Supplementary Information for

Acidobacteria are active and abundant members of diverse atmospheric H2-oxidizing communities detected in temperate soils

Andrew T. Giguere, Stephanie A. Eichorst^{*}, Dimitri Meier, Craig W. Herbold, Andreas Richter, Chris Greening, Dagmar Woebken

*Corresponding author: Stephanie A. Eichorst E-mail address for correspondence: stephanie.eichorst@univie.ac.at

This PDF file includes: Supplementary text Figures S1-S15 Tables S1-S5

Supplementary Information

SI-1. Benchmarking nearly full-length group 1h [NiFe]-hydrogenase libraries to parallel shorter read libraries.

Materials and Methods

Sister libraries using previously published group 1h [NiFe]-hydrogenase primers ([NiFe]-244F/568R and [NiFe]-1129F/1640R (1, 2) **Table S1**) were generated on the investigated soil samples. Briefly, DNA and RNA were extracted from ca. 0.4-0.5 g of soil using a modified bead-beating protocol in the presence of a CTAB buffer and phenol as previously described (3). Samples were purified using OneStep™ PCR Inhibitor Removal Kit (Zymo, Irvine, CA, USA) and quantified using the Qubit dsDNA BR Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA). For the generation of cDNA, extracts were purified with the Turbo DNA-free kit, quantified with the Qubit HS RNA Assay and reverse-transcribed using SuperScript IV Reverse Transcriptase, all according to the manufacturer's protocol. All reagents and kits were purchased at Thermo Fisher Scientific, Waltham, MA, USA.

The primers were adapted to include a universal 16 bp head sequence at their 5' end for subsequent barcoding, as described previously (4). The first-step PCR were performed in 25 µl volume containing the following components: 10-50 ng of DNA template, 2.5 µl of 10x DreamTaq Green Buffer, 1 mM MgCl₂ (only for [NiFe]-244F, [NiFe]-568R), 0.2 μM of each nucleotide dNTP mixture, 0.2 μg μl⁻¹ of BSA, 0.5 µM of each primer, and 1.25 U of DreamTaq Green DNA Polymerase (all from Thermo Fisher Scientific, Waltham, MA, USA). Primers were optimized for annealing temperature and cycle number using soil DNA. The final first-step PCR program for amplification was: 94 °C for 3 min followed by 45 ([NiFe]-244F, [NiFe]-568R) or 30 ([NiFe]-1129F, [NiFe]-1640R) cycles of 94 °C for 35 s, 63 °C ([NiFe]- 244F, [NiFe]-568R) or 68 °C ([NiFe]-1129F, [NiFe]-1640R) for 30 s and 72 °C for 45 s, and a single step of final elongation at 72 °C for 10 min.

Samples were purified using the ZR-96 DNA Clean-up Kit™ (Zymo, Irvine, CA, USA) and 3 µl from the purified sample was used for a second PCR reaction, which was 50 μ in volume and contained the following mixture: 5 µl of 10 × DreamTaq Green Buffer, 0.2 mM of each nucleotide dNTP mixture, 0.2 μ g μ l⁻¹ of BSA, 1.25 U of DreamTaq Green DNA Polymerase (all from Thermo Fisher Scientific, Waltham, MA, USA) and 0.8 μ M of a barcode primer, which also contained a universal 16 bp head sequence at the 5' end as described in (4). The final PCR program for amplification was: 94 °C for 4 min followed by 8 cycles of 94 °C for 30 s, 52 °C for 30 s and 72 °C for 45 s, and a single step of final elongation at 72 °C for 10 min. Second-step PCR products were purified using the ZR-96 DNA Clean-up Kit™ (Zymo, Irvine, CA, USA) and quantified using the Quant-iT™ PicoGreen® dsDNA Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA) on a Tecan Safire plate reader (Tecan, Männedorf, Switzerland).

Samples were pooled to equimolar amounts of 20 \times 10⁹ copies per sample, followed by library preparation and sequencing provided by Microsynth AG (Balgach, Switzerland). The library was prepared by adapter ligation and PCR using the TruSeq Nano DNA Library Prep Kit (Illumina, Cat FC-121-4001) according to the TruSeq nano protocol (Illumina, FC-121-4003), excluding the fragmentation step. Sequencing was performed on a MiSeq platform (Illumina, San Diego, CA, USA), in the 2 × 300 cycle configuration using the MiSeq Reagent kit v3 (Illumina, San Diego, CA, USA). Paired MiSeq reads were split into sample datasets as described previously (3). The 95% OTU cut-off was used based on the comparison between the sequence similarity scores of group 1h [NiFe]-hydrogenase large subunit gene and the 16S rRNA gene (5); very few sequences had a similarity above 90%, as such we felt that 95% was a conservative cut-off for this functional gene. This cut-off is also consistent with a recent publication where organisms of the same species had an average nucleotide identity of ≥95% (6). As such, we took this average identity value as a threshold to generate OTUs for each respective primer pair ([NiFe]-244F/568R and [NiFe]-1129F/1640R).

Results and Discussion

Comparison of the group 1h [NiFe]-hydrogenase DNA-based libraries

We evaluated the group 1h [NiFe]-hydrogenase community stemming from the newly designed long-read primers and the previously published primers ([NiFe]-244F/568R and [NiFe]-1129F/1640R (1, 7)) across beech forest soil, managed grassland soil and biological soil crusts. The [NiFe]-244F/568R primer pair generated 163,855 reads that clustered into 338 OTU95 (95% cut-off), whereas the [NiFe]- 1129F/1640R generated 39,453 reads that clustered into 66 OTU₉₅. Across both smaller fragment libraries, the soils were fairly well-sampled based on rarefaction curves (**Fig. S5a,b,** Shannon index). The group 1h [NiFe]-hydrogenase communities were distinct across the investigated soils based on

NMDS using the Bray-Curtis dissimilarity index (**Fig. S5c,d**), as observed in the long-read amplicon libraries (**Fig. 1d** (rarefied data), **S4b** (unrarefied data)). The diversity and richness were typically higher in the temperate soils (beech forest and managed grassland soils) relative to the biological soil crusts (**Fig. S5e,f**). These diversity and richness patterns were similar to the long-read amplicon libraries when data were rarefied and unrarefied (**Fig. S4c,d**).

Although similar alpha- and beta-diversity patterns amongst soils were observed across the 3 investigated primer pair sets, we observed clear differences in the phylogenetic affiliation of the OTU representatives using RAxML EPA trees (**Fig. S6** – comparison of the grey shaded clusters and dots). The OTU representatives stemming from the [NiFe]-244F/568R primer pair spanned numerous phylogenetic phyla based on RAxML EPA, such as *Actinobacteria*, *Proteobacteria*, *Verrucomicrobia*, *Acidobacteria*, *Planctomycetes* and *Chloroflexi* (**Fig. S6a**), while OTU representatives stemming from [NiFe]-1129F/1640R amplicon library were identified as *Actinobacteria*, except two OTU representative (**Fig. S6b**). The OTU representatives of the nearly full-length PacBio amplicon libraries also spanned numerous phylogenetic phyla based on RAxML EPA, such as *Actinobacteria*, *Proteobacteria*, *Verrucomicrobia*, *Acidobacteria*, *Planctomycetes*, *Chloroflexi* (**Fig. S6c**), but appeared to capture additional members of the group 1h [NiFe]-hydrogenase community not captured by the [NiFe]-244F/568R primer pair, including putative members of the *Deltaproteobacteria* and *Nitrospira*. This newly designed primer pair also covered an additional group of actinobacterial sequences (**Fig. S6c**, top-most Actinobacteria cluster), as well as an additional deep-branching cluster of sequences containing group 1h [NiFe]-hydrogenases of diverse origin, such as members of the *Actinobacteria*, *Chloroflexi, Bacteriodetes, Acidobacteria, Proteobacteria* and *Euryarchaeota* (**Fig. 6c**, "Distant group 1h cluster"). Although these sequences were all classified as a group 1h [NiFe]-hydrogenase based on the HydDB (8), they appear to be distantly related to the other, main branch of the tree (**Fig. 2b**). Furthermore, we detected numerous sequences that were placed in the tree without reference sequences (**Fig. S6c**, indicated by grey dots) in the long-read amplicon libraries, in comparison to 1 ([NiFe]-1129f/1640R) or 5 ([NiFe]-244f/568R) OTU representatives without reference sequences in the short-read libraries (**Fig. S6a-c** – comparison gray shading). Taken together, it appears that these new primers target a wider (and putative novel) phylogenetic diversity of group 1h [NiFe]-hydrogenases.

This increased diversity in the long-read amplicon libraries was further supported by assessing the average genetic distances. The average pairwise genetic distance among the nearly full-length sequences was ca. 48% (when trimmed to the [NiFe]-244F/568R region) or 50% (when trimmed to the [NiFe]-1129F/1640R region), compared to 18% ([NiFe]-1129F/1640R) or 25% ([NiFe]-244F/568R) from the shorter amplicon sequences. A similar pattern was observed for the genetic distance of the acidobacterial/plancomycetes sequences: nearly full-length sequences showed genetic distance of ca. 23% (when trimmed to the [NiFe]-244F/568R region), compared to sequences stemming from [NiFe]-

244F/568R amplicon libraries (distance of ca. 17%) suggesting more of the acidobacterial community was captured with this new primer pair.

Bidirectional nucleotide blast analyses were performed on the OTU representatives amongst the amplicon libraries to evaluate the similarity of the amplified group 1h [NiFe] hydrogenase sequences amongst the 3 primer pairs. Bidirectional nucleotide blast analysis allows one to ascertain the proportion of sequences that are homologous between the 2 investigated sequence sets. If the two investigated sequence sets are identical, a completely overlapping histogram will be observed. However, if the two investigated sequence sets are not identical, the histograms will appear shifted or skewed illustrating that different sequence representatives are present in each respective sequence set.

When OTU representatives of [NiFe]-244F/568R and [NiFe]-1129F/1640R Illumina-based sequences were mapped to long-read OTU representatives, the distribution of both histograms stemming from sequence sets of [NiFe]-244F/568R and [NiFe]-1129F/1640R were skewed to the right (**Fig. S7**, purple bars). The bit scores (reflection of similarity) were high. Taken together, this illustrates that more of the shorter sequences aligned to the long-read sequences. In contrast, when the longread OTU representatives were mapped to OTU representatives of the [NiFe]-244F/568R and [NiFe]- 1129F/1640R Illumina-based sequences, there was a bimodal distribution, skewed to the right upon (**Fig. S7**, blue bars). The bit scores (reflection of similarity) were lower. Taken together, these data indicate that less of the PacBio sequence representatives aligned to the smaller, Illumina-based representatives.

These data suggest that the long-read primers for the group 1h [NiFe]-hydrogenases not only captured the breadth of diversity one can attain with the primer sets, [NiFe]-244F/568R and [NiFe]- 1129F/1640R, but also captured additional diversity. This is further supported when searching for the primer regions of the [NiFe]-244F/568R and [NiFe]-1129F/1640R in the PacBio sequences. The majority of the long-read OTU representatives (ca. 46 to 85%) had more than 4 mismatches for the priming regions of [NiFe]-244F/568R and [NiFe]-1129F/1640R (**Table S3**). As these priming regions could not be found in the majority of long-read sequence representatives without numerous mismatches, it further illustrates that this newly designed primer set targets additional group 1h [NiFe]-hydrogenase diversity in these environmental samples.

cDNA libraries of the group 1h [NiFe]-hydrogenase

The [NiFe]-244F/568R primer pair generated 828,149 reads that clustered into 110 OTU95 (95% cutoff) stemming from cDNA-based libraries. The soils were fairly well-sampled based on rarefaction analysis (**Fig. S9a**). Unlike the DNA-based libraries stemming from the [NiFe]-244F/568R primer pair as well as the long-read amplicon libraries, the group 1h [NiFe]-hydrogenase communities did not separate based on soil type (**Fig. S9b**). This suggests that a diverse collection of microorganisms were expressing their *hhyL*, which was not (directly) influenced by the edaphic properties of the investigated soils. The diversity and richness were typically higher in the managed grassland soils relative to the beech forest soil (**Fig. S9c**). The rhizosphere soil diversity was lower than of the bulk soil from the managed grassland soil (**Fig. S9c**). This could suggest that bacteria with low-affinity hydrogenases might be more active in the rhizosphere, as these bacteria are believed to grow on the high concentrations of H₂ produced in microniches, such as N₂-fixing root nodules (9), which are found in rhizosphere soil. The OTU representatives of the cDNA libraries stemming from the [NiFe]-244F/568R primer pair spanned numerous phylogenetic phyla based on RAxML EPA, such as *Actinobacteria*, *Proteobacteria*, *Verrucomicrobia*, *Acidobacteria*, *Planctomycetes* and *Chloroflexi* (**Fig. S10**).

Summary

Although similar patterns in diversity and community composition were observed across the investigated primer pairs (**Figs. 1, S4, S5**), the long-read primers targeting nearly the full-length of the group 1h [NiFe]-hydrogenase (*hhyL*) gene captured a wider breath of phylogenetic diversity in soil samples compared to previously published primers. These targeted investigations revealed putatively novel sequences in our nearly full-length group 1h [NiFe]-hydrogenase libraries. The average pairwise genetic distance among the nearly full-length sequences was ca. 50% generating ca. 2403 OTUs, compared to 18% or 25% from the shorter amplicon sequences generating 325 or 61 OTUs. Previous investigations noted a discrepancy between the theoretical and measured estimates of H_2 -oxidizing bacteria in soil, and one of the discussed possible reasons were unidentified participants (5). Although it is unclear if we captured all of the group 1h [NiFe]-hydrogenase diversity in these soils, we submit that these newly developed primers capture more of the community, identifying some putative novel community members (**Fig. S6c**) for future investigations.

The goal of the cDNA libraries was to ascertain if acidobacterial group 1h [NiFe] hydrogenases were being expressed in the soil. No acidobacterial group 1h [NiFe]-hydrogenases were detected with [NiFe]-1129F/1640R primer pair, but were detected with our long-read primers (**Fig. 6c)** and in both the DNA- and cDNA-based libraries stemming from the primer pair [NiFe]-244F/568R (**Figs. S6a, S10**). Taken together this illustrates that the acidobacterial *hhyL* genes can be detected with the newly developed long-read primers and [NiFe]-244F/568R and appear to be expressed in the soil, as indicated in the cDNA-based libraries of the primer pair [NiFe]-244F/568R. Yet there was an increased genetic distance of acidobacterial/plancomycetes sequences amplified from the long-read primers (ca. 23%) compared to sequences stemming from [NiFe]-244F/568R (ca. 17%) indicating that this community was better captured with this new primer pair. This suggests that the additional sequence length,

primer location and/or degenerate nature of the primer pair are necessary to capture not only an increased acidobacterial group 1h [NiFe]-hydrogenase diversity, but also overall diversity in soils.

References

- 1. Constant P, Chowdhury SP, Pratscher J, Conrad R (2010) Streptomycetes contributing to atmospheric molecular hydrogen soil uptake are widespread and encode a putative highaffinity [NiFe]-hydrogenase. *Environ Microbiol* 12:821–829.
- 2. Constant P, Chowdhury SP, Hesse L, Conrad R (2011) Co-localization of atmospheric H₂ oxidation activity and high affinity H_2 -oxidizing bacteria in non-axenic soil and sterile soil amended with *Streptomyces* sp. PCB7. *Soil Biol Biochem* 43:1888–1893.
- 3. Angel R (2012) Total Nucleic Acid Extraction from Soil. *Protocol Exchange*. doi:10.1038/protex.2012.046.
- 4. Herbold CW, *et al*. (2015) A flexible and economical barcoding approach for highly multiplexed amplicon sequencing of diverse target genes. *Front Microbiol* 6:731.
- 5. Constant P, Chowdhury SP, 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 H2-oxidizing bacteria. *Appl Environ Microbiol* 77:6027– 6035.
- 6. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S (2018) High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Com* 9:5114.
- 7. Piché-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. *Microb Ecol* 74:771-775.
- 8. Søndergaard D, Pedersen CNS, Greening C (2016) HydDB: A web tool for hydrogenase classification and analysis. *Sci Reports*. 6:1–8.
- 9. Maimaiti J, Zhang Y, Yang J, Cen YP, Layzell DB, Peoples M, *et al*. (2007) Isolation and characterization of hydrogen-oxidizing bacteria induced following exposure of soil to hydrogen gas and their impact on plant growth. *Environ Microbiol*. 9:435–444.

SI -2. Investigating the *hhyS* **homolog in KBS 83 and** *E. aggregans*

A homologous gene to *hhyS* was detected upstream (ca. 3600 bp) of the structural and maturation genes of the group 1h [NiFe]-hydrogenase in *Acidobacteriaceae* bacterium KBS 83 (G002DRAFT_03432) and *E. aggregans* (Q363DRAFT_01016). These homologues were distantly related to the *hhyS* gene in each respective genome (ca. 70.7% amino acid sequence similarity based on ca. 320 amino acids across the strains). In comparison, the *hhyS* of *Acidobacteriaceae* bacterium KBS 83 and *E. aggregans* had an amino acid sequence similarity of 87.7% (based on ca. 320 amino acids). As seen with *P. methylaliphatogenes*, the homolog appears to be a part of a putative twocomponent sensory system (KBS 83: G002DRAFT_03432-03434; *E. aggregans*: Q363DRAFT_01014- 01016).

The *hhyS* homolog in *Acidobacteriaceae* bacterium KBS 83 was expressed in both exponential (ca. 128 copies ng⁻¹) and stationary phase (ca. 2,153 copies ng⁻¹), with ca. 13-fold upregulation in stationary phase upon normalization to the 16S rRNA gene in KBS 83. The *hhyS* homolog was not expressed in neither exponential nor stationary phase for *E. aggregans*.

Figure S1. Establishment of the acidobacterial strain-specific group 1h [NiFe]-hydrogenase *hhyL* (**panels a-d**) and *hhyS* (**panels e-h**) qPCR assays. qPCR amplification (**panels a,e**), standards (**panels b,f**), melting curves (**panels c,g**) are depicted for a serially diluted DNA standard. Neighbor-joining phylogenetic tree of amplified products illustrating specificity of the reaction (**panels d, h**). Data from *Acidobacteriaceae* bacterium KBS 83 strain specific primers are shown for qPCR amplification, standards and melting curves. Similar patterns were observed for *E. aggregans*.

Figure S2. Oyxgen concentration in the headspace of *Acidobacteriaceae* bacterium
KBS 83 & E. aggregans cultures during H_rconsumption assay. KBS 83 & E. aggregans cultures during H_2 consumption assay.

Figure S3. H₂ consumption controls over 24 hours for *T. roseus* KBS 63 (group 1h [NiFe]-hydrogenase negative strain) and *Acidobacteriaceae* bacterium KBS 83 (group 1h [NiFe]-hydrogenase positive strain), along with uninoculated medium. No significant difference in H_2 concentrations was found among the three controls over time (P>0.73).

Figure S4. Diversity statistics for the long-read libraries of the group 1h [NiFe]-hydrogenase across the four soil types: managed grassland soil (red), rhizosphere soil (orange), beech forest soil (blue) and biological soil crust (purple). Rarefaction analysis based on observed and Shannon index (**panel a**); beta diversity (Bray-Curtis dissimilarity) of unrarefied data (**panel b**) and diversity statistics of unrarefied (**panel c**) and rarefied data (**panel d**). Data were unrarefied for **panels b & c**, but rarefied in **panel d**. Analysis of variance with a Tukey's HSD mean separation was performed across the soil types for the diversity statistics for the unrarefied and rarefied data; similar letters indicate that no significant difference was observed (P > 0.05),

Figure S5. Diversity statistics for the short-read fragment libraries of the *hhyL* gene for three soil types: managed grassland soil (red), beech forest soil (blue) and biological soil crust (purple). Rarefaction analysis for [NiFe]-244F/568R (**panel a**) and [NiFe]-1129F/1640R (**panel b**); NMDS based on Bray-Curtis for [NiFe]- -244F/568R (**panel c**) and [NiFe]-1129F/1640R (**panel d**); and diversity statistics for [NiFe]-244F/568R (**panel e**) and [NiFe]-1129F/1640R (**panel f**). Note - triplicate libraries stemming from [NiFe]-1129F/1640R are depicted in **panel d**, but the libraries from each soil type cluster very closely. Analysis of variance with a Tukey's HSD mean separation was performed across the soil types for the diversity statistics; similar letters indicate that no significant difference was obserbedf ($P > 0.05$), 'nsd' = no significant difference. Libraries stemming from [NiFe]244F/568R were rarified to 10,000 sequences & from [NiFe]-1129f/1640R to 2,000 sequences.

Figure S6. RAxML-EPA tree of the group 1h [NiFe]-hydrogenase *hhyL* based on amino acid sequences. The phylogenetic placements of OTU representatives (in gray collapsed clusters) are shown stemming from the [NiFe]-244F/568R (**panel a**), [NiFe]-1129F/1640R primer pairs (**panel b**) and the newly designed long-read primer pair (**panel c -** same as Figure 2b, here for comparison of primer pair sets). The number of placed short reads group 1h [NiFe]-hydrogenase and total number of sequences in each cluster are depicted. The proportion of sequences within each soil type are depicted to the right of the taxonomic groups: managed grassland (red), beech forest (blue), and biological soil crust (purple). Sequences from group 1g [NiFe]-hydrogenases were used as an outgroup. The scale bar indicates the number of substitutions per site.

Figure S7. Assessing the overlap of the group 1h [NiFe]-hydrogenase communities using bidirectional nucleotide blast analyses: [NiFe]-244F/568R to nearly full-length sequences (**panel a**) and [NiFe]-1129F/1640R to nearly full-length sequenes (**panel b**). Top bit scores were normalized to self-nucleotide blast of either [NiFe]-244F/568R (**panel a**) or [NiFe]-1129F/1640R (**panel b**).

Figure S9. Diversity statistics of the *hhyL* sequences amplified from cDNA with primer pair ([NiFe]-244F/568R) for three soil types: managed grassland soil (red), rhizosphere soil (orange) and beech forest soil (blue). Rarefaction analysis (**panel a**); NMDS based on Bray- -Curtis (**panel b**) and diversity statistics (**panel c**). Analysis of variance with a Tukey's HSD mean separation was performed across the soil types for the diversity statistics; similar letters indicate that no significant difference was observed (P > 0.05), 'nsd' = no significant difference. Libraries were rarified to 2,300 sequences.

Figure S10. RAxML-EPA tree of the group 1h [NiFe]-hydrogenase (*hhyL*) based on amino acid sequences. The phylogenetic placements of OTU representatives (in gray collapsed clusters) from cDNA amplified with primer pair ([NiFe]-244F/568R) are shown. The number of placed short read group 1 [NiFe]-hydrogenase gene (*hhyL*) sequences and total number of sequences in each cluster are depicted. The proportion of sequences within each soil are depicted to the right of the taxonomic groups: managed grassland (red), rhizosphere (orange) and beech forest (blue). The outgroup was *Thermoproteus neutrophilus* (WP_012349775.1) and *Pyrobaculum islandium* (group 1g; WP_011761956.1). Scale bar indicates the number of substitu tions per site.

Figure S11. Defining carbon-limiting conditions across the investigated acidobacterial strains, *Acidobacteriaceae* bacterium KBS 83 (**panel a**) and *E. aggregans* (**panel b**). To determine carbon-limiting conditions, strains were grown in the VSB-6 defined medium (see Materials and Methods for more details on the medium composition) in differing glucose concentrations (5 mM and 10 mM glucose). Carbon-limiting conditions were defined when the cellular yield (optical density 600 nm) was proportional to the amount of carbon provided. Growth was monitored by measuring the optical density at 600 nm.

Acidobacteriaceae bacterium KBS 83 (SD 1)

2 kb

(b)

Figure S12. Identification of predicted promoter regions (depicted with red arrows) across the structural genes for the group 1h [NiFe]-hydrogenase operon in strain *Acidobacteriaceae* bacterium KBS 83 (**panel a**). Predicted promoter position and transcription factors, along with the respective scores (**panel b**). Data were generated using convolutional neural network models (https://arxiv.org/abs/1610.00121).

Figure S13. H_2 consumption assays of cells in exponential phase. Cells were harvested for both strains in exponential phase (**panel a**); arrows depict the time points in which the cells were harvested from each respective strain. Controls for this assay can be found in Figure S3. H_2 consumption of the exponential phase cells of both strains was monitored over time (**panel b**). No significant H₂ consumption was observed over time for neither *E. aggregans* (P > 0.71), nor *Acidobacteriaceae* bacterium KBS 83 $(P > 0.064)$.

Figure S14. Growth curve of *E. aggregan*s grown on an undefined medium (R2A medium, without agar) under aerobic conditions (**panel a**). Arrow depicts time point where H_2 consumption was measured over a 24-hour period (**panel b**). No significant H_2 consumption was detected over a 24-hour period (P > 0.18). The controls for this experiment can be found in Figure S3.

Figure S15. Distribution of the (**a**) *Edaphobacter aggregans* (group 1h [NiFe]-hydrogenase containing strain) (**b**) *Acidobacteriaceae* bacterium KBS 83 (group 1h [NiFe]-hydrogenase containing strain), (**c**) *Acidobacterium ailaaui* (group 1f [NiFe]-hydrogenase containing strain) and (**d**) *Pyrinomonas methylaliphatogenes* K22 (group 1h [NiFe]-hydrogenase containing strain) in the environment. Full-length 16S rRNA gene sequences for each respective strain (*E. aggregans*, DQ528761.1, 1,425 bp; *Acidobacteriaceae* bacterium KBS 83, FJ870383.1, 1,418 bp; *A. ailaaui*, KX306477.1, 1,258 bp; *Pyrinomonas methylaliphatogenes* K22, AM749787.1, 1,432bp) were blasted in NCBI using Megablast; a total of 59 (*Acidobacterium ailaaui*), 58 (*Edaphobacter aggregans)*, 26 (*Acidobacteriaceae* bacterium KBS 83) and 2 (Pyrinomonas methylaliphatogenes K22) sequences were recovered from the public database (NCBI) having a similiarity of \geq 97%. These were classified manually as orginating from "temperate soils", "thermophilic environments (such as soils, bagasse and volcanic deposits, all with temperatures >40˚C), "acid mine drainage", "aquatic" (such as deep-sea octacoral samples), "insect-associated" (such as the termite hindgut and gut of New Zealand's endemic Huhubeetle larvae), "microbial mats" "various" (such as "hot water", "leaves", "skin" and "waste pile") and "unknown" (no description was given in the sequence entries at NCBI, along with no associated manuscript). The percent of these sequences found in each respective environment is depicted.

Table S1. Primers used in this study.

- Constant, P., Chowdhury, S. P., Hesse, L., & Conrad, R. (2011). Co-localization of atmospheric H₂ oxidation activity and high affinity H₂-oxidizing bacteria in non-axenic soil and sterile soil amended with *Streptomyces sp*. PCB7. *Soil Biol.Biochem.*, *43*(9), 1888–1893. http://doi.org/10.1016/j.soilbio.2011.05.009
- Constant, P., Chowdhury, S. P., Pratscher, J., & Conrad, R. (2010). Streptomycetes contributing to atmospheric molecular hydrogen soil uptake are widespread and encode a putative high-affinity [NiFe]-hydrogenase. *Environ. Microbiol.*, *12*(3), 821–829. http://doi.org/10.1111/j.1462-2920.2009.02130.x
- Steger, D. (2010). Community structure and distribution of functional microbial groups within two complex environments. PhD Thesis. University of Vienna. http://doi.org/10.25365/thesis.10359

Table S2. Summary statistics of nearly full-length group 1h [NiFe]-hydrogenase sequences.

Table S3. Number of sequences (% of total sequences) containing the respective primer region of the listed primers, along with different mismatch cases. Nearly-full length group 1h [NiFe]-hydrogenase sequences stemming from the PacBio sequencing run were used (n=2403; *except for region 1640R, where n=2242 sequences were used as there were sequences without sequence data at this region).

Table S4. Details on publicly available MAG seqeunces depicted in Figure 3.

Table S5. Determining apparent kinetic parameters (average \pm standard error) for H₂ oxidization on whole cells for strain *Acidobacteriaceae* bacterium KBS 83 and *E. aggregans* using two methods (nonlinear regression and Hanes-Woolf plot).

^aRates given as µmols mg⁻¹ protein h⁻¹

 $^{\text{b}}$ Rates given as nmols mg⁻¹ protein h⁻¹

c Concentrations given in nM