

- Тема работы:

Эволюционный анализ ферментов триптофанзависимого пути синтеза ауксина у растений.

- Состав коллектива:

Турнаев Игорь Иванович*, Институт Цитологии и Генетики, без степени; Гунбин Константин Владимирович, Институт Цитологии и Генетики, к.б.н.; Суслов Валентин Владимирович, Институт Цитологии и Генетики, без степени; Афонников Дмитрий Аркадьевич, к.б. * - turn@bionet.nsc.ru.

- Работа поддержана бюджетным проектом 0324-2019-0040. «Генетические основы биотехнологий и биоинформатика», руководитель Колчанов Н.А. С 01.01.2019 по 31.12.2019.
- Постановка задачи

Белки flavin monooxygenases YUCCA – важные ферменты растений, которые участвуют в биосинтезе одного из ключевых гормонов – ауксина. Предполагается, что эти гены возникли в результате горизонтального переноса от бактерий. Однако ведутся споры, на каком этапе эволюции растений это произошло: у многоклеточных водорослей или наземных растений. Для уточнения ранней эволюции генов YUCCA мы реконструировали филогенетические отношения, между последовательностями flavin monooxygenases (FMO), суперсемейства белков, к которому относится и YUCCA.

- Современное состояние проблемы (на момент начала работы).

Исследование эволюции двух ферментов TAA и YUCCA, которые катализируют обе реакции двухстадийного пути синтеза ауксина у наземных растений группой Yue et. al., (Yue et. al., 2014) позволило им сформировать гипотезу, что оба эти фермента и, соответствующий путь биосинтеза ауксина появились у растений в результате горизонтального переноса генов от бактерий к первым наземным растениям общим предкам споровых и сосудистых растений. Сразу после этого группа Wang et. al., (Wang et. al., 2014) основываясь на новых данных по последовательностям недавно секвенированного генома харовой водоросли *Klebsormidium nitens* выдвинула альтернативную гипотезу, что оба белка TAA и YUCCA и рассматриваемый путь биосинтеза ауксина появились в эволюции растений раньше, уже у Харовых водорослей, которые в настоящее время рассматриваются как наиболее вероятные водные предки наземных растений. В результате по этому вопросу завязалась оживлённая научная дискуссия авторы которой высказывали аргументы в поддержку либо против первой или второй гипотезы [Turnaev et. al., 2015; Huang et al., 2014; Ke et.al., 2015; Wang et. al., 2016; Romani F., 2017; Bowman et. al., 2017; Nishiyama et.al., 2018].

- Научное содержание работы:

Нами был проведён комплексный теоретический анализ белков суперсемейства флавинзависимых монооксигеназ (FMOs): YUCCA и близкородственных к ним семейств

белков, включающий кроме филогенетического, также анализ активных сайтов, доменов белка и распространённости белков в разных таксонах.

Известно, что суперсемейство флавинзависимых монооксигеназ (FMOs) делится, основываясь на структуре и их ферментативных свойствах, на 8 классов (A-H) [van Berkel W.J. et al., 2006; Huijbers M.M. et al., 2014, 2–17]. Флавинзависимые монооксигеназы класса В (EC 1.14.13), анализ белков которого мы и проводили, включают три подкласса N-hydroxylating monooxygenases (NMOs), Baeyer–Villiger monooxygenases (BVMOs), flavin-containing monooxygenases (FMOs - эта группа является одноимённой с названием всего суперкласса) [Fraaije et al., 2002]. YUCCAs в свою очередь являются семейством, относящимся к подклассу FMOs [Schlauch, 2007].

Также нужно отметить, что недавно группой Riebel et al., [Riebel et al. 2013] была предложена определённая модификация классификации подклассов в классе В. В работе Riebel et al. 2013 установил, что белки бактерии *Rhodococcus jostii RHA1* FMO-A-G относящихся к подклассу FMOs класса В, выделяются по филогении, а некоторые по биохимическим свойствам (FMO-G-H), в новую группу белков, названную им type II FMO. Таким образом подкласс FMO было предложено разделить на две группы: ранее описанные белки подкласса FMOs – type I FMO и открытая им новая группа – type II FMO.

В нашей работе мы установили, что классификация, предложенная Riebel et al. требует дальнейшего углубления. Нами было установлено, что выделенная Riebel et al. группа type II FMO чётко разделяется по структуре белков и свойствам на две подгруппы: первая включает белки сходные по структуре и свойствам с белками type I FMO, также, как и с белками YUCCA, её мы назвали – type IIA FMO, вторая, названная нами type IIB FMO, включает белки значительно отличающиеся по структуре и, предположительно, свойствам от остальных представителей подкласса FMOs, в том числе и от YUCCA. Так в эту подгруппу вошли три белка FMO-G-H *R. jostii RHA1*, которые, как было ранее установлено Riebel et al., [Riebel et al., 2013], имеют уникальную способность кроме стандартной для белков FMOs способности окислять гетероатомы, катализировать также характерную для белков BVMO, но не для белков FMOs, способность катализировать реакцию Baeyer-villiger окисления, а также использовать в качестве кофактора кроме NADPH еще и NADH, что тоже нехарактерно для FMOs. Кроме того, белки этого кластера type IIB FMO широко представлены среди видов грибов, бактерий и умеренно у растений и простейших. Так белок GAQ82387.1 *K. nitens*, гомология которого с белками YUCCA наземных растений послужим из оснований для выдвижения гипотезы группы Wang et al., [Wang et al., 2014] о происхождении (ТАА-YUCCA) пути биосинтеза ауксина уже у харовых водорослей, по нашим данным оказался принадлежащим к type IIB FMO. Что с высокой вероятностью предполагает у него функции сходные с функциями белков FMO-G-H *R. jostii RHA1*, то есть значительно отличающиеся от функций YUCCA. Что в свою очередь является значительным аргументом в пользу гипотезы группы Yue et al. о происхождении (ТАА-YUCCA) пути биосинтеза ауксина у первых наземных растений.

- Подробное описание работы, включая используемые алгоритмы.

Анализ филогении суперсемейства FMO

Для того чтобы уточнить эволюционное происхождение белков семейства YUCCA мы провели филогенетический анализ белков суперсемейства FMO. С этой целью программой BLAST с порогом распознавания $e\text{-value} = 1 \cdot 10^{-5}$ были выявлены гомологи белков YUCCA

среди последовательностей в базе данных аминокислотных последовательностей из NCBI. В выборку были дополнительно включены белки из статьи Riebel et. al., (2013): семь аминокислотных последовательностей генов FMO-A-G бактерии *Rhodococcus jostii RHA1* и гена FMO-X *Stenotrophomonas maltophilia*. Эти белки были использованы для дальнейшей классификации гомологов YUCCA на группы, предложенные ранее Riebel et. al. Далее выборка выравнивалась методом Promals 2D. Для последовательностей полученной выборки было построено филогенетическое дерево белков суперсемейства FMO методом maximum likelihood (fig. 1A). На филогенетическом дереве (рис 1A) хорошо выделяются три основные группы белков. Первый кластер включает белки Baeyer-Villiger монооксигеназы (BVMO) и является аутгруппой по отношению к белкам FMO I и II типа (рис 1A, сиреневый фон). Белки этого кластера представлены у грибов, простейших и бактерий. Второй кластер представлен белками FMO I типа (по Riebel et.al., 2013; рис 1A, голубой фон). Белки этого кластера представлены у plant, animal, bacteria. Третий кластер включает ферменты type II FMO. В этот кластер входят белки, представленные у бактерий, растений, грибов и простейших (на рис. 1A) группы белков, входящих в этот кластер показаны зелёным, желтым и розовым цветом). В свою очередь, в кластере type II FMO можно выделить 3 группы белков. К первой относятся белки YUCCA растений, они формируют кластер, заметно выделяющийся от остальных белков type II FMO (Рис 1A, зелёный цвет). Ко второй группе относятся белки подкластеров of type II FMO, обозначенные здесь как a-d (Рис 1A, жёлтый цвет). Все эти последовательности встречаются только у бактерий. При этом, если в подкластерах a, c и d представлены последовательности из бактерий различных групп, то в подкластер b входят только последовательности цианобактерий, при этом этот подкластер выделяет наличие аномально длинной прикорневой ветки. Следует отметить, что последовательности type II FMO подкластера a имеют относительно невысокий уровень бутстрэп поддержки, в то время как остальные подкластеры из этой группы имеют бутстрэп-поддержку выше 90. К третьей группе белков относятся последовательности, обозначенные на дереве как type II FMO подкластер e (Рис 1A, В розовый цвет). В отличие от групп type II FMO подкластеров a-d группа type II FMO подкластер e содержит кроме бактериальных ещё и белки растений, грибов и простейших (fig. 1B). В подкластер e вошли последовательности растительных организмов, lycophyte *S. moellendorffii*, а также последовательность GAQ82387.1 харовой водоросли *K. nitens*, которая является ближайшим гомологом YUCCA в геноме *K. nitens* [Wang, 2014].

Последовательности белков FMO-E,G,F *Rhodococcus jostii RHA1* из работы Reiebel et. al. [Reiebel et. al., 2013] (Рис. 1B) относятся к кластеру e. Остальные белки из этой же работы, FMO -B,C,D *Rhodococcus jostii RHA1* и FMO-X *Stenotrophomonas maltophilia* расположены в type II FMO subcluster d. Белок FMO-A *Rhodococcus jostii RHA1* располагается в подкластере кластера c of type II FMO (рис 1B).

Результаты позволяют сделать вывод, что белки подкластера e в кластере type II FMO представляют особую группу, так как: а) выделяются на дереве в обособленную кладу с высоким уровнем бутстрэп-поддержки (fig. 1A, B); б) эта кладка единственная в кластере type II FMO включает белки разных царств (fig. 1B); в) она содержит последовательности ферментов FMO-E, -G, -F *Rhodococcus jostii RHA1* ферментативные свойства которых отличаются от таковых у других белков type II FMO [Riebel et al., 2013; Riebel et al., 2014] (fig. 1B). Чтобы подчеркнуть эти значительные структурно-функциональные различия белков type II FMO субкластеров a-d от белков type II FMO субкластера e мы предлагаем следующее переобозначение, которое будем использовать дальше по тексту: группу белков “type II FMO субкластеров a-d” мы переобозначаем как “type IIa FMO”, а группу белков “type II FMO субкластер e” – “type IIb FMO”.

Сравнительный анализ функциональных сайтов и доменов FMO-like белков

Нами был проведен анализ последовательностей трех функциональных мотивов (рис. 2А) в пяти кластерах последовательностей, выделенных на филогенетическом дереве FMO/BVMO: YUCCA, type IIa FMO, type IIb FMO, type I FMO, BVMO. Эти группы оказались однородными по последовательностям представленных в них мотивов.

Схема расположения FAD-binding, FMO identifying и NADPH binding мотивов в последовательности FMO [Schlaich N.L., 2007] на примере последовательности YUC2 AT4G13260 *A. thaliana* показана на рисунке 2А. Отображение мотивов в формате LOGO [Crooks G. E., et. al., 2004]) для 5 групп последовательностей показано на рис 2В.

Из рисунка 2В видно, что FAD-binding мотив у последовательностей всех групп белков является сходным и содержит 3 высококонсервативных глицина. Для FMO identifying мотива в группах YUCCA, type IIa FMO и type I FMO характерен консенсус FxGxxxHxxx^{Y/F/R}, причем в предпоследней позиции преобладает тирозин (Y). Для BVMO в этом мотиве предпоследний символ представлен консервативным триптофаном (W). Для группы последовательностей type IIb FMO характерен консенсус FxGxxxHxxx(H/y/f)P предпоследний символ мотива у 70% последовательностей представлен гистидином (H), однако среди последовательностей встречаются замены на Y и F, характерные для type I FMO/IIa FMO (в последовательностях остальных кластеров гистидин в этой позиции встречается лишь дважды в кластере type I FMO). В последней позиции этого мотива у белков type IIb FMO расположен пролин (P), тогда как у type I FMO/IIa FMO это лизин (K) либо аргинин (R).

Что касается NADPH binding мотива, то для ферментов YUCCA характерны все три консервативных глицина, для type IIa FMO третий глицин часто заменяется на аланин (A). В белках type IIb FMO вместо второго глицина в 75% случаев встречается аспарагин (N) (в других кластерах эта аминокислота не встречается ни разу). Для всех белков группы type II FMO также характерны высококонсервативные N и S в 4 и 5 позициях мотива, соответственно. Аминокислоты NADPH binding мотива в группе type I FMO аналогичны таковым в группе type IIa FMO. Для BVMO характерно наличие высококонсервативного аланина в последней позиции.

Таким образом, по набору аминокислот в последовательностях FMO identifying и NADPH binding мотивов большинства последовательностей кластера type IIb FMO имеются аминокислоты, нехарактерные для белков type IIa FMO. Отметим, что последовательность FMO identifying мотива белка GAQ82387.1 *Klebsormidium nitens* в предпоследней позиции имеет аминокислоту H (последняя глицин - G), которая также наблюдается в этой позиции для белка FMO-F *R. jostii* RHA1 [Riebel et al, 2013].

Мы провели сравнительный анализ консервативных доменов этих трёх групп белков (рис 3).

Из рисунка видно, что последовательности type IIb FMO имеют N-терминальный домен длиной около 160 а.о., который отсутствует у других белков type I/IIa FMO и BVMOs. Этот домен присутствует во всех последовательностях подкластера и идентифицируется как PFAM12680 – Snoal2 (Snoal-like domain), который принадлежит суперсемейству cl09109 (Nuclear transport factor 2 (NTF2-like) superfamily); PRK08241 - PRK08241 (RNA polymerase factor sigma-70) расположенных в районе 3-110 аа. Отметим также, что белки BVMOs имеют значительно более протяжённый около 180 а.о. С-терминальный район, по сравнению с белками FMO-like. В то же время центральный участок последовательностей, соответствующий домену типа «Pur redox 3 of NADB Rossmann superfamily» присутствует во всех трёх группах ферментов (fig. 3).

Таким образом, последовательности type IIb FMO отличает от других белков FMO-like наличие дополнительного домена на N- конце.

Распределение последовательностей, гомологичных type IIb FMO в основных таксонах организмов.

Мы провели анализ представленности гомологов белков этих трёх групп среди основных таксонов про- и эукариот в базе данных NCBI. В качестве query были взяты

следующие белковые последовательности: (1) for type IIb FMO: *Halomonas lutea* WP_019017022 FMO protein; (2) for **FMO-like**: растительный *Populus trichocarpa* FMO protein - PT09G14380 (Plaza plant Id)/ XP_002312911.2 (NCBI Id) and бактериальный *Actinobacteria bacterium OK074* FMO protein - WP_054213635 (NCBI Id); (3) for **BVMO**: *Bartonella henselae* WP_012854150 protein. Поиск проводился программой PHI-blast, с учетом мотивов консервативных сайтов с e-value = $1 \cdot 10^{-10}$. Для каждой из трёх групп белков при поиске гомологов были использованы соответствующие консенсусы FMO identified сайта: (1) для type IIb FMO – FxGxxxHxxxH; (2) для FMO-like proteins (это белки type I FMO и type Ia FMO, которые не разделяются при PHI-blast поиске, так как консенсусы их FMO identified sites значимо не отличаются между этими группами белков) - FxGxxxHxxxY/F и (3) для BVMOs - FxGxxxHxxxW.

Распределение количества обнаруженных гомологов по таксонам, с учётом уровня сходства гомологов, представлено в таблице 2.

Таблица 2. Распределение количества гомологов растительных белков трех семейств, представленных в базе данных NCBI по таксонам в зависимости от уровня сходства.

	Animal	Fungi	Plant	Bacteria	Archeae
type IIb FMO					
0 — e-70	0	419	4	1387	1
e-70 — e-40	0	80	0	2	0
e-40 — e-05	0	8	3	19	0
all homologs	0	507(285)	7(4)	1408(1041)	1(1)
BVMO					
0 — e-70	8	918	1	3155	1
e-70 — e-40	4	3475	2	1090	9
e-40 — e-05	4	1198	1	1605	1
all homologs	16(6)	5591(549)	4(2)	5850(1656)	11(9)
FMO-like_plant*					
0 — e-70	0	0	512	0	0
e-70 — e-40	0	0	16	4	0
e-40 — e-05	2486	417	302	2052	0
all homologs	2486(409)	417(244)	831(106)	2056(1306)	0
FMO-like_bacteria*					
0 — e-70	0	0	0	329	0
e-70 — e-40	2	0	612	537	0
e-40 — e-05	2917	1305	993	8660	0
all homologs	2919(187)	1305(433)	1605(114)	9526(837)	0

В таблице приведено количество распознанных в Phi-BLAST гомологов, отдельно, по каждой группе белков, для порогов распознавания e-value = 0 – e-70; e-70 – e-40; e-40 – e-05 и, также, в строке “all homologs” суммарное количество распознанных последовательностей для e-value 0 – e-05 с указанием в скобках числа видов которым принадлежат распознанные последовательности. Сокращения в таблице: type IIb FMO - type IIb FMO proteins by PHIBlast search homologs for *K. nitens* protein (GAQ82387.1); **BVMO_** - BVMOs proteins by PHIBlast search homologs for *G. obscurus* protein (WP_012947985.1); **FMO_plant** - FMO-like proteins by PHIBlast search homologs for *P. trichocarpa* FMO protein (XP_002312911.2); **FMO_bacteria** - FMO-like proteins by PHIBlast search homologs for bacteria *Actinobacteria bacterium OK074* FMO protein (WP_054213635).

Из таблицы 2 видно, что: 1) белки кластеров type IIb FMO и BVMO широко распространены среди бактерий и грибов, но практически отсутствуют у растений и не обнаруживаются у животных и архей, тогда как белки FMO-like оказались широко распространенными во всех исследованных таксонах кроме архей. 2) Уровень сходства query последовательностей с белками бактерий и грибов в кластерах type IIb FMO и BVMO, значительно выше, чем между белками различных таксонов FMO-like. На это указывает то,

что только между белками бактерий и грибов как для type IIb FMO, так и BVMO в нашем исследовании обнаружено значительное число близких гомологов ($e\text{-value} = 0 \text{ — } e\text{-70}$).

Анализ последовательностей растений, представленных в проекте 1KPlants.

С целью более полной идентификации последовательностей суперсемейства FMO в растительных организмах, мы провели дополнительное сравнение белков FMO с аминокислотными последовательностями, полученными в транскриптомных проектах 1000 plant genomes [<https://sites.google.com/a/ualberta.ca/onekp/>] и GreeAlgae Tree of Life (интернет адрес, по которому доступны последовательности, указан в подписи к рис. 4) [Cooper and Delwiche, 2016]. Филогенетическое дерево, построенное для последовательностей, гомологов FMO из этих баз данных показано на рис 4А. На дереве присутствуют все основные клады семейства FMO, которые были отмечены нами на рисунке 1: YUCCA, type IIa FMO, type IIb FMO, type I FMO, Baeyer-villiger monoоxygenases. Клада белков type IIb FMO приведена более детально на рисунке 4Б. В ней кроме белков бактерий и грибов представлены также белки красных водорослей, зелёных водорослей как из Chlorophyta, так и из Streptophyta (семейство Klebsormidiaceae), всех основных таксонов наземных растений (moss, liverworts, hornworts, clubmosses, ferns, conifera and angiospermae: как monocots, так и eudicots).

Результаты идентификации гомологов *Klebsormidium nitens* GAQ82387.1 и *Arabidopsis thaliana* YUC2 AT4G13260 в базах данных 1000 plant genomes и NCBI представлены в таблице 3.

Таблица 3. Поиск гомологов последовательностей *Klebsormidium nitens* GAQ82387.1 (группа type IIb FMO) и *Arabidopsis thaliana* YUC2 AT4G13260 (группа YUCCA) среди последовательностей из баз данных 1000 plant genomes и NCBI. В столбце 1 таблицы приведены названия таксонов и в скобках число видов каждого таксона представленных в базе 1000 plant genomes. В 2-5 столбиках таблицы в ячейках приведено число гомологичных последовательностей, в скобках - число видов в которых обнаружены гомологи.

*- база данных Algae Tree of Life [Cooper and Delwiche, 2016]

Таксоному			Гомологи GAQ82387 .1 в базе 1000 plant genomes	Гомологи GAQ82387 .1 в базе NCBI	Гомолог и YUCCA в базе 1000 plant genomes	Гомолог и YUCCA в базе NCBI	
Streptophyta	Streptophytina	Embryophyta	Eudicots (596)	4(3)	-	434(333)	1680(114)
			Monocots (104)	1(1)	-	47(35)	444(26)
			Conifers (73)	1(1)	-	4(4)	-
			Cycadales (4)	-	-	2(2)	-
			Leptosporangiate Monilophytes (65)	90(33)	-	25(19)	-
			Eusporangiate Moliniiformes (12)	1(1)	-	-	-
			Lycophytes (22)	5(4)	6(1)	-	10(1)
			Hornworts (9)	9(5)	-	3(3)	-

		Liverworts (28)	17(9)	2(1)	10(10)	6(2)
		Bryophyta (41)	1(1)	-	24(18)	6(1)
		Zygnemophyceae (5)	-	-	-	-
		Coleochaetophyceae (3)	-	-	-	-
		Charophyceae (1)	-	-	-	-
		Mesostigmatophyceae (1)	-	-	-	-
		Chlorokybophyceae (1)	-	-	-	-
		Klebsormidiophyceae (2)	2(2)	1(1)	-	-
Chlorophyta		Green algae (152)	5(4)	1(1)	-	-
Rodophyta		Red algae* (28)	2(2)	-	-	-

Из таблицы 3 мы можем видеть, что по данным базы 1000 геномов прежде всего можно видеть резкие различия по представленности гомологов генов YUCCA и GAQ82387 в организмах семенных растений. Если число обнаруженных гомологов YUCCA у семенных растений свыше 300 у двудольных и более 4 десятков у однодольных, то гомологи GAQ82387 этой базе данных представлены единичными последовательностями, при том, что количество геномов составляет почти 600, а однодольных превышает сотню.

Вторым интересным результатом является то, что гомологи GAQ82387 встречаются в отдельных геномах водорослей: в 4 геномах низших зеленых водорослей (Chlorophyta), в геномах красных водорослей, и у харовых (только у Klebsormidiophyceae). При этом, гомологи YUCCA у водорослей не были обнаружены как по данным 1000 plant геномов, так и по данным базы nr NCBI.

Относительно высокая представленность гомологов как YUCCA, так и GAQ82387 наблюдается в одном из двух таксонов папоротников, Leptosporangiate Monilophytes: у 19 и 33 представителей из 65, соответственно (база данных 1000 plant genomes).

У низших наземных растений по данным базы 1000 plant genomes картина выглядит мозаичной: оба гомолога обнаруживаются в геномах представителей Hornworts и Liverworts, у Lycophytes обнаруживаются гомологи GAQ82387, но не YUCCA, а у представителей Bryophyta часто встречаются гомологи YUCCA, а гомолог GAQ82387 лишь в одном случае.

Частая встречаемость гомологов обоих генов у папоротниковых и низших наземных растений вызывает вопрос: встречаются ли гомологи двух генов в геномах одного и того же вида одновременно? Мы исследовали этот вопрос и представили результаты в таблице 4. В ней приведены виды, для которых одновременно обнаружены гомологи GAQ82387 и YUCCA.

Таблица 4 Виды растений для которых обнаружены как гомологи *Klebsormidium nitens* GAQ82387.1, так и *Arabidopsis thaliana* YUC2 AT4G13260 (в базах 1000 plant genomes и NCBI). В 3 и 4 столбиках в ячейках приведено число обнаруженных гомологов для каждого вида.

Идентификаторы вида в базе 1000 plant genomes	taxon	GAQ82387.1	YUCCAs
TVSH_201823_Bituminaria_bituminosa	Core Eudicots/Бобоцветные	1	2
WWQZ_211706_Medinilla_magnifica	Core Eudicots/Миртоцветные	1	1

OCWZ_200432_Dioscorea_villosa	Monocots/Диоскорицветные	1	2
AFPO_201018_Blechnum_spicant	Leptosporangiate Monilophytes	4	1
BMJR_200209_Adiantum_tenerum	Leptosporangiate Monilophytes	2	2
DCDT_207190_Cheilanthes_arizonica	Leptosporangiate Monilophytes	3	1
FLTD_200266_Pteris_ensigormis	Leptosporangiate Monilophytes	1	2
GANB_201380_Cyathea_spinulosa	Leptosporangiate Monilophytes	4	1
KIIX_201108_Pilularia_globulifera	Leptosporangiate Monilophytes	2	1
KJZG_200972_Asplenium_platyneuron	Leptosporangiate Monilophytes	5	1
NDUV_201591_Vittaria_appalachiana	Leptosporangiate Monilophytes	1	2
NOKI_201577_Lindsaea_linearis	Leptosporangiate Monilophytes	5	1
PNZO_215202_Culcita_macrocarpa	Leptosporangiate Monilophytes	1	1
RICC_200988_Cystopteris_reevesiana	Leptosporangiate Monilophytes	3	1
UFJN_208949_Diplazium_wichurae	Leptosporangiate Monilophytes	3	1
UOMY_200602_Osmunda_sp	Leptosporangiate Monilophytes	1	1
WQML_200900_Cryptogramma_acrostichoides	Leptosporangiate Monilophytes	1	2
YLJA_207326_Polypodium_amorphum	Leptosporangiate Monilophytes	2	1
RXRQ_201835_Phaeoceros_carolinianus	Hornworts	3	1
TCBC_200001_Megaceros_vincentianus	Hornworts	1	1
HMHL_201008_Marchantia_paleacea	Liverworts	2	1
ILBQ_200700_Conocephalum_conicum	Liverworts	2	1
TXVB_207470_Lunularia_cruciata	Liverworts	3	1

Из таблицы 4 видно, что у ряда видов растений из базы 1000 plant genomes гомологи белков YUCCA и GAQ82387 наблюдаются у 3-х видов Liverworts, 2-х - Hornworts, 13-ти - Leptosporangiate Monilophytes, 1 - monocots и у двух видов Eudicots.

- Полученные результаты.

Нами был проведён комплексный теоретический анализ белков суперсемейства флавинзависимых монооксигеназ (FMOs): YUCCA и близкородственных к ним семейств белков, включающий кроме филогенетического, также анализ активных сайтов, доменов белка и распространённости белков в разных таксонах.

Известно, что суперсемейство флавинзависимых монооксигеназ (FMOs) делится, основываясь на структуре и их ферментативных свойствах, на 8 классов (A-H) [van Berkel

W.J. et. al., 2006; Huijbers M.M. et. al., 2014]. Флавинзависимые монооксигеназы класса В (EC 1.14.13), анализ белков которого мы и проводили, включают три подкласса N-hydroxylating monooxygenases (NMOs), Baeyer–Villiger monooxygenases (BVMOs), flavin-containing monooxygenases (FMOs - эта группа является одноимённой с названием всего суперкласса) [Fraaije et al., 2002]. YUCCAs в свою очередь являются семейством, относящимся к подклассу FMOs (Schlauch, 2007).

Также нужно отметить, что недавно группой Riebel et. al., (Riebel et. al. 2013) была предложена определённая модификация классификации подклассов в классе В. В работе Riebel et. al. 2013 установил, что белки бактерии *Rhodococcus jostii RHA1* FMO-A-G относящихся к подклассу FMOs класса В, выделяются по филогении, а некоторые по биохимическим свойствам (FMO-G-H), в новую группу белков, названную им type II FMO. Таким образом подкласс FMO было предложено разделить на две группы: ранее описанные белки подкласса FMOs – type I FMO и открытая им новая группа – type II FMO.

В нашей работе мы установили, что классификация, предложенная Riebel et. al., требует дальнейшего углубления. Нами было установлено, что выделенная Riebel et. al. группа type II FMO чётко разделяется по структуре белков и свойствам на две подгруппы: первая включает белки сходные по структуре и свойствам с белками type I FMO, также, как и с белками YUCCA, её мы назвали – type IIА FMO, вторая, названная нами type IIВ FMO, включает белки значительно отличающиеся по структуре и, предположительно, свойствам от остальных представителей подкласса FMOs, в том числе и от YUCCA. Так в эту подгруппу вошли три белка FMO-G-H *R. jostii RHA1*, которые, как было ранее установлено Riebel et. al., (Riebel et. al., 2013), имеют уникальную способность кроме стандартной для белков FMOs способности окислять гетероатомы, катализировать также характерную для белков BVMO, но не для белков FMOs, способность катализировать реакцию Baeyer-villiger окисления, а также использовать в качестве кофактора кроме NADPH еще и NADH, что тоже нехарактерно для FMOs. Кроме того, белки этого кластера type IIВ FMO широко представлены среди видов грибов, бактерий и умеренно у растений и простейших. Так белок GAQ82387.1 *K. nitens*, гомология которого с белками YUCCA наземных растений послужим из оснований для выдвижения гипотезы группы Wang et.al., (Wang et.al., 2014) о происхождении (ТАА-YUCCA) пути биосинтеза ауксина уже у харовых водорослей, по нашим данным оказался принадлежащим к type IIВ FMO. Что с высокой вероятностью предполагает у него функции сходные с функциями белков FMO-G-H *R. jostii RHA1*, то есть значительно отличающиеся от функций YUCCA. Что в свою очередь является значительным аргументом в пользу гипотезы группы Yue et. al. о происхождении (ТАА-YUCCA) пути биосинтеза ауксина у первых наземных растений.

- Иллюстрации, визуализация результатов.

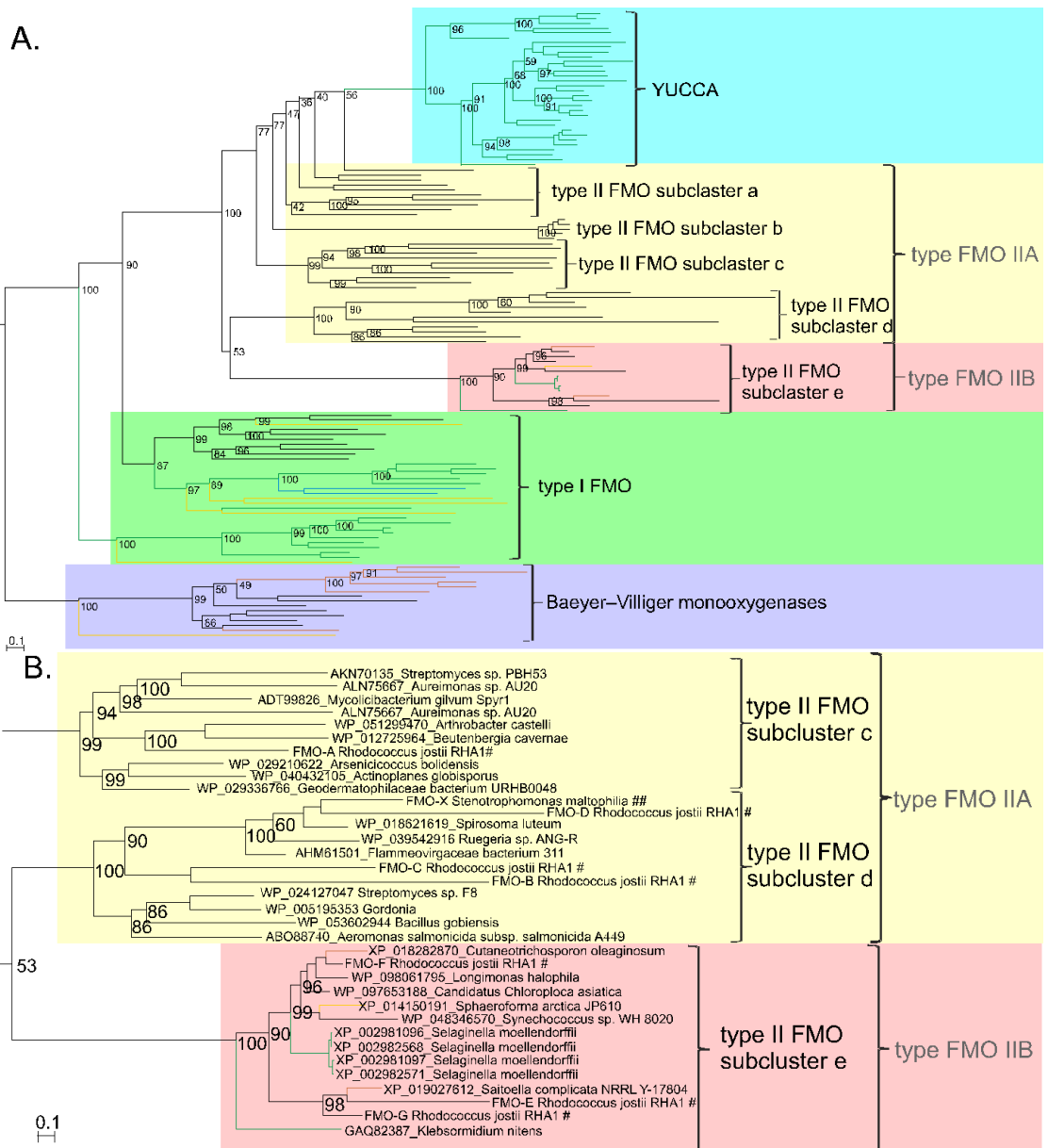
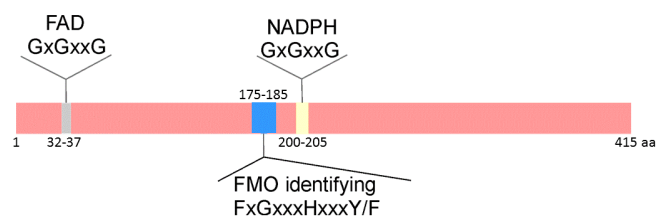


Рисунок 1. Филогения суперсемейства FMO-like белков. А) Филогенетическое дерево FMO-like белков. В качестве аутгруппы взяты белки BVMO - Baeyer-Villiger монооксигеназа - выделенные на рисунке фиолетовым цветом. Зелёным цветом обозначены белки растений, коричневым – грибов, жёлтым – простейших, синим – животных и чёрным – бактерий. Голубым прямоугольником выделена клада YUCCA растений; жёлтым - клада type II FMO subclaster a, b, c и d; розовым - type II FMO subclaster e; зелёным – type I FMO и сиреневым – BVMO (Baeyer villiger монооксигеназы). В) Фрагмент дерева филогении FMO-like белков включающий две клады: type II FMO subclasters c, d (из клады type IIa FMO) и type IIb FMO (обозначения аналогичны таковым на рис 1А). Кроме того, # - 7 последовательностей *Rhodococcus jostii RHA1* взяты из работы Riebel et. al., 2013; ## - Последовательность *Stenotrophomonas maltophilia* взятая из той же работы.

A)



B)

Conservative sites names	FAD binding site	FAD identified site	NADPH binding site
Conservative sites in FMO	GxGxxG	FxGxxxHxxx ^{Y/F^K/R}	GxGxxG/A
i. YUCs			
ii. type FMO IIA			
iii. type FMO IIB			
iv. type I FMO			
v. BVMO			

Рисунок 2. Распределение и выравнивание ключевых функциональных сайтов среди белков гомологов FMO-like. А) Схематически изображена локализация мотивов вдоль FMO proteins (позиции на схеме указаны for AT4G13260 - *Arabidopsis thaliana* YUC2 protein). В) В таблице показаны во 2-4 столбцах консенсусы трёх мотивов в FMO-like proteins: FMO-binding, FMO identifying and NADH-binding sites.

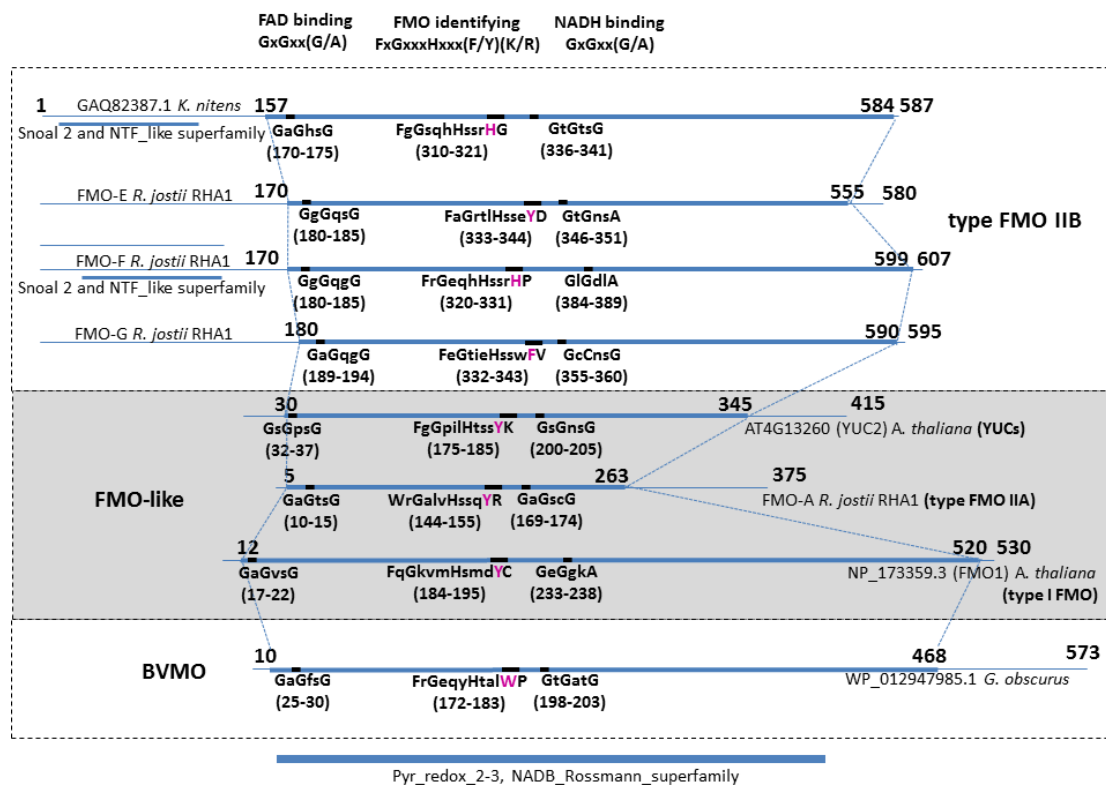
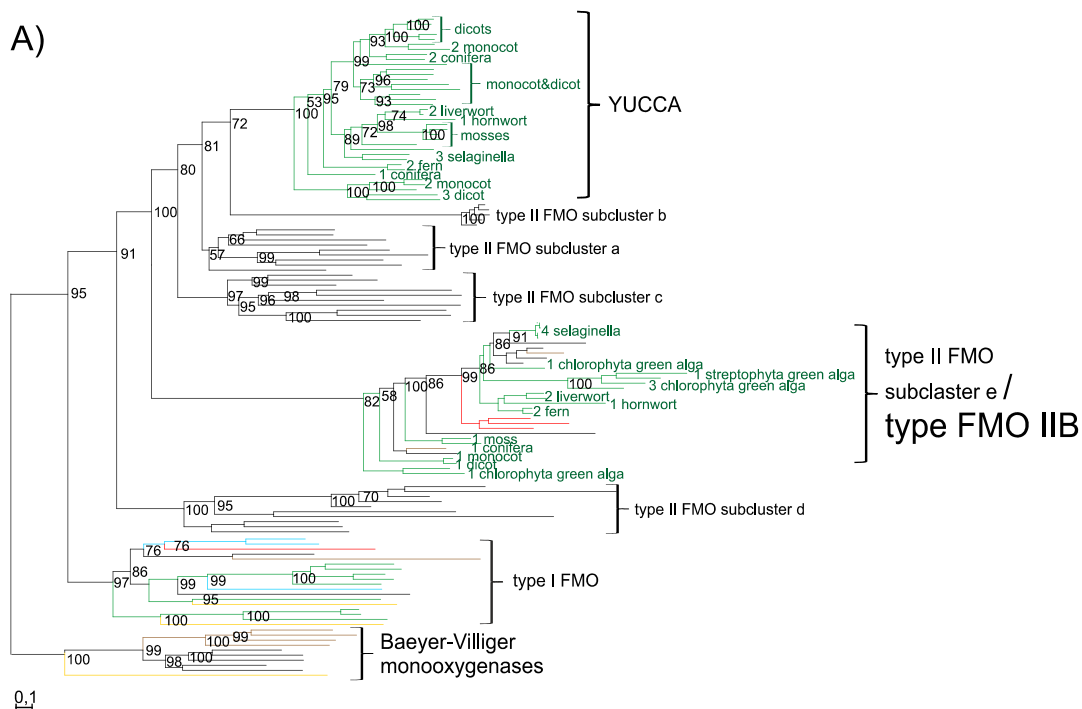


Рисунок 3. Сравнение консервативных доменов трёх групп белков: type IIb FMO, FMO-like (YUCs, type I FMO and type IIa FMO) and BVMO.



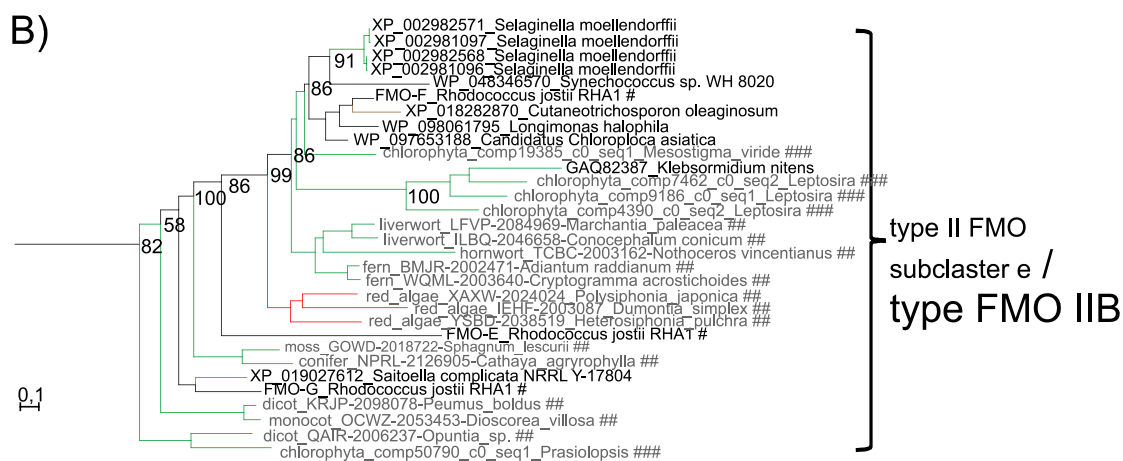


Рисунок 4. Общая филогения суперсемейства FMO-like белков, включающая как последовательности белков из полногеномных данных, так и RNAseq последовательности. А) Филогенетическое дерево FMO-like белков. В качестве аутгруппы взяты белки BVMO. Зелёным цветом обозначены белки растений, коричневым – грибов, жёлтым – простейших, синим – животных и чёрным – бактерий. В) Фрагмент дерева филогении FMO-like (полноразмерные белки + RNAseq последовательности) включающий type II FMO (обозначения аналогичны таковым на рис 4А). Кроме того, чёрным цветом текста выделены названия последовательностей белков из NCBI, а также 3 последовательности белков отмеченные на рисунке значком # взяты из исследования Riebel et. al., 2013; Серым цветом текста выделены последовательности mRNA: ## - взятые из базы 1000 plant genomes и #### - взятые из проекта Algae Tree of Life [Cooper and Delwiche, 2016]. Green Algal Transcriptomes for Phylogenetics and Comparative Genomics. Available online at: https://figshare.com/articles/Green_algal_transcriptomes_for_phylogenetics_and_comparative_genomics/1604778).

- Список литературы:
1. Yue J., Hu X., Huang J. Origin of plant auxin biosynthesis, *Trends Plant Sci.*, 2014, 19(12), 764-770.
 2. Wang C., Liu Y., Li S.-H., Guan-Zhu Han G.-Z. Origin of plant auxin biosynthesis in charophyte algae. *Trends in Plant Science*, 2014, 19(12), 741-743.
 3. Turnaev I. I., Gunbin K. V., Afonnikov D.A. Plant auxin biosynthesis did not originate in charophytes. *Trends Plant Sci.*, 2015, 20(8), 463-465.
 4. Huang J., Yue J., Hu X.. Origin of plant auxin biosynthesis in charophyte algae: a reply to Wang et al.. *Trends in Plant Science*, 2014, 19(12), 743.
 5. Wang, C., Li, S.-S., and Han, G.-Z. (2016). Commentary: plant auxin biosynthesis did not originate in charophytes. *Front. Plant Sci.* 7,158, doi: 10.3389/fpls.2016.00158.
 6. Ke, M., Zheng, Y., and Zhu, Z. (2015). Rethinking the origin of auxin biosynthesis in plants. *Front. Plant Sci.* 6, 1093. doi: 10.3389/fpls.2015.01093.
 7. Romani F. Origin of TAA Genes in Charophytes: New Insights into the Controversy over the Origin of Auxin Biosynthesis, *Frontiers in Plant Science*, 2017, 8, 1616. doi: 10.3389/fpls.2017.01616. eCollection 2017.
 8. Bowman J. L., Kohchi T., Yamato K. T., Jenkins J., Shu S.,, Yazaki K., Yokoyama R., Yoshitake Y., Yotsui I., Zachgo S., Schmutz J. Insights into Land Plant Evolution Garnered from the *Marchantia polymorpha* Genome. *Cell.*, 2017, 171(2), 287-304.
 9. Nishiyama T., Sakayama H., de Vries J., Buschmann H., Saint-Marcoux D.,, Harholt J., Dunand C., Zachgo S., Langdale J., Maumus F., Van Der Straeten D., Gould S.B., Rensing S.A., The Chara Genome: Secondary Complexity and Implications for Plant Terrestrialization, *Cell*, 2018, 174(2), 448-464.

10. Huijbers M. M. E., Montersino S., Westphal A. H., Tischler D., van Berkel W. J. H., Flavin dependent monooxygenases. *Archives of Biochemistry and Biophysics*, 2014, v. 544, 2-17.
11. van Berkel W. J. H., Kamerbeek N. M., Fraaije M. W., Flavoprotein monooxygenases, a diverse class of oxidative biocatalysts. *J. Biotechnol.* 2006, 124, 670–689
12. Fraaije M. W., Kamerbeek N. M., van Berkel W. J., Janssen D.B., Identification of a Baeyer-Villiger monooxygenase sequence motif. *FEBS Lett.* 2002, 518(1-3), 43-47.
13. Schlaich N.L., Identification of a Baeyer-Villiger monooxygenase sequence motif. *TRENDS in Plant Science*, 2007, 12(9), 412-417.
14. Riebel A., de Gonzalo G., Fraaije M. W., Expanding the biocatalytic toolbox of flavoprotein monooxygenases from *Rhodococcus jostii* RHA1. *Journal of Molecular Catalysis B: Enzymatic*, 2013, 88, 20-25.
15. Riebel A., Fink M. J., Mihovilovic M. D. Fraaije M. W. Type II Flavin-Containing Monooxygenases: A New Class of Biocatalysts that Harbors Baeyer–Villiger Monooxygenases with a Relaxed Coenzyme Specificity. *ChemCatChem*, 2014, 6, 1112-1117.
16. Croors G. E., Hon G., Chandonia J. M., Brenner S. E. WebLogo: A sequence logo generator, *Genome Research*, 2004, 14, 1188-1190.

- Эффект от использования кластера в достижении целей работы.

Кластер использовался для реконструкции филогенетических деревьев методом методом maximum likelihood.

- Перечень публикаций, содержащих результаты работы (если есть). Если имеется, указать импакт-фактор журнала (Thomson Reuters, РИНЦ,...).

Turnaev I.I., Suslov V.V., Afonnikov D.A., Gunbin K.V. Phylogenetic relationships between fmo classes and the origin of YUCCA. 217. Сборник тезисов конференции PlantGen2019. 2019. The Fifth International Scientific Conference. Стр. 217.

- Опционально: ваши впечатления от работы вычислительной системы и деятельности ИВЦ НГУ, а также предложения по их совершенствованию.

Для наших задач система работает достаточно быстро и без сбоев.