

FMO SUPERFAMILY PROTEIN PHYLOGENY AND THE ORIGIN OF YUCCA FAMILY.

Turnaev I.I.*¹, Suslov V.V.¹, Gunbin K.V.¹, Afonnikov D.A.¹.

¹ Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia.

* Corresponding author e-mail: turn@bionet.nsc.ru

In order to clarify the origin of YUCCAs family proteins, we constructed six variants of the phylogenetic tree for class B flavoprotein monooxygenases proteins to 2 samples (only proteins – “FMO_prot” and proteins + transcriptomic sequences – “FMO_prot&transcr”) using three programs IQ-TREE, RAxML and mrBAYES. Samples were constructed using p-BLAST search in NCBI, Plaza plant 2.5 databases and 1000 plant genomes, Green Algal Tree of Life transcriptome projects. AT4G13260 in *Arabidopsis thaliana* and GAQ82387.1 in *Klebsormidium nitens* were selected as the initial sequences for searching for homologues. Tree names are described in table 1.

Table 1. Names of phylogenetic trees for class B flavoprotein monooxygenases.

программы	выборки	выборка белки+транскрипты – FMO_prot&transcr	выборка белки – FMO_prot
IQ-TREE		FMO_prot&transcr-IQ-TREE	FMO_prot-IQ-TREE
RAxML		FMO_prot&transcr- RAxML	FMO_prot- RAxML
mrBAYES		FMO_prot&transcr- mrBAYES	FMO_prot- mrBAYES

Simplified tree cladograms are shown in fig. 1.

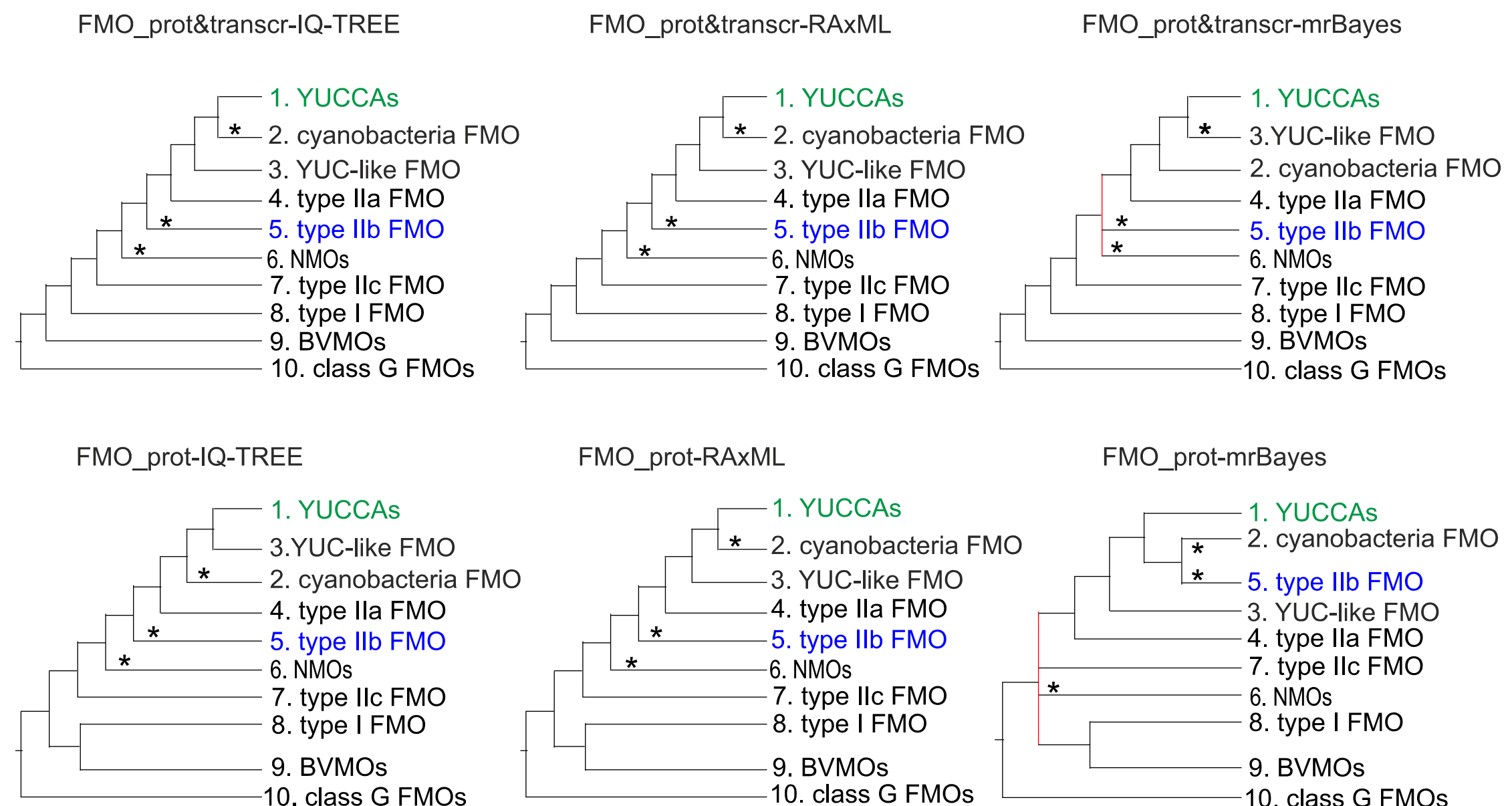


Figure 1. Simplified cladograms for class B flavoprotein monooxygenases proteins. The type IIb FMO and YUCCAs families are highlighted in blue and green text, respectively. * - long branches included in ancestral nodes of clade families are highlighted. Red line highlighted polyfurcations.

The FMO_prot&transcr-IQ-TREE phylogenetic tree is shown in figure 2A. A fragment of this tree – only the type IIb FMO family is shown in fig. 2B.

Class B flavoprotein monooxygenase proteins include three subclasses: FMOs – №1-5, 7-8 protein families in fig. 2A and fig. 1, NMOs – group № 6 of proteins – in the same place, and BVMOs – group №9. Class G FMO proteins (№ 10 in fig. 1 and 2A) were taken as an outgroup. Especially note that in the type II FMO protein group, first described by Riebel et al., [Riebel et al., 2013], we have identified three subgroups: type IIa FMO, type IIb FMO, and type IIc FMO (№4, №5, №7 groups, respectively, in fig. 1 and 2A).

A new family named type IIb FMO.

In this case, the type IIb FMO proteins (#5 in fig. 2A and fig. 1; all proteins in this family are shown in fig. 2B) differ significantly in a number of properties, both from other type II FMO proteins (№4 and №7 groups) and from the rest of the FMOs subclass proteins (№1-3, 6, 8, 9).

- This group (№5) contains proteins from red, green algae, land plant taxa, fungi, protozoa, and bacteria, whereas type IIa FMO (№4) and type IIc FMO (№7) contain only bacterial proteins.
- Type IIb FMO proteins have an N-terminal domain of 160 a. o., which is absent in type IIa FMO (#4) and type IIc FMO (#7), as in other of the class B flavoprotein monooxygenases proteins (№1-3, 6, 8, 9).
- There are also significant differences in the cosenses of FMO identified site (FxGxxxHxxx**H**[70%]/y[20%]/f[10%]) and FAD binding site (Gx**N**[75%]/g[17%]/c[13%]xxA/G) for type IIb FMO (№5) proteins compared to FMO identified site (FxGxxxHxxx**Y/F/W**) and FAD binding site (Gx**G**xxA/G) for the rest proteins of the class B flavoprotein monooxygenases (№1-3, 6, 8, 9).

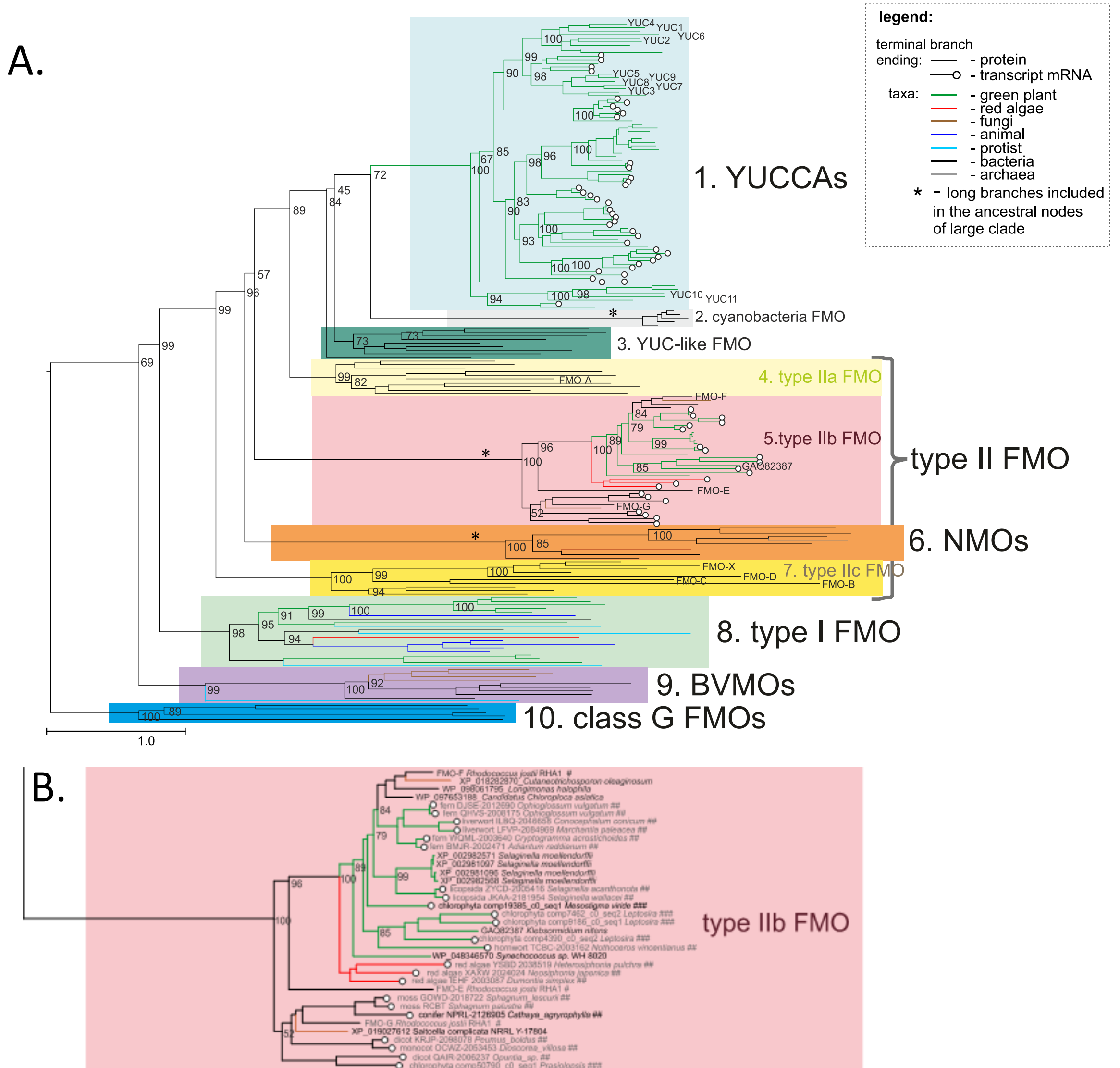


Figure 2. The FMO_prot & transcr-IQ-TREE phylogenetic tree is shown in figure A. B) A fragment of the FMO_prot & transcr-IQ-TREE phylogenetic tree that includes only the type IIb FMO family (№ 5 to figure 2A). The names of protein sequences taken from NCBI and from the article Riebel et al., 2013 are highlighted in black text; transcriptomic sequences are highlighted in gray text. The three protein sequences marked with the # icon are taken from the Riebel et. al. study, 2013; ## - taken from the 1000 plant genomes database (<https://sites.google.com/a/ualberta.ca/onekp/>) and ### - from the Green Algal Tree of Life project [<https://algae.eeb.uconn.edu/research-interests/green-algae-tree-of-life-project/>].

Facts from literary sources:

Riedel et. al. (Riedel et. al., 2013) in a study of actinobacteria *Rhodococcus jostii* RHA1 proteins identified subclass from the class B flavoprotein monooxygenases, which they named type II FMOs. Of the eight proteins (FMS ag proteins) identified by them in *R. jostii* RHA1, three (FMOe, FMOf, FMOg) had special catalytic properties not typical for FMOs subclass proteins. They were able to catalyze not only the reaction of sulfoxidations (the ability of FMOs and BVMOs), but also Baeyer–Villiger oxidations (the ability of BVMOs). In addition, FMOe, FMOf, FMOg proteins can use both NADPH and, much cheaper, NADH as a coenzyme. Although all other FMOs proteins use only NADPH. This makes these three proteins promising targets for their use in biotechnology [Riebel et al., 2013].

Based on their research, Yue. et al. (2014) proposed a hypothesis about the origin of {1} YUCCAs in the most recent common ancestor (MRSA) of the main land plant taxa as a result of HGT (horizontal gene transfer) from bacteria. Later, Wang et al. (Wang et al., 2015) based on the presence of {1} YUCCAs homologs of the GAQ82387.1 protein in Charophyta *K. nitens*, it was suggested that YUCCAs proteins appeared in evolution already in Charophyta (Fig. 1A, GAQ82387.1 protein).

Thus, the type Iib FMO family that we have identified includes the FMOe, FMOf, and FMOg proteins of *R. jostii* RHA1 (fig. 2A, 2B) described by Riebel et al., 2013, and the Charophyta (GAQ82387.1 *K. nitens*) proteins described by Wang et al., 2014.

A comparative analysis of the topologies of 6 trees for class B flavoprotein monooxygenases proteins.

The topology of FMO_prot&transcr-IQ-TREE, FMO_prot&transcr-RAXML, FMO_prot-IQ-TREE, FMO_prot-IQ-TREE trees is conservative. However, the topologies of FMO_prot-mrBayes to a greater extent than FMO_prot&transcr-mrBayes have important differences from the topology of the aforesaid trees: 1) the FMO_prot-mrBayes and FMO_prot&transcr-mrBayes trees are not up to resolved – they have tetrafurcation and trifurcation, respectively; 2) in FMO_prot-mrBayes, the {5} type Iib FMO clade (highlighted in blue text to fig. 2) approaches with {1} YUCCAs, so they become sister. so that they become sisterly.

This topology shift can be caused by the LBA effect, since the trees have three long branches that entering in ancestral nodes of the following clades: {2} cyanobacteria FMO, {5} type Iib FMO, and {6} NMOs (in fig. 1A, fig. 2 these long branches are marked *). Based on a comparative analysis of the phylogenies that we received, we assume that the Bayesian method, the mrBayes program, was more sensitive to LBA effect than the Maximum Likelihood method (implemented in IQ-TREE and RaxML programs), in our study. Our assumption is consistent with the conclusions of Kolaczkowski and Thornton [Kolaczkowski and Thornton, 2009], which showed that the Bayesian inference (BI) of phylogenetic relationships, in contrast to Maximum Likelihood methods (ML), leans in favor of topologies that group long branches together, which leads to an artifact topology of trees.

Thus, we tend to favor the phylogeny topology of class B flavoprotein monooxygenases, where the clade {5} type Iib FMO is remote from the clade {1} YUCCAs (fig. 1C, 4C), since it is found in all trees except FMO_prot-mrBayes (fig. 1C). In addition, only trees built by mrBayes are under-resolved (fig. 1, red lines).

Conclusion: 1) We have identified a new group of type Iib FMO proteins, including plant proteins (among them Charophyta GAQ82387.1 proteins from *K. nitens* described earlier by Wang et al., 2014) and FMOe, FMOf, FMOg bacteria proteins from *R. jostii* RHA1 (described earlier by Riebel et al., 2013).

2) Our data suggest that the type Iib FMO and YUCCAs proteins occurred in evolution independently of two different HGTs. This supports the hypothesis of Yue et al. about the origin of YUCCAs as a result of HGT from bacteria to land plants, but not the hypothesis of Wang et al. about the appearance of YUCCAs in evolution already in Charophyta.

Methods and Algorithms: The sequences recognition for forming samples was performed by the BLASTP program. Multiple sequence alignment was performed using Promals and Mafft. IQ-TREE 1.6.12, RAXML 8.1.24, and mrBayes 3.2.7 programs were used for phylogenetic tree reconstruction.

Acknowledgements: The work was supported by project 0324-2019-0040.

References:

1. Riebel A., de Gonzalo G., Fraaije M. F., Expanding the biocatalytic toolbox of flavoprotein monooxygenases from *Rhodococcus jostii* RHA1. *Enzymatic*, 2013, 88, 20–25.
2. Wang C., Liu Y., Li S.-S., and Han G.-Z., Origin of plant auxin biosynthesis in charophyte algae. *Trends in Plant Science*, Vol. 19, No. 12 pp. 741-743.
3. Yue J., Hu X., and Huang J. Origin of plant auxin biosynthesis. *Trends in Plant Science*, Vol. 19, No. 12 pp. 764-770.
4. Kolaczkowski B., Thornton J.W.. Long-Branch Attraction Bias and Inconsistency in Bayesian Phylogenetics. *PLoS One.*, 2009, 4(12), e7891.