Design of a degenerate primer pair to target a bacterial functional community: The hppd bacterial gene coding for the enzyme targeted by herbicides, a study case

C. Thiour-Mauprivez, Marion Devers-Lamrani, A. Mounier, Jérémie Beguet, A. Spor, C. Calvayrac, L. Barthelmebs, Fabrice Martin-Laurent

Design of a degenerate primer pair to target a bacterial functional community: the *hppd* bacterial gene coding for the enzyme targeted by herbicides, a study case

Thiour-Mauprivez C.[1,2], Devers-Lamrani M.[2], Mounier A.[2], Beguet J.[2], Spor A.[2], Calvayrac C.[1], Barthelmebs L.[1], Martin-Laurent F.[2]

[1]Univ. Perpignan Via Domitia, Biocapteurs-Analyses-Environnement, 66860 Perpignan, France; Laboratoire de Biodiversité et Biotechnologies Microbiennes, USR 3579 Sorbonne Universités (UPMC) Paris 6 et CNRS Observatoire Océanologique, 66650 Banyuls-sur-Mer, France.

[2]AgroSup Dijon, INRA, Univ. Bourgogne, Univ. Bourgogne Franche-Comté, F-21065 Dijon, France.

* Corresponding author: fabrice.martin@inra.fr

Abstract: The present work aimed to design a degenerate primer pair to target a large part of the *hppd* soil bacterial community, possibly affected by herbicides. We validated these primers by qPCR and high-throughput sequencing analysis of soil samples.

Keywords: *hppd*, degenerate primers, soil bacterial community.

Highlights:

- A novel way to design a degenerate primer pair targeting a functional community
- Soil bacterial *hppd* community abundance and diversity can be monitored
- The effects of HRAC F2 herbicides on soil communities can now be assessed

In recent scientific opinions, the European Food Safety Authority proposed to include soil ecosystemic services as a specific protection goal for risk assessment of pesticides and to protect soil microorganisms at the level of functional groups (EFSA, 2010). Another scientific opinion underlines the lack of standardized methods to measure the effects of pesticides on structural and functional endpoints of the microbial community (EFSA, 2017). To fill this knowledge gap, microbial genes encoding enzymes inhibited by substances with herbicidal activity could be ideal biomarkers of the direct effects of these molecules on soil microorganisms (Thiour-Mauprivez *et al*., 2019).

Screening of these active substances according to their mode of action led to the identification of those that affect the cell metabolism. A list of targeted enzymes was drawn up by the HRAC (https://hracglobal.com/tools/world-of-herbicides-map). Among them, 4-hydroxyphenylpyruvate dioxygenase (4-HPPD; EC 1.13.11.27) appeared as a possible biomarker sensitive to the herbicides of the HRAC (Herbicide Resistance Action Committee) F2 group (list of 13 active substances with herbicidal activity, **supplementary file 1**). This enzyme transforms tyrosine into fumarate and acetoacetate. These two compounds enter the central metabolism and lead to the formation of pyomelanine, a brown pigment that has antioxidant properties and protects fungi and bacteria from reactive oxygen species (ROS) (Langfelder et al., 2003; Turick et al., 2009). To our best knowledge, 2,869 HPPD protein sequences from various bacterial species are available in the Pfam database (http://pfam.xfam.org/family/PF14696) (**figure 1**). One could hypothesize that among the numerous populations harbouring 4-HPPD within the bacterial community, several of them might be sensitive to HRAC F2 herbicides. From this point of view, both *a priori* and *a posteriori* risk assessment of HRAC F2 herbicides on soil microorganisms might be assessed by monitoring parameters of the *hppd* bacterial community.

In this context, the aim of this study was to develop a molecular tool to monitor the abundance and diversity of this community in the soil. In order to target as many bacterial *hppd* sequences as possible, we developed a pipeline (**supplementary file 2**) to design a degenerate primer pair targeting a broad spectrum of bacterial 4-HPPD genes. This was challenging since the known bacterial 4-HPPD proteins available on Pfam presented both length and sequence polymorphisms. Sequence lengths varied from 237 to 491 amino acids (95% of the 4-HPPD sequences were 325 to 360 amino acids in length), and the most distant protein sequences showed only 22% similarity (79/354 identical amino acids). The corresponding nucleic acid sequences were retrieved from GenBank, aligned with Muscle

(Edgar, 2004), and too long, too short or too different sequences were removed manually. The 2,506 remaining sequences were converted into protein sequences using Mega7 (Kumar et al., 2016), resulting in a clean bacterial 4-HPPD database (see taxonomic affiliation of these sequences in **table 1**). The latter was submitted to j-CODEHOP (Boyce *et al*., 2009) to identify conserved motifs candidate for primer design. Two motifs were identified: HIA (harboured by 1,957 out of 2,506 sequences) and FFE (found in 2,361 sequences) (in pink and violet in **figure 1**). Twenty-base degenerate primers were then designed so that the nine bases in 3'-terminal position perfectly annealed the HIA or FFE motifs shared by 1,812 of the 2,506 analysed sequences (**figure 2**).

To validate our primer pair, three soil samples were collected from the surface layer (0-20 cm) of an arable field (University of Perpignan, France) (see properties in **supplementary file 3**). Soil samples were prepared as previously described (Romdhane *et al*., 2016). Nucleic acids were extracted with an RNeasy PowerSoil DNA Elution Kit (Qiagen) and the *hppd* gene was successfully amplified by PCR from these DNA extracts as described in **supplementary file 4**. The affiliation of the amplicons to the *hppd* gene sequence was checked by cloning and sequencing. Ninety-two percent of the 24 sequenced amplicons were related to *hppd* sequences (sequences submitted and released upon manuscript acceptance). The *hppd* community relative abundance was assessed by qPCR as described before (Romdhane *et al*., 2016) except for the HIA-F/FFE-R primer pair which was used at 10µM in our final reaction mixture with an annealing temperature of 54°C. A calibration curve was generated in triplicate using serial dilutions of the standard ranging from 10^2 to 10^7 copies *per* reaction. Three no-template controls were included for each qPCR assay. The calibration curve showed an acceptable efficiency of 79% (R^2 = 0.99). The detection limit was around 10^2 copies of *hppd* copies *per* reaction (data not shown). Relative abundance was 421±43 *hppd* sequences

per 1,000 sequences of 16S rDNA. *In silico* observations confirmed that *hppd* gene sequences were quite abundant among the soil bacterial community.

hppd sequence diversity of the amplicons obtained with a two-step PCR was assessed by Illumina 2*250 base pair MiSeq sequencing run (BioEnvironnement, University of Perpignan, France). The first PCR involved 27 cycles of amplification. It was performed in duplicate for each of the three soil DNA samples, and one no-template control using the NGS_HIA_F (5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNARGGBATYCAGCAYATYGCN-3') and NGS_FFE_R (5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNCGYTGDATRAWYTCRAARAA-3') primer pair with a specific tail added to match the second PCR primer pair. Six microliters of the pooled first PCR products were used as a template for the second PCR that involved eight cycles of amplification with barcoded primers in duplicate. After pooling and amplicon size verification by electrophoresis, *hppd* amplicons were normalized using a SequalPrep™ Normalization Plate Kit (Invitrogen). Ten microliters of each sample were used for Illumina MiSeq sequencing. After assembly using PEAR (Zhang *et al*., 2014) and quality check using the QIIME pipeline (Caporaso *et al*., 2010), a total of 744,084 hppd nucleic sequences were obtained.. These sequences were corrected and converted into proteins using framebot (Wang *et al*., 2013), then clustered using cd-hit (Li and Godzik, 2006) (threshold: 100%). The resulting 62,409 clusters were compared to the 4-HPPD database using blast, and affiliated to the closest relative sequence (minimum similarity: 57%). The designed primers allowed us to retrieve *hppd* sequences affiliated to most phylogenetic groups from our database, as shown in **figure 1**. However, only a few sequences related to Actinobacteria were obtained. This might be due not only to primer efficiency regarding this group but also to various biases such as DNA extraction and/or low abundance of Gram-positive bacteria in this

soil. A small number of *hppd* sequences that contained only one (5.6%) or no motif (2.7%) were also amplified, suggesting that they may correspond to sequences not yet described.

This primer pair appears to be suitable for measuring the abundance of the *hppd* bacterial community and estimating its diversity in an arable soil. Further studies will aim to monitor the evolution of the *hppd* bacterial community exposed to herbicides of the HRAC F2 group. It is noteworthy that this primer design method could also be applied to amplify other functional genes, paving the way for the analysis of other bacterial subcommunities.

Acknowledgements. The authors thank the Région Occitanie, European Funds for Regional Development (FEDER) and Perpignan University Via Domitia (BQR 2018) for their financial support. The authors also thank Miss Annie Buchwalter for reading the manuscript and providing suggestions.

References

Boyce, R., Chilana, P., Rose, T.M., 2009. iCODEHOP: a new interactive program for designing COnsensus-DEgenerate Hybrid Oligonucleotide Primers from multiply aligned protein sequences. Nucleic Acids Res. 37, W222–W228. https://doi.org/10.1093/nar/gkp379

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R., 2010. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods 7, 335–336. https://doi.org/10.1038/nmeth.f.303

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797. https://doi.org/10.1093/nar/gkh340

EFSA, 2010, 2010. EFSA J. 8, 1821. https://doi.org/10.2903/j.efsa.2010.1821

EFSA, 2017, 2017. EFSA J. 15, e04982. https://doi.org/10.2903/j.efsa.2017.4982

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol. Biol. Evol. 33, 1870–1874. https://doi.org/10.1093/molbev/msw054

Langfelder, K., Streibel, M., Jahn, B., Haase, G., Brakhage, A.A., 2003. Biosynthesis of fungal melanins and their importance for human pathogenic fungi. Fungal Genet. Biol. FG B 38, 143–158.

Letunic, I., Bork, P., 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 47, W256–W259. https://doi.org/10.1093/nar/gkz239

Li, W., Godzik, A., 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinforma. Oxf. Engl. 22, 1658–1659. https://doi.org/10.1093/bioinformatics/btl158

Price, M.N., Dehal, P.S., Arkin, A.P., 2009. FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. Mol. Biol. Evol. 26, 1641–1650. https://doi.org/10.1093/molbev/msp077

Romdhane, S., Devers-Lamrani, M., Barthelmebs, L., Calvayrac, C., Bertrand, C., Cooper, J.-F., Dayan, F.E., Martin-Laurent, F., 2016. Ecotoxicological Impact of the Bioherbicide Leptospermone on the Microbial Community of Two Arable Soils. Front. Microbiol. 7. https://doi.org/10.3389/fmicb.2016.00775

Thiour-Mauprivez, C., Martin-Laurent, F., Calvayrac, C., Barthelmebs, L., 2019. Effects of herbicide on non-target microorganisms: Towards a new class of biomarkers? Sci. Total Environ. 684, 314–325. https://doi.org/10.1016/j.scitotenv.2019.05.230

Turick, C.E., Beliaev, A.S., Zakrajsek, B.A., Reardon, C.L., Lowy, D.A., Poppy, T.E., Maloney, A., Ekechukwu, A.A., 2009. The role of 4-hydroxyphenylpyruvate dioxygenase in enhancement of solid-phase electron transfer by Shewanella oneidensis MR-1. FEMS Microbiol. Ecol. 68, 223–225. https://doi.org/10.1111/j.1574-6941.2009.00670.x

Wang, Q., Quensen, J.F., Fish, J.A., Lee, T.K., Sun, Y., Tiedje, J.M., Cole, J.R., 2013. Ecological Patterns of nifH Genes in Four Terrestrial Climatic Zones Explored with Targeted Metagenomics Using FrameBot, a New Informatics Tool. mBio 4, e00592-13. https://doi.org/10.1128/mBio.00592-13

Wheeler, T.J., Clements, J., Finn, R.D., 2014. Skylign: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models. BMC Bioinformatics 15, 7. https://doi.org/10.1186/1471-2105-15-7

Zhang, J., Kobert, K., Flouri, T., Stamatakis, A., 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinforma. Oxf. Engl. 30, 614–620. https://doi.org/10.1093/bioinformatics/btt593

175

176 **Figure legends**

177 **Figure 1.** Phylogenetic tree of the 4-HPPD database derived from Pfam, constructed with
178 FastTree (Price *et al*., 2009) (Neighbour-Joining) and visualized with iTol (Letunic and Bork,
179 2019). The inner circle represents the affiliation of the 4-HPPD sequences to the main
180 microbial groups indicated by different colours. The middle circle depicts the presence of the
181 HIA and/or FFE motifs in the HPPD sequences, indicated by different colours. The outer
182 circle shows the affiliation of the OTUs found in Perpignan soil samples to known HPPD
183 sequences.

184

185

186

187 **Figure 2.** HMM logos constructed with Skylign (Wheeler *et al*., 2014) showing the
188 occurrence of the amino acids which composed our primer pair. R, A or G; B, C or G or T; Y,
189 C or T; N, any base; D, A or G or T; W, A or T. The amino acid position and the amplified
190 product length were determined in the *hppd* sequence of *Pseudomonas fluroescens* F113
191 (NC_016830.1).

192

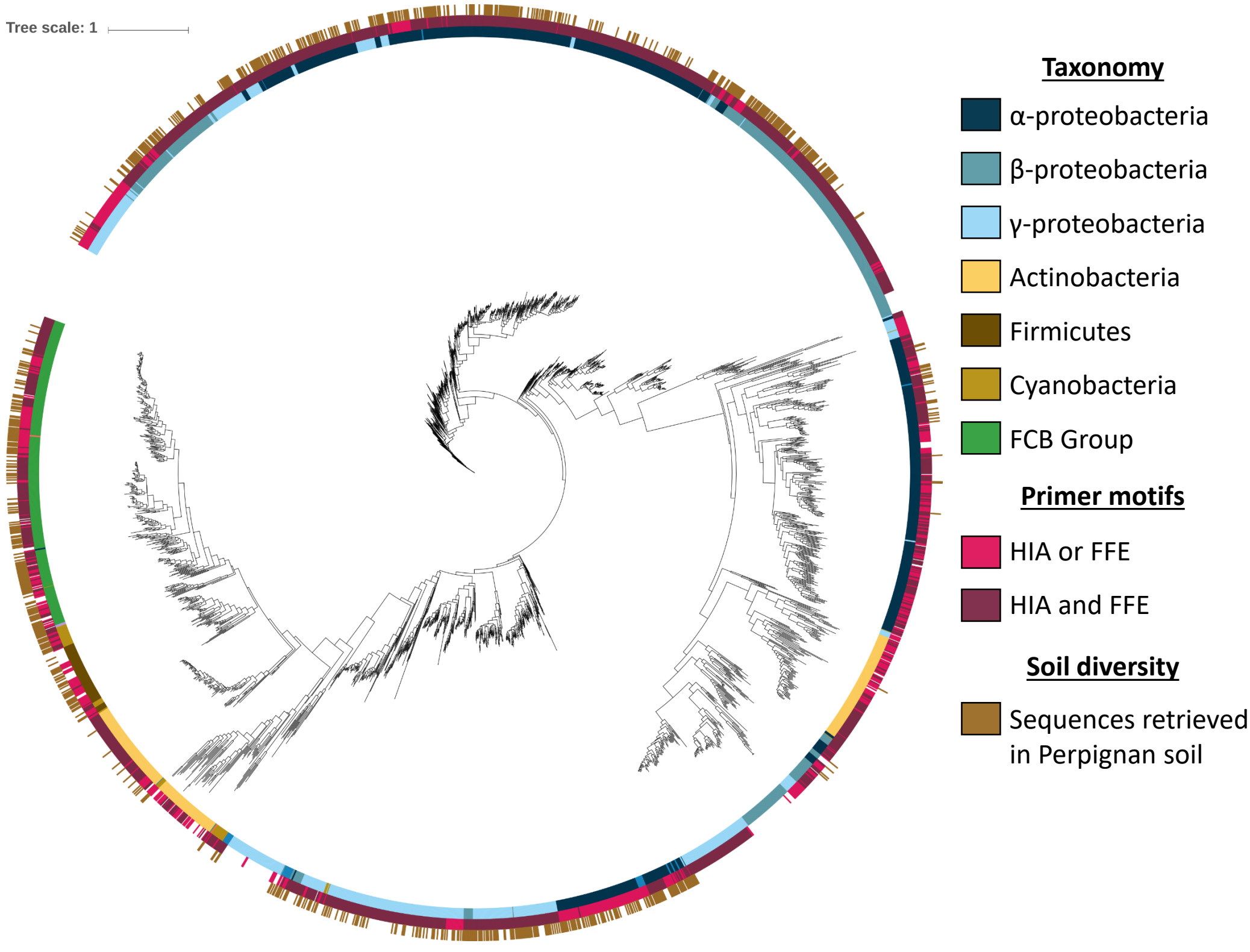

Figure 1. Phylogenetic tree of the 4-HPPD database derived from Pfam, constructed with FastTree[12] (Neighbor-Joining) and visualized with iTol[13].

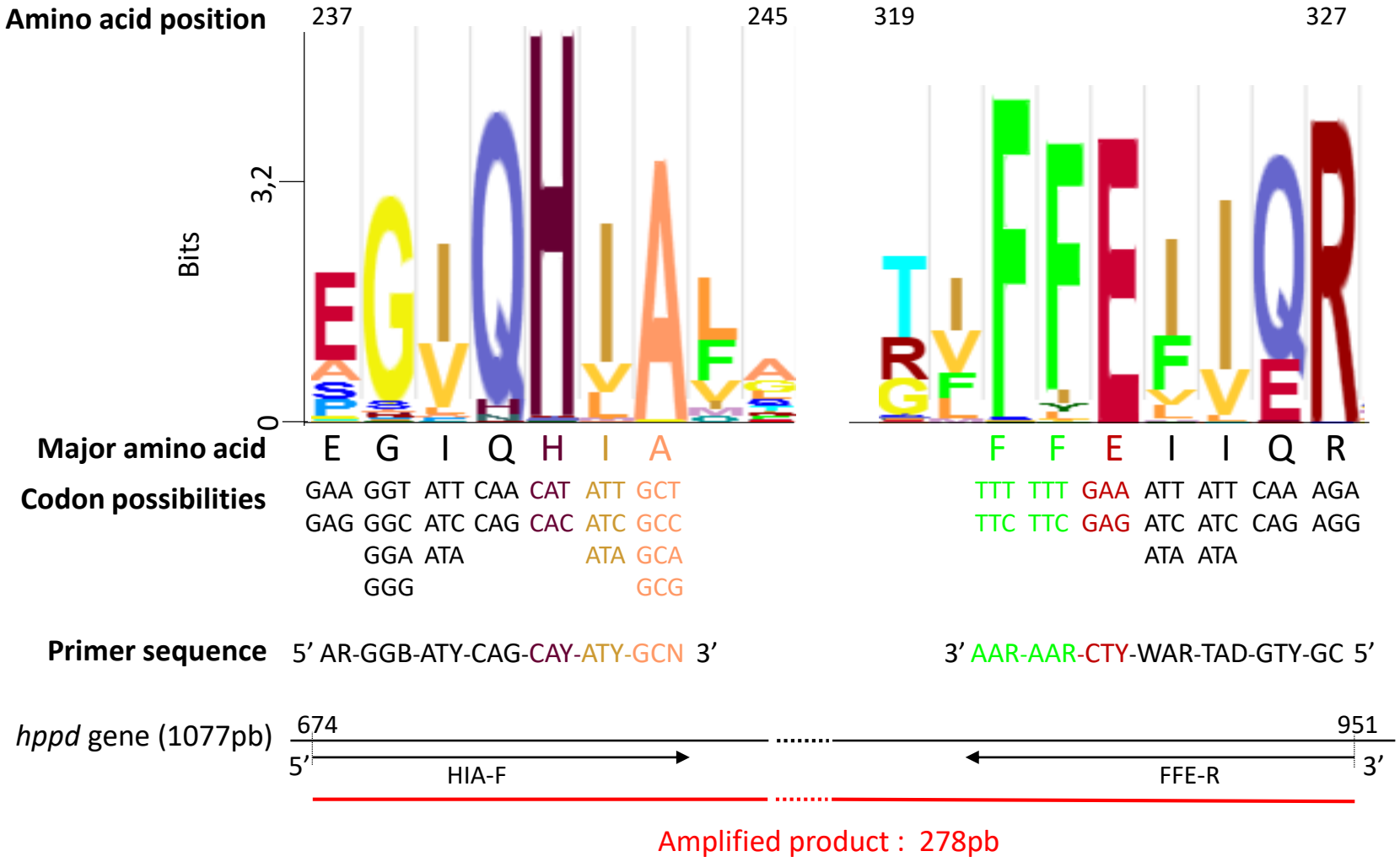

Figure 2. HMM logos constructed with Skylign[14] showing the occurrence of the amino acids which composed our primer pair.

Table 1. Taxonomic affiliation of our *hppd* sequence database.

Phylum	Class	Number of sequences
Proteobacteria	α-proteobacteria	822
	γ-proteobacteria	529
	β-proteobacteria	484
Terrabacteria group	Actinobacteria	265
	Firmicutes	67
	Cyanobacteria	37
FCB group	Bacteroidetes	298
	Gemmatimonadetes	3
	Acidobacteria	1