

# Journal of Visualized Experiments

## Quantification of the potential impact of glyphosate-based products on microbiomes --Manuscript Draft--

<b>Article Type:</b>	Methods Article - JoVE Produced Video
<b>Manuscript Number:</b>	JoVE63109R2
<b>Full Title:</b>	Quantification of the potential impact of glyphosate-based products on microbiomes
<b>Corresponding Author:</b>	Pere Puigbo University of Turku: Turun Yliopisto Turku, South-west Finland FINLAND
<b>Corresponding Author's Institution:</b>	University of Turku: Turun Yliopisto
<b>Corresponding Author E-Mail:</b>	pepuav@utu.fi
<b>Order of Authors:</b>	Suni Anie Mathew Anne Muola Kari Saikkonen Irma Saloniemi Marjo Helander Pere Puigbo
<b>Additional Information:</b>	
<b>Question</b>	<b>Response</b>
Please specify the section of the submitted manuscript.	Environment
Please indicate whether this article will be Standard Access or Open Access.	Open Access (\$3900)
Please indicate the <b>city, state/province, and country</b> where this article will be <b>filmed</b> . Please do not use abbreviations.	Turku, Finland
Please confirm that you have read and agree to the terms and conditions of the author license agreement that applies below:	I agree to the <a href="#">Author License Agreement</a>
Please confirm that you have read and agree to the terms and conditions of the video release that applies below:	I agree to the <a href="#">Video Release</a>
Please provide any comments to the journal here.	* Pere Puigbo and Marjo Helander are co-corresponding authors

**TITLE:**

Quantification of the Potential impact of Glyphosate-Based Products on Microbiomes

**AUTHORS AND AFFILIATIONS:**

Suni Anie Mathew<sup>1</sup>, Anne Muola<sup>1</sup>, Kari Saikkonen<sup>1</sup>, Irma Saloniemi<sup>2</sup>, Marjo Helander<sup>2,\*</sup>, Pere Puigbò<sup>2,3,4,\*</sup>

<sup>1</sup>Biodiversity Unit, University of Turku FI-20014 Turku, Finland

<sup>2</sup>Department of Biology, University of Turku FI-20014 Turku, Finland

<sup>3</sup>Nutrition and Health Unit, Eurecat Technology Centre of Catalonia, Reus, Catalonia

<sup>4</sup>Department of Biochemistry and Biotechnology, Rovira i Virgili University, Tarragona, Catalonia

Email addresses of co-authors:

Suni Anie Mathew (suni.mathew@utu.fi)

Anne Muola (anne.muola@utu.fi)

Kari Saikkonen (karisaik@utu.fi)

Irma Saloniemi (irma.saloniemi@utu.fi)

Marjo Helander (helander@utu.fi)

Pere Puigbò (pepuav@utu.fi)

\*Corresponding authors:

Marjo Helander (helander@utu.fi)

Pere Puigbò (pepuav@utu.fi)

**SUMMARY:**

Glyphosate-based products (GBP) are the most common broad-spectrum herbicides worldwide. In this article, we introduce general guidelines to quantify the effect of GBP on microbiomes, from field experiments to bioinformatics analyses.

**ABSTRACT:**

Glyphosate-based products (GBP) are the most common broad-spectrum herbicides worldwide. The target of glyphosate is the enzyme 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) in the shikimate pathway, which is virtually universal in plants. The inhibition of the enzyme stops the production of three essential amino acids: phenylalanine, tyrosine, and tryptophan. EPSPS is also present in fungi and prokaryotes, such as archaea and bacteria; thus, the use of GBP may have an impact on the microbiome composition of soils, plants, herbivores, and secondary consumers. This article aims to present general guidelines to assess the effect of GBP on microbiomes from field experiments to bioinformatics analyses and provide a few testable hypotheses. Two field experiments were presented to test the GBP on non-target organisms. First, plant-associated microbes from 10 replicated control and GBP treatment plots simulating no-till cropping are sampled and analyzed. In the second experiment, samples from experimental plots fertilized by either poultry manure containing glyphosate residue or non-treated control manure were obtained. Bioinformatics analysis of EPSPS protein sequences is utilized to determine the potential sensitivity of microbes to glyphosate. The first step in

estimating the effect of GBP on microbiomes is to determine their potential sensitivity to the target enzyme (EPSPS). Microbial sequences can be obtained either from public repositories or by means of PCR amplification. However, in the majority of field studies, microbiome composition has been determined based on universal DNA markers such as the 16S rRNA and the internal transcribed spacer (ITS). In these cases, sensitivity to glyphosate can only be estimated through a probabilistic analysis of EPSPS sequences using closely related species. The quantification of the potential sensitivity of organisms to glyphosate, based on the EPSPS enzyme, provides a robust approach for further experiments to study target and non-target resistant mechanisms.

## INTRODUCTION:

The heavy use of pesticides in modern agriculture is clearly a major contributor to the decline of biodiversity<sup>1</sup>. This paper focuses on glyphosate because glyphosate-based products (GBPs) have become the most widely used pesticides globally due to their efficiency and affordable price<sup>2,3</sup>. In addition to killing weeds in agricultural fields, GBPs are commonly used in silviculture, urban environments, and home gardens; additionally, they have been proclaimed as nontoxic to non-target organisms if used in accordance with the manufacturer's instructions. However, an increasing number of recent studies have revealed that residues of glyphosate and its degradation products may be retained and transported in soils, thereby having cascading effects on non-target organisms<sup>4-8</sup>. The effects of glyphosate are not limited only to plants—the shikimate pathway is present in many fungi and prokaryotes as well. Glyphosate targets the enzyme 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) in the shikimate pathway, also known as *aroA*<sup>9</sup>. This enzyme is at the center of the shikimate pathway in the synthesis of the three essential aromatic amino acids (phenylalanine, tyrosine, and tryptophan), and it is present in most prokaryotes, plants, and fungi<sup>10,11</sup>. Some microbial species have developed partial or absolute resistance to glyphosate by means of several mechanisms, including mutations in the EPSPS sequences. Thus, it has been suggested that the use of GBPs may have a direct effect on plant and animal microbiomes, including the human gut microbiome<sup>12-14</sup>. Nevertheless, the use of GBP may have an adverse impact on virtually any ecosystem function and service relying on microbes and microbe-facilitated processes. The consequent threats may concern biochemical soil processes, pollination biology, and animal and human wellbeing. This calls for a more comprehensive understanding of how glyphosate affects shikimate pathways and methods to assess the sensitivity of microbes to glyphosate.

In this protocol, we present a pipeline to test the effect of glyphosate and GBP on the microbiome, from field experiments to bioinformatics analyses. We describe in detail a recently published bioinformatics method that can be used to determine the potential sensitivity of organisms to glyphosate<sup>12</sup>. To the researchers' knowledge, this is the first and so far, the only bioinformatics tool to assess the intrinsic sensitivity of the enzyme EPSPS to the active component of GBPs. This bioinformatics method is based on the detection of known amino acid markers in the glyphosate target enzyme (EPSPS)<sup>12</sup>. The pipeline is divided into five main working phases (**Figure 1**): 1) a brief introduction to two field experiments to test the effect of GBPs, 2) a brief summary of microbiome analyses (16S rRNA, ITS, and *EPSPS* gene), 3) gathering EPSPS sequences from public repositories, 4) determining the potential sensitivity of organisms

to glyphosate, and 5) assessing the EPSPS class from universal microbial markers (16S rRNA and ITS).

## PROTOCOL:

### 1. Two field experiments to test the effect of GBPs

NOTE: This protocol presents two examples of field experimental designs to test the effect of GBPs on plant-associated microbes. Both experiments were conducted in set-aside fields with no previous history of herbicide or agricultural uses at the University of Turku Ruissalo Botanical Garden in Finland (60°26'N, 22°10'E). The soil is sandy clay with a high proportion of organic matter.

#### 1.1. Experiment 1

NOTE: This experiment was designed to simulate the general agricultural practices of no-till agriculture with GBP applications before and after the growing season to combat weeds.

1.1.1. Divide the experimental field into 10 replicated control and GBP treatment plots (23 m x 1.5 m) with buffer strips of vegetation between the plots (in this study in spring 2014<sup>15</sup>) (Figure 2).

1.1.2. Ensure that the plots are tilled with a rotary tiller to a depth of 5 cm and treated twice a year. Here, the plots were treated in the beginning (May) and the end (October) of the growing season.

1.1.3. Treat the control plots with tap water (5 L/plot) and the GBP plots with commercial GBP (glyphosate concentration 450 g·L<sup>-1</sup>, application rate 6.4 L·ha<sup>-1</sup> in 5 L of tap water per plot) to mimic the maximal permitted glyphosate dosage in agricultural practices (3 kg·ha<sup>-1</sup>).

1.1.4. Apply the treatments with a hand-operated pressure tank that has a manual sprayer. Two weeks after the GBP application, sow oats (*Avena sativa*), faba beans (*Vicia faba*), turnip rapes (*Brassica rapa* subsp. *oleifera*), and plant potatoes (*Solanum tuberosum*) in the plots according to agricultural practices.

1.1.5. During the growing season, hand-weed the plots to keep the plant competition and soil structure as similar as possible in the control and GBP-treated plots.

1.1.6. Sample the microbiota from experimental plants. In this study, the microbiota were sampled consecutively from 2017 through 2020 in both the GBP-treated and control treatment plots once per growing season over the course of the study.

1.1.7. Collect ten replicates of the plant samples (root and leaf) from the field, immediately place them on ice, and bring them to the laboratory for further processing, as described in

section 2.1.

## 1.2. Experiment 2

NOTE: This experiment was designed to test risks associated with the circular food economy; more precisely, it was designed to examine the consequences of GBP residues in manure applied as fertilizer to crop plants<sup>2</sup> (**Figure 3**).

1.2.1. Collect beddings, including wood shavings, feces, and some spilled feed, from quails fed on GBP-contaminated or control feeds in a 12-month aviary experiment<sup>16,17</sup>.

NOTE: The GBP-contaminated feed consisted of organic feed for laying chickens combined with an equivalent of 160 mg glyphosate/kg, that corresponds to a daily intake of 12–20 mg glyphosate per kilogram body mass in adult Japanese quails<sup>16,17</sup>.

1.2.2. For validation, send the samples to an accredited laboratory for glyphosate concentration measurements in six batches of feed.

1.2.3. In addition, measure the glyphosate residues in quail excreta samples after 12 months of exposure. The control group was fed the same organic feed with no GBP addition<sup>16,17</sup>.

1.2.4. During the aviary experiment, change the beddings bi-weekly. Collect the used beddings regularly from 8–12 months of exposure from GBP treatment and controls, pool per treatment, and store in closed containers in a dry, dark storage room at 6 °C before using as fertilizer.

1.2.5. Spread 12 L of the beddings manually on each of the 18 GBP and 18 control plots (size 1 m x 1 m) in a 6 x 6 chessboard grid in the experimental field at two time points. In this study, the beddings were spread in August 2018 and in May 2019.

1.2.6. Send the bedding samples to an accredited laboratory for the glyphosate concentration measurements directly after it is spread (in this study in May 2019).

1.2.7. Plant perennial grass and strawberry plants in each plot to study their root and leaf microbiota.

NOTE: In this study four perennial grass plants (*Festuca pratensis*), and two strawberry plants (*Fragaria x vesca*) were planted to each plot and studied for their root and leaf microbiota.

## 2. Microbiome analyses (16S rRNA, ITS and EPSPS gene)

NOTE: Most of the microbiome studies are based on the analysis of the 16S rRNA gene for bacterial and internal transcribed spacer (ITS) regions for fungal communities using next-generation sequencing technologies. Thus, the paper does not have information on the type of EPSPS. EPSPS sequences from thousands of species are available in public repositories (Protocol

section 3) (**Figure 4**).

## 2.1. 16S rRNA gene

2.1.1. From the detached leaf and root samples collected in the experiments described above, identify the endophytic microbes (i.e., microbes living inside plant tissues).

2.1.2. Wash the plant samples with tap water and then sterilize them to remove the epiphytic microbes (i.e., microbes on the surface of plant tissues). Sterilize using 3% bleach for 3 min, followed by a 70% ethanol solution for 1 min, and wash three times with autoclaved ultrapure water for 1 min each.

2.1.3. Freeze the samples at -80 °C until genomic DNA extraction.

2.1.4. Perform genomic DNA extraction using a commercially available plant DNA extraction kit following the manufacturer's protocol.

2.1.5. Target the variable regions V6-V8 of the 16S rRNA gene from extracted DNA samples using a nested approach with discriminating primers that bind specifically to bacterial DNA<sup>18</sup>, thus minimizing the amplification from host plant DNA.

2.1.6. After three rounds of polymerase chain reaction (PCR), tag the target gene with barcode and adapter sequences to prepare it as the template for sequencing. Follow steps 2.1.7– 2.1.11 for PCR amplification

2.1.7. Prepare the PCR master mix for the required number of samples so that each reaction has 30 µL of total volume and consists of 30 ng of DNA, 1x PCR buffer, 0.2 mM dNTPs, 0.3 µM of each primer, and 2000 U/mL DNA polymerase. Follow the same step for the second and third rounds of PCR.

2.1.8. For the first round of PCR, use primers 799F<sup>19</sup> and 1492R (modified from<sup>20</sup>) (**Table 1**). Set up the amplification profile on the thermocycler (3-min initial denaturation at 95 °C, followed by 35 cycles of denaturing at 95 °C for 45 s, annealing at 54 °C for 45 s, and extension at 72 °C for 1 min). Carry out the final extension at 72 °C for 5 min.

2.1.9. As the template for the second round of PCR, verify amplification by electrophoresis (5 µL of the PCR products on 1.5% agarose gel) and then dilute the rest 25 µL of the PCR product in autoclaved ultrapure water in the ratio of 1:10.

2.1.10. Repeat PCR with the diluted PCR templates and primers uni-1062F<sup>21</sup> and uni-1390R<sup>22</sup> (**Table 1**). Maintain the same PCR reaction conditions and amplification profile (step 2.1.8) except reducing the number of cycles to 25.

2.1.11. Dilute the resulting PCR products in autoclaved ultrapure water in the ratio of 1:1. Carry

out the third round of PCR to tag the products with the barcodes and P1-adapter sequence with 8 cycles of the same PCR profile as mentioned in step 2.1.8.

## 2.2. Library preparation

2.2.1. Verify the concentration and quality of PCR products on a bioanalyzer and pool the volumes constituting 30 ng of DNA of each sample in a 1.5 mL tube to prepare an equimolar library.

2.2.2. Select the amplicons of size 350–550 bp by size fractionation using an automated DNA size selection system on an agarose gel cassette. Note that this also eliminates non-specific amplicons and PCR reagents from the library. Collect the elute consisting of amplicons of the specified size into a vial in the cassette, resulting in a purified 16S rRNA gene library.

2.2.3. Pipette out the elute into a 1.5 mL tube and verify the purity and concentration on the bioanalyzer. Dilute the DNA library using autoclaved ultrapure water to a final concentration of 26 pM; the sample is ready for sequencing.

## 2.3. ITS

NOTE: The *ITS* region is amplified using *ITS*-specific primers (**Table 1**), and the resulting PCR product is labeled with barcodes and a P1-adapter sequence for sequencing.

2.3.1. Prepare the PCR master mix according to the same protocol as mentioned in section 2.1 with *ITS* primers.

2.3.2. Set the amplification profile on the thermocycler as 5-min initial denaturation at 95 °C followed by 35 cycles of denaturing, annealing, and extension at 95 °C for 30 s, 55 °C for 30 s, 72 °C for 1 min, respectively, and final extension 72 °C for 7 min.

2.3.3. Analyze 5 µL of the PCR product on 1.5 % agarose gel and dilute the remaining 25 µL to 1:10 using autoclaved ultrapure water. Use the diluted PCR product as the template for the second round of PCR.

2.3.4. Prepare PCR master mix for the required number of samples (refer to step 2.1.6) with barcode-tagged forward primers and P1-adapter tagged reverse primers. Amplify with the same amplification profile as in step 2.3.2, except using 8 cycles.

2.3.5. Prepare the resulting PCR products for sequencing as per the protocols mentioned in section 2.2.

## 2.4. EPSPS gene

2.4.1. Sequence and analyze the *EPSPS* genes of the microbes.

NOTE: In the interest of finding whether GBP exposure changed the composition of glyphosate-sensitive and resistant microbes in the community, the *EPSPS* genes of the microbes need to be sequenced and analyzed. Thus, 353 sequences of *EPSPS* genes from a wide collection of microbial taxa in the Alignable Tight Genomic Clusters (ATGC) database were gathered, and all protein sequences were aligned<sup>22</sup>. These alignments are available at ATGC database<sup>23</sup> and can be utilized to generate primers from conserved regions. An easy-to-use bioinformatics tool is designed to identify conserved regions from a multiple sequence alignment, and this is available at Pere Puigbo research page<sup>24</sup>. However, it is out of the scope of this publication to provide a detailed description of this web server. Nevertheless, a prospective protocol to utilize these primers for amplification of the *EPSPS* gene to find the sensitivity of the microbiome to glyphosate is provided in **Figure 4**.

### **3. Gathering EPSPS protein sequences from public repositories**

#### **3.1. EPSPS sequences to be used in macroevolution studies**

3.1.1. Gather EPSPS proteins from public repositories such as PFAM<sup>23</sup> (a database of protein families<sup>25</sup>), GenBank<sup>24</sup> (a database of genes, genomes and proteins<sup>26</sup>), COG<sup>25</sup> (Clusters of Orthologous Groups<sup>27</sup>; a database of orthologous proteins from archaea and bacteria); and PDB<sup>26</sup> (Protein Data Bank<sup>28</sup>; a database of protein structures).

NOTE: A recent study conducted by the researchers showed that these proteins could be utilized to perform microevolutionary and comparative analyses of the potential effect of glyphosate on organisms having the shikimate pathway<sup>12</sup>. The authors have developed a user-friendly website that gathers information on tens of thousands of EPSPS protein sequences<sup>29</sup>, including a manually curated dataset of proteins from the human gut microbiome<sup>12</sup>. The information in these pre-computed datasets includes the current EPSPS classification into putative sensitive and resistant to glyphosate, taxonomic information on species, annotations of the EPSPS active site, and links to PDB and NCBI databases. Moreover, the webserver includes the ID codes of the EPSPS and links to several external databases (**Table 2**).

#### **3.2. EPSPS sequences to be used in microevolution studies ATGC**

NOTE: General repositories of protein sequences are useful for conducting comparative studies among relatively distant organisms; however, the potential effect of glyphosate is relatively recent from an evolutionary point of view. Thus, in some studies, it is necessary to compare closely related species (e.g., different strains from the same bacterial species) to determine the effect of glyphosate<sup>14</sup>. In these cases, the database of Alignable Tight Genomic Clusters (ATGC)<sup>30</sup>, which contains a comprehensive list of closely related archaeal and bacterial genomes, is a more suited resource. The ATGC database contains information of several million proteins from thousands of genomes organized into hundreds of clusters<sup>30</sup>. Each genome cluster is alignable (genomes share synteny over ≥85% of their lengths) and tight (having a synonymous substitution rate below saturation). The researchers used the ATGC dataset in a



recent study to analyze microevolutionary changes in EPSPS proteins<sup>14</sup>. The following steps are needed to identify EPSPS protein sequence in the ATGC:

3.2.1. Download the whole database of ATGCs from the link<sup>31</sup> and all proteins of the COG0128 (code corresponding to EPSPS proteins in the database)<sup>32</sup> into a local project.

NOTE: If the researchers/experimenters are based in Finland, the CSC–IT Center for Science<sup>33</sup> provides storage and software facilities. It is important to gather all sequences in FASTA format.

3.2.2. Built a blast database of the COG0128 that contain orthologs of the EPSPS protein in a representative set of species of prokaryotes. The CSC has the blast program<sup>34</sup> preinstalled, allowing the use of the command ***makeblastdb -in COG0128.fa -dbtype prot*** to create a reference database of EPSPS sequences.

3.2.3. Map the ATGC database onto the COG0128.fa (EPSPS proteins) using an iterative blast search with the command ***blastp -query [ATGC\_X.fa] -db [COG0128.fa] -max\_target\_seqs 1 -outfmt 6 -out tmpfile -evalue 1e-150***.

3.2.4. As a result, it creates a dataset of EPSPS protein sequences within each. A pre-computed dataset of closely related EPSPS protein sequences from the ATGC database is available<sup>29</sup>.

#### **4. Algorithm to determine the potential sensitivity of organisms to glyphosate (EPSPClass web server: inputs, processing, and outputs)**

NOTE: The researchers have implemented an easy-to-use server that is freely available at <sup>29</sup> to determine the class of EPSPS protein sequences<sup>12,35</sup>. The server only requires an input of protein sequence in FASTA format to determine the identity percentage to each of the EPSPS classes and their potential sensitivity to glyphosate. Moreover, users can utilize the webserver to test their own reference sequences and amino acid markers. First, the algorithm (**Figure 5**) aligns query sequences and reference sequences using a multiple sequence alignment program<sup>35</sup> to determine amino acid positions. Then, it searches for the presence of amino acid markers to identify the EPSPS class (I, II, III, or IV) of the query sequence.

4.1. Introduce an EPSPS protein sequence in FASTA format into the input text box to identify the class of the enzyme (**Figure 6A**), and press **Send**.

4.2. Assess the potential sensitivity of the query sequence to glyphosate (**Figure 6B–E**) from the server provided outputs:

Output 1: Fraction of amino acid markers (i.e., identity) present in the query sequences (class I, II, and IV), and the number of motifs (class III).

Output 2: Alignments of the query and reference sequences based on marker residues.

Output 3: Full pairwise alignments of the query and reference sequences.

Output 4: EPSPS reference sequences: *Vibrio cholerae* (vcEPSPS, class I), *Coxiella burnetii* (cbEPSPS, class II), *Brevundimonas vesicularis* (bvEPSPS, class III), *Streptomyces davawensis*

(sdEPSPS, class IV).

4.3. At the end of the output page, find links to external tools such as blastp and conserved domains to further analyze the query EPSPS sequence (Figure 6F).

## 5. Assessing the EPSPS class from universal microbial markers (16S rRNA and ITS)

NOTE: Most microbiome studies are based on the analysis of the 16S rRNA and/or ITS<sup>36</sup>. In such cases, it is not possible to perform a direct analysis of the EPSPS sequence. Thus, a probabilistic approach to estimate the potential sensitivity of organisms to glyphosate is necessary. This analysis is straightforward and provides a reasonable estimate of the type of EPSPS sequences in a microbiome project. The process is divided into 3 steps (Figure 7 and Figure 8):

5.1. Identify EPSPS sequences from public repositories. The EPSPS class of a comprehensive dataset of representative sequences has been compiled and pre-computed from PFAM<sup>37</sup>, GenBank<sup>38</sup>, COG<sup>39</sup>, PDB<sup>40</sup>, ATGC<sup>30</sup>. Access these datasets from the EPSPSClass server's main page, containing taxonomical information and the EPSPS class of over 50,000 sequences (Figure 7).

5.2. Measure the height of the experimental plants biweekly during the growing season, and weigh the aboveground biomass of the plants at the end of the field season to compare the growth of the plants in GBP and control plots.

NOTE: The microbiota analyses from the field experiments have not yet been fully analyzed.

5.3. Use a spreadsheet to map the bacterial OTUs (based on 16S rRNA or ITS) from microbiome experiments into pre-computed datasets.

NOTE: Previous studies have shown that the EPSPS class (i.e., the intrinsic sensitivity to glyphosate) is highly conserved within a phylogenetic group<sup>14</sup>. Thus, it is relatively safe to assume that closely related species from highly conserved taxons may have similar EPSPS responses to glyphosate (Figure 8).

5.4. In the same spreadsheet, calculate the intrinsic sensitivity to glyphosate, based on a probabilistic score ( $S = s / (s + r + u)$  where S: Sensitivity Score; s: number of potentially sensitive sequences; r: number of potentially resistant sequences; u: number of unclassified sequences) calculated from known EPSPS sequences in public databases.

NOTE: This score ranges from 0 (no sensitive EPSPS sequences are found in a taxon) to 1 (all sequences in a taxon are sensitive to glyphosate) (Figure 8). Moreover, there are in-between values—i.e., species with sensitive, resistant or unknown strains.

## REPRESENTATIVE RESULTS:

The aim of this protocol is to provide a general pipeline, from field experiments to

bioinformatics analyses, that quantifies the potential sensitivity of organisms to the herbicide glyphosate. In the Experiment 2 the average glyphosate concentration in the quail feed was 164 mg/kg and the average glyphosate concentration of the excreta samples (urine and fecal matter combined) was 199 mg/kg. Beddings collected from quails fed with GBP-contaminated feed had, on average, 158 mg/kg and control beddings measuring 0.17 mg/kg of glyphosate. In the field experiments, plant species differed in their sensitivity to glyphosate residues in the soils (section 1). Biomass of oat and turnip rape was greater in control soils compared to GBP-treated soils. However, faba beans and potato appeared to benefit from GBP treatment at the end of the growing season<sup>15</sup>. Glyphosate in poultry manure decreased plant growth in grass (*Festuca pratensis*) and strawberry (*Fragaria x vesca*) (section 1). The microbiota analyses from the field experiments have not yet been fully analyzed and are not presented here (section 2). The results of this protocol, when read either directly (as shown in sections 3 and 4) or indirectly (section 5), provide a measure of the proportion of potentially sensitive and resistant organisms to glyphosate in a dataset (**Figure 9**). The use of this method was tested with a collection of EPSPS protein sequences from microbial species of the core human gut microbiome that were obtained from public repositories<sup>12</sup>. In the study, 890 strains from the 101 most abundant bacterial species were analyzed with the EPSPSClass method to quantify the proportion of sensitive and resistant bacteria. The results showed that 54% of the species in the core human gut microbiome are potentially sensitive to glyphosate<sup>12</sup>. This trend is also observed in most of the prokaryotic world; additionally, in eukaryotes (mainly plants and fungi), the proportion of potentially sensitive species is even higher<sup>12</sup>. Moreover, we have utilized this method to quantify changes in sensitivity in the EPSPS protein at a microevolutionary level<sup>14</sup>. We identified changes in sensitivity status in 12 out of 32 closely related groups of prokaryotes analyzed<sup>14</sup>. Thus, the continuous use of the GBPs may produce microbial dysbiosis (i.e., an imbalance of sensitive and resistant bacterial species) in plant, animal, and soil microbiomes. Moreover, it has been hypothesized that an increase in glyphosate-resistant bacteria may promote multidrug-resistant microbiomes<sup>14,41,42</sup>. Thus, this protocol sheds light on the interpretation of all these scenarios, as the EPSPS classification method provides a direct estimate of the intrinsic sensitivity of microbiomes to glyphosate.

#### FIGURE AND TABLE LEGENDS:

**Figure 1: General pipeline** This is a general pipeline to analyze sensitivity to GBP from field experiments to bioinformatics analysis.

**Figure 2: Field experiment 1 to test the effects of GBP residues on crop plant-associated microbes.** The experimental field consists of alternating 10 control plots and 10 GBP treatment plots (23 m x 1.5 m) with 1.5 m buffer strips between plots. Two times a year since 2014, the GBP plots were treated with commercial GBP (glyphosate concentration 450 g L<sup>-1</sup>, application rate 6.4 L ha<sup>-1</sup> in 5 L of tap water per plot) and the control plots with the same amount of tap water without glyphosate. The treatments were applied with a hand-operated pressure tank using a plastic hood in the sprinkler tip to protect GBPs from spreading outside the treatment plots. After a two-week safety period following the GBP application, oats (*Avena sativa*), faba beans (*Vicia faba*), and turnip rapes (*Brassica rapa* subsp. *oleifera*) were sown, and potatoes (*Solanum tuberosum*) were planted in the plots. Microbiota samples from the studied crop

plants, leaves and roots, were collected several times since the start of the experiment in 2014.

**Figure 3: Field experiment 2 tested the consequences of GBP residues in manure fertilizer for two perennial crops and their associated microbiota.** Beddings collected from a 12-month aviary experiment with Japanese quails fed with control or GBP-contaminated feed were used as manure fertilizer in a field experiment. The experimental field consisted of 18 control and 18 GBP plots (1 m x 1 m) arranged in a 6 x 6 chessboard grid. The beddings were spread on the experimental field twice, in August 2018 and May 2019 (25 L / plot). Control plots were fertilized with beddings collected from quails fed with control feed and GBP plots with beddings from quails fed with GBP-contaminated feed. Glyphosate residues in control beddings were 0.17 mg/kg of glyphosate and in GBP-bedding, the amount was 158 mg/kg of glyphosate. Two endophyte-symbiotic (E+), two endophyte-free (E-) *Festuca pratensis*, and two *Fragaria x vesca* were planted per plot in September 2018, approximately one month after the spread of the first beddings. Measurements of plant performance and fitness as well as sampling for root-and leaf-associated microbiota were conducted during two consecutive growing seasons (2019 & 2020).

**Figure 4: Analysis of the microbial taxa using 16S rRNA/ITS and sensitivity of microbiomes to glyphosate using the EPSPS gene.** (A) Analysis of 16S rRNA or ITS sequences to identify microbial taxa. (B) Analysis of EPSPS sequences to identify sensitivity of microbes to glyphosate (GS-glyphosate sensitive/GR-glyphosate resistant)

**Figure 5: Algorithm to identify the class of EPSPS protein sequences.** The input is an EPSPS protein sequence in FASTA format. The algorithm performs comparisons with known amino acid markers in reference protein sequences that determine the potential sensitivity to glyphosate. The algorithm was implemented at the freely accessible web server EPSPSClass<sup>29</sup>.

**Figure 6: Basic inputs and outputs of the EPSPSClass web server.** (A) Input: an EPSPS protein sequence in FASTA format. (B) Output 1 - identity: fraction of amino acid markers present in the query sequences (Classes I—IV) and motifs (Class III). (C) Output 2 - identity: alignments of the query and reference sequences. (D) Output 3 - pairwise alignments of the query and reference sequences. (E) Reference EPSPS sequences: *Vibrio cholerae* (vcEPSPS, class I), *Coxiella burnetii* (cbEPSPS, class II), *Brevundimonas vesicularis* (bvEPSPS, class III), *Streptomyces davawensis* (sdEPSPS, class IV). (F) Links to perform additional blastp searches and identification of conserved domains

**Figure 7: Access to pre-computed datasets of EPSPS sequences.** Follow the indications in the figure to access the pre-computed dataset of EPSPS sequences.

**Figure 8: Example of how to estimate the potential sensitivity in microbiome projects without EPSPS sequences.** The example uses values from the database of Alignable Tight Genomic Clusters<sup>30</sup>, which contains sequences from prokaryotic species. Hypothetical species from a microbiome project are *Staphylococcus aureus*, *Corynebacterium diphtheriae*, *Campylobacter jejuni*, *Chlamydia psittaci* and *Sulfolobus islandicus*. The sensitivity score to glyphosate is

calculated as Number\_Sensitive\_Sequences / Total\_Number\_Of\_Sequences.

**Figure 9: Scheme of the interpretation of the results from this protocol and hypothetical evolutionary scenarios.** (A) In a microbiome, the proportion of potential sensitivity (in green) and resistance (in red) bacteria is approximately 50:50. Black dots denote microbial species unclassified; thus, their sensitivity to glyphosate is unknown. In some microbiomes, the proportion of sensitive bacteria is slightly higher, as in the human gut microbiome<sup>12</sup>. (B) Over time, the use of glyphosate may lead to microbial dysbiosis (i.e., an imbalance in the proportion of sensitive and resistant bacteria) leading to different hypothetical scenarios. (C) Hypothetical case 1 (no selection): The use of glyphosate does not influence the microbiome; thus, the proportion of sensitive and resistant bacteria remains constant. (D) Hypothetical case 2: The use of glyphosate removes bacteria sensitive to glyphosate from the population. We speculate that this scenario may be dose-dependent. (E) Hypothetical case 3: Selection pressure from the use of glyphosate enhances mutations in the *EPSPS* gene that change the sensitivity status of bacteria. Thus, the entire microbial population becomes resistant to glyphosate. Moreover, in this scenario, there might be an increase in multidrug-resistant bacteria. (F) Hypothetical case 4: the use of glyphosate alters the composition of certain bacterial species, producing an imbalance towards resistant bacteria, whereas some bacterial species remain unaltered, possibly due to additional resistant mechanisms such as efflux pumps or by overexpression of the *EPSPS* gene<sup>13</sup>. This scenario may also lead to an increase in glyphosate-resistant bacteria, as well as an increase in bacterial resistance to additional antibiotics.

**Figure 10: Inputs and outputs of the EPSPSClass webserver to test user's own reference sequence.** (A) Input 1: query sequence. (B) Input 2: reference sequence. (C) Input 3: amino acid markers in the reference sequences. (D) Output: identity: fraction of amino acid markers in the query sequences (class I—IV and user's own reference sequences).

**Table 1: List of primers for PCR amplification of 16S rRNA and ITS in microbiome analysis**

**Table 2: Codes of the enzyme 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) in different databases**

**Table 3: Positions of the amino acid markers in the reference sequences**

## DISCUSSION:

This protocol provides general guidance on how to quantify the effect of GBP on microbiomes based on the analysis of the EPSPS protein. The protocol has three major critical steps: (i) Quantification of the EPSPS protein from microbiome data. This step is critical because EPSPS is the direct target enzyme of the herbicide. Thus, species that have a copy of the *EPSPS* gene may be impacted by the use of GBP. Nevertheless, even species that lack a copy of the *EPSPS* gene may be impacted by the herbicide through alternative non-target mechanisms<sup>43,44</sup>. (ii) If the analysis of the *EPSPS* gene is not included in the design of the study, it is possible to get a good estimate by analyzing the 16S rRNA (bacteria) or *ITS* (fungi). In this case, it is essential to rely on a comprehensive reference table (e.g., the ATGC database provides sequences of the EPSPS

protein from several closely related species). (iii) The EPSPS protein is divided into potentially sensitive or resistant to glyphosate depending on certain amino acid residues of the active site of the EPSPS. However, mutations affecting a single amino acid may alter this classification<sup>45</sup> and transitions among classes may occur in a relatively short period of time<sup>14</sup>.

The potential sensitivity of organisms to glyphosate can be determined by reference genomes, amino acid markers and sequence alignments. (i) Reference genomes: The EPSPS enzyme can be classified as potentially sensitive (class I [alpha or beta]<sup>46,47</sup>) or resistant (classes II<sup>48,49</sup>, III<sup>50</sup> and IV<sup>51</sup>) to glyphosate based on the presence of amino acid markers and motifs (in the case of class III). These amino acid markers and motifs are based on the location of amino acid residues in the EPSPS protein of *Vibrio cholerae* (vcEPSPS, class I), *Coxiella burnetii* (cbEPSPS, class II), *Brevundimonas vesicularis* (bvEPSPS, class III), and *Streptomyces davawensis* (sdEPSPS, class IV). (ii) Amino acid markers: Glyphosate interacts with the EPSPS enzyme and competes with phosphoenolpyruvate (PEP, the second substrate of the EPSPS enzyme)<sup>52,53</sup>. In certain species, small amino acid changes in the EPSPS sequence provide a higher affinity for the PEP and a resistance to glyphosate<sup>12, 14, 52, 54, 55</sup>. In other sequences, glyphosate binds the EPSPS sequence in a non-inhibitory conformation<sup>45</sup>. Although many resistant<sup>12,14,48,49,52,54,55</sup> and tolerant<sup>56,57</sup> EPSP sequences to glyphosate have been described, the current classification system for the EPSPS is divided into four major classes (I—IV)<sup>12</sup> (**Table 3**). (iii) Sequence alignments: In order to classify an EPSPS enzyme, we performed pairwise alignments, with a multiple sequence alignment program—default parameters<sup>35</sup>—, of the query sequence against each one of the reference sequences (vcEPSPS, cbEPSPS, bvEPSPS and sdEPSPS). These alignments are necessary to identify the positions of the amino acid markers in the query sequence. As a result, an enzyme is classified as described<sup>12</sup>—class I, II and/or IV based on the presence of amino acid markers and class III based motif markers.

The protocol is based on four known types of EPSPS: one type is sensitive, the other three are resistant). However, approximately 10% of EPSPS sequences in prokaryotes are yet unclassified (16% in archaea and 8% in bacteria)<sup>12</sup>. Thus, further research should analyze those sequences to determine glyphosate sensitivity. The EPSPSClass server provides an option to test new genetic markers. The identification of known classes of the EPSPS is straightforward, as shown in section 4.4. and **Figure 5**. Furthermore, in those cases where users want to compare their own query and reference proteins, the server provides an option to manually include a reference sequence and a set of amino acid markers (**Figure 10**). This option can be utilized to identify novel classes of the EPSPS, as well as to test other herbicides and target sequences.

The analysis of the EPSPS class is determined by sequence analysis and the presence/absence of amino acid markers. This is a preliminary estimate that can be used for hypothesis testing in the field. Amino acid markers have been determined in the literature based on empirical and observational studies<sup>46–51</sup>. However, reference protein sequences to determine EPSPS class have been tested only in a limited number of species and may occasionally fail to explain resistance to glyphosate. The effect of compensatory mutations, and EPSPS-associated domains (mostly in fungi) may also affect the sensitivity to glyphosate<sup>58</sup>. This paper's analysis is based on four EPSPS classes. A survey of bacteria in the human gut microbiome showed that around 30%

of them were unclassified (i.e., EPSPS proteins from these species do not belong to any of the known classes), and additional studies are needed to identify other EPSPS classes. Also, it should be noticed that the EPSPS protein sequence in bacteria and plants is unidomain, whereas fungal EPSPS proteins contain several domains<sup>59</sup>. Thus, a protein folding in fungi may lead to a different response of the EPSPS enzyme to glyphosate. Moreover, additional non-target mechanisms of resistance (e.g., efflux pumps and overexpression of the *EPSPS* gene<sup>13</sup>) or sensitivity to glyphosate (e.g., the effect of glyphosate on the mitochondrial transport chain<sup>12</sup>) are not considered.

Although GBPs have been around as a herbicide since 1974 and have been widely utilized since 1991, this is the first bioinformatics method to determine the potential sensitivity of organisms to glyphosate. The method is based on the identification of known amino acid residues in the target sequence. Thus, our method provides a baseline estimate of the potential effect of glyphosate on the species. In the near future, novel bioinformatics methods should include additional classes of the EPSPS protein to determine the potential sensitivity to glyphosate of unclassified sequences<sup>12,54,55</sup>. In addition, given that the exact behavior of the EPSPS enzyme may vary by single amino acid changes<sup>12,14,52,54,55</sup>, further in silico experiments should take into account small variations in the folding of the EPSPS protein, as well as the effect of the EPSPS-associated domains on the protein structure in fungi<sup>58</sup>. Moreover, it has been shown that tolerance to glyphosate may be produced by overexpression of the EPSPS protein<sup>56,57</sup>; thus bioinformatics analyses based on the *amelioration* of the codon usage<sup>60</sup> may be utilized to identify novel *EPSPS* sequences that maximize or minimize gene expression.

Farmers, politicians, and decision-makers urgently need a thorough understanding of the risks associated with the heavy use of pesticides. Thus, both bioinformatic tools revealing the potential sensitivity of organisms to pesticides and well-replicated, randomized, and field-realistic experimental studies conducted in different environments are necessary. The presented bioinformatic method designed to examine organisms' sensitivity to glyphosate can be modulated for other pesticides. Similarly, the methods of experimental ecology can be applied to study any related ecological questions. Together, the methods can be used to demonstrate casualties between field observations, genomic data, and pesticide use. All presented methods are invaluable in risk assessment. Bioinformatic methods can be used, for example, in monitoring microbial adaptations to agrochemicals and to provide a quantitative method to test the potential other associated risks, such as an increase in resistance of pathogens to agrochemicals, negative effects on microbes used as biological control agents in integrated pest management (IPM), and antibiotic resistance in bacteria.

#### **ACKNOWLEDGMENTS:**

This work was funded by the Academy of Finland (grant no. 311077 to Marjo Helander).

#### **DISCLOSURES:**

Conflicts of interest: none.

#### **REFERENCES**

1. Williams, G. M., Kroes, R., Munro, I. C. Safety evaluation and risk assessment of the herbicide Roundup and its active ingredient, glyphosate, for humans. *Regulatory Toxicology and Pharmacology*. **31** (2 Pt 1), 117–165 (2000).
2. Muola, A. et al. Risk in the circular food economy: Glyphosate-based herbicide residues in manure fertilizers decrease crop yield. *The Science of the Total Environment*. **750**, 141422 (2021).
3. Helander, M., Saloniemi, I., Saikkonen, K. Glyphosate in northern ecosystems. *Trends in Plant Science*. **17** (10), 569–574 (2012).
4. Fuchs, B., Saikkonen, K., Helander, M. Glyphosate-Modulated Biosynthesis Driving Plant Defense and Species Interactions. *Trends in Plant Science*. **26** (4), 312–323 (2021).
5. Helander, M., Saloniemi, I., Omacini, M., Druille, M., Salminen, J. -P., Saikkonen, K. Glyphosate decreases mycorrhizal colonization and affects plant-soil feedback. *The Science of the Total Environment*. **642**, 285–291 (2018).
6. Bai, S. H., Ogbourne, S. M. Glyphosate: environmental contamination, toxicity and potential risks to human health via food contamination. *Environmental Science and Pollution Research International*. **23** (19), 18988–19001 (2016).
7. Cuhra, M., Bøhn, T., Cuhra, P. Glyphosate: too much of a good thing? *Frontiers in Environmental Science*. **4** (2016).
8. de Brito Rodrigues, L. et al. Impact of the glyphosate-based commercial herbicide, its components and its metabolite AMPA on non-target aquatic organisms. *Mutation Research*. **842**, 94–101 (2019).
9. Steinrücken, H. C., Amrhein, N. The herbicide glyphosate is a potent inhibitor of 5-enolpyruvylshikimate-3-phosphate synthase. *Biochemical and Biophysical Research Communications*. **94** (4), 1207–1212 (1980).
10. Bentley, R. The shikimate pathway--a metabolic tree with many branches. *Critical Reviews in Biochemistry and Molecular Biology*. **25** (5), 307–384 (1990).
11. Richards, T. A. et al. Evolutionary origins of the eukaryotic shikimate pathway: gene fusions, horizontal gene transfer, and endosymbiotic replacements. *Eukaryotic Cell*. **5** (9), 1517–1531 (2006).
12. Leino, L. et al. Classification of the glyphosate target enzyme (5-enolpyruvylshikimate-3-phosphate synthase) for assessing sensitivity of organisms to the herbicide. *Journal of hazardous materials*. **408**, 124556 (2021).
13. Rainio, M.J., Ruuskanen, S., Helander, M., Saikkonen, K., Saloniemi, I., Puigbò, P. Adaptation of bacteria to glyphosate: a microevolutionary perspective of the enzyme 5-enolpyruvylshikimate-3-phosphate (EPSP) synthase. *BioRxiv*. biorxiv.2020.06.16.154005 (2020).
14. Rainio, M. J., Ruuskanen, S., Helander, M., Saikkonen, K., Saloniemi, I., Puigbò, P. Adaptation of bacteria to glyphosate: a microevolutionary perspective of the enzyme 5-enolpyruvylshikimate-3-phosphate synthase. *Environmental microbiology reports*. **13** (3), 309–316 (2021).
15. Helander, M., Pauna, A., Saikkonen, K., Saloniemi, I. Glyphosate residues in soil affect crop plant germination and growth. *Scientific Reports*. **9** (1), 19653 (2019).
16. Ruuskanen, S., Rainio, M. J., Uusitalo, M., Saikkonen, K., Helander, M. Effects of parental exposure to glyphosate-based herbicides on embryonic development and oxidative status: a long-term experiment in a bird model. *Scientific Reports*. **10** (1), 6349 (2020).



17. Ruuskanen, S. et al. female preference and adverse developmental effects of glyphosate-based herbicides on ecologically relevant traits in Japanese quails. *Environmental Science & Technology*. **54** (2), 1128–1135 (2020).
18. Mäki, A., Rissanen, A. J., Tirola, M. A practical method for barcoding and size-trimming PCR templates for amplicon sequencing. *Biotechniques*. **60** (2), 88–90 (2016).
19. Chelius, M. K., Triplett, E. W. The diversity of archaea and bacteria in association with the roots of *Zea mays* L. *Microbial Ecology*. **41** (3), 252–263 (2001).
20. Lane, D. J. 16S/23S rRNA sequencing. *Nucleic acid techniques in bacterial systematics*. John Wiley & Sons, New York, 115–175 (1991).
21. Ghyselinck, J., Pfeiffer, S., Heylen, K., Sessitsch, A., De Vos, P. The effect of primer choice and short read sequences on the outcome of 16S rRNA gene based diversity studies. *Plos One*. **8** (8), e71360 (2013).
22. Zheng, D., Alm, E. W., Stahl, D. A., Raskin, L. Characterization of universal small-subunit rRNA hybridization probes for quantitative molecular microbial ecology studies. *Applied and Environmental Microbiology*. **62** (12), 4504–4513 (1996).
23. JoVE Supplementary Material at <[https://ppuigbo.me/programs/EPSPSClass/JOVE\\_SM](https://ppuigbo.me/programs/EPSPSClass/JOVE_SM)> (2021).
24. Alignments Conserved Positions at <<https://ppuigbo.me/programs/primers>> (2021).
25. Protein Families at <<http://pfam.xfam.org>> (2021).
26. NCBI GenBank at <<https://www.ncbi.nlm.nih.gov/genbank>> (2021).
27. COG Database at <<https://www.ncbi.nlm.nih.gov/research/cog>> (2021).
28. Protein Data Bank at <<https://www.rcsb.org>> (2021).
29. EPSPSClass at <<https://ppuigbo.me/programs/EPSPSClass>> (2021).
30. Kristensen, D. M., Wolf, Y. I., Koonin, E. V. ATGC database and ATGC-COGs: an updated resource for micro- and macro-evolutionary studies of prokaryotic genomes and protein family annotation. *Nucleic Acids Research*. **45** (D1), D210–D218 (2017).
31. ATGC\_NCBI at <[http://ftp.ncbi.nlm.nih.gov/pub/kristensen/ATGC/atgc\\_home.html](http://ftp.ncbi.nlm.nih.gov/pub/kristensen/ATGC/atgc_home.html)> (2021).
32. COG2020 at <<https://ftp.ncbi.nih.gov/pub/COG/COG2020/data>> (2021).
33. CSC – IT CENTER FOR SCIENCE LTD at <<https://www.csc.fi>> (2021).
34. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J. Basic local alignment search tool. *Journal of Molecular Biology*. **215** (3), 403–410 (1990).
35. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*. **32** (5), 1792–1797 (2004).
36. Tkacz, A., Hortala, M., Poole, P. S. Absolute quantitation of microbiota abundance in environmental samples. *Microbiome*. **6** (1), 110 (2018).
37. El-Gebali, S. et al. The Pfam protein families database in 2019. *Nucleic Acids Research*. **47** (D1), D427–D432 (2019).
38. Benson, D. A. et al. GenBank. *Nucleic Acids Research*. **41** (Database issue), D36–42 (2013).
39. Galperin, M. Y., Makarova, K. S., Wolf, Y. I., Koonin, E. V. Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Research*. **43** (Database issue), D261–269 (2015).
40. Burley, S. K., Berman, H. M., Kleywegt, G. J., Markley, J. L., Nakamura, H., Velankar, S.

Protein data bank (PDB): the single global macromolecular structure archive. *Methods in Molecular Biology*. **1607**, 627–641 (2017).

41. Raoult, D., Hadjadj, L., Baron, S. A., Rolain, J. -M. Role of glyphosate in the emergence of antimicrobial resistance in bacteria? *The Journal of Antimicrobial Chemotherapy*. **76** (7), 1655–1657 (2021).

42. Liu, J., Gefen, O., Ronin, I., Bar-Meir, M., Balaban, N. Q. Effect of tolerance on the evolution of antibiotic resistance under drug combinations. *Science*. **367** (6474), 200–204 (2020).

43. Peixoto, F. Comparative effects of the Roundup and glyphosate on mitochondrial oxidative phosphorylation. *Chemosphere*. **61** (8), 1115–1122 (2005).

44. Peillex, C., Pelletier, M. The impact and toxicity of glyphosate and glyphosate-based herbicides on health and immunity. *Journal of Immunotoxicology*. **17** (1), 163–174 (2020).

45. Funke, T., Han, H., Healy-Fried, M. L., Fischer, M., Schönbrunn, E. Molecular basis for the herbicide resistance of Roundup Ready crops. *Proceedings of the National Academy of Sciences of the United States of America*. **103** (35), 13010–13015 (2006).

46. Light, S. H., Krishna, S. N., Minasov, G., Anderson, W. F. An unusual cation-binding site and distinct domain-domain interactions distinguish Class II Enolpyruvylshikimate-3-phosphate synthases. *Biochemistry*. **55** (8), 1239–1245 (2016).

47. Firdous, S., Iqbal, S., Anwar, S., Jabeen, H. Identification and analysis of 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) gene from glyphosate-resistant *Ochrobactrum intermedium* Sq20. *Pest Management Science*. **74** (5), 1184–1196 (2018).

48. Barry, G. F., Kishore, G. M., Padgett, S. R., Stallings, W. C. US5627061A - Glyphosate-tolerant 5-enolpyruvylshikimate-3-phosphate synthases. Patent Number: 5627061 (1997).

49. Priestman, M. A., Funke, T., Singh, I. M., Crupper, S. S., Schönbrunn, E. 5-Enolpyruvylshikimate-3-phosphate synthase from *Staphylococcus aureus* is insensitive to glyphosate. *FEBS Letters*. **579** (3), 728–732 (2005).

50. Carozzi, N., Carr, B., Hammer, P. US20060253921A1 - Identification of a new class of EPSP synthases. World Intellectual Property Organization Publ. of the Int. Appl. without Int. search REP. WO2006US13161 (2006).

51. Lira, J. M., Cicchillo, R. M., Nair, S. K. US20130217577A1 - Novel class of glyphosate resistance genes (2013).

52. Funke, T. et al. Structural basis of glyphosate resistance resulting from the double mutation Thr97 -> Ile and Pro101 -> Ser in 5-enolpyruvylshikimate-3-phosphate synthase from *Escherichia coli*. *The Journal of Biological Chemistry*. **284** (15), 9854–9860 (2009).

53. Schönbrunn, E. et al. Interaction of the herbicide glyphosate with its target enzyme 5-enolpyruvylshikimate 3-phosphate synthase in atomic detail. *Proceedings of the National Academy of Sciences of the United States of America*. **98** (4), 1376–1380 (2001).

54. Stalker, D. M., Hiatt, W. R., Comai, L. A single amino acid substitution in the enzyme 5-enolpyruvylshikimate-3-phosphate synthase confers resistance to the herbicide glyphosate. *The Journal of Biological Chemistry*. **260** (8), 4724–4728 (1985).

55. Eschenburg, S., Healy, M. L., Priestman, M. A., Lushington, G. H., Schönbrunn, E. How the mutation glycine96 to alanine confers glyphosate insensitivity to 5-enolpyruvyl shikimate-3-phosphate synthase from *Escherichia coli*. *Planta*. **216** (1), 129–135 (2002).

56. Achary, V. M. M. et al. Overexpression of improved EPSPS gene results in field level

747 glyphosate tolerance and higher grain yield in rice. *Plant Biotechnology Journal*. **18** (12), 2504–  
748 2519 (2020).

749 57. Huang, Z., Liu, Y., Zhang, C., Jiang, C., Huang, H., Wei, S. Molecular basis of natural  
750 tolerance to glyphosate in *Convolvulus arvensis*. *Scientific Reports*. **9** (1), 8133 (2019).

751 58. Tall, T. A census analysis of the 5-enolpyruvylshikimate-3-phosphate (EPSP) synthase  
752 and EPSP-associated domains (2020).

753 59. Tall, T., Puigbò, P. The glyphosate target enzyme 5-Enolpyruvyl Shikimate 3-Phosphate  
754 Synthase (EPSPS) contains several EPSPS-associated domains in fungi. *ATLA Summary of*  
755 *Proceedings*. **76** (1), 6 (2020).

756 60. Puigbò, P., Guzmán, E., Romeu, A., Garcia-Vallvé, S. OPTIMIZER: a web server for  
757 optimizing the codon usage of DNA sequences. *Nucleic Acids Research*. **35** (Web Server issue),  
758 W126–131 (2007).

Figure  
**Figure 1**

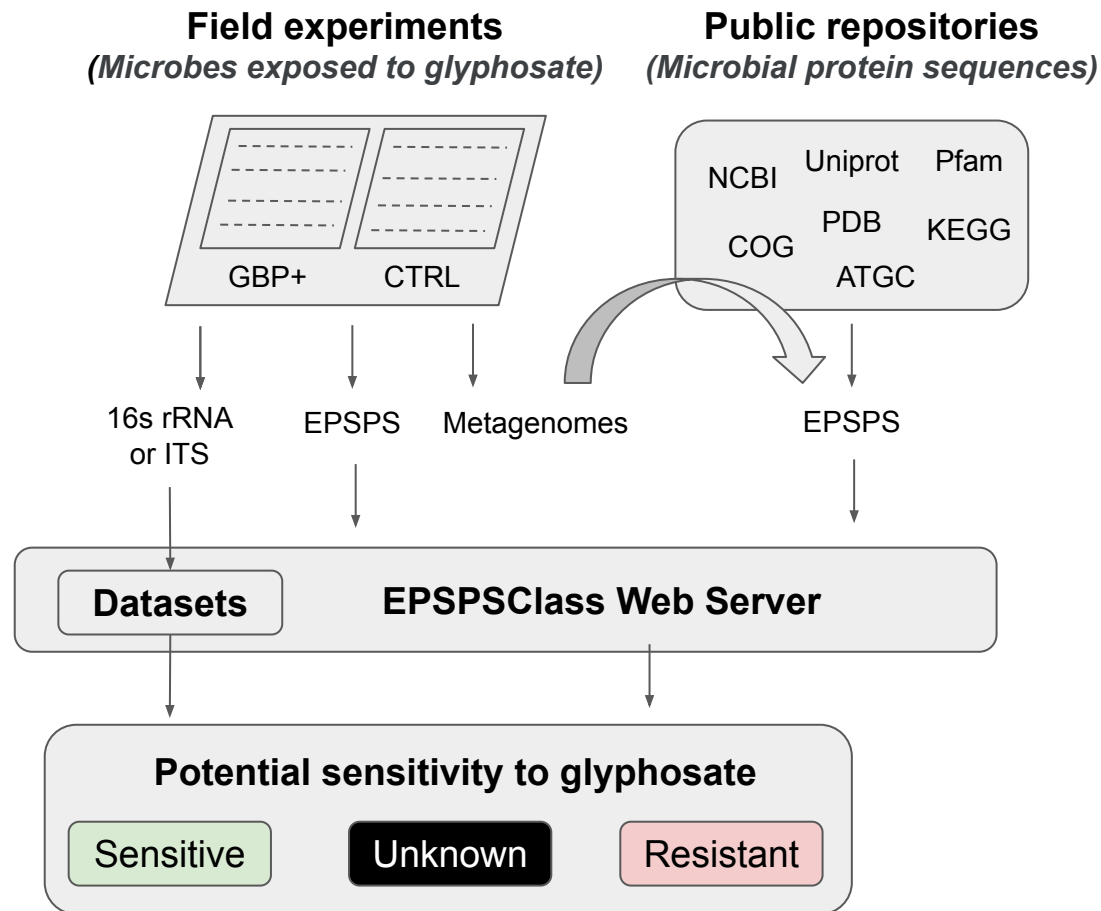


Figure  
**Figure 2**

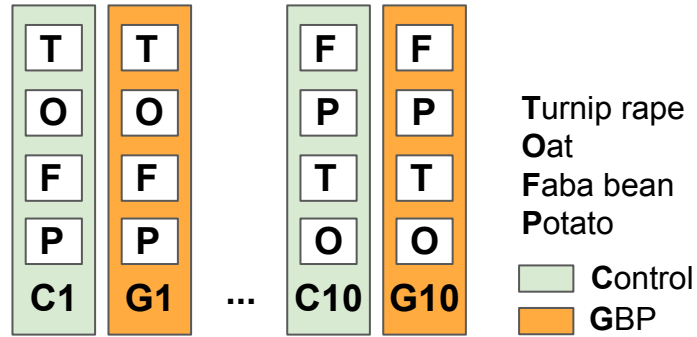


Figure  
**Figure 3**

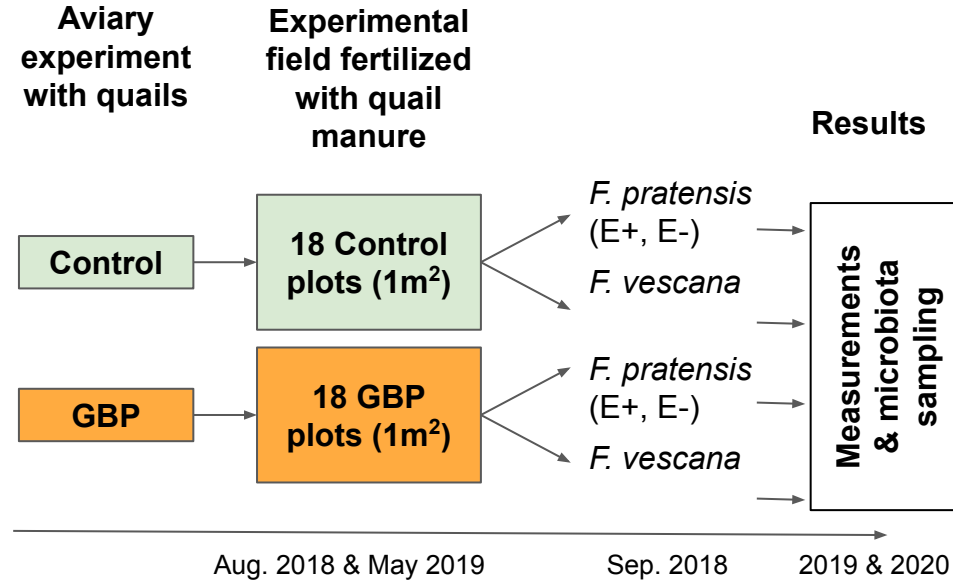
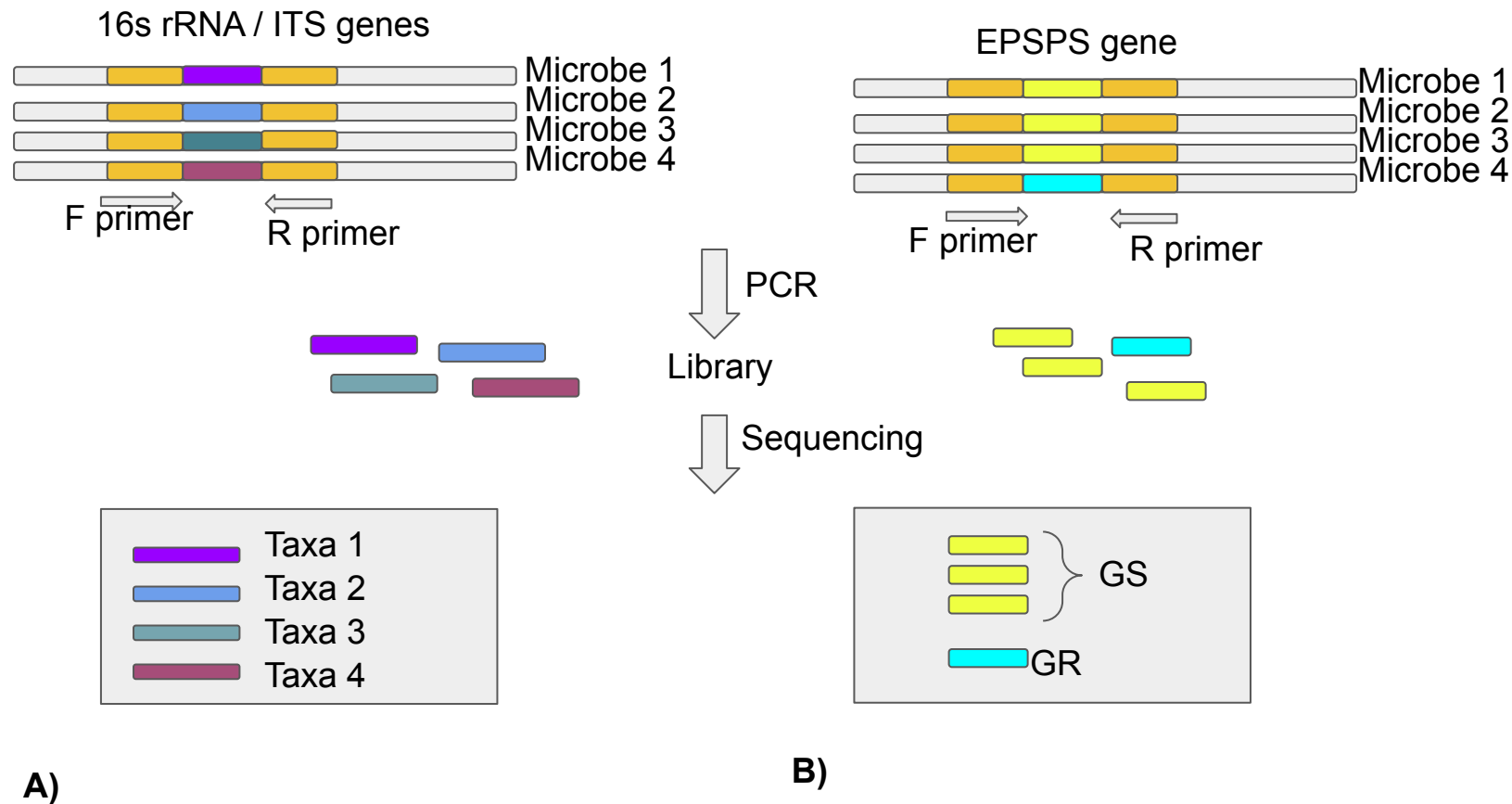
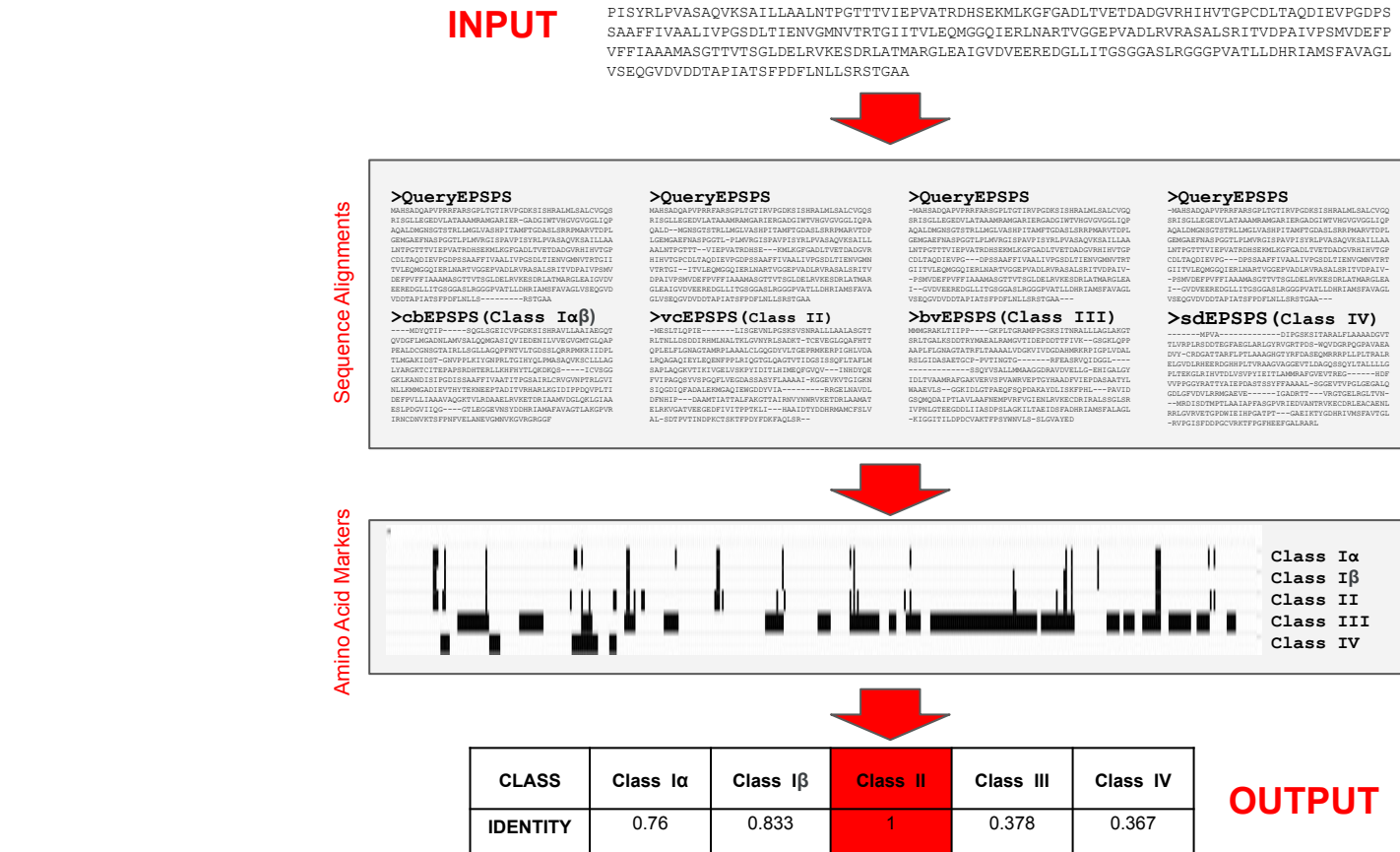


Figure  
**Figure 4**







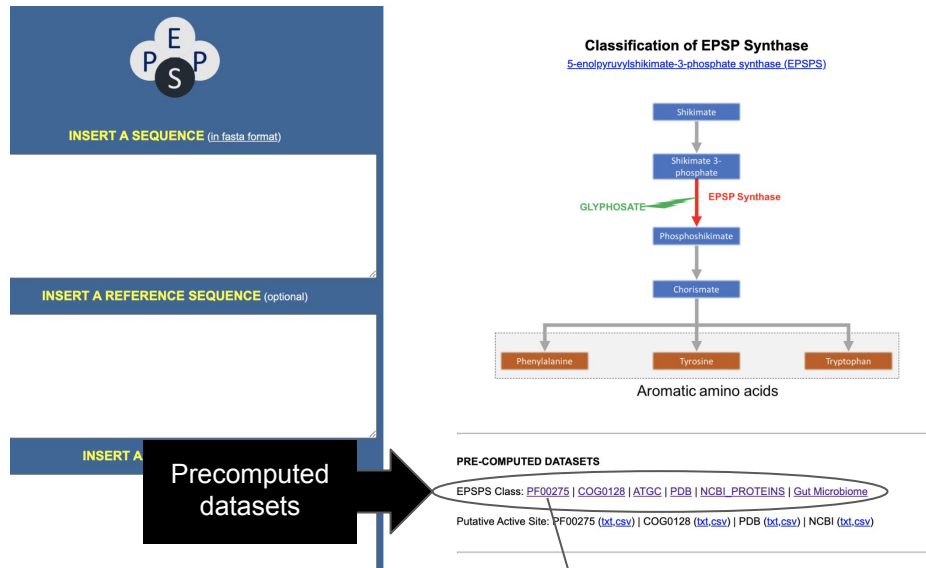
D)

C)

E)

F)

[Run BLAST](#) | [Conserved domains](#)

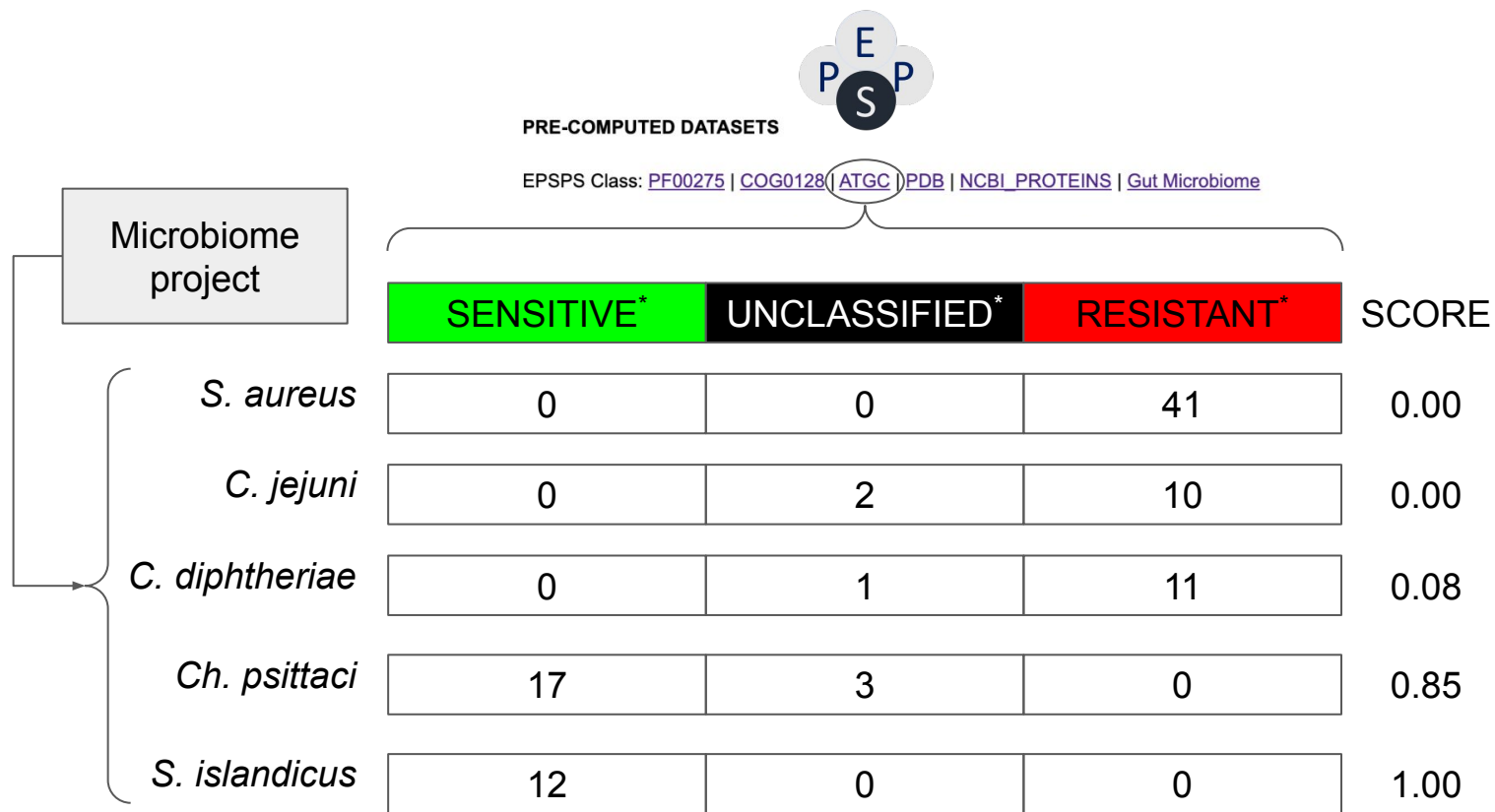


Id	ClassII (Identity)	ClassIII (Identity)	ClassII (Identity)	ClassIII (Identity)	Class IV (Identity)	Prediction	Length	Ch.Cb.CILCIV.N	Taxonomy
AD0N1BLJ2_9SPHN (AD0N1BLJ2)	0.760	0.833	1.000	0.18	0.367	ClassII	447	0.0,1.0,0.0	Blattariaceae sp. AAP25.
AD0A04G8T8_SPEZI (AD0A04G8T8)	1.000	1.000	0.605	0.18	0.333	ClassII	1581	1.1,0.0,0.0	Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales; Sphingomonadaceae; Blattariaceae.
ESL1B95_PUCGT (ESL1B95)	0.800	0.778	0.526	0.18	0.300	ClassII	1560	0.0,0.0,0.1	Microsporidia; Sclerotiniaceae; Sclerotium.
AR0A_NITOC (AR0A_NITOC)	0.760	0.833	1.000	0.18	0.233	ClassII	444	0.0,1.0,0.0	Puccinia graminis f. sp. tritici (75-36-700-3 / race SCCL); Black stem rust fungus).
WOR1M8_9BACT (WOR1M8)	0.760	0.833	1.000	0.18	0.367	ClassII	418	0.0,1.0,0.0	Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Microascales; Nitrososporaceae; Nitrososporaceae.
AD0A04P9B2_9ACTN (AD0A04P9B2)	0.960	0.889	0.500	0.18	0.367	ClassII	418	0.0,1.0,0.1	Bacteria; Proteobacteria; Gammaproteobacteria; Chromatiales; Chromatiaceae; Nitrososporaceae.
AD0A04P9B2_9ACTN (AD0A04P9B2)	0.960	0.889	0.500	0.18	0.367	ClassII	418	0.0,1.0,0.1	Bacteria; Proteobacteria; Gammaproteobacteria; Chromatiales; Chromatiaceae; Nitrososporaceae.
AD0A04P9B2_9ACTN (AD0A04P9B2)	0.960	0.889	0.500	0.18	0.367	ClassII	418	0.0,1.0,0.1	Bacteria; Proteobacteria; Gammaproteobacteria; Chromatiales; Chromatiaceae; Nitrososporaceae.
UZZ2449_PSEAC (UZZ2449)	0.360	0.556	0.342	0.18	0.367	ClassII	418	0.0,1.0,0.1	Bacteria; Proteobacteria; Gammaproteobacteria; Chromatiales; Chromatiaceae; Nitrososporaceae.
CTD9Q0_9RHOB (CTD9Q0)	0.360	0.500	0.316	0.18	0.367	ClassII	418	0.0,1.0,0.1	Bacteria; Proteobacteria; Gammaproteobacteria; Chromatiales; Chromatiaceae; Nitrososporaceae.
AR01_SCLS1 (AR01_SCLS1)	1.000	1.000	0.579	0.18	0.367	ClassII	418	0.0,1.0,0.1	Bacteria; Proteobacteria; Gammaproteobacteria; Chromatiales; Chromatiaceae; Nitrososporaceae.

EPSPS Class prediction

Species and taxonomic information

Figure 8



\* Number of sequences in the ATGC dataset

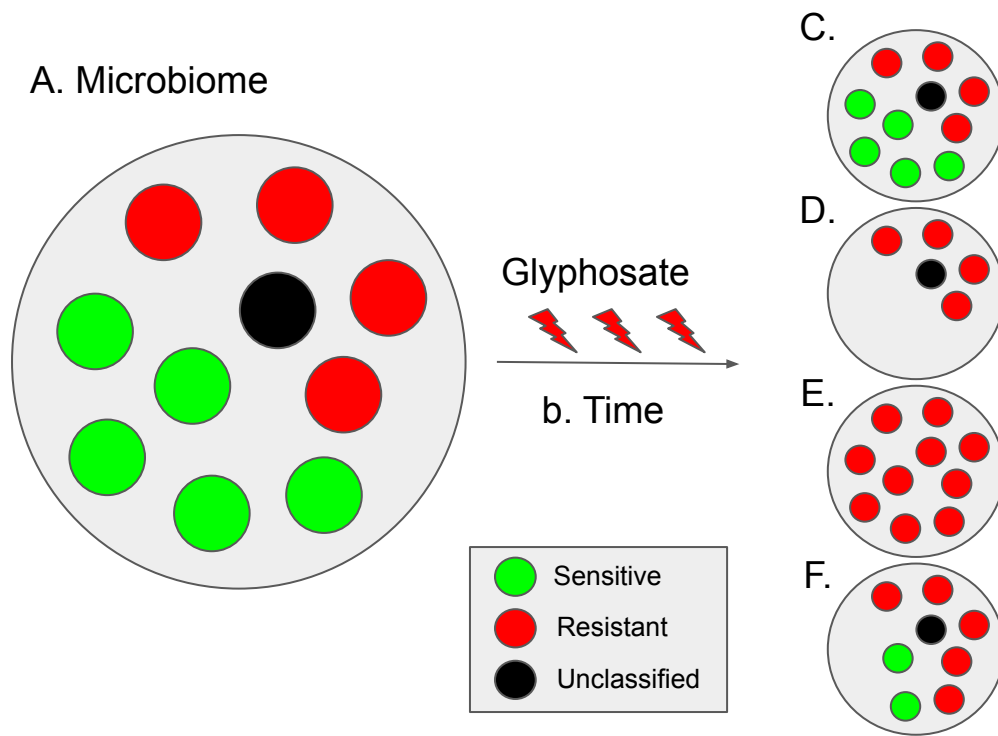




Figure  
**Figure 10**

[Click here to access/download;Figure;Figure10.pdf](#)

A) 

#### INSERT A SEQUENCE (in fasta format)

```
>sp|Q9KR80|AROA_VIBCH
MESLTLQPIELISGEVNLPGSKSVSNRALLAALASGTTRLTNLLSDDIRHMLNALTKL
GVNYRLSADKTTCEVEGLGQAFHTTQPLELFLGNAGTAMRPLAALCLGQGDYVLTGEPR
MKERPIGHLVDA LRQAGA QIEVLEQENFPPLRIQGTGLQACTVTIDGSISSQFLTAFLMS
APLAQGGKVTIKIVGELVSKPYIDITLHIMEQFGVQVINHDYQEFVIPAGQSVSPGQFLV
EGDASSASYFLAAAIKGGEVKVTGIGKNSIQGDIQFADALEKMGAI EWGDDYVIARRG
ELNAVDLDFNHPDAAMTIATTALFAKGTTAIRNVYNWRVKETDRLAAMATELRKVGATV
EEGEDFIVITPPTKLIHAAIDTYDDHRMAMCFSLVALSDTFVTINDPKCTSKTFPDYFDK
FAQLSR
```

B) 

#### INSERT A REFERENCE SEQUENCE (optional)


```
>Reference_Sequence
MESLTLQPIELISGEVNLPGSKSVSNRALLAALASGTTRLTNLLSDDIRHMLNALTKL
GVNYRLSADKTTCEVEGLGQAFHTTQPLELFLGNAGTAMRPLAALCLGQGDYVLTGEPR
MKERPIGHLVDA LRQAGA QIEVLEQENFPPLRIQGTGLQACTVTIDGSISSQFLTAFLMS
APLAQGGKVTIKIVGELVSKPYIDITLHIMEQFGVQVINHDYQEFVIPAGQSVSPGQFLV
EGDASSASYFLAAAIKGGEVKVTGIGKNSIQGDIQFADALEKMGAI EWGDDYVIARRG
ELNAVDLDFNHPDAAMTIATTALFAKGTTAIRNVYNWRVKETDRLAAMATELRKVGATV
EEGEDFIVITPPTKLIHAAIDTYDDHRMAMCFSLVALSDTFVTINDPKCTSKTFPDYFDK
FAQLSR
```

C) 

#### INSERT AA MARKERS (optional)

```
#AA,Position
S/A,21
K,22
S,23
R,27
D,49
G,96
T,97
R,100
R,124
```

SEND

D) 

#### EPSP Synthase Class

Query sequence ID: 378384

CLASS	IDENTITY	MOTIFS (n_positives / n_total)
Class I (alpha)	1.000	-
Class I (beta)	1.000	-
Class II	0.605	-
Class III	0.465	0 / 18
Class IV	0.233	-
Reference	1.000	-

Primers	Sequence
799F	AACMGGATTAGATACCCKG
1492R	GGYTACCTTGTTACGACTT
uni-1062F	ACGACGTTGTAAAAGTCAGCTCGTGYGTGA
uni-1390R	CATTAAGTTCCCATTAACGGGCGGTGTGTRCAA
uni-ITS7F	ACGACGTTGTAAAAGTGARTCATCGAATCTTTG
uni-ITS4R	CATTAAGTTCCCATTCCTCCGCTTATTGATATGC

Database	Code
KEGG	K00800
COG	COG0128
CDD	PRK02427
PFAM	PF00275
BRENDA	EC.2.5.1.19

Link
<a href="https://www.kegg.jp/dbget-bin/www_bget?K00800">https://www.kegg.jp/dbget-bin/www_bget?K00800</a>
<a href="https://www.ncbi.nlm.nih.gov/research/cog/cog/cog012">https://www.ncbi.nlm.nih.gov/research/cog/cog/cog012</a>
<a href="https://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=PRK02427">https://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=PRK02427</a>
<a href="http://pfam.xfam.org/family/PF00275">http://pfam.xfam.org/family/PF00275</a>
<a href="https://www.brenda-enzymes.org/enzyme.php?ecno=2.5.1.19">https://www.brenda-enzymes.org/enzyme.php?ecno=2.5.1.19</a>



Reference Genome
<p><b>&gt;Vibrio cholerae (vcEPSPS)</b></p> <p>MESLTQPIELISGEVNLPGSKSVSNRALLAALASGTTRLTNLLDSDDIRHMLNATKL  GVNYRLSADKTTCEVEGLGQAFHTTQPLELFLGNAGTAMRPLAAALCLGQG DYVLTGEPR  MKERPIGHLVDALRQAGAQIEYLEQENFPPLRIQGTGLQAGTVTIDGSISSQFLTAFLMS  APLAQGKVTIKIVGELVSKPYIDITLHIMEQFGVQVINHDYQEFVIPAGQSYVSPGQFLV  EGDASSASYFLAAAAIKGGEVKVTGIGKNSIQGDIQFADALEKMGAQIEWGDDYVIARRG  ELNAVDLDFNHIPDAAMTIATTALFAKGTAIRNVYNWRVKETDRLAAMATELRKVGATV  EEGEDFIVITPPTKLIHAAIDTYDDHRMAMCFSLVALSDTPVTINDPKCTSKTFPDYFDK  FAQLSR</p>
<p><b>&gt;Coxiella burnetii (cbEPSPS)</b></p> <p>MDYQTIPSQGLSGEICVPGDKSISHRAVLLAAIAEGQTQVDGFLMGADNLAMVSALQQMG  ASIQVIEDENILVVEGVGMTGLQAPPEALDCGNSGTAIRLLSGLLAGQPFNTVLTGDSSL  QRRPMKRIIDPLTLMGAKIDSTGNVPPLKIYGNPRLTGIHYQLPMASAQVKSCLLLAGLY  ARGKTCITEPAPSRDHTERLLKFHYTLQKDKQSICVSGGGKLANDISIPGDISSAAFF  IVAATITPGSAIRLCRVGNPTRLGVINLLKMMGADIEVTHYTEKNEEPTADITVRHARL  KGIDIPPDQVPLTIDFEPVLLIAAAVAQGKTVLRDAAELRVKETDRIAAMVDGLQKLGIA  AESLPDGVIIQGGTLEGGEVNSYDDHRIAMAFAVAGTLAKGPVRIRNCDNVKTSFPNFVE  LANEVGMNVKGVRGRGGF</p>
<p><b>&gt;Brevundimonas vesicularis (bvEPSPS)</b></p> <p>MMMGRAKLTIIPPGKPLTGRAMPPGSKSITNRALLLAGLAKGTSRLTGALKSDDTRYMAE  ALRAMGVTIDEPPDDTTFIVKGSGLQPPAAPLFLGNAGTATRFLTAAAALVDGKVIVDGD  AHMRKRPIGPLVDALRSLGIDASAETGCPPVTINGTGRFEASRVQIDGGLSSQYVSALLM  MAAGGDRAVDVELLGEHIGALGYIDLTVAAMRAFGAKVERVSPVAWRVEPTGYHAADFVI  EPDASAATYLVAAEVLSSGKIDLGTPAEQFSQPDAKAYDLISKFPHLPAVIDGSQMMDAI  PTLAVLAAFNEMPVRFVGIENLRVKECDRIRALSSGLSRIVPNLGTTEGDDLIASDPSL  AGKILTAEIDSFADHRIAMSFALAGLKIGGITILDPDCVAKTFPSYWNVLSSLGVAYED</p>

---

**>Streptomyces davawensis (sdEPSPS)**

MPVADIPGSKSITARALFLAAAADGVTTLVRPLRSDDTEGFAEGLARLGYRVGRTPDSWQ  
VDGRPQGPAVAEADVYCRDGATTARFLPTLAAAGHGTYRFDASEQMRRRPLLPLTRALRE  
LGVDLRHEERDGHHP LTVRAAGVAGGEVTL DAGQSSQYL TALLLGPLTEKGLRIHVTDL  
VSVPYIEITLAMMRAFGVEVTREGHDFVPPGGYRATTYAIEPDASTSSYFFAAAALSGG  
EVTVPGLGEGALQGD LGFVDVLRRMGAEVEIGADRTTVRGTGELRGLTVNMRDISDTMPT  
LAAIAPFASGPVRIEDVANTRVKECDRLEACAENLRRLGVRVETGPDWIEIHPGATPTGA  
EIKTYGDHRIVMSFAVTGLRVPGISFDDPGCVRKTFPGFHEEFGALRARL

---

---

**Class: Amino acid markers**

---

**Class I (alpha):** 21,S/A; 22,K; 23,S; 27,R; 49,D; 96,G; 97,T; 100,R; 124,R; 125,P; 150,P; 171,S; 172,Q; 205,T; 241,E; 243,D; 272,Q; 342,E; 345,R; 357,G; 385,D; 386,H; 387,R; 412,K; 415,P

**Class I (beta):** 21,S/A; 22,K; 27,R; 49,D; 124,R; 125,P; 171,S; 172,Q; 205,T; 243,D; 314,D; 341,K; 342,E; 345,R; 357,G; 385,D; 386,H; 387,R

---

**Class II:** 20,D; 21,K; 22,S; 25,H; 26,R; 48,D; 93,N; 99,R; 103,G; 116,G; 117,D; 123,R; 124,P; 127,R; 131,P; 132,L; 167,S; 168,A; 169,Q; 171,K; 194,R; 198,E; 231,P; 233,D; 235,S; 260,N; 263,R; 315,D; 316,E; 338,E; 342,K; 343,E; 346,R; 385,D; 386,H; 387,R; 414,S; 416,P

**Class III:** 39,L,I; 40,A,I; 41,K,I; 42,G,I; 43,K/T,I; 44,S,I; 45,R/H,I; 46,L,I; 47,S/T,I; 48,G,I; 49,A,I; 50,L,I; 51,K,I; 52,S,I; 53,D,I; 54,D,I; 55,T,I; 71,E,II; 72,P,II; 73,D,II; 74,D/A,II; 75,S/T,II; 76,T,II; 77,F,II; 78,V/I,II; 79,V,II; 80,T/E/K,II; 81,S/G,II; 82,Q/S/E/T,II; 83,G,II; 102,R,III; 103,F,III; 104,L,III; 105,T,III; 106,A,III; 107,A,III; 125,K,IV; 126,R,IV; 127,P,IV; 128,I,IV; 129,G/M/L,IV; 130,P,V; 146,T/S,V; 147,G,V; 148,C,V; 149,P,V; 150,P,V; 151,V,V; 198,I,VI; 199,G,VI; 200,A,VI; 201,R/K/L,VI; 202,G,VI; 203,Y,VI; 204,I/V,VI; 205,D,VI; 206,L,VI; 207,T,VI; 226,W,VII; 227,R/K,VII; 228,V,VII; 229,A/H/E/S,VII; 230,P/A,VII; 231,T,VII; 232,G,VII; 241,E,VIII; 42,P,VIII; 243,D,VIII; 244,A,VIII; 245,S,VIII; 246,A,VIII; 247,A,VIII; 248,T,VIII; 249,Y,VIII; 250,L,VIII; 251,W,VIII; 252,A/G,VIII; 253,A,VIII; 254,E/G,VIII; 255,V/L/A,VIII; 256,L,VIII; 261,I,IX; 262,D,IX; 263,I/L,IX; 264,G,IX; 270,F,X; 271,T/S,X; 272,Q,X; 273,P,X; 274,D,X; 275,A,X; 276,K,X; 277,A,X; 283,Q/K/S,XI; 284,F,XI; 285,P,XI; 286,N/H,XI; 287,M/L,XI; 288,P/Q,XI; 289,A,XI; 290,T/E/V,XI; 291,V/I,XI; 292,D/V,XI; 293,G,XI; 294,S,XI; 295,Q,XI; 296,M,XI; 297,Q,XI; 298,D,XI; 299,A,XI; 300,I,XI; 301,P,XI; 302,T,XI; 303,L/I,XI; 304,A,XI; 305,V,XI; 306,L/I,XI; 307,A,XI; 308,A,XI; 309,F,XI; 310,N,XI; 313,P,XII; 314,V,XII; 315,R,XII; 316,F,XII; 317,V/T,XII; 318,E/G,XII; 319,L/I,XII; 320,A/E,XII; 321,N,XII; 322,L,XII; 323,R,XII; 324,V,XII; 325,K,XII; 326,E,XII; 327,C,XII; 328,D,XII; 329,R,XII; 330,I/V,XII; 348,E,XIII; 349,G,XIII; 350,D,XIII; 351,D,XIII; 352,L,XIII; 353,L/I,XIII; 354,V/I,XIII; 357,D/N,XIV; 358,P,XIV; 359,A/S,XIV; 360,L,XIV; 361,A,XIV; 362,G,XIV; 367,A,XV; 368,L/S/E,XV; 369,I,XV; 370,D,XV; 371,T/S,XV; 372,H/F,XV; 373,A/S,XV; 374,D,XV; 375,H,XV; 376,R,XV; 381,F,XVI; 382,A,XVI; 383,L,XVI; 384,A,XVI; 385,G/A,XVI; 386,L,XVI; 387,K,XVI; 388,I/V,XVI; 389,S/G,XVI; 390,G,XVI; 391,I,XVI; 395,N/D,XVII; 396,P,XVII; 397,A/D,XVII; 398,C,XVII; 399,V,XVII; 400,A/G,XVII; 401,K,XVII; 411,A/S/P,XVIII; 412,S,XVIII; 413,L,XVIII; 414,G,XVIII; 415,V,XVIII

---

---

**Class IV:** 13,T; 14,A; 15,R; 16,A/G; 17,L; 18,F; 39,E; 40,G; 41,F; 42,T/A; 43,E; 44,G; 80,G; 81,A;  
82,T; 83,T; 84,A; 85,R; 86,F; 87,L; 88,P; 89,T/A; 90,L; 91,A/V; 92,A; 93,A; 100,F; 101,D; 102,A;  
103,S

---



[Click here to access/download](#)

**Table of Materials**

[Mathew\\_et\\_al\\_JoVE\\_Table\\_of\\_Materials\\_R1-63109.xlsx](#)



**Editorial comments:**

1. Please note that the manuscript has been formatted to fit the journal standard. Comments to be addressed are included within the manuscript. Please review and revise.

Thank you! The manuscript has been reviewed and revised following comments.

2. Please note that the step numbers have been revised to fit the JoVE format. Please check the step numbers and section numbers referenced within the steps of the Protocol.

Done.

3. Please reword the lines to avoid the issue of plagiarism: 140-141, 146-149, 555-557. Please refer to the iThenticate report attached.

These lines have been reworded to avoid any issue of plagiarism.

4. The protocol section of the manuscript must contain only steps that direct the readers to do something. Brief notes can be introduced. Section 1 has been modified to steps than paragraphs. Please check. Please move it to the Representative Results or Discussion section if you wish to retain them as paragraphs.

The Section 1 has been revised and modified.

5. Please ensure that all the figures are discussed in the Representative Results section (results mentioned in sections 1, 2, 5). However, for figures showing the experimental set-up, please reference them in the Protocol. Ensure to explain the Representative Results in the context of the technique you have described, e.g., how do these results show the technique, suggestions about how to analyze the outcome, etc. Data from both successful and sub-optimal experiments can be included.

All sections are included in the representative results.

6. Please consider removing the website links from the manuscript's main text and including them in the references. Cite the appropriate reference numbers in the respective lines.

Done. Websites have been included as references.

7. Please ensure that the Table of Materials includes all the supplies (chemicals, reagents, equipment, detection kits, software, instruments, etc.) used in this study. Please sort the Table in alphabetical order.

The table of materials has been updated to include all the supplies.

8. The highlighted content in the current version of the manuscript exceeds the 3-page limit. Please highlight up to 3 pages of the Protocol (including headings and spacing) that identifies the essential steps of the Protocol for the video, i.e., the steps that should be visualized to tell the most cohesive story of the Protocol. Remember that non-highlighted Protocol steps will remain in the manuscript, and therefore will still be available to the reader. Please ensure that the highlighted steps form a cohesive narrative with a logical flow from one highlighted step to the next. Please highlight complete sentences (not parts of sentences). Please ensure that the highlighted part of the step includes at least one action that is written in the imperative tense.

The highlighted content has been reduced.