response to $H_2$ exposure (Figure 1). In fact, land-use type exerted more ~~incidence~~ influence on hydrogenase gene or transcript distribution and gene expression than the $H_2$ treatment (Table S2).

Bacterial and fungal species richness along with edaphic factors were previously reported to contribute significantly in explaining $H_2$-oxidation rates

variations in structural equation models (Khdhiri et al., 2017). Here, we added a supplementary composite variable representing hydrogenase gene counts (Mg), transcript counts (Mt) or gene expression profiles (Mt/Mg) to these models to evaluate the contribution of meta-omic datasets in explaining HOB activity. The attempt to reduce hydrogenases' molecular profiles into principal components failed to reach appreciable explanatory power (Figure S4) and therefore, composite variables were computed through the sum of all gene counts, transcript counts and gene expression ratio for each sample. Added composite variables exerted no significant contribution in explaining low-affinity $H_2$ oxidation activity regardless of the analysis level (Figure 2A, C, E). On the other hand, high-affinity activity was partly explained by the composite variable Mt (Figure 2D). Under-representation of hydrogenases genes in the sequenced metagenomes caused by insufficient sequencing depth is likely a critical factor explaining the lack of a significant contribution of meta-omics datasets to explain measured activity of HOB. Regarding high-affinity HOB, our meta-omics datasets recovered only 7 genes encompassing group 5 [NiFe]-hydrogenases (Figure 1), while PCR-amplicon sequencing of *hhyL* gene (encoding the large subunit of hydrogenase belonging to group 5) led to the identification of 664 OTUs in the same soils (Piche-Choquette et al., 2017).

Gene distribution and expression captured through meta-omics surveys were not representative of the overall activity of HOB populations in soil. $H_2$ oxidation activity gradients observed in this study are comparable to extreme variations occurring in the environment when comparing different land-uses. For instance, potential high-affinity $H_2$ oxidation rates varied between 100 and 2000 pmol $g_{(dw)}^{-1}$ $h^{-1}$ in soil samples collected from spruce forest and arid desert,

respectively (Constant et al., 2011). Similarly, investigation of the impact of land-use change on $H_2$ dry deposition unveiled 32-64% lower $H_2$ soil uptake rate in a grazed pasture when compared to native rainforest in the dry season (Pendall et al., 2010). Considering the lack of significant covariation between hydrogenase molecular profiles and $H_2$ oxidation activities under the extreme gradient of activities we experienced, we do not expect better resolution of HOB activity in natural environments. Thus, we recommend the use of qPCR and qRT-PCR techniques in studies aimed at inferring the potential activity of HOB in soil. For instance, multiple linear regression parameterized with total carbon and the relative abundance of high-affinity HOB explained up to 76% variation in high-affinity $H_2$ oxidation rates measured in 26 soils (Khdhiri et al., 2015). Likewise, *hhyL* gene expression ratio analyses explained 45% variation of $H_2$ oxidation activity measured in soil (Piche-Choquette et al., 2017). With the increasing size of public (meta)genomic databases and the democratization of high-throughput sequencing techniques, observations reported in this study are relevant way beyond the scope of the case of the biological sink of $H_2$. Indeed, it is estimated that high-affinity HOB represent approximately 1% of soil microbial communities (Khdhiri et al., 2015), which is comparable to the abundance of other functional groups in soil including CO- and $CH_4$-oxidizing bacteria (Degelmann et al., 2010; Lalonde and Constant, 2016). In conclusion, this work exemplifies that translation of meta-omics profiles into potential ecosystem processes or metabolic pathways is hampered by the poor coverage of typical sequencing efforts and the uncoordinated response of individual members of the same functional guilds to environmental pressures.

## 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. M.K. and S.P.-C are grateful to the Fondation Universitaire Armand-Frappier INRS for their Ph.D. scholarship.

## Accession numbers

Raw sequence reads of the metatranscriptomic analysis were deposited in the Sequence Read Archive of the National Center for Biotechnology Information under BioProject PRJXXX.

## References

Constant, P., Chowdhury, S.P., Hesse, L., Pratscher, J., Conrad, R., 2011. 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, 6027-6035.

Degelmann, D.M., Borken, W., Drake, H.L., Kolb, S., 2010. Different atmospheric methane-oxidizing communities in European beech and Norway spruce soils. Applied and Environmental Microbiology 76, 3228-3235.

Graham, E.B., Knelman, J.E., Schindlbacher, A., Siciliano, S., Breulmann, M., Yannarell, A., Beman, J.M., Abell, G., Philippot, L., Prosser, J., Foulquier, A., Yuste, J.C., Glanville, H.C., Jones, D.L., Angel, R., Salminen, J., Newton, R.J., Burgmann, H., Ingram, L.J., Hamer, U., Siljanen, H.M., Peltoniemi, K., Potthast, K., Baneras, L., Hartmann, M., Banerjee, S., Yu, R.Q., Nogaro, G., Richter, A., Koranda, M., Castle, S.C., Goberna, M., Song, B., Chatterjee, A., Nunes, O.C., Lopes, A.R., Cao, Y., Kaisermann, A., Hallin, S., Strickland, M.S., Garcia-Pausas, J., Barba, J., Kang, H., Isobe, K., Papaspyrou, S., Pastorelli, R., Lagomarsino, A., Lindstrom, E.S., Basiliko, N., Nemergut, D.R., 2016. Microbes as Engines of Ecosystem Function: When Does Community Structure Enhance Predictions of Ecosystem Processes? Frontiers in Microbiology 7, 214.

Graham, E.B., Wieder, W.R., Leff, J.W., Weintraub, S.R., Townsend, A.R., Cleveland, C.C., Philippot, L., Nemergut, D.R., 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., Biswas, A., Carere, C.R., Jackson, C.J., Taylor, M.C., Stott, M.B., Cook, G.M., Morales, S.E., 2016. Genomic and metagenomic surveys of hydrogenase distribution indicate H2 is a widely utilised energy source for microbial growth and survival. ISME Journal 10, 761-777.

Khdhiri, M., Hesse, L., Popa, M.E., Quiza, L., Lalonde, I., Meredith, L.K., 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., Piche-Choquette, S., Tremblay, J., Tringe, S.G., 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.

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, 1324-1333.

Lenz, O., Friedrich, B., 1998. A novel multicomponent regulatory system mediates H2 sensing in Alcaligenes eutrophus. Proceedings of the National Academy of Sciences 95, 12474-12479.

Li, H., Durbin, R., 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26, 589-595.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078-2079.

Pendall, E., Schwendenmann, L., Rahn, T., Miller, J.B., Tans, P.P., White, J., 2010. Land use and season affect fluxes of CO2, CH4, CO, N2O, H2 and isotopic source signatures in Panama: evidence from nocturnal boundary layer profiles. Global Change Biology 16, 2721-2736.

Pérez-Valera, E., Goberna, M., Verdú, M., 2015. Phylogenetic structure of soil bacterial communities predicts ecosystem functioning. FEMS microbiology ecology 91, fiv031.

Piche-Choquette, S., Khdhiri, M., Constant, P., 2017. Survey of High-Affinity H2-Oxidizing Bacteria in Soil Reveals Their Vast Diversity Yet Underrepresentation in Genomic Databases. Microbial Ecology.

Pieterse, G., Krol, M., Batenburg, A., M Brenninkmeijer, C., Popa, M., O'doherty, S., Grant, A., Steele, L., Krummel, P., Langenfelds, R., 2013. Reassessing the variability in atmospheric H2 using the two-way nested TM5 model. Journal of Geophysical Research: Atmospheres 118, 3764-3780.

Powell, J.R., Welsh, A., Hallin, S., 2015. Microbial functional diversity enhances predictive models linking environmental parameters to ecosystem properties. Ecology 96, 1985-1993.

Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841-842.

Schuler, S., Conrad, R., 1991. Hydrogen oxidation in soil following rhizobial H2 production due to N2 fixation by a Vicia faba-Rhizobium leguminosarum symbiosis. Biology and Fertility of Soils 11, 190-195.

Todd-Brown, K.E.O., Hopkins, F.M., Kivlin, S.N., Talbot, J.M., Allison, S.D., 2012. A framework for representing microbial decomposition in coupled climate models. Biogeochemistry 109, 19-33.

Vignais, P.M., Billoud, B., 2007. Occurrence, classification, and biological function of hydrogenases: an overview. Chemical Reviews 107, 4206-4272.

Wieder, W.R., Bonan, G.B., Allison, S.D., 2013. Global soil carbon projections are improved by modelling microbial processes. Nature Climate Change 3, 909-912.

## Figure Legend

**Figure 1.** [NiFe]-hydrogenase~~s~~ gene expression ratio as a function of functional groups. The middle line in the boxes represents the median and the dots reflect outliers. Lower and upper lines delimit the 25% and 75% quantiles, respectively. Each boxplot is colored according to its affiliated phylum. Gene identifier labels are colored according to the land-use type where $H_2$ exposure led to a significant differential expression ratio (F; Farmland, L; Larch and P; Poplar). The shaded gray line marks the separation between over- and under-expression levels.

**Figure 2.** Structural equation models (SEM) to test the hypothesis of a significant contribution of microbial communities' molecular data from the three levels of analysis to the low- and high-affinity $H_2$ oxidation models. Hexagon variables represent statistical composites estimated based on partial least square linear regression analysis of the respective variables. The $R^2$ values are presented for both oxidation rates. The simple path (direct arrow) test~~s~~ the significance effect between the two variables while double directed arrows denote a covariation between variables. All path coefficients were standardized and were significant if $P < 0.05$. The "ns" represent a non-significant relationship. For each level of analysis both the low- and high-affinity $H_2$ oxidation activities are respectively shown as A and B for the metagenomics level; C and D for the metatranscriptomics level and E and F for the expression ratio level.

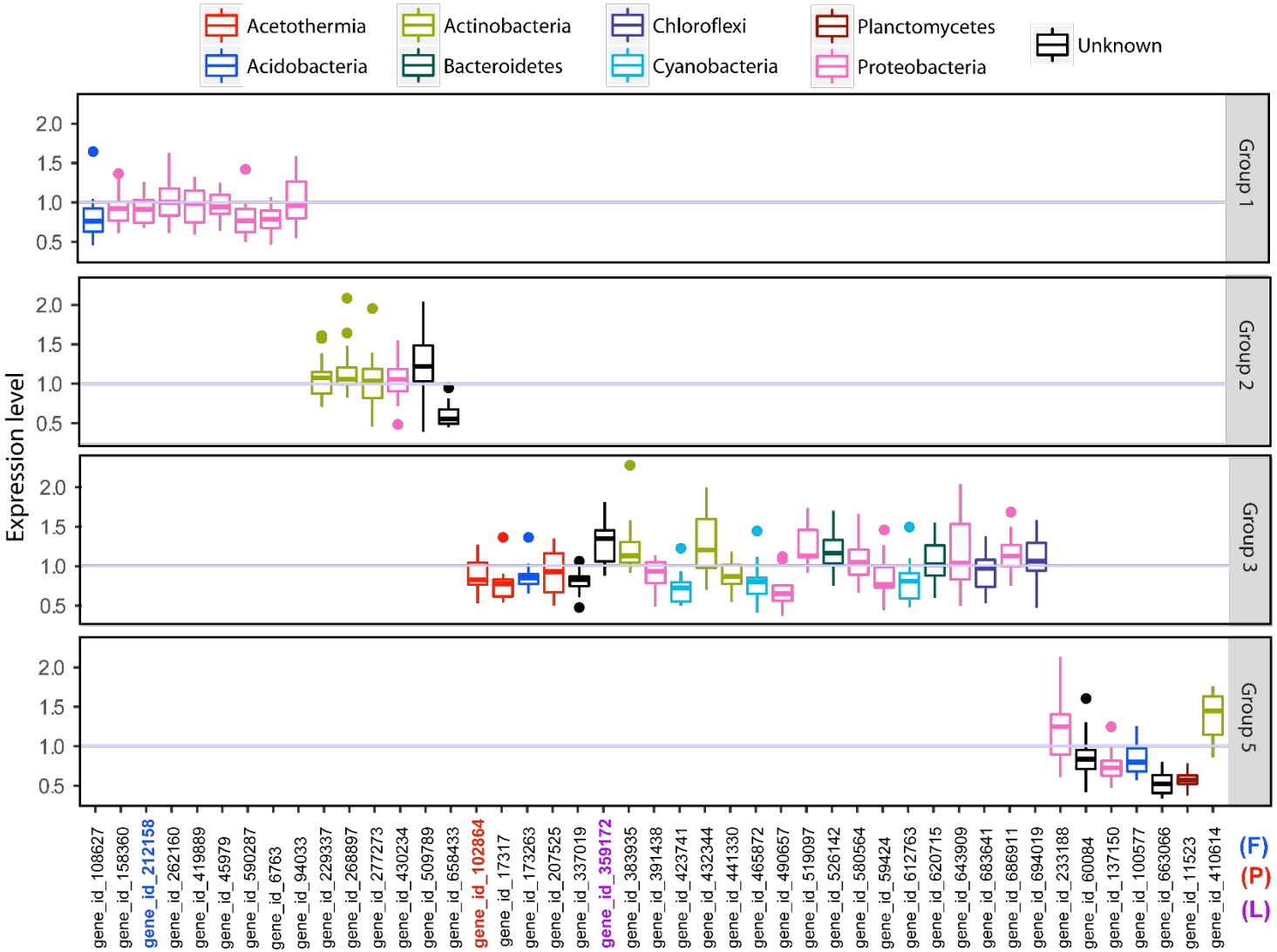
294

295

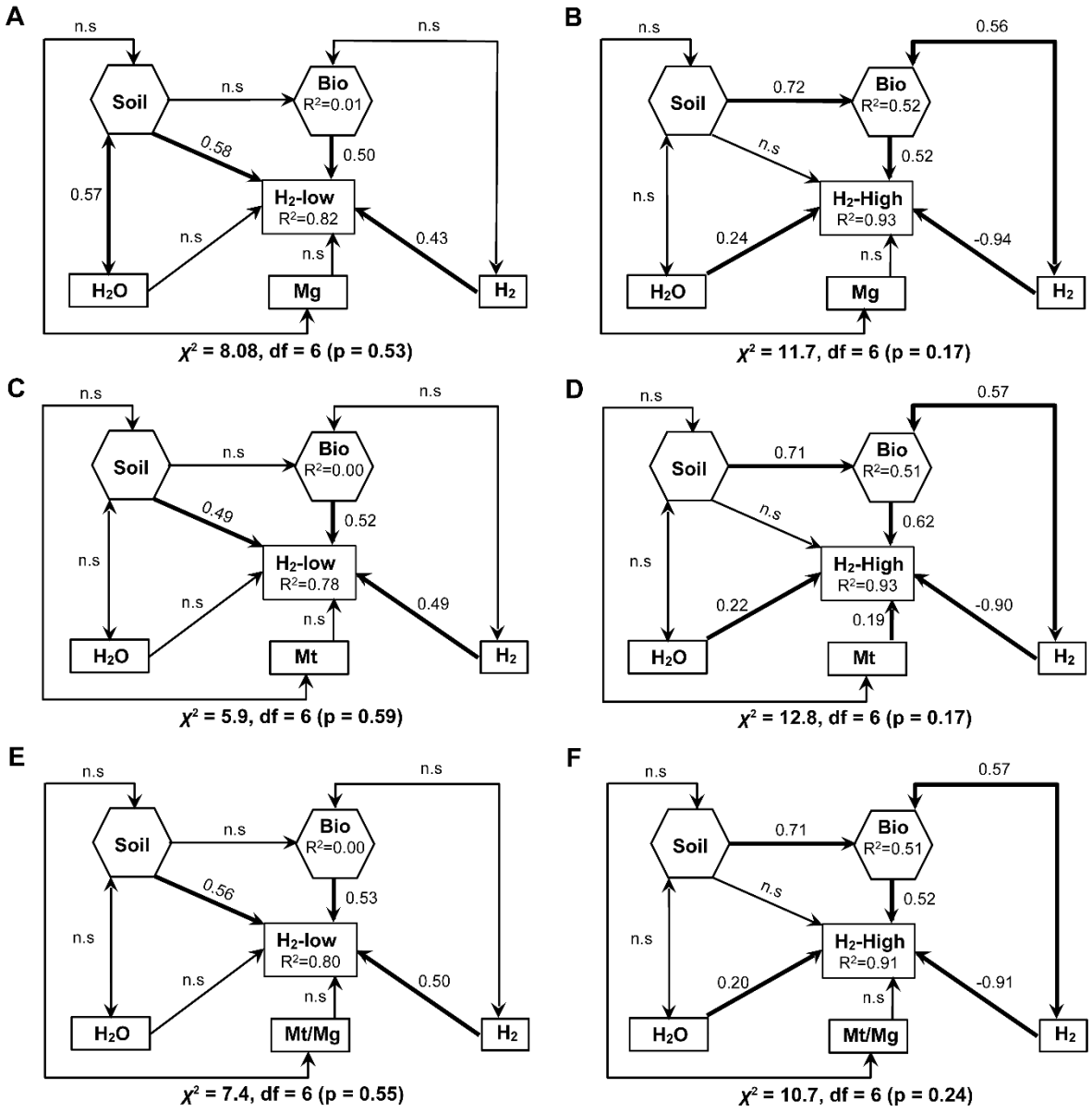
296

297

**Figure 1.**


Acetothermia
Acidobacteria
Actinobacteria
Bacteroidetes
Chloroflexi
Cyanobacteria
Planctomycetes
Proteobacteria
Unknown
Expression level
Group 1
Group 2
Group 3
Group 5
(F)
(P)
(L)

300**Figure 2.**




301

302

**Table S2.** PERMANOVA analysis displaying the proportion of variance in meta-omic profiles explained by land-use type, $H_2$ treatment or the combination land-use type and $H_2$ treatment.

| [NiFe]-hydrogenase dataset | Factors | $R^2$ (*p*-value) |
|---|---|---|
| Metagenome | $H_2$ | 0.03 (0.597) |
| | Land-use | **0.37 (0.001)** |
| | $H_2$ x Land-use | 0.08 (0.460) |
| Metatranscriptome | $H_2$ | 0.04 (0.353) |
| | Land-use | **0.40 (0.001)** |
| | $H_2$ x Land-use | 0.07 (0.529) |
| Expression ratio | $H_2$ | 0.05 (0.675) |
| | Land-use | 0.15 (0.066) |
| | $H_2$ x Land-use | 0.11 (0.678) |

* Significant $R^2$ are shown in bold

**Figure S2.** Distribution of [NiFe]-hydrogenase genes and transcripts at the phylum level. The plots represent the sum of relative gene counts for each phylum. The barplot shows the average value of genes counts from three replicates of the three soils (F; Farmland, L; Larch and P; poplar) exposed to elevated (H) or low (L) $H_2$ concentrations. 18 and 17 samples were considered for metagenomic and metatranscriptomic libraries, respectively.

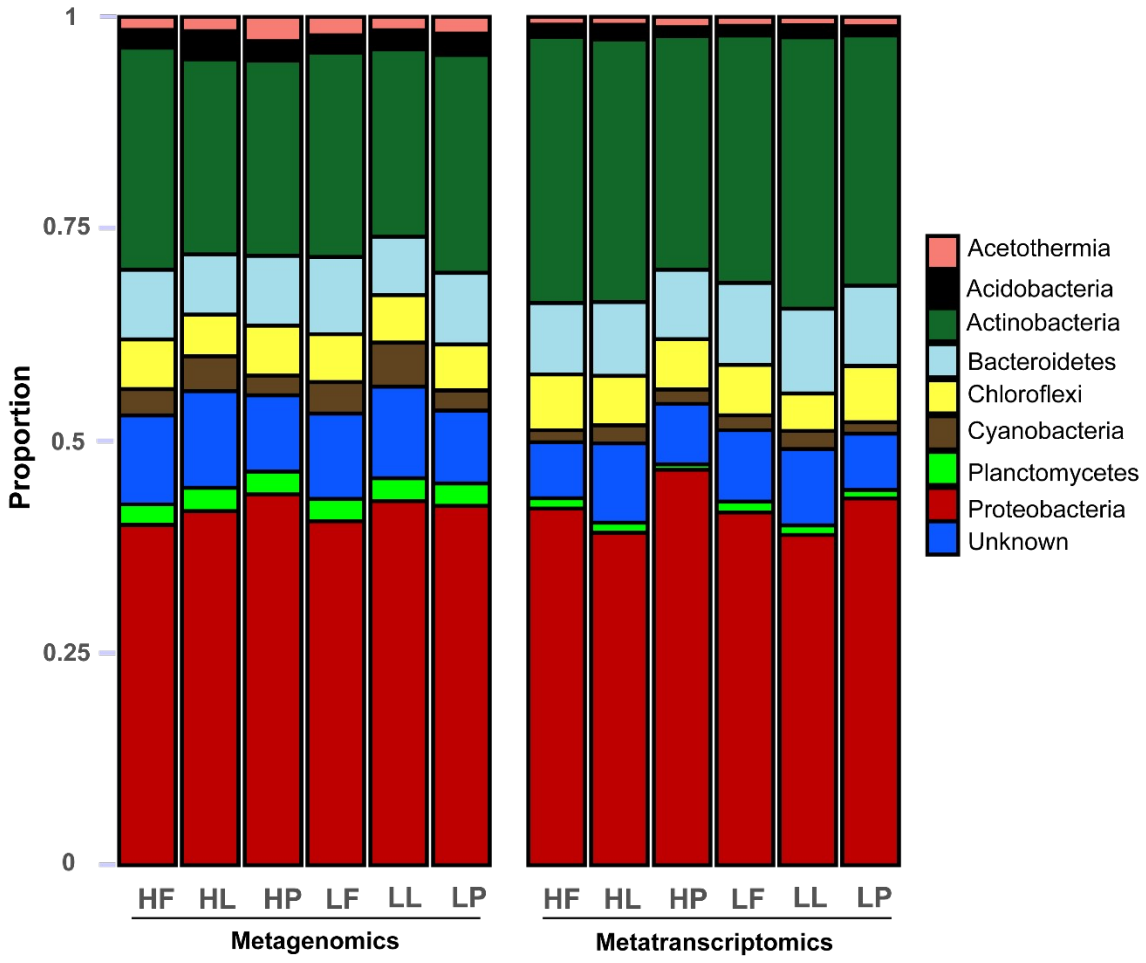


**Figure S3.** Heatmap of hydrogenases expression ratios at the phylum level. Each expression ratio value represents the ratio of the sum calculated from all genes of their respective phylum in the three replicates of each treatment condition. The first letter of soil nomenclature denotes the treatment condition (H and L for high and low $H_2$ treatment, respectively) and the second letter refers to the soil land-use (F; Farmland, L; Larch, P; Poplar).

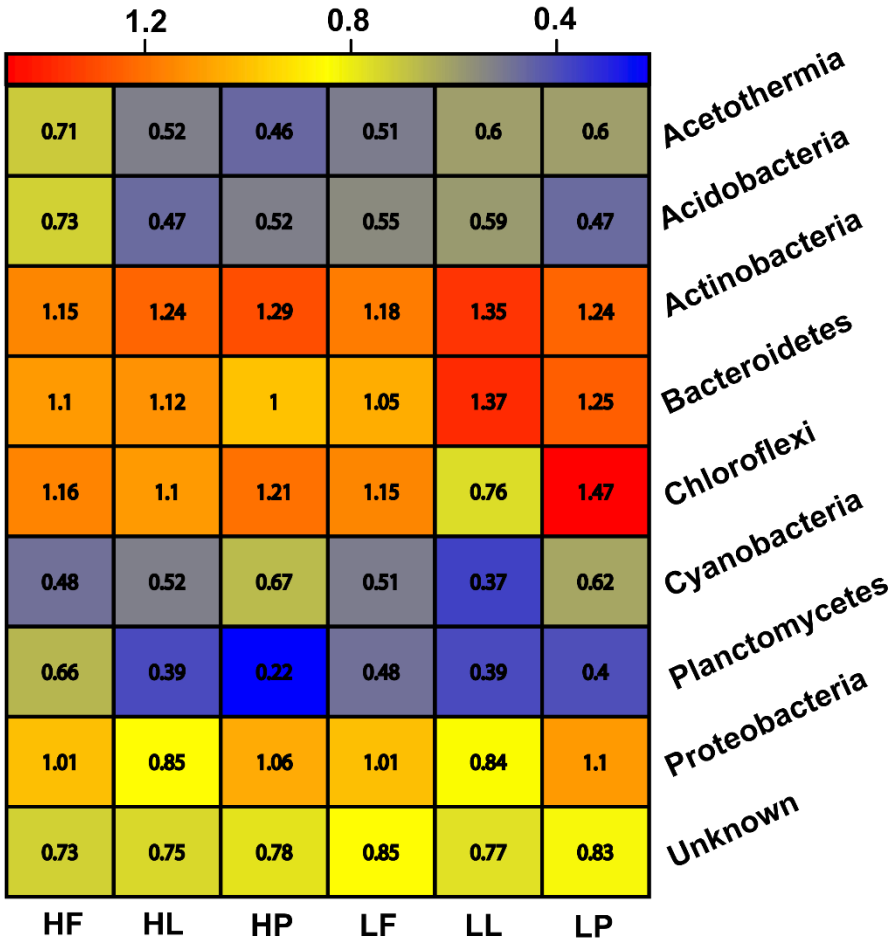


**Figure S4.** Principal component analysis showing the distribution of different land-use types based on the relative abundance of the 45 putative [NiFe]-hydrogenases on A) the metagenomics level, B) the metatranscriptomics level and C) the expression ratios level. Land-use types are represented by three different symbols (squares; Farmland, circles; Larch, triangles; Poplar). Black symbols refer to the high concentration treatment while white symbols indicate soils exposed to low concentration of $H_2$.