

Проект группы №8 “Nymphalidae”



Используемые инструменты

1. BioPython - для хранения и предобработки геномных данных
2. NCBI datasets - для скачивания генома
3. zhunt3 - для поиска Z-DNA (params: 12 8 12)
4. zdna-bert - для поиска Z-DNA (confidence: 0.5, min_len: 10)
5. hmmer



Эпигенетика Nymphalidae

Nymphalidae

Maniola

☐ Maniola jurtina

Nymphalis

☐ Nymphalis io

Melitaea

☐ Melitaea cinxia

Vanessa

☐ Vanessa tameamea

☐ Vanessa cardui

☐ Vanessa atalanta (red admiral)

Danaus

☐ Danaus plexippus

- Эпигенетика данного семейства изучена не очень хорошо. Модельным видом является *Danaus plexippus*.
- Для бабочек этого семейства характерен низкий уровень метилирования
- Наличие различных эпигенетических механизмов в основном связано с миграциями и приспособлением к холодным температурам

Таблица выбранных организмов

Участник	Организм	Геном	Длина генома	Число генов
Щукин Аркадий	<i>Vanessa cardui</i>	GCF_905220365.1	424.8 Mb	15 001
Кузнецов Игорь	<i>Vanessa atalanta</i>	GCF_905147765.1	370.4 Mb	14 188
Яковлев Иван	<i>Melitaea cinxia</i>	GCF_905220565.1	499.4 Mb	17 197
Хрущев Вячеслав	<i>Maniola jurtina</i>	GCF_905333055.1	402 Mb	17 227
Уткина Вика	<i>Nymphalis io</i>	GCF_905147045.1	384.2 Mb	13 823
Северов Владимир	<i>Vanessa tameamea</i>	GCF_037043105.1	363.4 Mb	13 538
Валерий Миронов	<i>Danaus plexippus</i>	GCF_018135715.1	245.2 Mb	14 790

Число квадруплексов

	Exons	Introns	Promoters	Downstream	Intergenic	Bcero
Vanessa cardui	958	1548	116	18	992	3632
Vanessa atalanta	434	722	83	15	1745	4432
Melitaea cinxia	1331	6423	504	141	7405	15804
Maniola jurtina	1290	3562	216	37	1899	7004
Nymphalis io	1267	3348	91	19	2359	4405
Vanessa tameamea	149	1950	95	22	1385	3431
Danaus plexippus	657	807	159	45	1921	5049

Доля квадруплексов

	Exons, %	Introns, %	Promoters, %	Downstream, %	Intergenic, %
Vanessa cardui	26.37	42.62	3.19	0.49	27.31
Vanessa atalanta	9.7	41.6	14.0	2.8	31.8
Melitaea cinxia	9.46	40.46	3.25	1.05	45.78
Maniola jurtina	18.42	50.86	3.08	0.53	27.11
Nymphalis io	19.64	51.90	1.41	0.29	39.36
Vanessa tameamea	4.34	56.83	2.76	0.64	40.3
Danaus plexippus	13.01	15.98	3.15	0.89	38.05

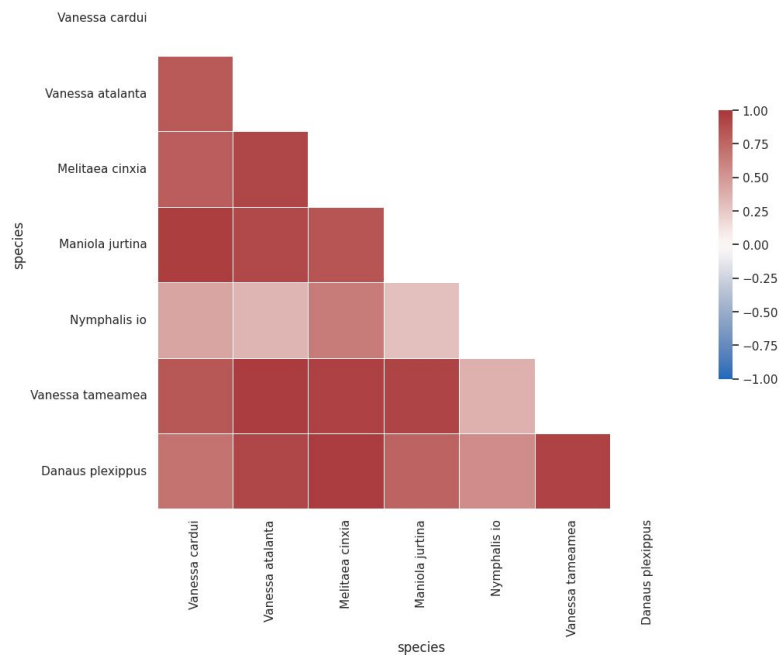
Число найденных Z ДНК (ZDNABERT)

	Exons	Introns	Promoters	Downstream	Intergenic	Bcero
Vanessa cardui	60052	370277	17242	2861	313148	763580
Vanessa atalanta	9269	5154	676	93	4033	18023
Melitaea cinxia	21715	35695	2688	556	30430	91084
Maniola jurtina	27218	42601	2797	430	21979	95025
Nymphalis io	10115	9705	813	121	5572	26326
Vanessa tameamea	6580	29962	1498	295	17231	48910
Danaus plexippus	7141	185	0	0	4315	15647

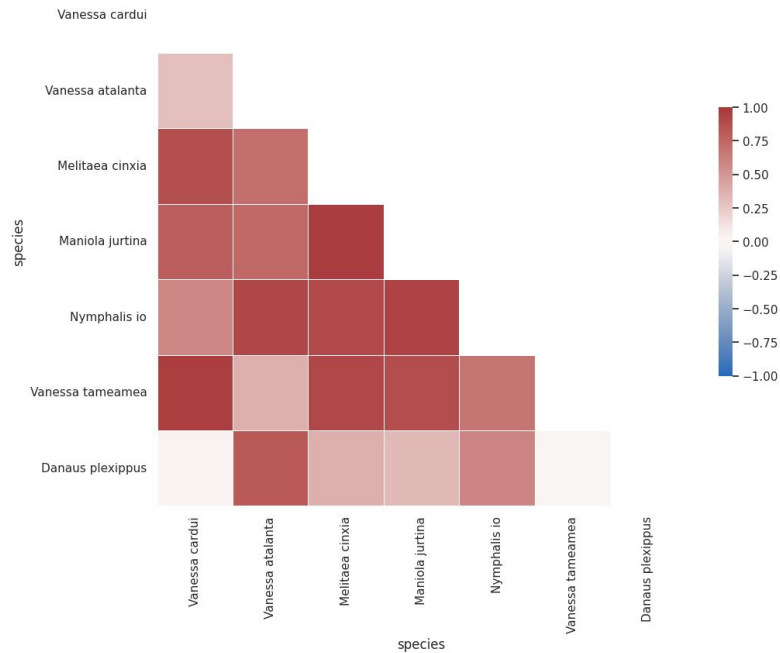
Доля найденных Z ДНК (ZDNABERT)

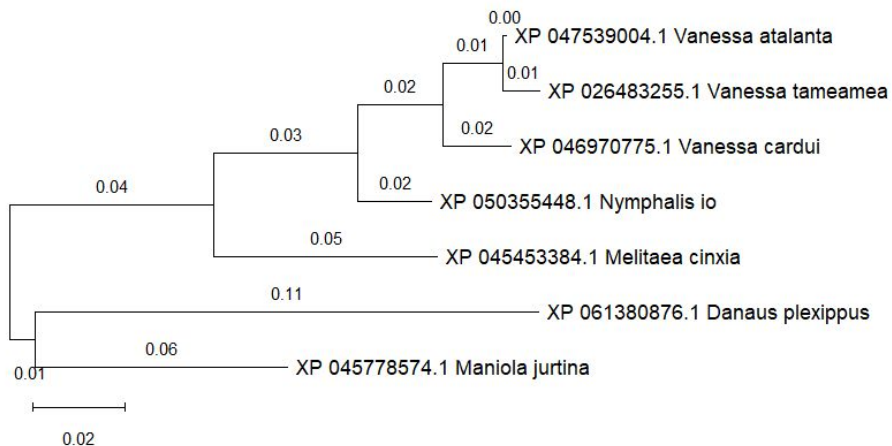
	Exons, %	Introns,%	Promoters, %	Downstream, %	Intergenic, %
Vanessa cardui	7.86	48.49	2.25	0.37	41.01
Vanessa atalanta	52.0	28.6	3.8	0.5	22.4
Melitaea cinxia	25.74	40.85	2.82	0.62	29.96
Maniola jurtina	28.64	44.83	2.95	0.45	23.13
Nymphalis io	41.33	39.65	3.32	0.49	22.77
Vanessa tameamea	13.4	61.25	3.06	0.6	35.23
Danaus plexippus	45.64	1.18	0	0	27.58

Квадруплексы:



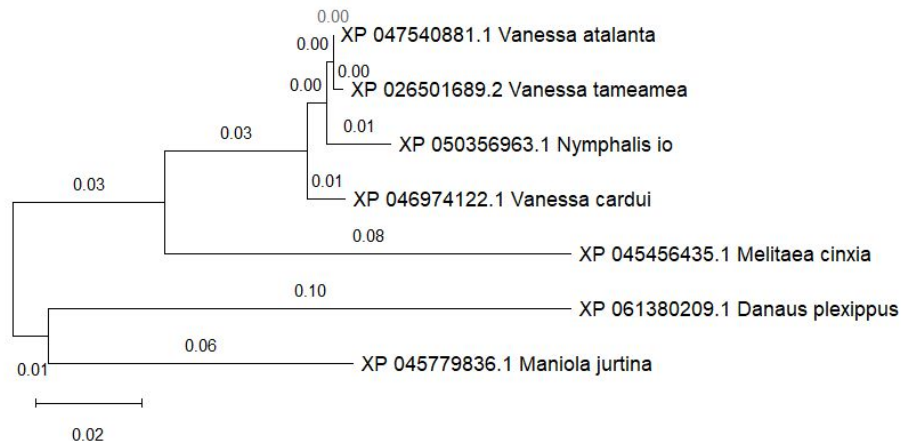
Z-ДНК:

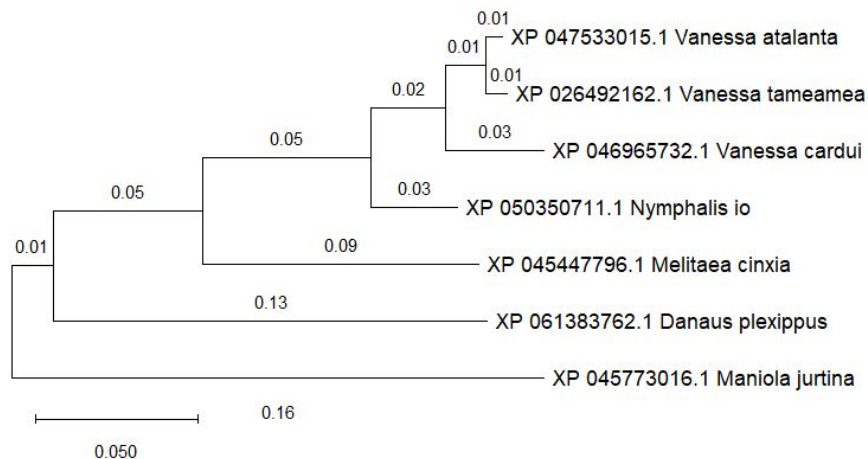




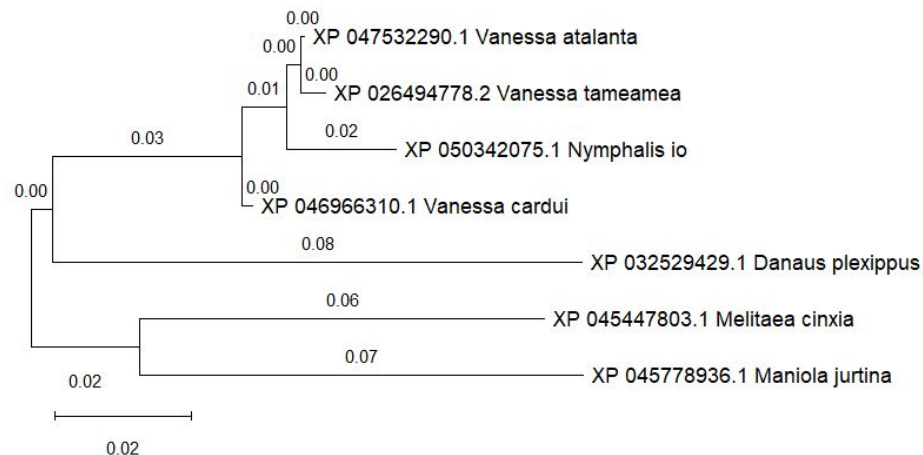
HGNC approved symbol	Pfam domain
GNAT_acetyltr_2	PF13718
This domain has N-acetyltransferase activity. It has a GCN5-related N-acetyltransferase (GNAT) fold	

HGNC approved symbol	Pfam domain
JmjC	PF02373
JmjC-domain proteins may be protein hydroxylases that catalyse a novel histone modification	



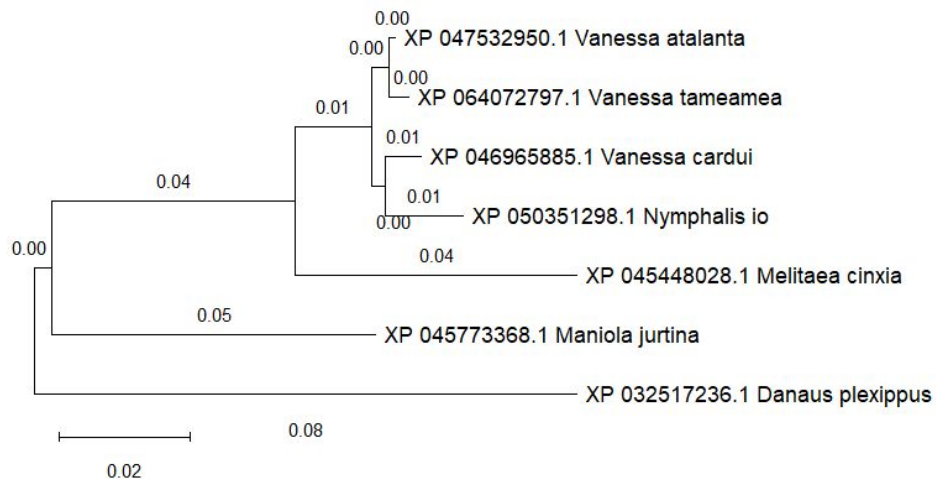


HGNC approved symbol	Pfam domain
DNA_methylase	PF00145
C-5 cytosine-specific DNA methylases	



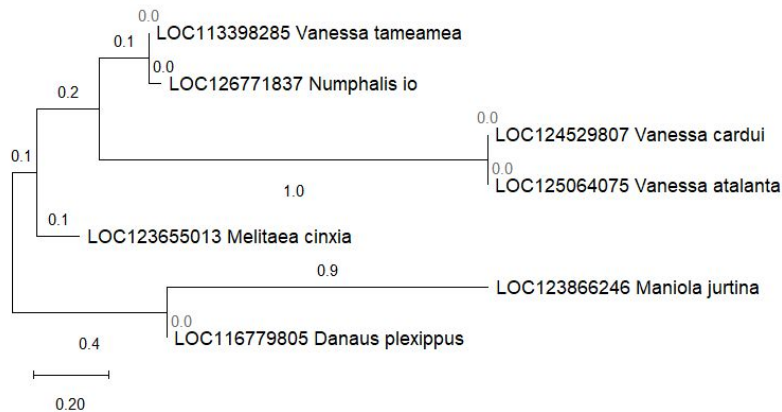
HGNC approved symbol	Pfam domain
DOT1	PF08123
Histone methyltransferase that specifically trimethylates histone H3 to form H3K79me3.	

HGNC approved symbol	Pfam domain
HDAC4_Gln	PF12203
Class II histone deacetylase	

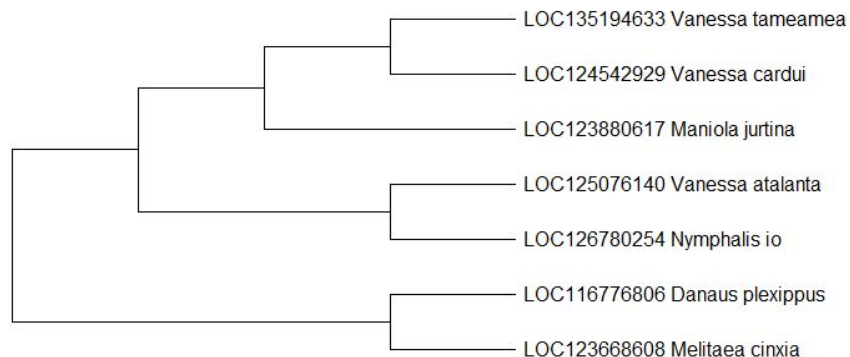


На филогенетических деревьях видно, что белковые последовательности группируются в соответствии с филогенетическим родством видов. Представители рода *Vanessa* (*atalanta*, *tameamea*, *cardui*) формируют тесно связанные кластеры, что свидетельствует о высокой степени консервативности исследуемых доменов у близкородственных видов.

Выравнивания для квадруплексов

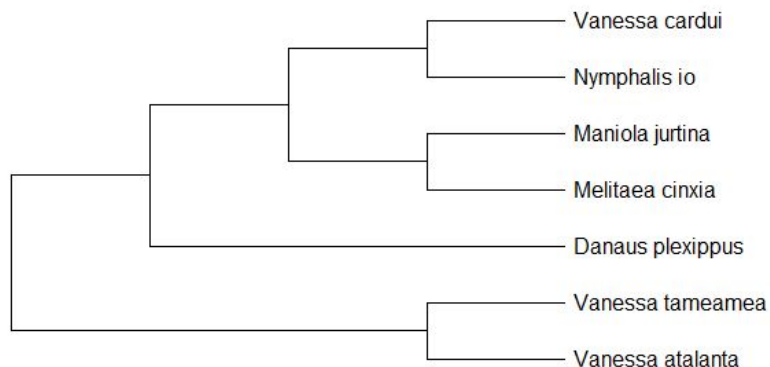
[illegible]

Выравнивания для квадруплексов



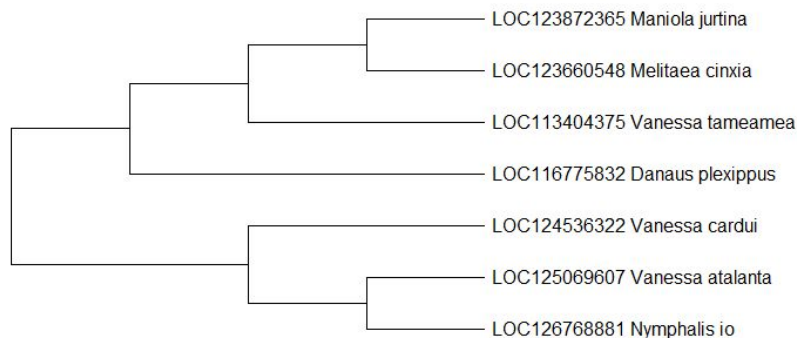
Species/Abbrev			*						*						*					*	*	*	*	*	
1. LOC135194633 <i>Vanessa tameamea</i>	C	C	C	A	T	C	-	C	C	T	C	T	G	C	C	C	T	C	A	C	A	C	C	C	-
2. LOC124542929 <i>Vanessa cardui</i>	C	C	C	C	C	G	-	C	C	C	T	C	G	C	C	C	T	C	G	-	-	C	C	C	-
3. LOC125076140 <i>Vanessa atalanta</i>	-	C	C	C	C	G	-	C	C	C	T	T	A	A	G	C	C	C	G	A	T	C	C	C	-
4. LOC126780254 <i>Nymphalis io</i>	-	C	C	C	C	G	-	C	C	C	T	T	A	A	G	C	C	C	G	A	T	C	C	C	-
5. LOC123880617 <i>Maniola jurtina</i>	-	-	C	C	C	G	-	A	C	C	C	G	G	C	A	C	C	C	G	-	-	C	C	C	-
6. LOC116776806 <i>Danaus plexippus</i>	C	C	C	T	T	A	C	C	C	C	T	C	C	C	A	C	C	T	A	A	G	C	C	C	-
7. LOC123668608 <i>Melitaea cinxia</i>	-	C	C	C	G	G	C	C	C	G	A	C	C	C	G	C	C	C	G	A	A	C	C	C	C

Выравнивания для квадруплексов



Species/Abbrv		*	*	*													*		*						*							
1. Vanessa tameamea	-	-	G	G	G	C	G	C	G	C	G	C	G	G	G	G	C	G	C	G	-	G	G	G	G							
2. Vanessa cardui	G	G	G	G	G	C	G	-	-	-	-	G	G	C	C	G	A	G	G	G	A	C	G	G	-	-	-	-				
3. Vanessa atalanta	-	G	G	G	A	C	G	-	-	-	-	G	G	T	A	T	G	G	G	A	C	T	G	C	C	C	G	G	G	G		
4. Nymphalis io	-	-	G	G	G	C	G	C	C	G	G	G	G	C	T	G	G	C	G	G	G	G	G	C	G	A	G	-	G	G	G	G
5. Maniola jurtina	-	-	G	G	G	C	G	-	-	-	-	G	A	G	G	G	C	G	G	A	G	G	C	G	G	G	-	-	-	-	-	
6. Danaus plexippus	-	-	G	G	G	C	G	G	C	G	G	G	G	G	A	C	T	G	G	C	G	G	G	G	G	C	G	C	G	G	G	G
7. Melitaea cinxia	-	G	G	G	G	C	A	-	-	-	-	-	-	-	A	T	G	G	G	G	G	G	T	G	G	C	C	G	G	G	G	

Выравнивания для квадруплексов



DNA Sequences	Translated Protein Sequences																							
Species/Abbrv	*	*	*															*						*
1. LOC113404375 <i>Vanessa tameamea</i>	C	C	C	C	A	C	C	C	C	A	C	G	C	G	C	C	C	T	C	C	C	A	C	C
2. LOC124536322 <i>Vanessa cardui</i>	C	C	C	C	T	C	T	A	C	C	C	T	C	-	C	G	A	C	C	C	T	C	T	C
3. LOC125069607 <i>Vanessa atalanta</i>	C	C	C	C	T	G	C	C	T	C	C	C	T	G	G	G	C	C	T	G	T	C	C	C
4. LOC126768881 <i>Nymphalis io</i>	C	C	C	C	T	T	T	A	C	C	C	C	-	-	-	-	-	T	C	C	T	C	T	C
5. LOC123872365 <i>Maniola jurtina</i>	C	C	C	T	-	-	-	-	-	-	-	-	-	-	-	-	G	C	C	C	A	A	C	C
6. LOC116775832 <i>Danaus plexippus</i>	C	C	C	C	C	C	C	A	C	A	C	C	C	-	-	-	C	T	T	A	A	G	C	C
7. LOC123660548 <i>Melitaea cinxia</i>	C	C	C	T	C	G	C	C	C	G	G	G	C	-	C	G	C	C	C	A	C	A	C	C

Что по квадруплексам?

В силу вычислительной невозможности. Произвести анализ генов-ортологов оказалось затруднительно. Поэтому квадруплексы были выбраны методом пристального взгляда, основываясь на схожести расположения в геномах (в частности, в 1-ой и в z хромосомах, на примерно одинаковых позициях). Так что на предыдущих слайдах квадруплексы предположительно в ортологичных генах. Выпетливания и повороты можно зафиксировать на выравниваниях.

*Все выравнивания были выполнены с помощью алгоритма Muscle, а деревья - Maximum likelihood

Выводы

- 1) Таблички и деревья в целом получились хорошие. Родство всех 7 бабочек не подвергается сомнению.
- 2) Геномы бабочек изучены достаточно плохо, поэтому нет удобных баз ортологичных генов
- 3) Геномы бабочек слишком велики (300-500Mb) для комфортного анализа, 12 часов непрерывной работы ни одна среда (и нервная система) не выдержит. А ZDNABERT вообще страшно запускать.