Classification of the glyphosate target enzyme (5-enolpyruvylshikimate-3-phosphate synthase)

Lyydia Leino ^{1,&}, Tuomas Tall ^{1,&}, Marjo Helander ¹, Irma Saloniemi ¹, Kari Saikkonen ², Suvi Ruuskanen ¹, Pere Puigbò ^{1,3,*}

¹ *Department of Biology, University of Turku, Turku, Finland*

² *Biodiversity Unit, University of Turku, Finland*

³ *Nutrition and Health Unit, Eurecat Technology Centre of Catalonia, Reus, Catalonia, Spain*

[&] *Equal contribution of the authors*

^{*} *Corresponding author: pepuav@utu.fi*

Keywords: shikimate pathway, epsps enzyme, glyphosate, herbicide, resistance, microbiome, bioinformatics resource, biomarkers

ABSTRACT

Glyphosate is the most common broad-spectrum herbicide. It targets the key enzyme of the shikimate pathway, 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS), which synthesizes three essential aromatic amino acids (phenylalanine, tyrosine and tryptophan) in plants. Because the shikimate pathway is also found in many prokaryotes and fungi, the widespread use of glyphosate may have unsuspected impacts on the diversity and composition of microbial communities, including the human gut microbiome. Here, we introduce the first bioinformatics method to assess the potential sensitivity of organisms to glyphosate based on the type of EPSPS enzyme. We have precomputed a dataset of EPSPS sequences from thousands of species that will be an invaluable resource to advancing the research field. This novel methodology can classify sequences from >90% of eukaryotes and >80% of prokaryotes. A conservative estimate from our results shows that 54% of species in the core human gut microbiome are sensitive to glyphosate.

INTRODUCTION

Glyphosate is the most efficient and widely used nonselective herbicide. Historically, it was commercialized in the 1970s and then became the most inexpensive herbicide after the patent expired in 2000¹⁻³. Since then, numerous generic glyphosate-containing herbicides have made glyphosate-based herbicides (GBHs) the most commonly used pesticides worldwide⁴. The dominance of GBHs in the pesticide market is mainly attributed to the use of transgenic crops such as soy, corn and canola, of which nearly 90% are glyphosate-resistant varieties⁵. In Europe, where transgenic crops are hardly grown, they are much used in no-till cropping, where weeds are eradicated by glyphosate prior to sowing. In addition, cereal, bean and seed crops are commonly desiccated by glyphosate before harvest.

The biochemical target enzyme for the herbicide glyphosate is 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS)⁶, the key enzyme of the shikimate pathway, which synthesizes the three essential aromatic amino acids (phenylalanine, tyrosine and tryptophan) in most prokaryotes, plants and fungi^{7,8}. Glyphosate is proclaimed safe for humans and other nontarget organisms because the shikimate metabolic pathway, inactivated by glyphosate, is not present in vertebrates. However, until recently, the presence of the shikimate pathway and diversity of EPSPS in many microbes have largely been ignored. As microbes are ubiquitous, associated with virtually all higher organisms, and essential in maintaining fundamental organismal functions⁹⁻¹¹, predicting the consequences of glyphosate use via its potential effects on the microbiome is challenging. The first step toward a more comprehensive understanding of how glyphosate affects higher organisms and biotic interactions involving microbes is to survey microbe susceptibilities to glyphosate.

The widespread use of glyphosate may have a strong impact on the diversity and community composition of both eukaryotes (mainly plants and fungi) and prokaryotes. Nearly all plant and most fungal populations are sensitive to glyphosate^{12,13}, but we are still only beginning to understand the effects of glyphosate on the gut microbiome. The first evidence has shown that glyphosate can affect the bee gut microbiota composition¹⁴. An association between the use of glyphosate and antibiotic resistance has been suggested^{15,16}, though it is not clear in which direction, and thus, further studies are needed.

Here, we propose the first bioinformatic method to predict the glyphosate sensitivity/resistance of organisms based on the type of EPSPS enzyme. We have used this methodology to perform a comprehensive classification of the glyphosate target enzyme, EPSPS, based on sensitivity/resistance to glyphosate, taxonomic distribution and domain architectures. Our methodology classifies EPSPS enzymes into four different classes with differential sensitivities to glyphosate based on the presence and absence of amino acid markers in the active site¹⁷⁻²⁴. The classification of organisms based on the type of EPSPS enzyme will be extremely useful to assess species that are putatively sensitive or resistant to glyphosate. We have precomputed a dataset of EPSPS enzymes from thousands of species that will be an invaluable resource to test hypotheses in the field. This dataset includes 890 sequences from species in the core human gut microbiome, of which 54% are putatively sensitive to glyphosate (in a conservative estimate).

RESULTS

Classification of EPSPS enzymes based on sensitivity to glyphosate

EPSPS enzymes can be classified into four groups based on differential sensitivity to the herbicide glyphosate. This classification is based on the presence and absence of amino acid markers in the active site of the protein^{17–24} (figure 1, supplementary figures 1–2). In general, species containing class I EPSPS sequences (alpha and beta) are sensitive to glyphosate, whereas species with class II sequences tend to be resistant^{19–22} (supplementary table 1). EPSPS proteins belonging to classes III and IV putatively result in resistance to glyphosate^{23,24} and are relatively rare in nature (<5% of the sequences), e.g., all class IV EPSPS sequences are found in one single actinobacteria clade (figure 2b and supplementary figure 3). In prokaryotes, the majority of species have class I enzymes and are thus sensitive to glyphosate (84% in archaea and 57% in bacteria), whereas class II enzymes (resistant to glyphosate) represent only 2% and 32% of archaeal and bacterial species, respectively. In eukaryotes, the majority of EPSPS proteins belong to class I, including 69% of viridiplantae species and 91% of fungi. Although a relatively large portion of eukaryotic species remain unclassified (especially in viridiplantae, 31%), the number of species potentially resistant to glyphosate based on amino acid markers is quite low (figure 2 and supplementary figures 4 and 5). As an example, we determined the EPSPS classes in symbiotic fungal endophytes living internally and asymptotically within many grass species (figure 3). Recent studies have shown that glyphosate may alter this symbiotic relationship²⁵. EPSPS domains from endophytes were aligned with four reference sequences of EPSPS enzymes that contain the key amino acid markers of classes I–IV (supplementary tables 1–2). Although there are variations in the numbers and lengths of EPSPS-associated domains in endophytes, all EPSPS sequences indicated sensitivity to glyphosate (class I).

A large portion of EPSPS proteins (especially in prokaryotes) do not belong to any of the four known classes and are thus hereinafter termed unclassified EPSPS (figure 4, supplementary tables 3–4). Moreover, an analysis of dipeptides of EPSPS domains shows taxonomic and class differences among sequences (figure 4, supplementary figures 4 and 5). Both eukaryotic and archaeal species are merged within the larger cluster of bacterial species, but it is possible to differentiate two independent clusters for fungi and another two clusters for viridiplantae. This pattern is also shown in certain bacterial groups, e.g., in Proteobacteria, Firmicutes, Actinobacteria and cyanobacteria. Further empirical studies are needed to identify novel amino acid markers that determine the potential sensitivity of unclassified EPSPS proteins to glyphosate. Therefore, our dataset of more than 50,000 EPSPS proteins obtained from public databases^{26,27} and classified according to the potential sensitivity to the herbicide glyphosate (classes I–IV or unclassified) will be highly useful for future studies. This dataset is open and freely available at <https://ppuigbo.me/programs/EPSPSClass>. Moreover, a web server is also offered to compute the EPSPS class of a protein from its raw amino acid sequence, which will be highly useful in identifying new classes of EPSPS proteins.

Taxonomic distribution of the EPSPS enzymes

EPSPS enzymes are present in three domains of life: archaea, bacteria and eukaryotes (mostly in plants and fungi)^{7,8}. A phylogenetic analysis of the EPSPS domain (>10,000 sequences) shows that none of bacteria, archaea or eukaryotes form a monophyletic group (Figure 5a and supplementary tables 5–6). In eukaryotes, EPSPS is present in fungi, viridiplantae, stramenopiles and rhodophyta in two independent paraphyletic clusters. The first cluster of eukaryotes (Eukaryotes 1) contains single-domain sequences corresponding to viridiplantae (main taxa), stramenopiles and rhodophyta, whereas the second cluster (Eukaryotes 2) contains multidomain proteins (including the EPSPS domain and associated domains) from fungi and a few stramenopiles (figure 5bc). The phylogenetic analysis is based on alignment of only the EPSPS domain (see the Materials and Methods); thus, the

independent eukaryotic clades may not be explained by different domain architectures. Moreover, there are additional scattered eukaryotic sequences within bacterial groups that may be the product of contamination during genome sequencing (e.g., a simple BLAST analysis showed that the EPSPS of the winter moth is the product of bacterial contamination –data not shown) and putative ancient horizontal gene transfers (e.g., a relatively large cluster including several rosids) from bacterial species (supplementary figure 6). An analysis of COG0128 from the Clusters of Orthologous Groups (COG) database²⁸, which contains precomputed orthologous genes of more than 700 prokaryotes²⁷, shows that 92% of archaea and 88% of bacteria have at least one copy of the EPSPS protein (figure 5a). Nevertheless, the enzyme is remarkably infrequent in some bacterial groups, such as Mollicutes (present in only 10% of species) and Thermotoga (present in only 57% of species).

EPSPS domain architectures and analysis of the EPSPS-associated domains

EPSPS proteins include (by definition) the EPSPS domain, but in several species, the EPSPS protein contains additional associated domains (e.g., fungal species) (figure 5c and supplementary table 5). We analyzed ~10,000 protein sequences from the Pfam database²⁶ that contain the domain PF00275 (EPSPS). The EPSPS domain is ~450 amino acids long and is present in 62 domain architectures and as a single domain in most prokaryotes and viridiplantae (Pfam data²⁶). In fungi, the EPSPS domain forms part of multidomain proteins (usually more than 5 domains) that are ~1,300 amino acids long (figure 5c and supplementary table 6). Moreover, in a small number of bacterial and viridiplantae species, the EPSPS protein contains the EPSPS domain and an additional associated domain. The most common EPSPS-associated domains are shikimate pathway (SKI, DHQ synthase, DHquinase I) and aromatic amino acid synthesis (Shikimate dh N, PDH) enzymes and promiscuous domains (HTH 3). Additional EPSPS-associated domains exhibit functions such as DNA modification, gene expression and other enzymatic activities (supplementary table 6). Overall, the EPSPS-associated domains can be classified into shikimate (shikimate pathway domains), enzymatic (domains with catalytic function), expression (domains whose products are needed in controlling gene expression) and structural functions (domains that do not have catalytic function, e.g., binding sites, histones and helix-turn-helix domains).

Survey of the resistance and sensitivity to glyphosate in the human gut microbiomes

To test this new resource in a real-world scenario, we analyzed EPSPS sequences of 890 strains from 101 common bacterial species in the human gut microbiome²⁹. Our results suggest that EPSPS sensitivity to glyphosate is quite conserved within bacterial species in the human gut microbiome (supplementary table 7). Fifty-four percent of species have an EPSPS in class I, i.e., potentially sensitive to glyphosate (table 1). The core gut microbiome contains 75 species together representing 22-47% of the total species abundance in the gut microbiome, which contains approximately 160 varying species²⁹. Therefore, with a conservative calculation, 12-26% of bacterial species in the human microbiome might be sensitive and thus affected by glyphosate. In addition, 29% of species in the core microbiome are likely to be resistant to glyphosate (class II or III), and 17% of the species are still unclassified or contain some unclassified strains (table 1). Class IV EPSPS enzymes are not found in the microbiome dataset²⁹, as they are almost exclusive to streptomyces and a few other actinobacteria (table 3).

Within the ten most frequent bacterial species in the core human gut microbiome²⁹, four species are resistant to glyphosate (*Dorea formicigenerans*; *Clostridium* sp. SS2-1; *Eubacterium hallii*; *Coprococcus comes* SL7 1), four are sensitive (*Faecalibacterium prausnitzii* SL3-3; *Bacteroides*

vulgatus ATCC 8482; *Bacteroides uniformis* and *Bacteroides* sp. 9 1 42FAA), and two are unclassified (*Roseburia intestinalis* M50 1 and *Eubacterium rectale* M104 1). The list of genera with strains sensitive to glyphosate includes (all sequences from the first 18 genera are sensitive to glyphosate) *Actinomyces*, *Anaerofustis*, *Anaerotruncus*, *Catenibacterium*, *Citrobacter*, *Collinsella*, *Desulfovibrio*, *Enterobacter*, *Faecalibacterium*, *Holdemania*, *Methanobrevibacter*, *Mollicutes*, *Parabacteroides*, *Prevotella*, *Subdoligranulum*, *Escherichia*, *Actinomyces*, *Anaerofustis*, *Bacteroides*, *Bifidobacterium*, *Clostridium*, *Eubacterium*, *Ruminococcus*, *Providencia*, and *Fusobacterium*. Moreover, in some genera, the sensitivity to glyphosate may vary widely, e.g., all EPSPS sequences from *Ruminococcus bromii* strains (n=14) belong to class I, whereas those from *Ruminococcus gnavus* strains (n=17) belong to class II. We also detected intraspecific variation in EPSPS sequences in 7% of all studied species (table 1).

DISCUSSION

A large proportion of bacteria in the gut microbiome²⁹ are susceptible to glyphosate (class I); thus, the intake of glyphosate may severely affect the composition of the human gut microbiome. Our analysis suggests that the proportion of species susceptible to glyphosate is at least 12-26% of the total species in the human gut. Thus, the use of glyphosate may provide a competitive advantage to bacteria resistant to glyphosate over sensitive bacteria. Although data on glyphosate residues in human gut systems are still lacking²⁹, our results suggest that glyphosate residues decrease bacterial diversity and modulate bacterial species composition in the gut. Nevertheless, other studies have shown the impact of glyphosate on microbiomes^{14,30,31}. We may assume that long-term exposure to glyphosate residues leads to the dominance of resistant strains in the bacterial community. Moreover, some sensitive strains may become resistant to glyphosate through accumulation of mutations in the EPSPS domain or acquisition of a resistance gene *via* horizontal gene transfer. In plants (*Chloris virgata*) and bacteria (*Escherichia coli* and *Salmonella typhimurium*), changing Pro106 to Ser reduces the organism's sensitivity to glyphosate^{32,33}. Although Pro106 does not molecularly directly interact with glyphosate, its substitution with different amino acids results in structural changes in the active site, inhibiting glyphosate action³⁴. In plants, Gly101, Thr102, Pro106, Gly144 and Ala192 mutations can provide diverse levels of resistance³⁴. In *E. coli* and comparable bacteria, the corresponding positions in EPSPS (supplementary figure 7) are Gly96, Thr97, Pro101, Gly137 and Ala183.

In vitro studies have revealed that species such as *Bacteroides vulgatus*, *Bifidobacterium adolescentis*, *Enterococcus faecalis*, *Staphylococcus aureus* and *Lactobacillus buchneri* treated with Roundup UltraMax® were moderate to highly sensitive to glyphosate (minimal inhibitory concentration (MIC) = 0.075 – 0.600 mg/ml), whereas *Escherichia coli* and *Salmonella typhimurium* were relatively unreactive (MIC values = 1.200 and 5.000 mg/ml)³¹. Our results indicate that *B. vulgatus*, *B. adolescentis*, *E. coli* and *S. typhimurium* are mostly sensitive to glyphosate according to their EPSPS (table 1). Results from Shehata et al. do not disprove this, even though the last two species did show high tolerance in vitro³¹. However, *E. faecalis*, *L. buchneri* and *S. aureus*, which we predict to be resistant according to the amino acid markers in EPSPS, are actually fairly sensitive to glyphosate in vitro³¹. This suggests that some other factors can modulate the sensitivity of species to glyphosate-related products. These can include surfactants in different products (the use of Roundup UltraMax® could have toxic effects on glyphosate-resistant bacteria)

The complexity of glyphosate's effects on the gut microbiota not only is dependent on the direct impacts of glyphosate blocking (or not) the EPSPS enzyme but also has an indirect effect on bacterial

interactions. Numerous bacteria express antagonistic activity *via* bacteriocins^{31,35}. The impact of glyphosate on antagonist species is beyond the scope of this article, but it may potentially disrupt microbial biofilms³⁶. Projecting the exact effects of glyphosate on individual microbes or interactions among them in complex communities and the potential cascading effects on higher trophic levels is not straightforward. Thus, the bioinformatics tool and the precomputed dataset presented in this article will be highly useful to elucidate the impact of glyphosate on human health. To determine the actual impact of glyphosate on the human gut microbiota and other organisms, further empirical studies are needed (1) to reveal glyphosate residues in food, (2) to determine the effects of pure glyphosate and commercial formulations on microbiomes and (3) to assess the extent to which our EPSPS amino acid markers predict bacterial susceptibility to glyphosate in *in vitro* and real-world scenarios.

CONCLUSIONS

We have introduced a comprehensive classification of EPSPS proteins based on their potential sensitivity to the broad-spectrum herbicide glyphosate. This classification of organisms based on the type of EPSPS enzyme can be utilized to test several hypotheses related to the use of glyphosate and related commercial products. Moreover, this novel resource that includes a precomputed dataset of more than 50,000 proteins and a web server to determine EPSPS protein classes will be highly useful in further research. A conservative estimate from our results shows that 54% of species in the core of the human gut microbiome are sensitive to glyphosate, which represents approximately a quarter of the total number of bacterial species in the gut.

MATERIAL AND METHODS

Protein sequences and domain architectures

We obtained 10,231 protein sequences annotated 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) enzymes from the Pfam database (PF00275)³⁷. Domain architectures and taxonomical annotations were obtained from the Pfam³⁷ and Uniprot³⁸ databases, respectively. We also obtained 37 protein sequences, putatively resistant to glyphosate, from the International Survey of Herbicide-Resistant Weeds³⁹ and 679 protein sequences of prokaryotes from the database of Clusters of Orthologous Groups (COG0128)²⁷.

Alignment and phylogenetic tree construction

A phylogenetic tree of the EPSPS synthase was built from protein sequences containing at least 350 amino acids to filter out severely truncated sequences. We aligned the sequences with the program muscle (-maxiters 2)⁴⁰ and refined the alignment with the program gblocks⁴¹ with the minimal length of a block set at 6 amino acid positions, and the maximum number of allowed contiguous nonconserved amino acid positions was set at 20. The final alignment contained 152 positions from 6 selected blocks. This alignment was used to reconstruct a phylogenetic tree with the program fasttree⁴² with default parameters.

Principal components analysis of dipeptides

We obtained the frequencies of pairs of consecutive amino acids (dipeptides) of each protein sequence and performed a principal components analysis (PCA), using the *prcomp* function of the *r* software package⁴³. The resulting PCA plot was analyzed based on taxonomy and EPSPS classification.

Classification of the EPSPS

EPSPS enzymes can be classified into classes I (alpha or beta) ^{19,20}, class II ^{21,22}, III ²³ and IV ²⁴ based on the presence of amino acid markers (classes I, II and IV) and motifs (class III). Markers and motifs used to classify EPSPS enzymes into one of the four classes are based on the amino acid positions of EPSPS enzymes from *Vibrio cholerae* (vcEPSPS, class I), *Coxiella burnetii* (cbEPSPS, class II), *Brevundimonas vesicularis* (bvEPSPS, class III), and *Streptomyces davawensis* (sdEPSPS, class IV) (supplementary table 5 and 6). To classify an EPSPS enzyme, we performed pairwise alignments of the query sequence and each reference sequence (vcEPSPS, cbEPSPS, bvEPSPS and sdEPSPS). An enzyme is classified as class I, class II and/or class IV if it contains all the amino acid markers from the respective reference sequence(s) ¹⁹ and classified as class III if it contains at least one complete motif (out of 18) from the sdEPSPS sequence ^{20,24}.

Web server and datasets

A web server to determine the EPSPS class is open and freely available at <http://ppuigbo.me/programs/EPSPSClass>. The *EPSPSClass* web server requires only a query protein sequence in FASTA format to determine the putative EPSPS class (I, II, III or IV) and to calculate an identity percentage for each class. However, the server is not limited to these classes, and users can easily test their own reference sequence and amino acid markers. Moreover, a series of protein datasets have been automatically precomputed from different databases, such as the UniProt ³⁸, COG ²⁷, PDB ⁴⁴ and NCBI ⁴⁵ databases. These datasets of predictions of resistance and sensitivity to glyphosate for more than 50,000 EPSPS proteins are available from the server home page.

Search for EPSPS sequences in the human gut microbiome

The target human gut bacterial species and strains were chosen according to supplementary tables 5, 8 and 12 provided by Qin et al. 2010 ²⁹. We obtained a list of 75 nonredundant human gut bacterial species with >1% genome coverage by Illumina reads in >50% of the cohort individuals (n=124) ²⁹ and >10% genome coverage in any number of individuals. Moreover, we added EPSPS sequences from additional strains to the dataset to determine the intraspecific diversity in EPSPS. We searched for additional human gut bacteria EPSPS sequences through protein BLAST searches ⁴⁶ using vcEPSPS (supplementary table 2) as a reference. Out of the thousands of putative EPSPS sequences from the BLAST results, we selected only putative complete EPSPS sequences from bacterial strains (i.e., multispecies or partial sequences were excluded) belonging to the gut microbiome. The resulting sequences were analyzed with our *EPSPSClass* web server, which compares a query sequence against reference sequences for classes I, II, III and IV (supplementary table 2), to determine an identity percentage to each EPSPS class. Our final dataset (available from the server main page) contains the EPSPS classification for 890 strains from 101 species of prokaryotes.

ACKNOWLEDGMENTS

This research is supported by funds from the Turku Collegium for Science and Medicine (PP, TT) and by the Academy of Finland (MH, grant #311077).

REFERENCES

1. Duke, S. O. The history and current status of glyphosate. *Pest Manag. Sci.* **74**, 1027–1034 (2018).

2. Helander, M., Saloniemi, I. & Saikkonen, K. Glyphosate in northern ecosystems. *Trends Plant Sci.* **17**, 569–574 (2012).
3. Myers, J. P. *et al.* Concerns over use of glyphosate-based herbicides and risks associated with exposures: a consensus statement. *Environ. Health* **15**, 19 (2016).
4. Duke, S. O. Summing up the past year for Pest Management Science. *Pest Manag. Sci.* **73**, 7–8 (2017).
5. *Global Status of Commercialized Biotech/GM Crops in 2017: Biotech Crop Adoption Surges as Economic Benefits Accumulate in 22 Years.* (ISAAA: Ithaca, NY., 2017).
6. Steinrücken, H. C. & Amrhein, N. The herbicide glyphosate is a potent inhibitor of 5-enolpyruvylshikimate-3-phosphate synthase. *Biochemical and Biophysical Research Communications* **94**, 1207–1212 (1980).
7. Richards, T. A. *et al.* Evolutionary origins of the eukaryotic shikimate pathway: gene fusions, horizontal gene transfer, and endosymbiotic replacements. *Eukaryotic Cell* **5**, 1517–1531 (2006).
8. Bentley, R. The shikimate pathway--a metabolic tree with many branches. *Crit. Rev. Biochem. Mol. Biol.* **25**, 307–384 (1990).
9. Gilbert, J. A. *et al.* Current understanding of the human microbiome. *Nat. Med.* **24**, 392–400 (2018).
10. Wang, B., Yao, M., Lv, L., Ling, Z. & Li, L. The human microbiota in health and disease. *ENG.* **3**, 71–82 (2017).
11. Braga, R. M., Dourado, M. N. & Araújo, W. L. Microbial interactions: ecology in a molecular perspective. *Braz. J. Microbiol.* **47 Suppl 1**, 86–98 (2016).
12. Funke, T., Han, H., Healy-Fried, M. L., Fischer, M. & Schönbrunn, E. Molecular basis for the herbicide resistance of Roundup Ready crops. *Proc Natl Acad Sci USA* **103**, 13010–13015 (2006).
13. Morjan, W. E., Pedigo, L. P. & Lewis, L. C. Fungicidal effects of glyphosate and glyphosate formulations on four species of entomopathogenic fungi. *Environ. Entomol.* **31**, 1206–1212 (2002).
14. Motta, E. V. S., Raymann, K. & Moran, N. A. Glyphosate perturbs the gut microbiota of honey bees. *Proc Natl Acad Sci USA* **115**, 10305–10310 (2018).
15. Kurenbach, B., Hill, A. M., Godsoe, W., van Hamelsveld, S. & Heinemann, J. A. Agrichemicals and antibiotics in combination increase antibiotic resistance evolution. *PeerJ* **6**, e5801 (2018).
16. Kurenbach, B. *et al.* Sublethal exposure to commercial formulations of the herbicides dicamba, 2,4-dichlorophenoxyacetic acid, and glyphosate cause changes in antibiotic susceptibility in *Escherichia coli* and *Salmonella enterica* serovar Typhimurium. *MBio* **6**, (2015).
17. 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*. *J. Biol. Chem.* **284**, 9854–9860 (2009).
18. Pollegioni, L., Schönbrunn, E. & Siehl, D. Molecular basis of glyphosate resistance-different approaches through protein engineering. *FEBS J.* **278**, 2753–2766 (2011).
19. 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**, 1239–1245 (2016).
20. 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 Manag. Sci.* **74**, 1184–1196 (2018).
21. Barry, G. F., Kishore, G. M., Padgett, S. R. & Stallings, W. C. US5627061A -

- Glyphosate-tolerant 5-enolpyruvylshikimate-3-phosphate synthases. (1997).
22. 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 Lett.* **579**, 728–732 (2005).
23. Carozzi, N., Carr, B. & Hammer, P. US20060253921A1 - Identification of a new class of EPSP synthases . (2016).
24. Lira, J. M., Cicchillo, R. M. & Nair, S. K. US20130217577A1 - Novel class of glyphosate resistance genes. (2013).
25. Helander, M. *et al.* Glyphosate decreases mycorrhizal colonization and affects plant-soil feedback. *Sci. Total Environ.* **642**, 285–291 (2018).
26. El-Gebali, S. *et al.* The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432 (2019).
27. 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 Res.* **43**, D261–9 (2015).
28. Tatusov, R. L., Koonin, E. V. & Lipman, D. J. A genomic perspective on protein families. *Science* **278**, 631–637 (1997).
29. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).
30. Gómez-Gallego, C. *et al.* Glyphosate-based herbicide affects the composition of microbes associated with Colorado potato beetle (*Leptinotarsa decemlineata*). *FEMS Microbiol. Lett.* **367**, (2020).
31. Shehata, A. A., Schrödl, W., Aldin, A. A., Hafez, H. M. & Krüger, M. The effect of glyphosate on potential pathogens and beneficial members of poultry microbiota in vitro. *Curr. Microbiol.* **66**, 350–358 (2013).
32. Heap, I. & Duke, S. O. Overview of glyphosate-resistant weeds worldwide. *Pest Manag. Sci.* **74**, 1040–1049 (2018).
33. *Glyphosate resistance in crops and weeds: history, development, and management.* (John Wiley & Sons, Inc., 2010). doi:10.1002/9780470634394
34. Sammons, R. D. & Gaines, T. A. Glyphosate resistance: state of knowledge. *Pest Manag. Sci.* **70**, 1367–1377 (2014).
35. Juturu, V. & Wu, J. C. Microbial production of bacteriocins: Latest research development and applications. *Biotechnol. Adv.* **36**, 2187–2200 (2018).
36. van Overbeek, L. S. & Saikkonen, K. Impact of Bacterial-Fungal Interactions on the Colonization of the Endosphere. *Trends Plant Sci.* **21**, 230–242 (2016).
37. Finn, R. D. *et al.* The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **44**, D279–85 (2016).
38. The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **45**, D158–D169 (2017).
39. Heap, I. The International Survey of Herbicide Resistant Weeds. at <<http://www.weedscience.com>>
40. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
41. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552 (2000).
42. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 — approximately maximum-likelihood trees

- for large alignments. *PLoS ONE* **5**, e9490 (2010).
43. R: The R Project for Statistical Computing. at <<https://www.r-project.org/>>
 44. Burley, S. K. *et al.* Protein data bank (PDB): the single global macromolecular structure archive. *Methods Mol. Biol.* **1607**, 627–641 (2017).
 45. Benson, D. A. *et al.* GenBank. *Nucleic Acids Res.* **41**, D36–42 (2013).
 46. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

FIGURE LEGENDS

Figure 1. Position of amino acid markers in the reference EPSPS sequences

Amino acid markers of EPSPS classes I (EPSPS from *Vibrio cholerae*; vcEPSPS), II (EPSPS from *Coxiella burnetii*; cbEPSPS), III (EPSPS from *Brevundimonas vesicularis*; bvEPSPS) and IV (EPSPS *Streptomyces davawensis*; sdEPSPS) are shown in black. The complete list of amino acids is found in supplementary table 1 and supplementary figure 2. This figure was made with the web server <https://matrix2png.msl.ubc.ca/>.

Figure 2. Taxonomic distribution and classification of EPSPS proteins

- a) Percentage of the EPSPS proteins in prokaryotic sequences within each taxon. EPSPS is present in 652 out of the 678 genomes analyzed. The data were obtained from the COG database. EPSPS proteins belong to COG0128 (category E; amino acid transport and metabolism).
- b) Percentages of classes I, II, III, and IV and unclassified (none) in prokaryotes (source of sequences: COG database).
- c) Percentages of classes I, II, III, and IV and unclassified (none) in eukaryotes, fungi and viridiplantae (source of sequences: UniProt database).

Figure 3. Identification of the EPSPS class in endophytes.

The EPSPS enzymes of endophytes are mapped onto the reference sequences (class I, vcEPSPS; class II, cbEPSPS; class III, bvEPSPS; and class IV, sdEPSPS) to determine the potential sensitivity to glyphosate. Endophyte EPSPS sequences were obtained from <http://www.endophyte.uky.edu>.

Figure 4. Principal components analysis (PCA) plot of dipeptides.

The plot shows the first and second components of the PCA from ~10,000 EPSPS proteins of archaea (blue), bacteria (green) and eukaryotes (red). Additional clusters from this plot, based on taxonomy and EPSPS class, are available in supplementary figures 3–4.

Figure 5. Phylogenetic tree of the EPSPS domain made from ~10,000 sequences (>350 AA) obtained from the Uniprot database.

- a) Phylogenetic tree including archaea (blue), bacteria (green) and eukaryotes (red)
- b) Phylogenetic tree of the fungal clade shows the conservation of the multidomain structure of the EPSPS protein.
- c) Example of a common multidomain architecture in a fungal EPSPS protein that includes the domains EPSPS (PF00275) and its associated domains PF01488 (Shikimate_DH), PF08501 (Shikimate_DH_N), PF01487 (DH_Quinase_I), PF01202 (SKI) and PF01761 (DHQS). The majority of bacteria and plants contain only the EPSPS domain (PF00275).

Table 1. List of common gut bacteria and their susceptibility to glyphosate in a taxonomic level of species. Species sensitive and resistant to glyphosate are highlighted in green and red respectively. The species that showed intraspecific variation are highlighted in blue and unclassified species are highlighted in black. If the EPSPS of a strain is not of any Class, the sensitivity is unknown. The total number of sequences is N=890.

Sensitive to glyphosate (Class I, 100%)		Resistant to glyphosate (Class II , 100%)	Resistant to glyphosate (Class III, 100%)
<i>Actinomyces odontolyticus</i> (1)	<i>Parabacteroides merdae</i> (8)	<i>Anaerobaculum hydrogeniformans</i> (1)	<i>Alistipes putredinis</i> (3)
<i>Anaerofustis stercorihominis</i> (1)	<i>Bifidobacterium breve</i> (28)	<i>Blautia hansenii</i> (3)	
<i>Anaerotruncus colihominis</i> (5)	<i>Bifidobacterium catenulatum</i> (8)	<i>Bryantella formatexigens</i> (1)	Some sensitive and resistant strains
<i>Bacteroides caccae</i> (5)	<i>Bifidobacterium dentium</i> (5)	<i>Clostridium asparagiforme</i> (2)	(Class I , 96% / Class II, 4%)
<i>Bacteroides capillosus</i> (1)	<i>Bifidobacterium longum</i> (102)	<i>Clostridium bartlettii</i> (2)	<i>Bifidobacterium adolescentis</i> (28): 27S + 1R
<i>Bacteroides cellulosilyticus</i> (4)	<i>Catenibacterium mitsuokai</i> (3)	<i>Clostridium scindens</i> (2)	
<i>Bacteroides coprocola</i> (2)	<i>Citrobacter portucalensis</i> (5)	<i>Clostridium</i> sp. SS2/1 (2)	Some sensitive and unclassified strains
<i>Bacteroides coprophilus</i> (2)	<i>Clostridium methylpentosum</i> (1)	<i>Clostridium symbiosum</i> (1)	(Class I, 94% / No Class, 6%)
<i>Bacteroides eggerthii</i> (3)	<i>Clostridium nexile</i> (1)	<i>Coprococcus comes</i> (9)	<i>Fusobacterium nucleatum</i> (2): 1S +1U
<i>Bacteroides finegoldii</i> (4)	<i>Collinsella aerofaciens</i> (9)	<i>Coprococcus eutactus</i> (6)	<i>Bifidobacterium pseudocatenulatum</i> (46): 45S + 1U
<i>Bacteroides fragilis</i> (56)	<i>Desulfovibrio piger</i> (4)	<i>Dorea formicigenerans</i> (18)	<i>Clostridium leptum</i> (5): 4S + 1U
<i>Bacteroides intestinalis</i> (11)	<i>Enterobacter cancerogenus</i> (1)	<i>Dorea longicatena</i> (11)	
<i>Bacteroides ovatus</i> (14)	<i>Escherichia coli</i> (112)	<i>Enterococcus casseliflavus</i> (2)	Some resistant and unclassified strains
<i>Bacteroides plebeius</i> (11)	<i>Eubacterium siraeum</i> (4)	<i>Eubacterium hallii/Anaerobutyrum hallii</i> (9)	(Class II, 79% / No Class, 21%)
<i>Bacteroides</i> sp. 2_2_4 (1)	<i>Faecalibacterium prausnitzii</i> (53)	<i>Helicobacter canadensis</i> (1)	<i>Clostridium</i> sp. L2/50 (2): 1R + 1U
<i>Bacteroides</i> sp. 3_2_5 (1)	<i>Fusobacterium mortiferum</i> (1)	<i>Helicobacter cinaedi</i> (2)	<i>Enterococcus faecalis</i> (4): 3R + 1U
<i>Bacteroides</i> sp. 4_3_47FAA (1)	<i>Fusobacterium ulcerans</i> (1)	<i>Helicobacter pullorum</i> (1)	<i>Ruminococcus torques</i> (8): 7R + 1U
<i>Bacteroides</i> sp. 9_1_42FAA (2)	<i>Parabacteroides</i> sp. 2_1_7 (1)	<i>Helicobacter winghamensis</i> (1)	
<i>Bacteroides</i> sp. D1 (1)	<i>Prevotella copri</i> (27)	<i>Lactobacillus brevis subsp. gravesensis</i> (1)	Unclassified strains (No Class, 100%)
<i>Bacteroides thetaiotaomicron</i> (13)	<i>Providencia alcalifaciens</i> (1)	<i>Lactobacillus buchneri</i> (1)	<i>Bacteroides pectinophilus</i> (1)
<i>Bacteroides uniformis</i> (10)	<i>Providencia rustigianii</i> (1)	<i>Lactobacillus hilgardii</i> (1)	<i>Butyrivibrio crossotus</i> (2)
<i>Bacteroides vulgatus</i> (6)	<i>Providencia stuartii</i> (1)	<i>Lactobacillus plantarum</i> (2)	<i>Clostridium spiroforme</i> (1)
<i>Bifidobacterium bifidum</i> (22)	<i>Ruminococcus bromii</i> (14)	<i>Lactobacillus ruminis</i> (2)	<i>Eubacterium rectale</i> (26)
<i>Fusobacterium varium</i> (1)	<i>Subdoligranulum variabile</i> (3)	<i>Listeria grayi</i> (1)	<i>Eubacterium ventriosum</i> (7)
<i>Holdemania filiformis</i> (1)		<i>Mitsuokella multacida</i> (3)	<i>Fusobacterium necrophorum</i> (1)
<i>Mollicutes bacterium</i> (3)		<i>Ruminococcus gnavus</i> (17)	<i>Fusobacterium periodonticum</i> (1)
<i>Parabacteroides distasonis</i> (10)		<i>Ruminococcus lactaris</i> (3)	<i>Proteus penneri</i> (1)
<i>Parabacteroides johnsonii</i> (4)		<i>Ruminococcus obeum</i> (18)	<i>Providencia rettgeri</i> (1)
			<i>Roseburia intestinalis</i> (9)

(n): Number of sequences; R: glyphosate resistant sequences; S: glyphosate sensitive sequences, U: unclassified.

Figure 1

Figure 2

Figure 3

Figure 4

Figure 5

a)

Eukaryotes 1
Viridiplantae (n = 78)
Stramenopiles (n = 5)
Rodophyta (n = 2)

Eukaryotes 2
Fungi (n = 364)
Stramenopiles (n = 19)
Other (n = 8)

b)

c)