

Evolution of the YUCCA protein family: phylogenetic relationships between the YUCCA and Type IIb FMO families

Turnaev I.I.^{1,2*}, Gunbin K.V.¹, Suslov V.V.¹, Akberdin I.R.^{1,3,4}, Kolchanov N.A.^{1,2,4}, Afonnikov D.A.^{1,2,4}

¹ *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

² *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

³ *BIOSOFT.RU, LLC, Novosibirsk, Russia*

⁴ *Novosibirsk State University, Novosibirsk, Russia*

* turn@bionet.nsc.ru

Key words: Auxin biosynthesis, YUCCA, flavin-dependent monooxygenase, phylogeny, charophytes, horizontal gene transfer

Motivation and Aim: YUCCA is the second of two enzymes in the auxin biosynthesis pathway (TAA/YUCCA) in terrestrial plants. The origin of the YUCCA family is the subject of intense debate: YUCCAs appeared as a result of horizontal gene transfer (HGT) from bacteria to the most recent common ancestor (MRCA) of land plants [2] or to MRSA of charophyte algae and land plants [3]. This necessitated an in-depth comparative analysis of the YUCCA proteins and their closest protein families.

Methods and Algorithms: Two samples of YUCCA homologues were created. The first sample includes only protein sequences. Protein sequences homologous to YUCCA were recognized in db PlazaPlant 2.5, NCBI, and the *Picea abies* genome project (Spruce Genome Project) using BlasP. Proteins from Riebel et al., [1], namely the seven amino acid sequences FMO-A to FMO-G of *Rhodococcus jostii* *RHA1* bacteria and FMO-X of *Stenotrophomonas maltophilia* bacteria were added. The second sample includes protein sequences from the first sample supplemented with transcriptome sequences. The transcriptome sequences of YUCCA homologues were recognized among the sequences of the 1000 Plant Genomes Project and Green Algal Tree of Life Project using the BlasP program. Multiple sequence alignment was performed using the Promals program. IQ-TREE 1.6.12, RAxML 8.1.24, and mrBayes 3.2.7 were used to reconstruct the phylogenetic tree.

Results: Phylogenetic trees for an extended sample including both protein and transcriptome sequences of YUCCA homologues were reconstructed by IQ-TREE (Fig. 1. FMO_prot&transcr-IQ-TREE tree cladogram), RAxML (Fig. 1. FMO_prot&transcr-RAxML) and mrBayes (Fig. 1. FMO_prot&transcr-RaxML). Phylogenetic trees from a sample of YUCCA homolog protein sequences were reconstructed using IQ-TREE (Fig. 1. FMO_prot-IQ-TREE), RAxML (Fig. 1. FMO_prot-RAxML) and mrBayes (see Fig. 1, FMO_prot-RaxML) programs (see Fig. 1). Class G flavoprotein monooxygenase proteins were used as an outgroup for all trees. In all 6 phylogenetic trees (see Fig. 1), 10 groups of protein or mRNA sequences (excluding the outgroup class G FMOs, a total of 9 groups) belonging to three subclasses of class B FMO are well distinguished. 1) FMO subclass, includes groups: YUCCAs, only plant sequences are represented; YUC-like FMOs (only bacteria); cyanobacterial FMOs (cyanobacteria); FMO type IIa and FMO type IIc (both bacteria only); FMO type IIb includes bacteria, plants, fungi and protists; FMO type I (bacteria, protists, plants and

animals sequences). 2) Subclass NMO contains one NMO group (bacteria and fungi). 3) Subclass, BVMOs (fungi, protists and bacteria).

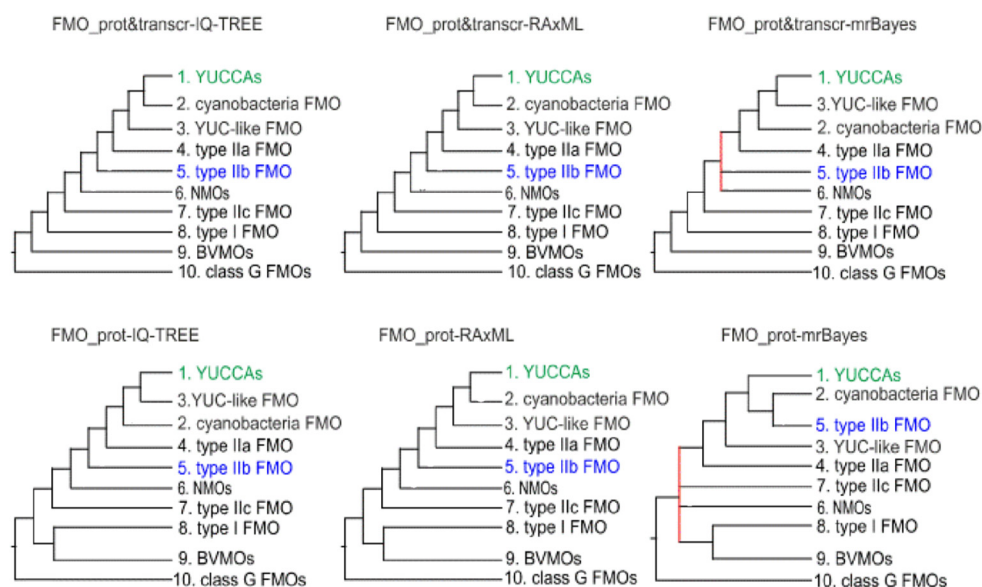


Fig. 1. Simplified cladograms for class B flavoprotein monooxygenases proteins. The type IIb FMO and YUCCAs families are highlighted in blue and green text respectively. Red line highlighted polyfurcations

The type IIb FMO clade (highlighted in blue text in Fig. 1) also contains sequences of plant organisms: four sequences of the lycophyte *Selaginella moellendorffii* and a YUCCA homologue from the charophyte alga *Klebsormidium nitens*, GAQ82387.1.

The type IIb FMO clade on all trees (see Fig. 1) except the FMO_prot-mrBayes tree is not sister to the YUCCA clade (highlighted in green text in fig. 1). The type IIb FMO clade in the FMO_prot-mrBayes tree is not uniquely defined as sister to the YUCCA clade. In addition, given the tetrafurcation in the FMO_prot-mrBayes tree, its phylogeny is much less reliable than that of the other 5 trees to Fig. 1. Additional analysis of the domain structure, active sites, and distribution in taxa also showed a significant difference between the clade type IIb FMO and YUCCA proteins. Thus, our results support the hypothesis of Yue et al. [2] on the origin of YUCCA as a result of horizontal transfer of the ancestral FMO gene from bacteria to MRSA of all land plants and does not support the hypothesis of Wang et al. [2] about an earlier origin of YUCCA in charophytic algae.

Acknowledgements: The study is supported by the Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS (075-15-2019-1662).

References

1. Riebel A., de Gonzalo G., Fraaije M.W. Expanding the biocatalytic toolbox of flavoprotein monooxygenases from *Rhodococcus jostii* RHA1. *J Mol Catal B Enzym.* 2013;88:20-25.
2. Yue J., Hu X., Huang J. Origin of plant auxin biosynthesis. *Trends Plant Sci.* 2014;19:764-770.
3. Wang C., Liu Y., Li S.-H., Guan-Zhu Han G.-Z. Origin of plant auxin biosynthesis in charophyte algae. *Trends Plant Sci.* 2014;19:741-743.