

УДК 577.212.2

UDC 577.212.2

03.00.00 Биологические науки

Biological sciences

СТРУКТУРНЫЙ И ФИЛОГЕНЕТИЧЕСКИЙ АНАЛИЗ ГЕНОВ VVMYBA1, VVMYBA2 И VVMYBA3**STRUCTURAL AND PHYLOGENETIC ANALYSIS OF VVMYBA1, VVMYBA2 AND VVMYBA3 GRAPEVINE GENES**

Милованов Александр Валериевич
к.б.н., старший преподаватель

Milovanov Alexander Valerievich
Cand.Biol.Sci., Senior Lecturer

Звягин Андрей Сергеевич
к.б.н.

Zvyagin Andrey Sergeevich
Cand.Biol.Sci.

Трошин Леонид Петрович
д.б.н., профессор
Кубанский государственный аграрный университет, Краснодар, Россия

Troshin Leonid Petrovich
Dr.Sci.Biol., Professor
Kuban State Agrarian University, Krasnodar, Russia

Статья представляет структурный и филогенетический анализ генов VIT_02s0033g00410, VIT_02s0033g00390 и VIT_02s0033g00450 генома винограда и близкородственного ортологичного гена MYB114 генома арабидопсиса. Данные гены ответственны за биосинтез антоциана в органах модельных растений и представляют интерес не только для практического производства и селекции, но и для фундаментальных работ по генетике. Данные гены были проанализированы на GC-состав нуклеотидов, наличие cis-регуляторных элементов и промотерных регионов. ДНК и протеиновые последовательности были выровнены для поиска схожих элементов, что позволило далее проанализировать ультраконсервативные домены четырех генов. По результатам поиска и выявления консервативных регионов было построено кластерное древо, позволившее выявить отделение побочных линий генов от, предположительно, главного. При этом, построение консенсусных деревьев, основанных на ДНК и протеиновых последовательностях, выявило абсолютное их сходство. «Древо минимальной эволюции» позволило подсчитать примерные даты появления мутаций и расхождения ветвей генов между собой. При этом за точку отсчета времени бралось само появление семейства *Vitis*, его отделение от порядка *Rosales*. В конце был произведен поиск гомологичных метаболических путей у винограда и арабидопсиса, который выявил наличие в виноградном протеоме гомологичных протеинов. В свою очередь, это уже подтверждает наличие сходных путей биосинтеза и, как следствие, взаимодействий типа «ДНК-белок» и «белок-белок»

The article presents the structural and phylogenetic analysis of VIT_02s0033g00410, VIT_02s0033g00390 and VIT_02s0033g00450 genes of the grapevine genome and the closely related orthologous gene MYB114 of the Arabidopsis genome. These genes are responsible for the biosynthesis of anthocyanin in the organs of model plants and are of interest not only for practical production and breeding, but also for fundamental research. These genes were analyzed for GC-composition of nucleotides, the presence of cis-regulatory elements and promoter regions. DNA and protein sequences were aligned to look for similar elements, which allowed further analysis of the ultraconservative domains of four genes. Based on the results of search and identification of the conservative regions, a cluster tree was constructed, which made it possible to identify the separation of gene sidelines from, presumably, the main one. At the same time, the construction of consensus trees based on DNA and protein sequences revealed their absolute similarity. "The Minimal Evolution Tree" allowed calculating the approximate dates of the appearance of the mutations and the divergence times of the gene branches between each other. At the same time, the appearance of the *Vitis* genus and its separation from the *Rosales* was taken as the time first divergence point. In the end, homologous metabolic pathways were searched between grapevine and Arabidopsis, which revealed the presence of homologous proteins in the grape proteome. In this turn, it already confirms the existence of similar biosynthetic pathways and, as a consequence, interactions such as "DNA-protein" and "protein-protein"

Ключевые слова: АНТОЦИАН, ДНК, ГЕН, ПРОТЕОМ, ЭВОЛЮЦИЯ, *VITIS*, *ARABIDOPSIS*, ГОМОЛОГИЧНЫЕ ГЕНЫ, РЕГУЛЯТОРНЫЕ ДНК-ЭЛЕМЕНТЫ

Keywords: ANTHOCIAN, DNA, GENE, PROTEOM, EVOLUTION, *VITIS*, *ARABIDOPSIS*, HOMOLOGOUS GENES, REGULATORY DNA ELEMENTS

Doi: 10.21515/1990-4665-134-026

Введение

Антоциановая окраска ягод винограда является одной из самых отличительных характеристик данной ценной сельскохозяйственной культуры. И метаболические пути этого признака кодируются тремя генами *VvmybA1*, *VvmybA2* и *VvmybA3* (7, 11). В свою очередь известно, что *VvmybA1* в большинстве сортов не транскрибируется нигде кроме кожицы ягод, в то время как в сорте Бэйли Аликан. А эта особенность была утеряна (8). В норме все три гена располагаются друг за другом во второй хромосоме ядерного генотипа винограда (16), но случается и так, что ввиду огромной делеции они теряются из генома, как это случилось, например, с сортом Пино нуар (23), что привело к появлению ныне популярного сорта винограда Пино блан.

Как известно, отбор человеком различных гибридов и клонов привел к созданию огромного разнообразия сортов винограда. Большинство из них принадлежит к европейским видам *Vitis vinifera* L., происходящих из множественных центров одомашнивания дикорастущего лесного винограда *Vitis silvestris* Gmel. (3). Параллельный отбор клонов и спонтанные скрещивания породили еще большее разнообразие в фенотипах, были отобраны основные белые, розовые и красные сорта.

В виду того, что изначально и далее человеком отбирались гибриды по свойствам ягод, а именно окраске, это создало определенное генетическое разнообразие среди сортов именно в строении генов *VvmybA1*, *VvmybA2* и *VvmybA3*. Но, несмотря на это, строение данных генов является довольно консервативным при их сравнении, что говорит об их значимости и для филогенетического анализа при сравнении культурных с дикими формами. Таким образом, мы проведем небольшой

структурный и филогенетический анализ генов, ответственных за синтез антоциана в ягодах.

Материалы и методы

Для исследований использовались сиквенсы генов VIT_02s0033g00410, VIT_02s0033g00390 и VIT_02s0033g00450 генома винограда и MYB114 генома арабидопсиса (как наиболее близкий по протеиновой последовательности), загруженные с сайта EnsemblPlants (10).

Анализ GC-нуклеотидов производился в GC-Profile (5). Описание регуляторных cis-мотивов в сиквенсах производился с помощью Possum (14) и Nsite (19). Поиск промотерных регионов осуществляли с помощью веб-ресурса Tfsitescan (<http://www.ifti.org/cgi-bin/ifti/Tfsitescan.pl>).

Уравнение ДНК и протеиновых последовательностей производилось с помощью ClustalO (18). Анализ наличия консервативных последовательностей проводился при помощи NCBI CD-search (13). Гомологичные сиквенсы рода *Vitis* и близкородственных растений осуществлялись с помощью NCBI BLAST (1). Филогенетическое древо создавалось при помощи MEGA7 (12).

Для изучения общих свойств между растениями, был выбран ортологичный ген MYB114 арабидопсиса, визуализированный при помощи программы VAS ePlant (21).

Результаты и обсуждение

Изучение GC состава. Сиквенсы, загруженные с EnsemblPlants, были проанализированы на AT/GC-состав для составления первичного представления о качественной структуре генов и полученные данные отображены в таблице 1 и на рисунках 1–4.

Таблица 1. – АТ/GC-состав изученных последовательностей

Название	АТ (%)	GC (%)
VIT_02s0033g00410	58.33	41.67
VIT_02s0033g00390	58.24	41.76
VIT_02s0033g00450	58.80	41.20
MYB114	66.64	33.36

Как мы можем видеть из таблицы 1, содержание GC-нуклеотидов у генов винограда примерно одинаковое и почти на 10% выше, чем у арабидопсиса, что говорит о более высокой термостабильности генов в целом. Тем не менее, как мы можем видеть из изображений 1–4, общая тенденция повышения-понижения количества GC-нуклеотидов сохраняется не только между генами у винограда, но и при сравнении с геном арабидопсиса.

Рисунок 1 – Изменение GC-состава у гена VIT_02s0033g00410

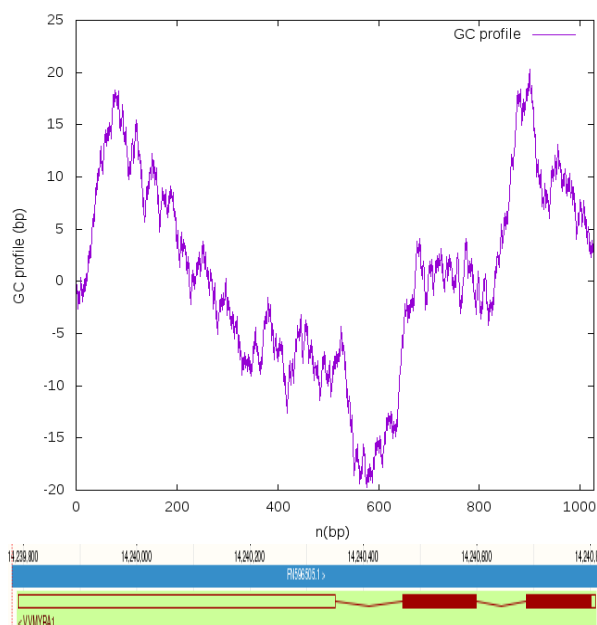


Рисунок 3 – Изменение GC-состава у гена VIT_02s0033g00450

Рисунок 2 – Изменение GC-состава у гена VIT_02s0033g00390

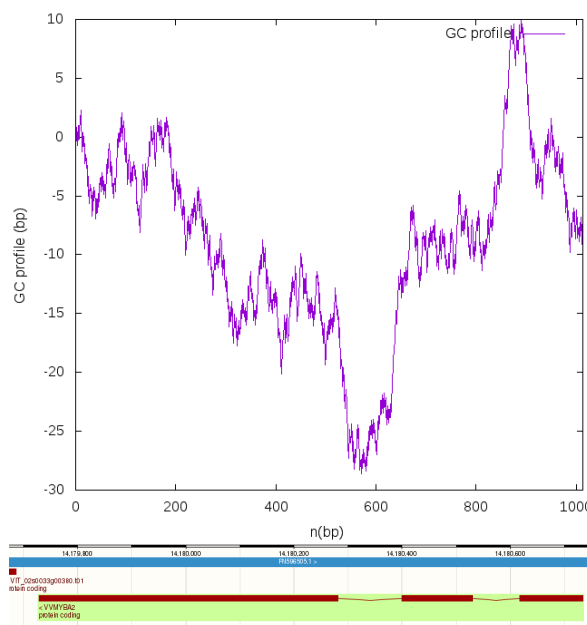
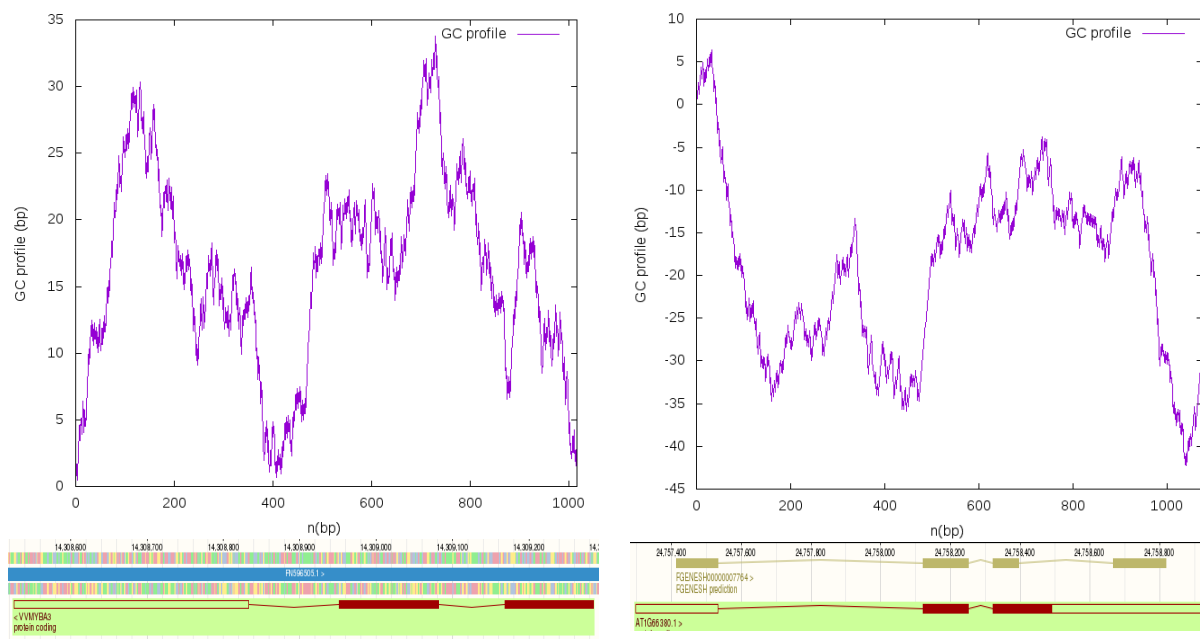


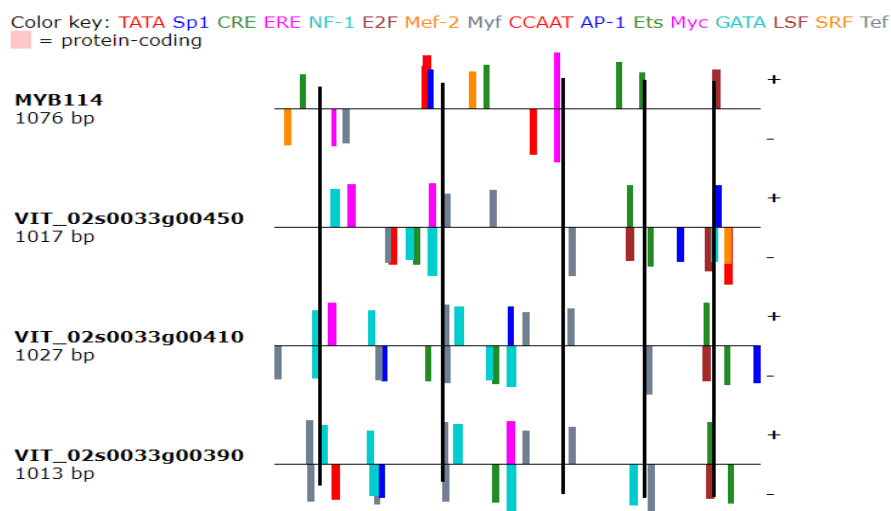
Рисунок 4 – Изменение GC-состава у гена MYB114



При этом важно отметить, что повышение содержания GC совпадает с положением экзонных последовательностей генов у обоих видов. В то время как положение интронных последовательностей совпадает с повышенным содержанием АТ, что не противоречит современным представлениям о строении генома (4, 17). С другой стороны, с 200 по 400 нуклеотид у арабидопсиса имеется повышение количества GC-нуклеотидов, несмотря на то, что там расположен интрон. Но это объяснимо тем, что общее содержание GC-нуклеотидов в гене MYB114 арабидопсиса значительно ниже, чем у винограда и поэтому небольшое повышение их количества так отразилось на диаграмме, во-первых, и наложением нескольких сиквенсов нескольких генов (по данным EnsemblPlants), во-вторых. Тем не менее, к сожалению, до сих пор неизвестно, какие эволюционные силы формируют создание такого разнообразия в ходе эволюции.

Анализ регуляторных мотивов. Как уже было сказано, поиск Cis-элементов проводили при помощи программы Possum. Результаты представлены на рисунке 5 и в таблице 2.

Рисунок 5 – Схематическое изображение Cis-элементов в изученных генах



Черные линии на рисунке 5 указывают на сходные места расположения изученных элементов. Как мы можем видеть, изученные виноградные гены имеют сходные паттерны Cis-элементов, которые расположены примерно в одних и тех же местах. Например, в первом случае – это NF-1-элемент, во втором и третьем – Myf, и в пятом – LSF. Интересно, что в четвертом полном совпадении паттернов выявлено у VIT_02s0033g00410 и VIT_02s0033g00390 (Myf), в то время как у VIT_02s0033g00450 – Ets, но тоже на обратной цепи ДНК. Хотя с другой стороны, такой же элемент был выявлен у MYB114, который является гомологичным этим генам, но уже на «передней» цепи ДНК. Сравнение с геном MYB114 арабидопсиса показывает относительное сродство с генами винограда в расположении элементов, но при том, что у него они расположены зачастую на «обратной» цепи ДНК и они сами по себе другие. Так, в первом случае – это Myc, во втором – Sp1, в третьем – Myc, в четвертом – Ets и в пятом случае – LSF, также как и у генов винограда. Более детальное описание Cis-элементов приведено в таблице 2.

Таблица 2. – Описание Cis-элементов, обнаруженных у четырех изученных генов

МУВ114				VIT_02s0033g00450			
Название	Позиция	Нить ДНК	Сиквенс	Название	Позиция	Нить ДНК	Сиквенс
SRF	24 – 36	-	tctatctttggtc	NF-1	120 - 137	+	tggtgggaaaatcccaga
Ets	59 – 69	+	tgagggaagg	ERE	155 - 168	+	aaatcacctcacc
Myc	128 – 137	-	ggcaaatggc	Tef	233 - 244	-	cagaggcttgcg
Tef	154 – 165	-	agctggtatgta	TATA	241 - 255	-	tgcgagactttatag
TATA	329 – 343	+	acatatagaccgg	GATA	278 - 290	-	gggtttatctctc
TATA	331 – 345	+	atatatagaccggg	Ets	293 - 303	-	ttcttctgga
Sp1	340 – 352	+	cgggggttggfcc	ERE	324 - 337	+	ttgtgacctgcca
SRF	434 – 446	+	gaccaactattgg	NF-1	323 - 340	-	gtggtgacctgccaata
Ets	465 – 475	+	tagaggaagat	Tef	356 - 367	+	agcagctctccc
CCAAT	566 – 581	-	tctacgattggtttgt	Myf	452 - 463	+	gaacagcttcag
Myc	622 – 631	+	aacacgtgcg	Myf	617 - 628	-	ctgcagcttttt
Myc	622 – 631	-	aacacgtgcg	Ets	739 - 749	+	gagggggaacca
Ets	757 – 767	+	tgagggaag	LSF	737 - 751	-	cggagggggaaccaga
Ets	810 – 820	+	gagaggaaaat	Ets	782 - 792	-	catttctcag
LSF	971 – 985	+	acctggtcggaccgc	AP-1	845 - 855	-	atcgagtcaac
				LSF	903 - 917	-	caacctcgaacctgt
				GATA	915 - 927	-	tggttctatctcat
				AP-1	925 - 935	+	catgactcggga
				SRF	943 - 955	-	cccctaattggtt
				CCAAT	943 - 958	-	cccctaattggttgat
VIT_02s0033g00410				VIT_02s0033g00390			
Название	Позиция	Нить ДНК	Сиквенс	Название	Позиция	Нить ДНК	Сиквенс
Tef	2 – 13	-	agcatgaatgca	Myf	68 - 79	+	ccacagctgtag
NF-1	81 – 98	+	tggtgggtgtatcccaga	Myf	71 - 82	-	cagctgtagttt
NF-1	82 – 99	-	gttgggtgtatcccagaa	NF-1	93 - 110	+	ccttggcaaggctttgga
ERE	116 – 129	+	aaatcacctcacc	TATA	122 - 136	-	ggtgaggtttatag
GATA	201 – 213	+	ttgcgataagcat	GATA	194 - 206	+	ttgcgataagcat
Myf	216 – 227	-	ctccagaagccg	NF-1	199 - 216	-	ataagcatctctccagaa
AP-1	227 – 237	-	gaaaagtcagt	Myf	209 - 220	-	ctccagaagccg
Ets	320 – 330	-	ggcttctgga	AP-1	220 - 230	-	gaaaagtcagt
Myf	357 – 368	+	ccacagctgtag	Myf	350 - 361	+	ccacagctgtag
Myf	360 – 371	-	cagctgtagttt	Myf	353 - 364	-	cagctgtagttt
NF-1	382 – 399	+	ccttggcaaggctttgga	NF-1	375 - 392	+	ccttggcaaggctttgga
GATA	448 – 460	-	gggtttatctctc	Ets	456 - 466	-	ttcttctgga
Ets	463 – 473	-	ttcttctgga	ERE	487 - 500	+	tggtgacctgcca
AP-1	495 – 505	+	ggtgactatgc	NF-1	486 - 503	-	gtggtgacctgccaata
NF-1	493 – 510	-	gtggtgactatgccaata	Tef	519 - 530	+	agcagctctccc
Tef	526 – 537	+	agcagctctccc	Myf	615 - 626	+	gaacagcttcag
Myf	622 – 633	+	gaacagcttcag	GATA	743 - 755	-	cttgatctggc
Myf	787 – 798	-	ctgcagcttttt	Myf	780 - 791	-	ctgcagcttttt
Ets	909 – 919	+	gagggggaacca	Ets	903 - 913	+	gagggggaacca
LSF	907 – 921	-	cggagggggaaccaga	LSF	901 - 915	-	cggagggggaaccaga
Ets	952 – 962	-	catttctcag	Ets	946 - 956	-	catttctcag
AP-1	1015 - 1025	-	atcgagtcaac				

Из данной таблицы видно, что всего было обнаружено 15, 20, 22 и 21 элемент для каждого сиквенса: МУВ114, VIT_02s0033g00450, VIT_02s0033g00410 и VIT_02s0033g00390 соответственно. Анализ присутствия/отсутствия элементов представлен в таблице 3 ниже.

Таблица 3. – Таблица встречаемости Cis-элементов у изученных генов

Название	МУВ114	VIT_02s0033g00450	VIT_02s0033g00410	VIT_02s0033g00390
TATA	-	+	-	+
ССААТ	+	+	-	-
Sp1	+	-	-	-
AP-1	-	+	+	+
CRE	-	-	-	-
Ets	+	+	+	+
ERE	-	+	-	+
Myc	+	-	-	-
NF-1	-	+	+	+
GATA	-	+	+	+
E2F	-	-	-	-
LSF	+	+	+	+
Mef-2	-	-	-	-
SRF	+	-	-	-
Myf	-	+	+	+
Tef	+	+	+	+

Как мы можем увидеть из таблицы 3, были обнаружены уникальные и распространенные элементы, в то время как некоторые не были обнаружены вообще. С другой стороны, некоторые элементы были обнаружены, например, только у двух сиквенсов, что вызывает сомнения в их функциональной значимости, несмотря на их наличие.

Таким образом, у всех сиквенсов присутствуют такие элементы как Ets, LSF и Tef, что говорит о их ультраконсервативной природе в виду распространения между видами, а также в виду того, что Ets и LSF располагаются примерно в одинаковых местах, чего нельзя сказать о Tef, который находится в сходных позициях у сиквенсов VIT_02s0033g00410 и VIT_02s0033g00390.

Далее, поиск функционально значимых транскрипционных факторов проводили при помощи Nsite. Всего было обнаружено 34 транскрипционных фактора, результаты с уровнем гомологии сиквенсов не ниже 80% отображены в таблице 4.

Таблица 4. – Обнаруженные транскрипционные факторы

VIT_02s0033g00410			
Название вида	Транскрипционный фактор	Сиквенс	
Spinaciaoleracea	PetH/TFBS: CT-B /BF: unknown nuclear factor	317 catccacttc	308
Arabidopsisthaliana	AVP1/TFBS: Q-motif /BF: homeodomain transcription factor	505 gcatagtcaccac	493
Spinaciaoleracea	rps22/TFBS: rGC /BF: unknown nuclear factor	155 tccatgggtccctttt	170
Phaseolusvulgaris	grp1.8/TFBS: VSF-1 BS1 /BF: VSF-1	116 ttccatttgatgtgg	102
Arabidopsisthaliana	MYB44 (At5g67300)/TFBS: VRE1 /BF: VIP1	359 acagctgta	367
Arabidopsisthaliana	ANAC019/TFBS: HB05/HB12 BS2 /BF: HB05; HB12	774 caattattt	766
Arabidopsisthaliana	Synthetic oligonucleotides/TFBS: R2R3-MYB TFs BS	307 accacaacc	299
Всего 7 транскрипционных мотивов			
VIT_02s0033g00390			
Названиевида	Транскрипционный фактор	Сиквенс	
Arabidopsisthaliana	Synthetic oligonucleotides/TFBS: Hex motif /BF: TGA1; GBF1;	20 tgacgtgg	27 300 tgacgtgg 307
Triticumaestivum	H3/TFBS: Hex /BF: HBP-11(17); HBP-1b(c38);	307 ccacgtca	300/27 ccacgtca 20
Lycopersiconesculentum	rbcS3A/TFBS: GT-1/16 J1 /BF: unknown nuclear factor	869 attaattgtgt	858
Arabidopsisthaliana	Synthetic oligonucleotides/TFBS: Hex motif /BF: OBF4; OBF5	18 ggtgacgtgga	28/298 ggtgacgtgga 308
Arabidopsisthaliana	MYB44 (At5g67300)/TFBS: VRE1 /BF: VIP1	70 acagctgta	78/352 acagctgta 360
Arabidopsisthaliana	ANAC019/TFBS: HB05/HB12 BS2 /BF: HB05; HB12	767 caattattt	759
Arabidopsisthaliana	Synthetic oligonucleotides/TFBS: R2R3-MYB TFs BS	20 accacaacc	12/300 accacaacc 292
Всего 12 транскрипционных мотивов			
VIT_02s0033g00450			
Названиевида	Транскрипционный фактор	Сиквенс	
Spinaciaoleracea	rps22/TFBS: rGC /BF: unknown nuclear factor	194 tccatgggtccctttg	209
Nicotianatabacum	Adh2/TFBS: D-box /BF: D factor	807 ctgggtcca	816
Phaseolusvulgaris	grp1.8/TFBS: VSF-1 BS1 /BF: VSF-1	155 ttccatttgatgtgg	141
Arabidopsisthaliana	YUC4 (At5g11320)/TFBS: STY1 BS (YUC4) /BF: STY1	45 acacttactc	35
Arabidopsisthaliana	ANAC019/TFBS: HB05/HB12 BS2 /BF: HB05; HB12	604 caattattt	596
Arabidopsisthaliana	Synthetic oligonucleotides/TFBS: P8 (ext) /BF: GT-4	3 gattaacaacttggaaacttaaat	27
Всего 6 транскрипционных мотивов			
MYB114			
Названиевида	Транскрипционный фактор	Сиквенс	
Hordeumvulgare	HVA22/TFBS: G-box /BF: unknown nuclear factor	631 cgcacgtgtt	622
Lycopersiconesculentum	rbcS3C/TFBS: AT-rich V /BF: unknown nuclear factor	506 tagaacattttaaaaaaaaaaaaaaaaaaaaaatctt	471
Arabidopsisthaliana	STK/TFBS: GA-2 /BF: BPC1	691 agaagaga	699
Arabidopsisthaliana	STK/TFBS: GA-4 /BF: BPC1	809 agagaggaa	817
Solanumtuberosum	St4cl-1/TFBS: FP2 /BF: unknown nuclear factor	354 tcggaccaacccc	342
Arabidopsisthaliana	BiP2/TFBS: TL1 /BF: TBF1	850 gaagaaga	842
Arabidopsisthaliana	Synthetic oligonucleotides/TFBS: SEQ20 /BF: WRKY70	563 taaacgactttt	552
Oryzasativa	OsYSL2/TFBS: IDE2 (1) /BF: IDEF2	798 gccaaagtatca	808
Arabidopsisthaliana	Synthetic oligonucleotides/TFBS: O-box (oligo1) /BF: ORA47	429 ataccgaccac	440
Всего 9 транскрипционных мотивов			

Как мы можем видеть из таблицы 4, различные транскрипционно значимые факторы, принадлежащие к разным видам растений, были найдены. Тем не менее, в виду того, что они являются ультраконсервативными и распространены среди всего живого, это не удивительно, а вполне ожидаемо. Помимо этого, как мы видим, у виноградных генов ярко выражена гомология с арабидопсисом в виду того, что абсолютное большинство выявленных транскрипционных факторов

принадлежит именно к классу факторов этого растения, что дополнительно указывает на их в прошлом общую эволюционную филогению.

Интересно, что у всех трех изученных генов нашлись сиквенсы, которые относятся к генам регулирования синтеза антоциана арабидопсиса, в то время как у самого гена MYB114 прямых совпадений не нашлось. Тем не менее, большинство совпадений с транскрипционными факторами арабидопсиса у MYB114 показали совпадения с факторами клеточного и органного развития, что также указывает на его значимость в регулировании метаболических путей. В любом случае это было неожиданным, так как протеиновая последовательность арабидопсиса имеет достаточно высокую гомологию, хотя и короче, чем у винограда.

Несмотря на высокую гомологию генов, достаточно сложно найти у них одинаковые транскрипционные факторы, тем не менее, некоторые совпадают. Например, ttccatttgatgtgg присутствует у VIT_02s0033g00410 и VIT_02s0033g00450; acagctgta у VIT_02s0033g00410 и VIT_02s0033g00390; caattattt у всех трех виноградных генов, но не у арабидопсиса; ассасаасс у VIT_02s0033g00410 и VIT_02s0033g00390. Домены tgacgtgg, tgacgtgg, ссacgtca и ggtgacgtgga широко распространены внутри гена VIT_02s0033g00390, в то время как не могут быть обнаружены ни у одного другого гена.

Таким образом, мы видим, что, несмотря на то, что данные четыре гена кодируют гомологичную протеиновую последовательность, имеются структурные различия, что говорит о том, что гены разнесены в эволюционной дистанции. Также важно отметить, что у генов винограда имеются сходства в строении транскрипционных факторов с другими генами арабидопсиса типа MYB, что указывает нам на гомологичность и с другими генами.

Поиск промотерных регионов. Далее, мы провели поиск гомологичных промотерных регионов у всех четырех выбранных последовательностей (таблица 5).

Таблица 5. – Список обнаруженных промотерных регионов

VIT_02s0033g00410	VIT_02s0033g00390	VIT_02s0033g00450	MYB114
8-13 DAF-12	16-9 HvGAMYB	12- 5 SBF-1	68-73 DAF-12
30-25 DAF-12	17-24 TGA1	21-16 DAF-12	79-74 DAF-12
183-178 DAF-12	20-25 ASF-1	47-52 DAF-12	105-110 DAF-12
303-296 HvGAMYB	20-25 HBP-1	69-64 DAF-12	142-147 DAF-12
310-315 DAF-12	57-63 GT-1	222-217 DAF-12	199-194 DAF-12
346-352 GT-1	85-90 DAF-12	306-301 DAF-12	246-241 DAF-12
374-379 DAF-12	115-110 DAF-12	340-345 DAF-12	309-304 DAF-12
404-399 DAF-12	176-171 DAF-12	446-451 DAF-12	330-324 GT-1
476-471 DAF-12	296-289 HvGAMYB	467-462 DAF-12	376-381 DAF-12
496-502 GT-1	297-304 TGA1	499-505 GT-1	464-459 DAF-12
510-515 DAF-12	300-305 ASF-1	563-568 DAF-12	547-542 DAF-12
616-621 DAF-12	300-305 HBP-1	686-692 GT-1	604-609 DAF-12
637-632 DAF-12	339-345 GT-1	778-783 DAF-12	643-648 DAF-12
669-675 GT-1	367-372 DAF-12	785-780 DAF-12	662-667 DAF-12
733-738 DAF-12	397-392 DAF-12	822-817 DAF-12	755-760 DAF-12
856-862 GT-1	469-464 DAF-12	975-980 DAF-12	823-828 DAF-12
948-953 DAF-12	503-508 DAF-12	982-977 DAF-12	910-904 Zmhoxla
955-950 DAF-12	609-614 DAF-12	1002-997 DAF-12	930-924 GT-1
992-987 DAF-12	630-625 DAF-12		1002-997 DAF-12
	662-668 GT-1		1060-1066 MybSt1
	726-731 DAF-12		1075-1069 GT-1
	849-855 GT-1		
	942-947 DAF-12		
	949-944 DAF-12		
	986-981 DAF-12		

Как мы можем видеть, всего было выявлено 83 промотерных региона, 19 (VIT_02s0033g00410), 18 (VIT_02s0033g00450), 25 (VIT_02s0033g00390) и 21 (MYB114) для каждой изученной последовательности. При этом очевидно наличие у всех генов таких регионов как DAF-12 и GT-1. При этом, только у VIT_02s0033g00410 и VIT_02s0033g00390 были обнаружены промотеры HvGAMYB, в то время как у VIT_02s0033g00450 ничего подобного не было. Но у MYB114 арабидопсиса был обнаружен фактор MybSt1, родственник HvGAMYB

винограда. Также был обнаружен ряд уникальных для каждой последовательности промотеров, которые отображены в таблице 5. Интересно отметить, что в отличие от винограда у арабидопсиса MybSt1 располагается в другом конце гена, что вполне ожидаемо в виду того, что часть его cis-элементов расположена на обратной цепи.

Мы полагаем, что появление такой последовательности как DAF-12 является ошибкой или просто совпадением, поэтому мы не принимаем ее во внимания. Несмотря на это, стоит отметить, что Plant Transcription Factor Database (9) при поиске указывает на последовательности, родственные MYB фактору пшеницы. Тем не менее, не вызывает сомнения наличие GT-1, MybSt1 и HvGAMYB в связи с тем, что они являются последовательностями, сходными по их назначению. Также, по данным Plant Transcription Factor Database, все эти три домена являются родственными MYB-like ДНК связывающими доменами, выполняющими сходные функции, такие как регуляция транскрипции.

Сравнение ДНК и протеиновых последовательностей. Для сравнения ДНК и кодируемых трансляционных протеиновых последовательностей мы провели их «уравнение», используя программу ClustalO. Результаты сравнения последовательностей приведены на рисунках 6 и 7.

Рисунок 5 – Сравнение ДНК последовательностей

MYB114 VIT_02s0033g00450 VIT_02s0033g00410 VIT_02s0033g00390	ATTC AATATACCCGACCAACTATTGAAACTTTAATGTGAACTT--AGAGGAAGATTTTT TTTATCTCCCTTCTCCGGAACCTGAACTCTCTTTTCAAGTGGTGACCATGCCAATA TTTATCTCCCTTCTCCGGAACCTGAACTCTCTTTTCAAGTGGTGACCATGCCAATA TTTTTTTCCCTTCTCCGGAACCTGAACTCTCTTTTCAAGTGGTGACCATGCCAATA * * * * *	47 34 51 50
MYB114 VIT_02s0033g00450 VIT_02s0033g00410 VIT_02s0033g00390	TTTTTTTTTTTTTTTAAATGTTCTATGTTTATCATCGTTACATATTTCAAAAAA GTTCTTGACATCATTAGCAGTCTCCCTGGAAGCCTACCCGCAATCAAGGACCATCTATA GTTCTTGACATCATTAGCAGTCTCCCTGGAAGCCTACCCGCAATCAAGGACCATCTATA * * * * *	53 40 57 56
MYB114 VIT_02s0033g00450 VIT_02s0033g00410 VIT_02s0033g00390	TGGG---AACTTCCGAAAGTGGTTAATCTACAGATTGGTTGTCTCTAATTCATAAA CAGCACAACATCAGACAAGTA-TCAACTGAACAGCT----CTTTGTAAAGTTCTGAA CAGCACAACATCAGACAAGTA-TCAACTGAACAGCT----CTTTGTAAAGTTCTGAA CAGCACAACATCAGACAAGTA-TCAACTGAACAGCT----CTTTGTAAAGTTCTGAA * * * * *	59 45 62 61
MYB114 VIT_02s0033g00450 VIT_02s0033g00410 VIT_02s0033g00390	TGAAATTTAGTAAA-----GTTTTATACGAACACGCTGATGTGTATAGTAC--A CAGCTTCAGAACTAATAAATCATTATCAAGTAGTACTTGAGTTATTATAGACTTGC CAGCTTCAGAACTAATAAATCATTATCAAGTAGTACTTGAGTTATTATAGACTTGC CAGCTTCAGAACTAATAAATCATTATCAAGTAGTACTTGAGTTATTATAGACTTGC * * * * *	64 51 68 67
MYB114 VIT_02s0033g00450 VIT_02s0033g00410 VIT_02s0033g00390	TTGTT--TCTATTGGTGGCATAGATTCTTTATGATAAAATATAGA----- CTGTCCCAACAATTTGGAAGCTAATCATGAGATCAACCTCGTCTAATGCAAACTCT CTGTCCCAACAATTTGGAAGCTAATCATGAGATCAACCTCGTCTAATGCAAACTCT CTGTCCCAACAATTTGGAAGCTAATCATGAGATCAACCTCGTCTAATGCAAACTCT * * * * *	69 57 74 73
MYB114 VIT_02s0033g00450 VIT_02s0033g00410 VIT_02s0033g00390	---AAGACAGCAATCTCTTAGGTTAACTAATGCAATATTGGTTTTGTAGGGCT CCTCTTGTATCCGGCTTCAAATAATTGAGCCATCTCAACTGCAGCTTTTTCGGCAT CCTCTTGTATCCGGCTTCAAATAATTGAGCCATCTCAACTGCAGCTTTTTCGGCAT CCTCTTGTATCCGGCTTCAAATAATTGAGCCATCTCAACTGCAGCTTTTTCGGCAT * * * * *	74 63 80 79
MYB114 VIT_02s0033g00450 VIT_02s0033g00410 VIT_02s0033g00390	AAATCGGTGCGAGAAAGTTGAGACTAAG-ATGGTTAACTATTTGAAGCCAAAGTA--- CTATTCAA-----CCCTGCAAACTCAAGAAAACGAACTCTTT-AAGCAAAACAGAA CTATTCAA-----CCCTGCAAACTCAAGAAAACGAACTCTTT-AAGCAAAACAGAA CTATTCAA-----CCCTGCAAACTCAAGAAAACGAACTCTTT-AAGCAAAACAGAA * * * * *	80 68 85 84

Рисунок 6 – Сравнение протеиновых последовательностей

MYB114 VIT_02s0033g00450 VIT_02s0033g00410 VIT_02s0033g00390	MEGSSKGLRKGANTAEEDSLRQCIYKYGEGKHWQVPLRAGLNRCRKSCLRRLNLYKPS --MESLGVRRKGAHQEEDVLLRKCIEKYEGEKHMLVPLRAGLNRCRKSCLRRLNLYKPD --MESLGVRRKGAHQEEDVLLRKCIEKYEGEKHMLVPLRAGLNRCRKSCLRRLNLYKPD --MKS LGVRRKGAHQEEDVLLRKCIEKYEGEKHMLVPLRAGLNRCRKSCLRRLNLYKPD * * * * *	60 58 58 58
MYB114 VIT_02s0033g00450 VIT_02s0033g00410 VIT_02s0033g00390	IKRGEFSSDEVLDLRLHLKLLGNRWSLIAGRLPGRANDVKNYINHLKXKHEPCCKTKI IKRGEFALDEVLDLRLHLKLLGNRWSLIAGRLPGRANDVKNYINHLKXKHEPCCKTKI IKRGEFALDEVLDLRLHLKLLGNRWSLIAGRLPGRANDVKNYINHLKXKHEPCCKTKI IKRGEFALDEVLDLRLHLKLLGNRWSLIAGRLPGRANDVKNYINHLKXKHEPCCKTKI * * * * *	126 118 118 118
MYB114 VIT_02s0033g00450 VIT_02s0033g00410 VIT_02s0033g00390	KRIN-----IITPPNPAQIV-----DIF----- DKPQTHSKTKAIKSRPLDRKNC-----NITDKGTHQVY DKPQTHSKTKAIKSRPHKFKSALPRFELKTTAVDTFDTQVSTSRKPSSTSPQNDIILHW KKPQTHSKTKAIKSRPHKFKSALPRFELKTTAVDTFDTQVSTSRKPSSTSPQNDIILHW * * * * *	138 151 178 178
MYB114 VIT_02s0033g00450 VIT_02s0033g00410 VIT_02s0033g00390	GRINPGR----- ESLLAEHAQMDQETDFASAGEMLIASLRETEATQKGPMDGVIETQIGGEGDFPFDVGF ESLLAEHAQMDQETDFASAGEMLIASLRETEATQKGTSHSKTK-AIKSRPHKFKSALPR * * * * *	138 158 238 237
MYB114 VIT_02s0033g00450 VIT_02s0033g00410 VIT_02s0033g00390	----- NDTPNTQVHILI----- FELKTTAVDTFDTQVSTSKLHVITTE	139 158 250 265

Для наглядности гомологичности и идентичности последовательностей ДНК приведена только часть сравнения середины нуклеотидных цепей, в то время как 5' и 3' концы имеют разнородное нуклеотидное строение, а также укорочены. Тем не менее, суммируя приведенные факты с прошлым разделом, мы можем сделать вывод, что они не отличаются функционально. Что же касается укорочения генов, то это вполне нормальное эволюционное явление, которое встречается довольно часто и служит для удаления «лишней» ДНК (2, 6, 20).

Сравнение протеиновых последовательностей показало высокую консервативность, несмотря на то, что само строение генов у винограда и арабидопсиса сильно отличается даже в центральной части. Тем не менее, как мы можем видеть на картинке 6, часть аминокислот у MYB114 отсутствует, также как и у VIT_02s033g00450, но это, видимо, не влияет на их функциональность. При этом важно отметить, что концы остальных двух протеиновых последовательностей винограда сильно отличаются по набору аминокислот, хотя и выполняют схожие функции.

В конце стоит отметить на расположение названия сиквенсов. В обоих случаях они расположены в одинаковом порядке, который нам указывает на гомологию между ними. Следовательно, наиболее гомологичными сиквенсами являются MYB114 и VIT_02s033g00450, в то время как остальные два отделены филогенетически. В связи с этими данными мы провели поиск консервативных регионов и филогенетический анализ.

Поиск консервативных регионов. В виду того, что выбранные для изучения гены являются близкородственными, следовательно, они должны обладать сходными консервативными последовательностями. Для подтверждения данного предположения мы провели анализ в NCBI CD-search. Результаты представлены ниже на рисунках 7 – 10.

Рисунок 7 – Консервативные домены VIT_02s0033g00410

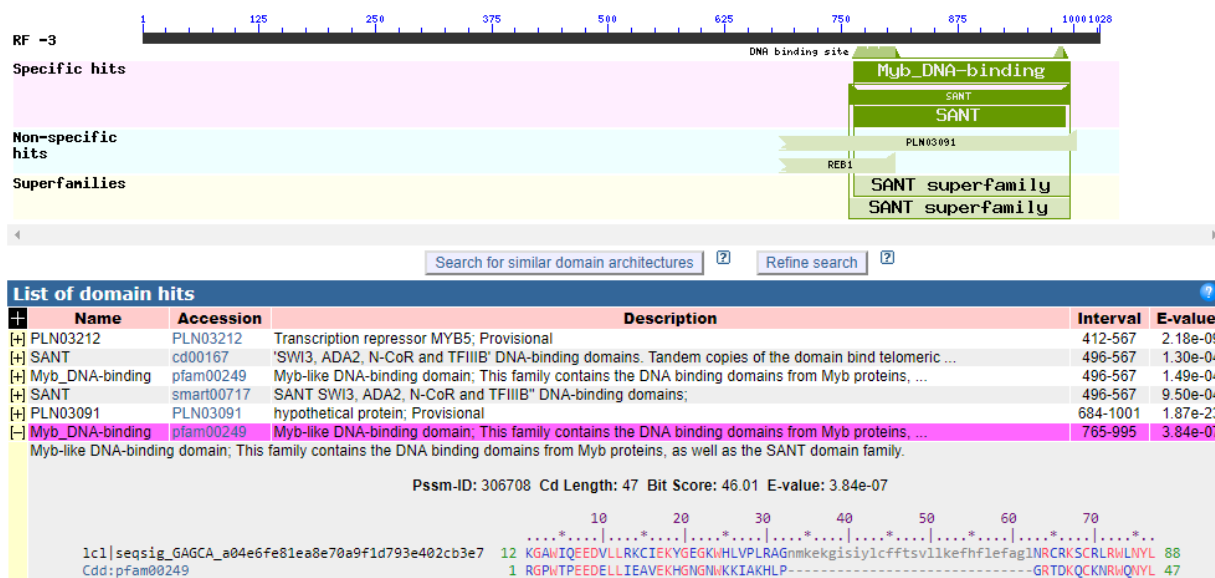


Рисунок 8 – Консервативные домены VIT_02s0033g00390

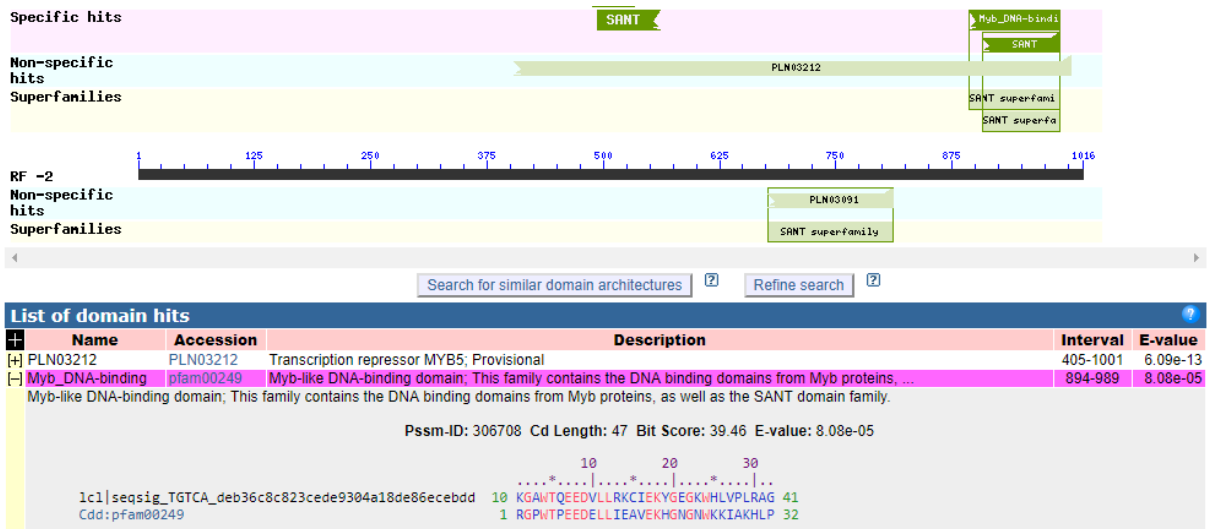


Рисунок 9 – Консервативные домены VIT_02s0033g00450

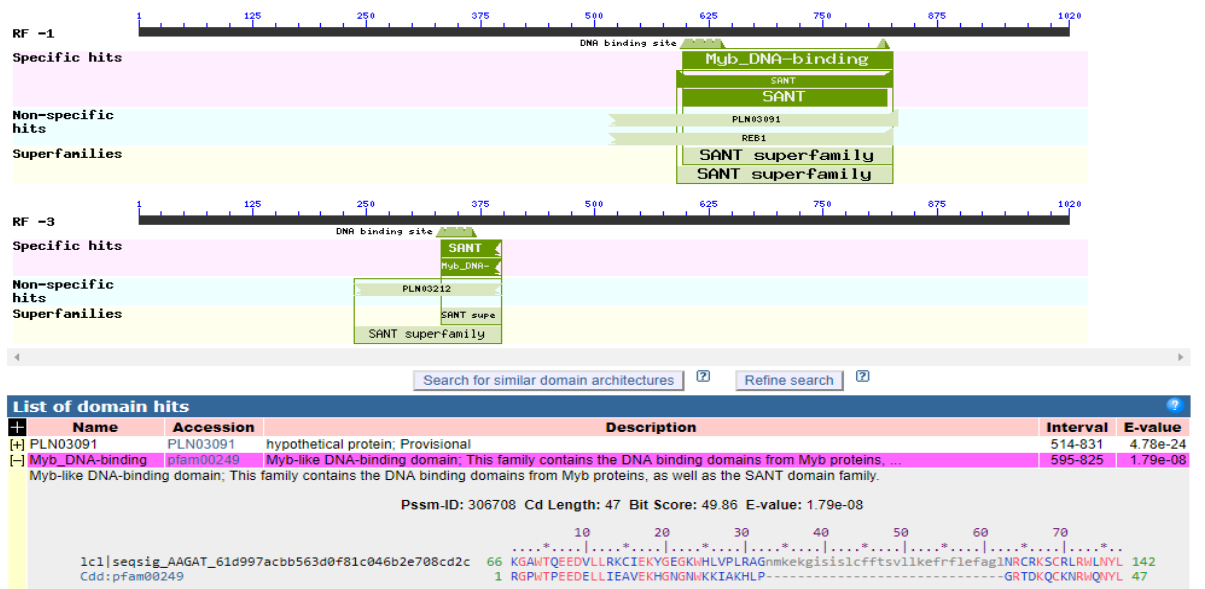
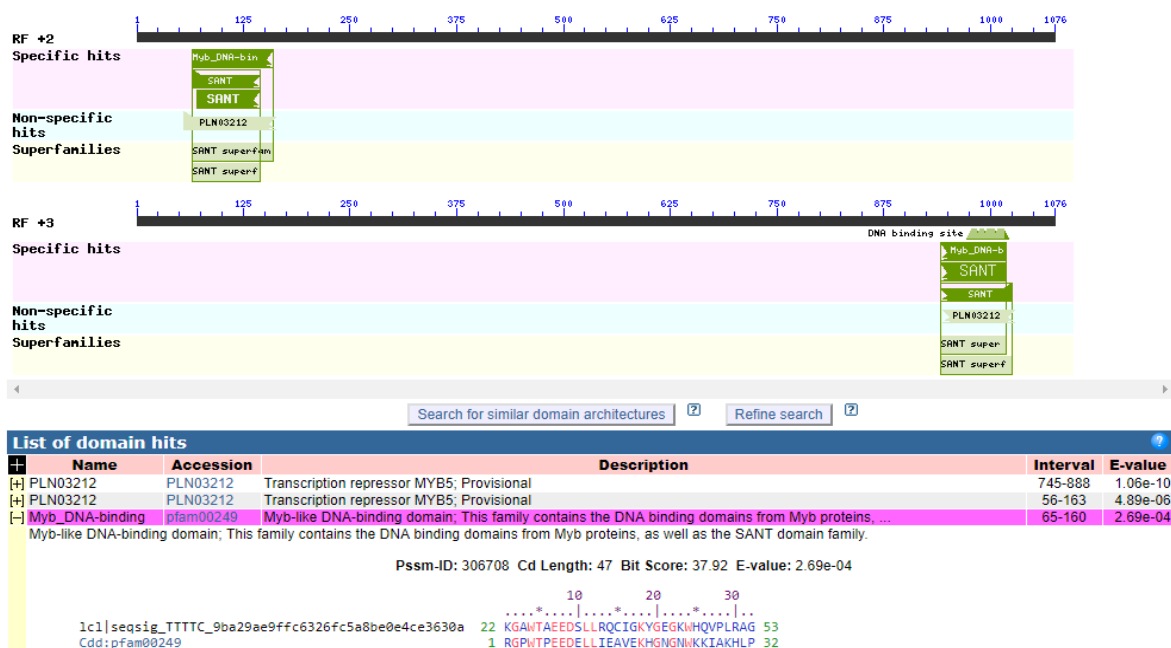


Рисунок 10 – Консервативные домены MYB114



Как мы можем видеть, все изученные гены содержат Myb-like DNA binding domain, как и предполагалось. При этом стоит отметить, что E-value, определяющее достоверность, показало высокую точность результата. При этом все выявленные консервативные последовательности принадлежат к семейству SANT доменов, ответственных за связывание Myb протеинов.

Также стоит обратить внимание, что у всех изученных последовательностей одна и та же часть идентифицировалась как консервативный домен: KGAWIQEEDVLLRKCIEKYGEGKWHLVPLRAG. При сравнении в ClustalO было выявлено, что данные участки практически идентичны у обоих видов, за исключением изменений шести аминокислот, что, очевидно, не влияет на функциональность.

Таким образом, мы видим, что изученные гены обладают и ультраконсервативными доменами связывания ДНК с Myb протеинами. Это еще раз говорит об их ортологичной взаимосвязи и чтобы установить, как именно она проявляется в рамках эволюционного родства, мы провели филогенетический анализ.

Филогенетический анализ. Для проведения филогенетического анализа взаимосвязей между изученными генами мы выбрали две основные стратегии: 1) сравнение ДНК сиквенсов; 2) сравнение протеиновых сиквенсов. Результаты сравнения отображены на рисунках 11 и 12.

Рисунок 11 – Филогенетическое сравнение ДНК последовательностей

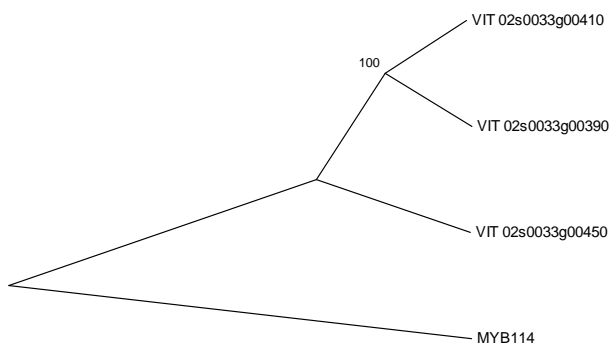
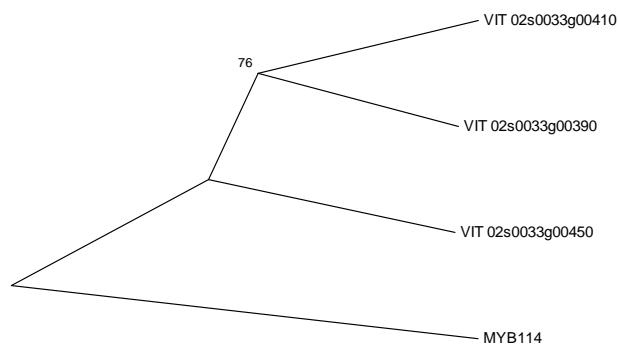


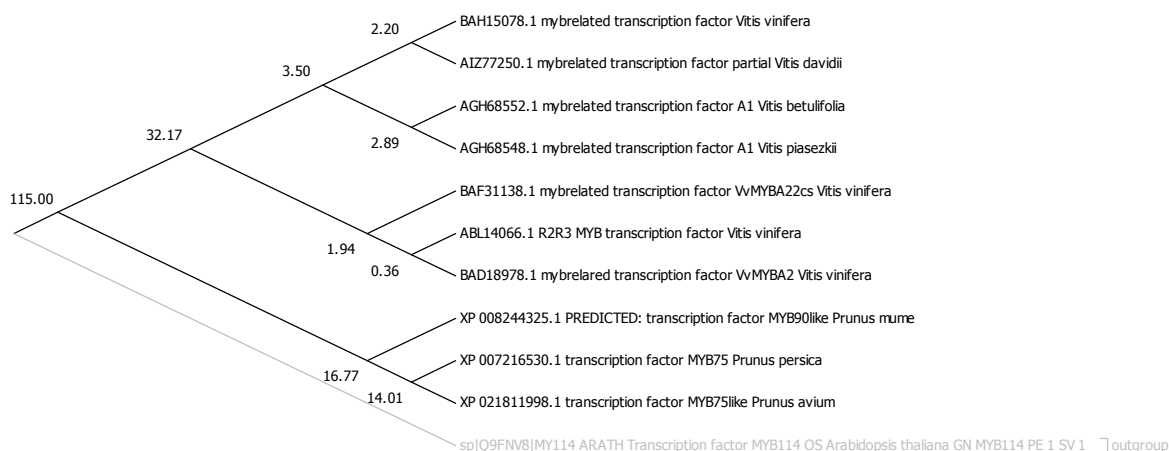
Рисунок 12 – Филогенетическое сравнение протеиновых последовательностей



Как мы можем видеть из рисунков, представленных выше, филогенетическое расстояние между семействами *Brassicaceae* и *Vitaceae* достаточно большое, при этом явно прослеживается отделение от, предположительно, основного гена (VIT_02s0033g00410) побочных (VIT_02s0033g00390 и VIT_02s0033g00450).

Для представления более точного временного промежутка появления мутаций мы построили временное древо в программе MEGA7 (с использованием параметров «древа минимальной эволюции»), получив дополнительные сиквенсы генов из баз данных NCBI. Чтобы древо получилось наиболее консервативным, мы использовали протеиновые последовательности.

Рисунок 13 – Филогенетическое временное древо, основанное на протеиновых последовательностях трех изученных генов



В связи с палеонтологическими остатками представителей рода *Vitis* известно, что выделение его из подкласса Розид произошло примерно 110-120 млн. лет назад (22). Поэтому размер шкалы – это каждый номер, который мы видим на узлах расхождения, умноженный на миллион. Таким образом, нами были взяты три семейства растений для вычисления времени произошедших мутаций: арабидопсис, розация и витацее. При этом стоит отметить, что в оригинальном древе арабидопсис располагается между розацией и витацией, точно также как и в классификации Вилкстрема и др., но так как в нашем древе он был выбран аут группой, то поэтому он вынесен за древо. Тем не менее, мы можем видеть четкое разделение групп по генам, как и в прошлый раз.

Таким образом, мы можем видеть, что мутации, приведшие к разделению известных генов, произошли примерно 32,17 и 3,5 млн. лет назад, пока что по неизвестным причинам. Также было обнаружено, что выявленные внутригенные различия у винограда появились примерно 2 и 3 млн. лет назад, при этом стоит отметить, что различия, появившиеся между видами при видообразовании старше, чем различия, обнаруженные внутри рода *Vitis*.

Таким образом, мы видим, что ортологичные гены, кодирующие схожие белки, имеются у многих двудольных растений. Также важно, что не смотря на высокую консервативность, эти протеиновые последовательности возможно использовать для филогенетических исследований, так как они дают четкое разграничение между исследуемыми семействами, родами и видами.

Изучение интеракций протеиновой последовательности гена МУВ114. К сожалению, пока что еще не существует столь подробного описания взаимодействий белок-белок и белок-ДНК для винограда. Поэтому, для лучшего понимания строения и взаимодействия протеиновых последовательностей и их конформаций у винограда, мы провели исследование аналогичных функций у их ортолога МУВ114 арабидопсиса.

Рисунок 14 – Карта интеракций МУВ114 с взаимно экспрессирующимися белками



This image was generated with the Interaction Viewer at bar.utoronto.ca/eplant by Waese et al. 2017

Как мы можем видеть из рисунка 14, ген МУВ114 (AT1G66380 по европейской классификации) арабидопсиса ко-экспрессируется с тремя другими генами (AT1G63650, AT4G00480 и AT4G09820). При этом, коэффициент ко-экспрессии составляет от 0,25 до 0,4, что является

достаточно сильным показателем. При этом важно отметить, что данные три белка принадлежат к одному суперсемейству, которое включает в себя ДНК связывающие белки.

Далее, для того чтобы найти сходные метаболические пути у винограда, мы использовали EnsemblPlants, UniProt и NCBI BLAST. Стоит уточнить, что ввиду большого размера генов и дальности генетического расстояния между виноградом и арабидопсисом, мы использовали только протеиновые сиквенсы для поиска гомологичных последовательностей.

При сравнении с протеиновым сиквенсом, который кодируется геном AT1G63650, у винограда был обнаружен «musc anthocyanin regulatory protein», что говорит о наличии ортолога. При этом идентичность составила только 44%, в то время как Evalue – $3.1e-167$, что говорит о том, что это гомологи.

При сравнении с протеиновым сиквенсом, который кодируется геном AT4G00480, у винограда не было обнаружено существенных сходств в протеоме. Тем не менее, «musc anthocyanin regulatory protein isoform X1» и «transcriptionfactor EGL1» показали существенные результаты в плане частичной гомологии функциональных участков. Поэтому вполне возможно сделать предположение о наличии у винограда ортолога данного протеина.

При сравнении с протеиновым сиквенсом, который кодируется геном AT4G09820, у винограда были обнаружены не охарактеризованные протеины. При этом важно, что, несмотря на низкую идентичность (49%), Evalue показал хорошие результаты ($5e-105$). Это также как и в прошлый раз говорит о том, что у винограда в протеоме есть гомолог данного сиквенса, хотя и достаточно отдаленно лишь его напоминающий.

Очень важен тот факт, что все три обнаруженных у винограда белка относятся к «MYC1b HLH-like DNA binding protein», так как уже было упомянуто, что к этому же суперсемейству относятся и протеины

арабидопсиса. Суммируя эти факты, можно заключить и то, что они должны выполнять схожие функции по интеракции ДНК типа «спираль-петля-спираль» (15).

Заключение

Как мы видим, изученные гены – важный элемент генома винограда и его метаболические пути еще не до конца изучены. Тем не менее, используя гомологичные структуры и модельные растения, мы можем узнать больше. Также использование современных ресурсов позволяет нам изучать структуру, филогенетику и метаболизм испытуемых растений, более того, предсказывать возможные белок-белок и белок-ДНК взаимодействия.

Суммируя вышеизложенное, содержание GC-элементов в изученных генах является стандартным для кодирующих последовательностей, они обладают сходными функциональными cis-элементами и промотерными регионами. Уравнение последовательностей ДНК и протеинов показало, что они обладают схожими фрагментами, из чего было сделано (и подтверждено) предположение, что они включают в себя ультраконсервативные домены. Более того, было выявлено, что эти домены относятся к одному суперсемейству ДНК-связывающих фрагментов.

В связи с наличием ультраконсервативных регионов, филогенетический анализ показал четкое разделение в процессе эволюции генов между собой. При этом, первым отделился от основной ветви VIT_02s0033g00450 и потом VIT_02s0033g00390. Также было выявлено, что предположительно значительные изменения в структуре протеиновых последовательностей, вызванных мутациями (и возможно видообразованием) произошли 32 и 3,5 млн. лет назад.

В конце, в виду плохой изученности метаболических путей, мы провели поиск и сравнение гомологичных последовательностей у арабидопсиса и винограда. Было установлено, что продукты гена MYB114 производят интеракции с тремя другими продуктами генов. Далее, при сравнении с протеомом винограда были выявлены три возможных ортолога этих же генов у винограда. Более того, все изученные последовательности содержали ультраконсервативные домены одного суперсемейства с функцией интеракции ДНК типа «спираль-петля-спираль».

Таким образом, мы установили некоторые особенности генома и протеома винограда. Тем не менее, данные вопросы нуждаются в дальнейшей разработке в виду того, что еще накоплено мало знаний о сортах и видах винограда, а также сиквенсов их генов для полномасштабного сравнения.

Список литературы

1. Altschul S. F. et al. Basic local alignment search tool // *Journal of molecular biology*. – 1990. – Т. 215. – №. 3. – С. 403-410.
2. Aminetzach Y. T., Macpherson J. M., Petrov D. A. Pesticide resistance via transposition-mediated adaptive gene truncation in *Drosophila* // *Science*. – 2005. – Т. 309. – №. 5735. – С. 764-767.
3. Arroyo□García R. et al. Multiple origins of cultivated grapevine (*Vitis vinifera* L. ssp. *sativa*) based on chloroplast DNA polymorphisms // *Molecular ecology*. – 2006. – Т. 15. – №. 12. – С. 3707-3714.
4. Elhaik E. et al. Identifying compositionally homogeneous and non homogeneous domains within the human genome using a novel segmentation algorithm // *Nucleic acids research*. – 2010. – Т. 38. – №. 15. – С. 158-158.
5. Gao F., Zhang C. T. GC-Profile: a web-based tool for visualizing and analyzing the variation of GC content in genomic sequences // *Nucleic acids research*. – 2006. – Т. 34. – №. suppl_2. – С. 686-691.
6. Hao W., Golding G. B. Does gene translocation accelerate the evolution of laterally transferred genes? // *Genetics*. – 2009. – Т. 182. – №. 4. – С. 1365-1375.
7. Jaillon O. et al. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla // *Nature*. – 2007. – Т. 449. – №. 7161. – С. 463-467.
8. Jeong S. T. et al. Expression of VvmybA1 gene and anthocyanin accumulation in various grape organs // *American journal of enology and viticulture*. – 2006. – Т. 57. – №. 4. – С. 507-510.

9. Jin J. et al. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants // *Nucleic acids research*. – 2017. – Т. 45. – №. D1. – С. 1040-1045.
10. Kersey P. J. et al. Ensembl Genomes 2018: an integrated omics infrastructure for non-vertebrate species // *Nucleic Acids Research*. – 2017. – Т. 10. – №. 12. – С. 2809-2816.
11. Kobayashi S. et al. Myb-related genes of the Kyoho grape (*Vitis labruscana*) regulate anthocyanin biosynthesis // *Planta*. – 2002. – Т. 215. – №. 6. – С. 924-933.
12. Kumar S., Stecher G., Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets // *Molecular biology and evolution*. – 2016. – Т. 33. – №. 7. – С. 1870-1874.
13. Marchler-Bauer A. et al. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures // *Nucleic acids research*. – 2016. – Т. 45. – №. D1. – С. 200-203.
14. Mitchell E. M. et al. Use of techniques derived from graph theory to compare secondary structure motifs in proteins // *Journal of Molecular Biology*. – 1990. – Т. 212. – №. 1. – С. 151-166.
15. Murre C. et al. Structure and function of helix-loop-helix proteins // *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression*. – 1994. – Т. 1218. – №. 2. – С. 129-135.
16. Pelsy F. Molecular and cellular mechanisms of diversity within grapevine varieties // *Heredity*. – 2010. – Т. 104. – №. 4. – С. 331-340.
17. Pozzoli U. et al. Both selective and neutral processes drive GC content evolution in the human genome // *BMC evolutionary biology*. – 2008. – Т. 8. – №. 1. – С. 99.
18. Sievers F., Higgins D. G. Clustal omega // *Current protocols in bioinformatics*. – 2014. – С. 3131-31316.
19. Solovyev V. V., Shahmuradov I. A., Salamov A. A. Identification of promoter regions and regulatory sites // *Computational Biology of Transcription Factor Binding*. – 2010. – С. 57-83.
20. Velayudhan B. T. et al. Glycoprotein gene truncation in avian meta pneumovirus subtype C isolates from the United States // *Virus genes*. – 2008. – Т. 37. – №. 2. – С. 266-272.
21. Waese J. et al. ePlant: Visualizing and exploring multiple levels of data for hypothesis generation in plant biology // *The Plant Cell Online*. – 2017. – С. 73-79.
22. Wikström N., Savolainen V., Chase M. W. Evolution of the angiosperms: calibrating the family tree // *Proceedings of the Royal Society of London B: Biological Sciences*. – 2001. – Т. 268. – №. 1482. – С. 2211-2220.
23. Yakushiji H. et al. A skin color mutation of grapevine, from black-skinned Pinot Noir to white-skinned Pinot Blanc, is caused by deletion of the functional VvmybA1 allele // *Bioscience, biotechnology, and biochemistry*. – 2006. – Т. 70. – №. 6. – С. 1506-1508.

References

1. Altschul S. F. et al. Basic local alignment search tool // *Journal of molecular biology*. – 1990. – Т. 215. – №. 3. – С. 403-410.
2. Aminetzach Y. T., Macpherson J. M., Petrov D. A. Pesticide resistance via transposition-mediated adaptive gene truncation in *Drosophila* // *Science*. – 2005. – Т. 309. – №. 5735. – С. 764-767.
3. Arroyo-García R. et al. Multiple origins of cultivated grapevine (*Vitis vinifera* L. ssp. *sativa*) based on chloroplast DNA polymorphisms // *Molecular ecology*. – 2006. – Т. 15. – №. 12. – С. 3707-3714.

4. Elhaik E. et al. Identifying compositionally homogeneous and non homogeneous domains within the human genome using a novel segmentation algorithm // *Nucleic acids research.* – 2010. – Т. 38. – №. 15. – С. 158-158.
5. Gao F., Zhang C. T. GC-Profile: a web-based tool for visualizing and analyzing the variation of GC content in genomic sequences // *Nucleic acids research.* – 2006. – Т. 34. – №. suppl_2. – С. 686-691.
6. Hao W., Golding G. B. Does gene translocation accelerate the evolution of laterally transferred genes? // *Genetics.* – 2009. – Т. 182. – №. 4. – С. 1365-1375.
7. Jaillon O. et al. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla // *Nature.* – 2007. – Т. 449. – №. 7161. – С. 463-467.
8. Jeong S. T. et al. Expression of VvmybA1 gene and anthocyanin accumulation in various grape organs // *American journal of enology and viticulture.* – 2006. – Т. 57. – №. 4. – С. 507-510.
9. Jin J. et al. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants // *Nucleic acids research.* – 2017. – Т. 45. – №. D1. – С. 1040-1045.
10. Kersey P. J. et al. Ensembl Genomes 2018: an integrated omics infrastructure for non-vertebrate species // *Nucleic Acids Research.* – 2017. – Т. 10. – №. 12. – С. 2809-2816.
11. Kobayashi S. et al. Myb-related genes of the Kyoho grape (*Vitis labruscana*) regulate anthocyanin biosynthesis // *Planta.* – 2002. – Т. 215. – №. 6. – С. 924-933.
12. Kumar S., Stecher G., Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets // *Molecular biology and evolution.* – 2016. – Т. 33. – №. 7. – С. 1870-1874.
13. Marchler-Bauer A. et al. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures // *Nucleic acids research.* – 2016. – Т. 45. – №. D1. – С. 200-203.
14. Mitchell E. M. et al. Use of techniques derived from graph theory to compare secondary structure motifs in proteins // *Journal of Molecular Biology.* – 1990. – Т. 212. – №. 1. – С. 151-166.
15. Murre C. et al. Structure and function of helix-loop-helix proteins // *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression.* – 1994. – Т. 1218. – №. 2. – С. 129-135.
16. Pelsy F. Molecular and cellular mechanisms of diversity within grapevine varieties // *Heredity.* – 2010. – Т. 104. – №. 4. – С. 331-340.
17. Pozzoli U. et al. Both selective and neutral processes drive GC content evolution in the human genome // *BMC evolutionary biology.* – 2008. – Т. 8. – №. 1. – С. 99.
18. Sievers F., Higgins D. G. Clustal omega // *Current protocols in bioinformatics.* – 2014. – С. 3131-31316.
19. Solovyev V. V., Shahmuradov I. A., Salamov A. A. Identification of promoter regions and regulatory sites // *Computational Biology of Transcription Factor Binding.* – 2010. – С. 57-83.
20. Velayudhan B. T. et al. Glycoprotein gene truncation in avian meta pneumovirus subtype C isolates from the United States // *Virus genes.* – 2008. – Т. 37. – №. 2. – С. 266-272.
21. Waese J. et al. ePlant: Visualizing and exploring multiple levels of data for hypothesis generation in plant biology // *The Plant Cell Online.* – 2017. – С. 73-79.
22. Wikström N., Savolainen V., Chase M. W. Evolution of the angiosperms: calibrating the family tree // *Proceedings of the Royal Society of London B: Biological Sciences.* – 2001. – Т. 268. – №. 1482. – С. 2211-2220.

23. Yakushiji H. et al. A skin color mutation of grapevine, from black-skinned Pinot Noir to white-skinned Pinot Blanc, is caused by deletion of the functional VvmybA1 allele // Bioscience, biotechnology, and biochemistry. – 2006. – Т. 70. – №. 6. – С. 1506-1508.