The Glyphosate Target Enzyme 5-Enolpyruvyl Shikimate 3-Phosphate Synthase (EPSPS) Contains Several EPSPS-Associated Domains in Fungi †

Tuomas Tall [1,*] and Pere Puigbò [2,3]

[1] Department of Biology, University of Turku; pepuav@utu.fi or pere.puigbo@eurecat.org or ppuigbo@urv.cat

[2] Nutrition and Health Unit, Eurecat Technology Center of Catalonia

[3] Department of Biochemistry and Biotechnology, Rovira i Virgili University

[*] Correspondence: teltal@utu.fi

† Presented at the 1st International Electronic Conference on Genes: Theoretical and Applied Genomics, 2–30 November 2020; Available online: https://iecge.sciforum.net/.

Abstract: The 5-enolpyruvylshikimate 3-phosphate synthase (EPSPS) is the central enzyme of the shikimate pathway to synthesize three aromatic amino acids in fungi, plants and prokaryotes. This enzyme is the target of the herbicide glyphosate. In most plants and prokaryotes, the EPSPS protein is constituted by a single domain, whereas in fungi contains several EPSPS-associated domains. Here, we perform a comprehensive analysis of 390 EPSPS proteins of fungi to determine the distribution and the evolution of the EPSPS-associated domains. The results of this study will be useful to determine the potential differential impact of glyphosate on alternative domain architectures in fungi.

1. Introduction

Glyphosate, the most used herbicide against weeds, and glyphosate-based products (GBP) target the enzyme 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) [1,2]. The EPSPS (a.k.a. as aroA) is the central enzyme in the shikimate pathway for the synthesis of three essential amino acids [3]. The enzyme is present in plants and prokaryotes as single-domain protein, and in fungi as multi-domain protein [5]. Due to the enzyme is not found in animals, the use of the glyphosate is supposed to be safe for human health. Even if this is the case, the herbicide may still affect the biodiversity of free-living and host-associated microorganisms [7–10]. Thus, the study of the glyphosate target enzyme will provide clues about the potential effect of the herbicide. In recent studies, we analyzed the distribution of the EPSPS protein in organisms and we estimated the differential sensitivity of organisms to the herbicide [7,8].

In this study, we analyze the distribution and evolution of the EPSPS and EPSPS-associated domains in fungi. In plants and prokaryotes, the EPSPS is a single-domain, thus the potential sensitivity to the herbicide glyphosate depends, in addition to levels of gene expression and other cell factors [11–13], on the type of amino acids in the EPSPS protein. However, the effect of the herbicide on the multi-domain structure of the EPSPS in fungi is yet unclear. This survey of the EPSPS and EPSPS-associated domains will be helpful to determine the potential effect of glyphosate on fungal species.

2. Material and Methods

We obtained protein data of the EPSPS and EPSPS-associated domains from the PFAM database (http://pfam.xfam.org) [14]. First, we analyzed frequencies of EPSPS-associated domains across 390 fungal species, and then we summarized the main domain function in a Venn's diagram. We also analyzed the distribution of the associated domains by taxonomical groups in fungi. A phylogenetic tree of the EPSPS domain was built with the program FastTree2 [15] after aligning the protein sequences with the programs MUSCLE [16] and Gblocks [17]. The program Cytoscape [18] was used to reconstruct a bipartite network of protein domains and fungal species. The resultant network was used to visualize presence and absence of domains and the distribution of the different architectures of the EPSPS multi-domain. The analysis of the evolution of the EPSPS-associated domains in fungi was performed by Dollon parsimony with the program Count [19].

3. Results

3.1. Functional and Domain Characterization

The EPSPS domain is approximately 450 amino acids long (~1350 nucleotides) and is present (by definition) in all EPSPS proteins [7]. The EPSPS-associated domains can be classified in four partially overlapping groups (Figure 1): shikimate (shikimate pathway proteins), enzymes (proteins with catalytic function), expression (domains whose products are needed in controlling gene expression) and structural function (proteins that do not have a catalytic). The function of the EPSPS enzyme is to transfer alkyls aryl (other than methyl) groups. It transfers enolpyruvate from phosphoenolpyruvate to 3-phosphoshikimate [20]. The analysis of the EPSPS-associated domains across fungal species shows that are mostly involved in the shikimate pathway (e.g., SKI, DHQ_synthase, DHquinase_I) and in the synthesis of aromatic amino acids (e.g., Shikimate_dh_N, PDH). There are also some promiscuous domains (e.g., HTH_3) associated to the EPSPS in some fungal species. The least frequent domains, mainly involved in DNA modification and gene expression, are amply distributed across fungal species.
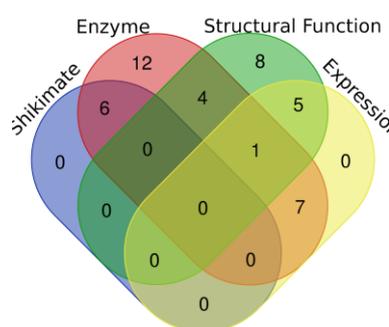

Figure 1. A Venn's diagram of the EPSPS-associated domains. Blue is Shikimate pathway, red is enzymes, green is structural function and yellow is gene expression. The diagram was made with the following tool: http://bioinformatics.psb.ugent.be/webtools/Venn/.

3.2. A Bipartite Network of Species and EPSPS-Associated Domains

A total of 22 domains are associated with the EPSPS protein in fungal EPSPS protein architectures, across 390 fungal species. The most frequent EPSPS-associated domains, which are involved in the shikimate pathway, are: SKI (Shikimate kinase) (n = 374); DHQ synthase (3-dehydroquinate synthase) (n = 374); DHquinase (3-dehydroquinate dehydratase) (n = 367); Shikimate dh N (Shikimate dehydrogenase substrate binding domain) (n = 366); and Shikimate DH (Shikimate/quinate 5-dehydrogenase) (n = 136). There are also 16 domains that are only present one or two proteins.

The bipartite network (Figure 2) connects fungal species divided by order (ascomycota in the left and basidiomycota, mucoromycota, chytridiomycota, blastocladiomycota, zoopagomycota and

unknown in the right). The EPSPS and most common EPSPS-associated domains are located in the center of the figure in the same order as in domains. Other domains were set on the outside of the network. Ascomycota seems to be the most variable phylum, in terms of EPSPS-associated domains and contains several unfrequent domains. Moreover, it is the phylum that contains the least number of Shikimate DH domains, e.g.,: the domain is infrequent in Eurotiomycetes (APE), Dothideomycetes (APD) and Leotiomycetes (APL). Majority of multi-domain architectures (n = 220) are five domains long, which are involved in the shikimate pathway and approximatelly 1/3 of the sequences (n = 134) have all six domains of the shikimate pathway. Overall, the bipartite network shows that the EPSPS domain in fungi is mostly associated to other domains of the shikimate pathway. However, there are few exceptions (e.g., Dothideomycetes).
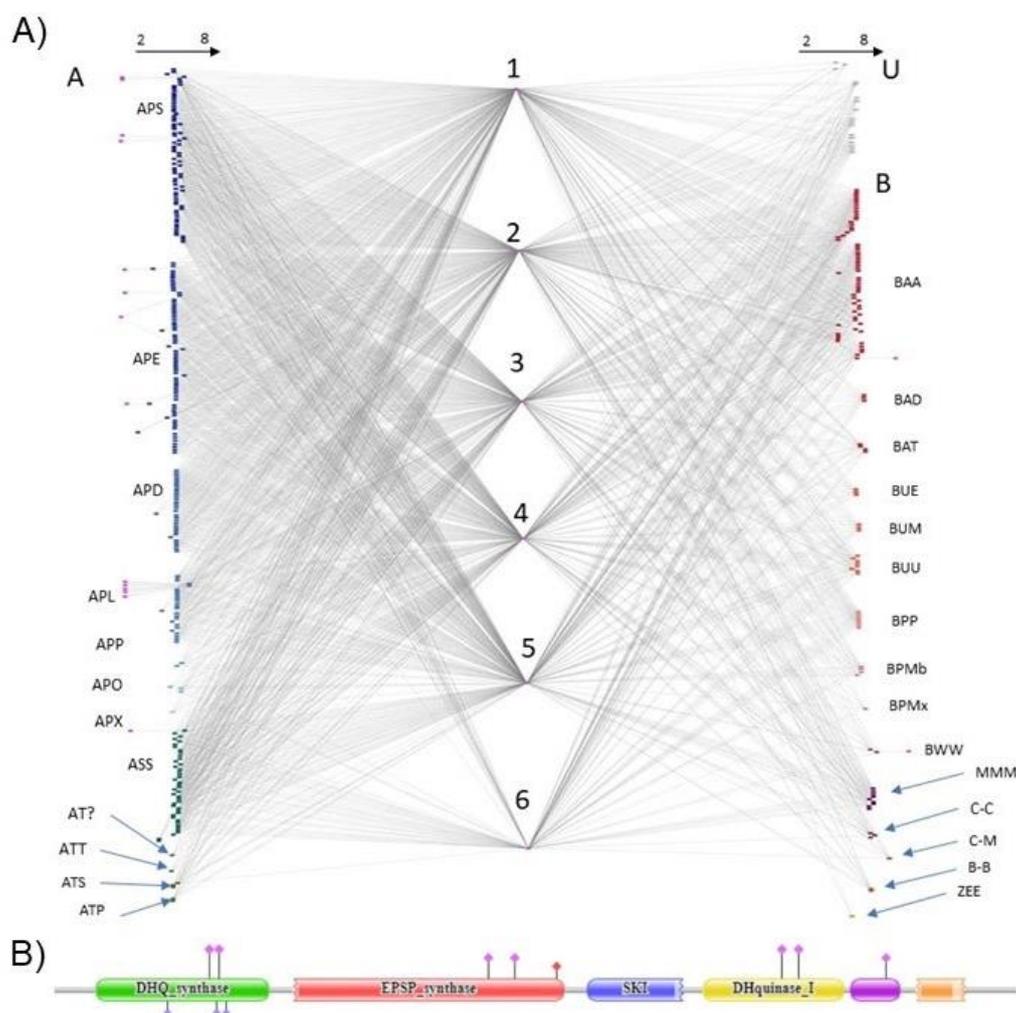

Figure 2. A bipartite network of fungal species and EPSPS-associated domains. (**A**) A bipartite network of fungal samples. Letters stands for different phylums and classes of fungi. On the left side there are fungi of the phylum ascomycota (A) and on the right side there are fungi of phylum Basidiomycota (B) Mucoromycota (M), Chytridiomycota (C), Blastocladiomycota (Bl) and Zoopagomygota (Z), Unknown taxa (U). Numbers correspond to fungal multi-domain: (1) DH synthase (2) EPSP synthase (3) SKI (4) DH quinase type 1 (5) Shikimate dh N and (6) Shikimate DH. Arrows on the top of the figure note the number of domains in multi-domain, which ranges between 2 to 8 domains. Subphylum and class starting from upper left are Pezizomycotina: Sordariomycetes (APS), Eurotiomycetes (APE), Dothideomycetes (APD), Leotiomycetes (APL), Pezizomycetes (APP), Orbiliomycetes (APO) and Xylonomycetes (APX); Saccharomycotina: Saccharomycetes (ASS); Taphrinomycotina: incertae sedis (AT?), Taphrinomycetes (ATT), Schizosaccharomycetes (ATS) and Pneumocystidomycetes (ATP); Agariomycotiina: Agaricomycetes (BAA), Dacrymycetes (BAD) and Tremellomycetes (BAT); Ustilaginomycotina: Exobasidiomycetes (BUE), Malasseziomycetes (BUM)

and Ustilaginomycetes (BUU); Pucciniomycotina: Pucciniomycetes (BPP), Microbotryomycetes (BPMb) and Mixiomycetes (BPMx); Wallemiomycotina: Wallemiomycetes (BWW); Mucoromycotina: Mucoromycetes (MMM); Chytridiomycota: Chytridiomycetes (C-C) and Monoblepharidomycetes (C-M); Blastocladiomycota: Blastocladiomycetes (B-B); Entomophthoromycotina: Entomophthoromycetes (ZEE). A detailed analysis of the bipartite network can be found in appendix III. (**B**) Example of a multi-domain EPSPS protein of 1616 amino acids (protein A0A060S9A7_PYCCI from *Pycnoporus cinnabarinus*). In this example, the SKI and Shikimate DH are fractured.

3.3. Maximum Parsimony Analysis of EPSPS-Associated Domains

Given the Shikimate DH domain is found in most of species, it is likely that it was present in the common ancestor of fungi and lost in some lineages. Moreover, rarer domains are late inclusions to the multi-domain structure. We analyzed evolution of all EPSPS-associated domains with Dollon parsomony (Figure 3).
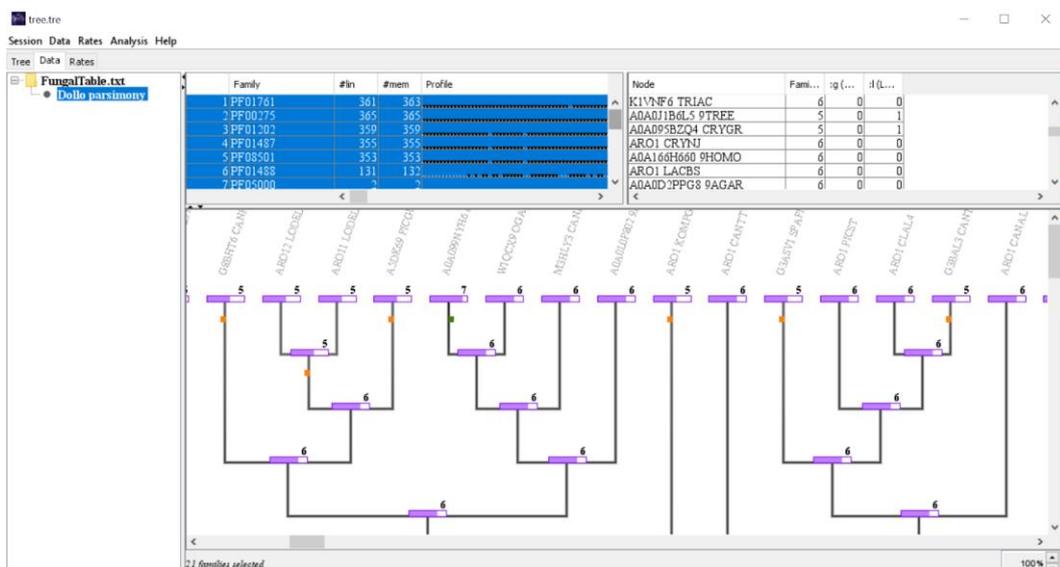

Figure 3. Screenshot of the output of the program COUNT [19]. The image shows a part of phylogenetic tree (center); domain family profile (top left) and node information (top right). The multi-domain structure of the EPSPS in fungi is ancestral (6 domains) and there are recent domain gains and losses.

The Shikimate DH domain has been lost multiple times in different branches of the phylogenetic tree (Figure 3). In addition, many domains have been lost during the evolution of fungi and some new domains have been acquired recently, and independently, in few species, and other were acquired in internal lineages and subsequently lost in few branches. Notice that our analysis does not take into account domain duplications. The conclusion from this Dollon parsimony analysis is that a six- multi-domain struture (i.e., a protein sequence with all six most abundant domains) was the original EPSPS sequence in fungi (Table 1). Moreover, we suggest the hypothesis that sequences with the five multi-domains originated by loss of shikimate DH independently in different branches of the evolutionary tree. These losses occurred at early and late stages in the evolution of. In some EPSPS proteins, the shikimate DH (located at the C-term of the protein) is truncated. It is possible that this domain was lost in a crossing over event without affecting the shikimate pathway, however we do not know if the fragment is found in any other part of genome. Also, smaller multi-domains originated from loss of other shikimate domains, usually in higher branches. Moreover, inclusion of rare domains also occurred in higher branches.

Table 1. EPSPS and most frequent EPSPS-associated domains.

Gene	Frequency	Species	Function
EPSPS	1448	8833	Sixth step of **shikimate pathway (SP)**, EPSP Synthase
SKI	424	7716	First step of the SP pathway, phosphorylates shikimate
DHQ_synthase	420	7429	Second step of the SP pathway, removes a phosphate from DHAP
DHquinase_I	416	2073	Third step of the SP pathway, 3-dehydroquinate dehydratase
Shikimate_dh_N	402	7658	The substrate binding domain of the shikimate dehydrogenase
HTH_3	218	9670	Helix-turn-helix, a major structural motif capable of binding DNA
Shikimate_DH	160	6719	Fourth step of the SP pathway, quinate 5-dehydrogenase
PDH	127	7136	Prephenate dehydrogenases are part of Tyrosine biosynthesis.
Cytidylate_kin	88	6874	Kinase of cytidine 5'-monophosphate
PF13193	17	8031	AMP-binding enzyme C-terminal domain for PF00501

Gene: gene names; Frequency: frequency of the gene in the sample; Species: number of species with at least one copy of the domain; and function: description of the product.

4. Conclusions

The EPSPS multi-domain structure in fungi ranges between two to eight domains. An evolutionary analysis shows that the ancestral state of the EPSPS protein included six domains (DHQ synthase, EPSPS, SKI, DHquinase I, Shikimate DH N and Shikimate DH). The current multi-domain structure of the EPSPS in fungi is the product of multiple independent domain gains and losses throughout evolution. Further analyses should determine the importance of the EPSPS-associated domains on the sensitivity/resistance of the EPSPS enzyme to glyphosate.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Bentley, R.; Haslam, E. The shikimate pathway—A metabolic tree with many branche. *Crit. Rev. Biochem. Mol. Biol.* **1990**, *25*, 307–384, doi:10.3109/10409239009090615.

2. Richards, T.A.; Dacks, J.B.; Campbell, S.A.; Blanchard, J.L.; Foster, P.G.; McLeod, R.; Roberts, C.W. Evolutionary origins of the eukaryotic shikimate pathway: Gene fusions, horizontal gene transfer, and endosymbiotic replacements. *Eukaryot. Cell*. **2006**, *5*, 1517–1531, doi:10.1128/EC.00106-06.

3. Maeda, H.; Dudareva, N. The Shikimate Pathway and Aromatic Amino Acid Biosynthesis in Plants. *Annu. Rev. Plant. Biol*. **2012**, *63*, 73–105, doi:10.1146/annurev-arplant-042811-105439.

4. Steinrücken, H.C.; Amrhein, N. The herbicide glyphosate is a potent inhibitor of 5-enolpyruvylshikimic acid-3-phosphate synthase. *Biochem. Biophys. Res. Commun*. **1980**, *94*, 1207–1212, doi:10.1016/0006-291X(80)90547-1.

5. Lumsden, J.; Coggins, J.R. The Subunit Structure of the Arom Multienzyme Complex of Neurospora Crassa. *Biochem. J.* **1997**, *161*. Available online: https://sciwheel.com/work/item/8971159/resources/8495092/pdf (accessed on 9 October 2020).

6. Basu, M.K.; Carmel, L.; Rogozin, I.B.; Koonin, E.V. Evolution of protein domain promiscuity in eukaryotes. *Genome Res*. **2008**, *18*, 449–461, doi:10.1101/gr.6943508.

7. Leino, L.; Tall, T.; Helander, M.; Saloniemi, I.; Saikkonen, K.; Ruuskanen, S.; Puigbò, P. Classification of the glyphosate target enzyme (5-enolpyruvylshikimate-3-phosphate synthase) Equal contribution of the authors. *bioRxiv* **2020**, doi:10.1101/2020.05.27.118265.

8. Rainio, M.J.; Ruuskanen, S.; Helander, M.; Saikkonen, K.; Saloniemi, I.; Puigbo, P. Adaptation of bacteria to glyphosate: A microevolutionary perspective of the enzyme 5-enolpyruvylshikimate-3-phosphate (EPSP) synthase. *bioRxiv* **2020**, doi:10.1101/2020.06.16.154005.

9. Balbuena, M.S.; Tison, L.; Hahn, M.L.; Greggers, U.; Menzel, R.; Farina, W.M. Effects of sublethal doses of glyphosate on honeybee navigation. *J. Exp. Biol*. **2015**, *218*, 2799–2805, doi:10.1242/jeb.117291.

10. Vandenberg, L.N.; Blumberg, B.; Antoniou, M.N.; Benbrook, C.M.; Carroll, L.; Colborn, T.; Everett, L.G.; Hansen, M.; Landrigan, P.J.; Lanphear, B.P.; et al. Is it time to reassess current safety standards for glyphosate-based herbicides? *J. Epidemiol. Community Health*. **2017**, *71*, 613–618, doi:10.1136/jech-2016-208463.

11. Burchfield, S.L.; Bailey, D.C.; Todt, C.E.; Denney, R.D.; Negga, R.; Fitsanakis, V.A. Acute exposure to a glyphosate-containing herbicide formulation inhibits Complex II and increases hydrogen peroxide in the model organism Caenorhabditis elegans. *Environ. Toxicol. Pharmacol*. **2019**, *66*, 36–42, doi:10.1016/j.etap.2018.12.019.

12. Nerozzi, C.; Recuero, S.; Galeati, G.; Bucci, D.; Spinaci, M.; Yeste, M. Effects of Roundup and its main component, glyphosate, upon mammalian sperm function and survival. *Sci. Rep*. **2020**, *10*, 11026, doi:10.1038/s41598-020-67538-w.

13. Gomes, M.P.; Le Manac'h, S.G.; Hénault-Ethier, L.; Labrecque, M.; Lucotte, M.; Juneau, P. Glyphosate-dependent inhibition of photosynthesis in willow. *Front. Plant Sci*. **2017**, *8*, 207, doi:10.3389/fpls.2017.00207.

14. Sonnhammer, E.L.L.; Eddy, S.R.; Durbin, R. Pfam: A comprehensive database of protein domain families based on seed alignments. *Proteins Struct. Funct. Genet*. **1997**, *28*, 405–420, doi:10.1002/(SICI)1097-0134(199707)28:3<405::AID-PROT10>3.0.CO;2-L.

15. Price, M.N.; Dehal, P.S.; Arkin, A.P. FastTree 2—Approximately maximum-likelihood trees for large alignments. *PLoS ONE*. **2010**, *5*, e9490, doi:10.1371/journal.pone.0009490.

16. Edgar, R.C. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinform*. **2004**, *5*, 113, doi:10.1186/1471-2105-5-113.

17. Castresana, J. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Mol. Biol. Evol*. **2000**, *17*, 540–552, doi:10.1093/oxfordjournals.molbev.a026334.

18. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Res*. **2003**, *13*, 2498–2504, doi:10.1101/gr.1239303.

19. Csurös, M. Ancestral reconstruction by asymmetric Wagner parsimony over continuous characters and squared parsimony over distributions. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Berlin/Heidelberg, Germany, 2008; Volume 5267 LNBI, pp. 72–86, doi:10.1007/978-3-540-87989-3_6.

20. Jaworski, E.G. Mode of action of N-phosphonomethylglycine. Inhibition of aromatic amino acid biosynthsis. *J. Agric. Food Chem*. **1972**, *20*, 1195–1198, doi:10.1021/jf60184a057.