

Страменопилы

(финальная презентация)

	Название	Идентификатор	Длина генома	Число генов
Екатерина Гришина	<i>Aureococcus anophagefferens</i>	GCF_000186865.1	56.7 Mb	11522
Полина Шайдурова	<i>Phytophthora sojae</i>	GCF_000149755.1	82.6 Mb	28142
Илья Герман	<i>Blastocystis hominis</i>	GCF_000151665.1	18.8 Mb	6020
Ян Аникиев	<i>Nannochloropsis gaditana</i>	GCF_000240725.1	34 Mb	3465
Михаил Сизов	<i>Aphanomyces invadans</i>	GCF_000520115.1	71.4 Mb	15416
Ульяна Ключникова	<i>Aphanomyces astaci</i>	GCF_000520075.1	75.8 Mb	19 584
Георгий Караваев	<i>Thalassiosira pseudonana</i>	GCF_000149405.2	32.4 Mb	11771
Денис Михайлов	<i>Phytophthora nicotianae</i>	GCF_000247585.1	82.4 Mb	23240
Егор Попов	<i>Phaeodactylum tricornutum</i>	GCF_000150955.2	27.5 Mb	10398

Число G квадруплексов

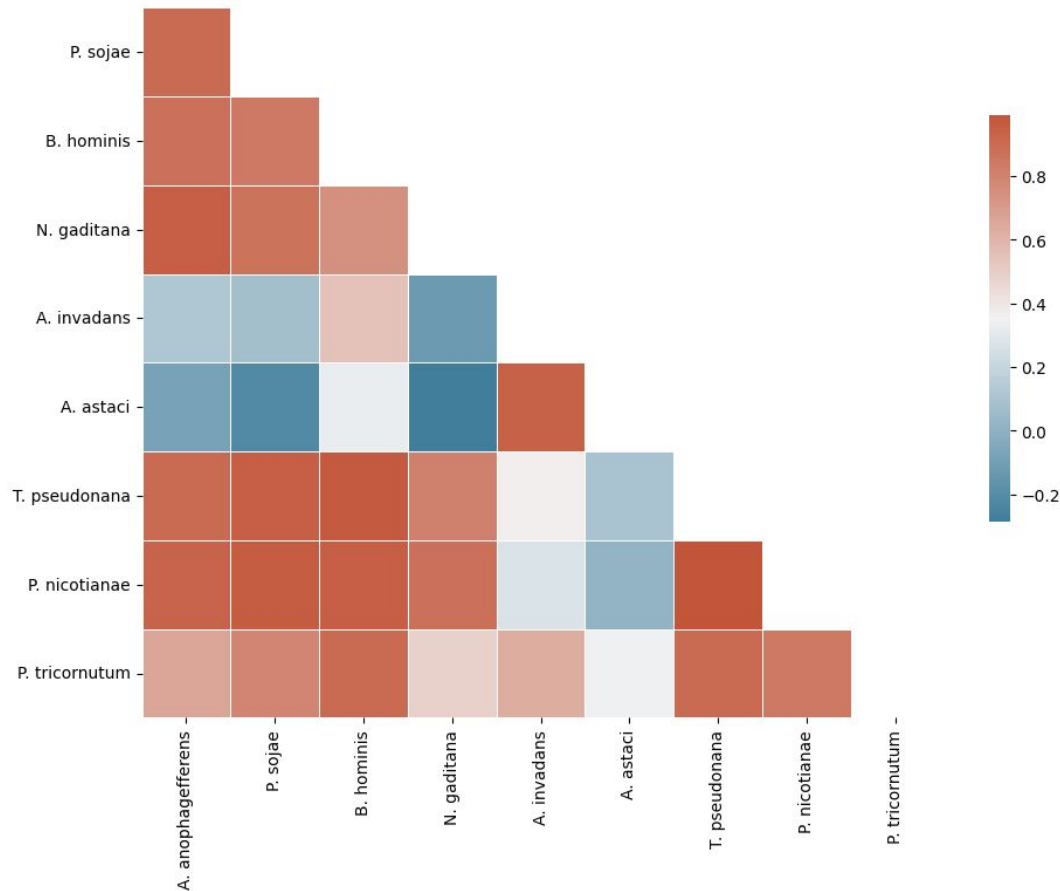
	Всего	Экзоны	Интроны	Промотеры	Downstream	Межгенники
A. anophagefferens	415	101	73	105	33	256
P. sojae	2426	504	256	862	100	704
B. hominis	105	42	8	34	9	59
N. gaditana	18495	1908	1550	1756	573	15349
A. invadans	33	31	1	14	3	8
A. astaci	170	127	11	22	2	8
T. pseudonana	1197	393	60	474	131	769
P. nicotianae	41	12	1	14	5	28
P. tricornutum	45	21	0	25	4	25

Доля G квадруплексов

	Экзоны	Интроны	Промотеры	Downstream	Межгенники
A. anophagefferens	0.24	0.18	0.25	0.08	0.62
P. sojae	0.21	0.11	0.35	0.04	0.29
B. hominis	0.4	0.08	0.32	0.09	0.56
N. gaditana	0.1	0.08	0.09	0.03	0.83
A. invadans	0.94	0.03	0.42	0.01	0.24
A. astaci	0.75	0.06	0.13	0.02	0.05
T. pseudonana	0.33	0.05	0.39	0.11	0.64
P. nicotianae	0.29	0.02	0.34	0.12	0.68
P. tricornutum	0.47	0.0	0.56	0.09	0.56

G-Quadruplex fractions

A. anophagefferens -



Видим, что почти все организмы коррелируют по долям квадруплексов, кроме A. invadans и A. astaci, представителей одного рода

Число Z-ДНК (ZDNABERT)

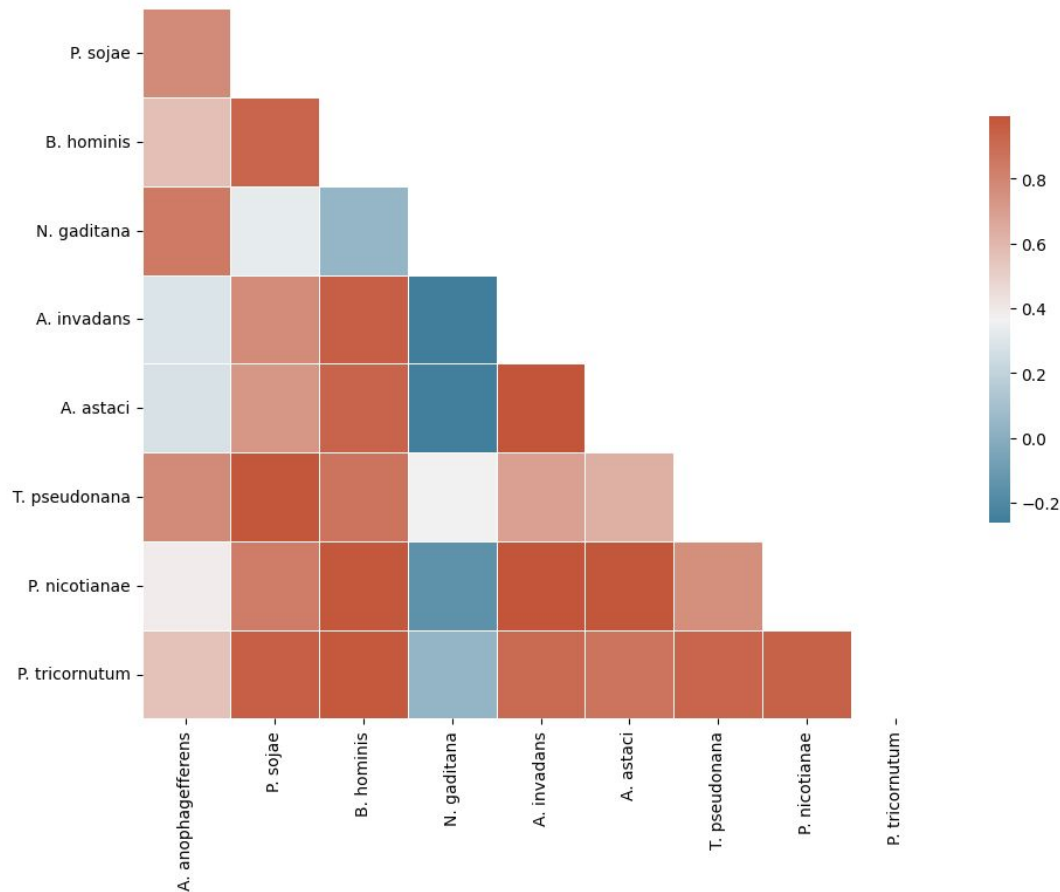
	Всего	Экзоны	Интроны	Промотеры	Downstream	Межгенники
A. anophagefferens	19737	7074	1500	4006	777	11408
P. sojae	104869	57643	7218	18127	2138	21743
B. hominis	1178	836	49	336	48	328
N. gaditana	30287	4262	970	2775	599	25249
A. invadans	2717	3538	303	989	174	351
A. astaci	5661	4306	189	846	109	211
T. pseudonana	3132	1618	197	1306	281	1362
P. nicotianae	2084	2273	77	792	144	467
P. tricornutum	1113	741	46	471	105	340

Доля Z-ДНК (ZDNABERT)

	Экзоны	Интроны	Промотеры	Downstream	Межгенники
A. anophagefferens	0.36	0.08	0.20	0.04	0.58
P. sojae	0.53	0.07	0.17	0.02	0.21
B. hominis	0.71	0.04	0.29	0.04	0.28
N. gaditana	0.14	0.03	0.09	0.02	0.83
A. invadans	1.3	0.11	0.36	0.06	0.13
A. astaci	0.76	0.03	0.15	0.02	0.04
T. pseudonana	0.51	0.06	0.42	0.09	0.43
P. nicotianae	1.09	0.04	0.38	0.07	0.22
P. tricornutum	0.67	0.04	0.42	0.09	0.31

ZDNABERT fractions

A. anophagefferens -



Здесь хорошо коррелируют все, кроме A. anophagefferens и N. gaditana, оомицет и водоросль (живущие в воде)

Параметры

- Выравнивая строились при помощи **ClustalW** со стандартными параметрами
- Деревья строились при помощи **Minimum Evolution Tree** со стандартными параметрами
- Параметры для ZHUNT были взяты с прошлого года, то есть запуск команды выглядел так: **!./zhunt3 12 8 12 genomic.fna** и порог был равен **300**
- Параметры для ZDNABERT были стандартными, то есть:

model = 'HG kouzine'
model_confidence_threshold = 0.5
minimum_sequence_length = 10

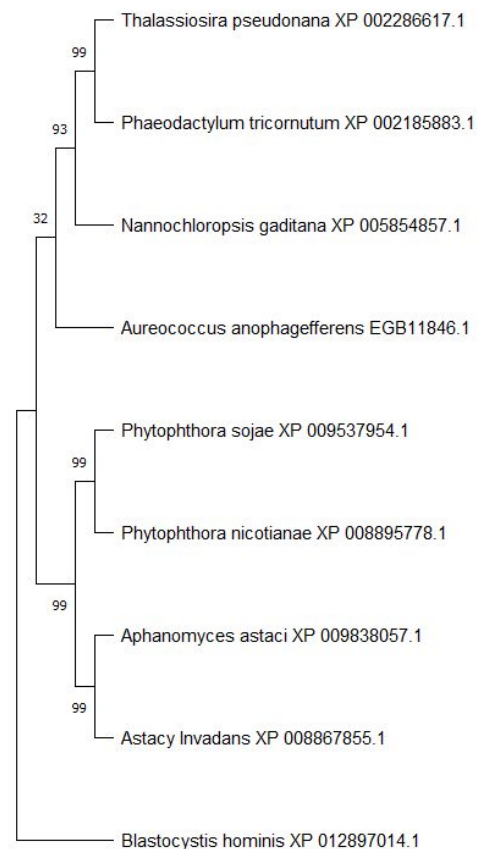
Pfam Domain	HGNC approved symbol	Function
PF00004	ATAD2	Chromatin remodeling
PF00012	HSPA1A	Histone modification write cofactor
PF00022	ACTB	Chromatin remodeling, Histone modification read
PF00063	MYO1C	Chromatin remodeling cofactor
PF00069	AURKA	Histone modification write

1. Aureococcus anophagefferens EGB11846.1	D R D L A Q Y S F A T T L H Q Q R - - - S Q I G N - - - T S F A F P N - - - - - A A P G A A P A A A G G G F A A A A D D D E E D L Y S *
2. Thalassiosira pseudonana XP_002286617.1	R D L L A Q Y A S F A Q T L Q Q S R - - - A A V T G A A G G S L A T F A F P - - - - - Q G G G M A A A G G D A A A E D D E E D L Y S *
3. Phytophthora sojae KP_009537954.1	R D L L A Q Y S T F A Q T L Q Q A R - - - S H V T A - G G T S L A N F S F P - - - - - G R N V N A G A A G G A T D A A A D E D E E D L Y S *
4. Blastocystis hominis XP_012897014.1	D A D L L K Y E S F S Q K M K Q Q R G N M G S G V A N F S F G D A Q N I N A P A D D L D G L Y D D V P A R N P V H P P N N R A A T S Q G Q T Q P Q P P A P E S G R G D E D F E O D L Y *
5. Aphanomyces astaci XP_009838057.1	R D L L H Q Y S T F A Q T L Q Q S R - - - A Q V G A - S G S S L S S F S F P - - - - - S R N V V G G - - G G G T T A A V D E E E E D L Y S *
6. Astacy Invadans XP_008867855.1	R D L L H Q Y S T F A Q T L Q Q S R - - - A Q V G A - T G T S L S T S F F P - - - - - S R N V V G G - - C A A T A A V D E E E E D L Y S *
7. Phaeodactylum tricornutum XP_002185883.1	R D L L A Q Y A S F A Q T L Q Q S R - - - A A V S G S T G G S L A T F A F P - - - - - D A N A A V G V G A - - A A E D D D D E E D L Y S *
8. Phytophthora nicotianae XP_008895778.1	R D L L H Q Y S T F A Q T L Q Q A R - - - S H V T A - G G T S L A N F S F P - - - - - G R N V N A V P G G A A D A A A D E D E E D L Y S *
9. Nanmochloropsis gaditana XP_005854857.1	R D L L Q Q Y S S F A Q T L Q Q A R - - - S Q I G - P G G S L A A F R F P - - - - - E G G E A A A G G A G - - A A A E E E D I - - D L Y S

PF00004

Видно, что последовательности очень похожи, мы действительно нашли ортологов. На дереве организмы из одного семейства *Phytophthora* и *Arphanomyses* группируются в одну кладу. Также 4 водоросли группируются в одну кладу.

*У *invadans* перепутано название, должно быть *Aphanomyces invadans*.

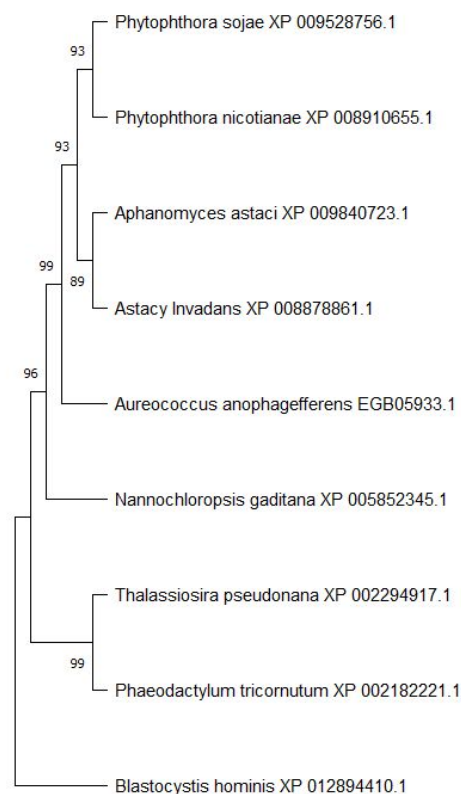


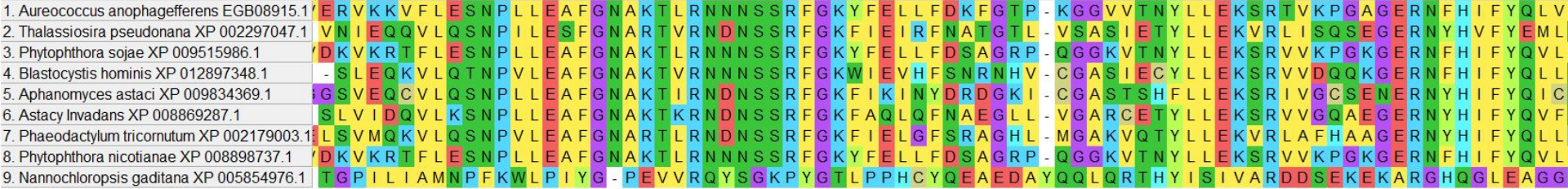
[illegible]

1. <i>Aureococcus anophagefferens</i> EGB05933.1	C	D	V	D	I	R	K	D	L	Y	S	N	I	V	L	S	G	G	T	T	M	Y	P	G	I	G	E	R	M	T	K	E	L	T	A	L	A	P	S	T	M	K	I	K	V	V	A	P	P	E	R	K	Y	S	V	W	I	G	G	S	I	L	S	S	L	S	T	F	Q	S	M	W	I	S	K
2. <i>Thalassiosira pseudonana</i> XP_002294917.1	C	D	V	D	I	R	K	D	L	Y	A	N	I	V	L	S	G	G	T	T	M	F	P	G	I	S	E	R	M	S	K	E	I	T	A	L	A	P	A	S	I	K	V	K	I	V	A	P	P	E	R	K	Y	S	V	W	I	G	G	S	I	L	A	S	L	S	T	F	Q	S	M	W	I	S	K
3. <i>Phytophthora sojae</i> XP_009528756.1	C	D	V	D	I	R	K	D	L	Y	C	N	I	V	L	S	G	G	T	T	M	Y	P	G	I	G	E	R	M	T	K	E	L	T	A	L	A	P	S	T	M	K	I	K	V	V	A	P	P	E	R	K	Y	S	V	W	I	G	G	S	I	L	S	S	L	S	T	F	Q	Q	M	W	I	S	K
4. <i>Blastocystis hominis</i> XP_012894410.1	C	D	V	D	L	R	R	D	L	Y	N	N	V	V	L	S	G	G	T	T	F	F	D	G	M	G	D	R	M	Q	K	E	L	T	A	L	V	P	S	S	I	K	V	R	V	I	A	S	P	D	R	K	Y	M	V	W	I	G	G	G	M	L	A	Q	L	S	S	F	Q	D	C	W	I	T	K
5. <i>Aphanomyces astaci</i> XP_009840723.1	C	D	V	D	I	R	K	D	L	Y	C	N	I	V	L	S	G	G	T	T	M	Y	P	G	I	S	E	R	M	T	K	E	L	T	A	L	A	P	S	T	M	K	I	K	V	V	A	P	P	E	R	K	Y	S	V	W	I	G	G	S	I	L	S	S	L	S	T	F	Q	Q	M	W	I	S	K
6. <i>Astacy Invadans</i> XP_008878861.1	C	D	V	D	I	R	K	D	L	Y	C	N	I	V	L	S	G	G	T	T	M	Y	P	G	I	S	E	R	M	T	K	E	L	T	A	L	A	P	S	T	M	K	I	K	V	V	A	P	P	E	R	K	Y	S	V	W	I	G	G	S	I	L	S	S	L	S	T	F	Q	Q	M	W	I	S	K
7. <i>Phaeodactylum tricornutum</i> XP_002182221.1	C	D	V	D	I	R	K	D	L	Y	A	N	I	V	L	S	G	G	T	T	M	F	P	G	I	S	E	R	M	S	K	E	I	T	A	L	A	P	A	S	I	K	V	K	I	V	A	P	P	E	R	K	Y	S	V	W	I	G	G	S	I	L	A	S	L	S	T	F	Q	S	M	W	I	S	K
8. <i>Phytophthora nicotianae</i> XP_008910655.1	C	D	V	D	I	R	K	D	L	Y	C	N	I	V	L	S	G	G	T	T	M	Y	P	G	I	G	E	R	M	T	K	E	L	T	A	L	A	P	S	T	M	K	I	K	V	V	A	P	P	E	R	K	Y	S	V	W	I	G	G	S	I	L	S	S	L	S	T	F	Q	Q	M	W	I	S	K
9. <i>Nannochloropsis gaditana</i> XP_005852345.1	C	D	V	D	I	R	K	D	L	Y	S	N	I	V	L	S	G	G	T	T	M	F	P	G	I	G	E	R	M	T	K	E	I	S	A	L	A	P	P	T	M	K	V	K	V	V	A	P	P	E	R	K	Y	S	V	W	I	G	G	S	I	L	S	S	L	S	T	F	Q	S	M	W	I	S	K

PF00022

Здесь почти полное совпадение последовательностей. Заметно отличается только паразит *B. hominis*. На дереве организмы из одних семейств попадают в одну кладу.

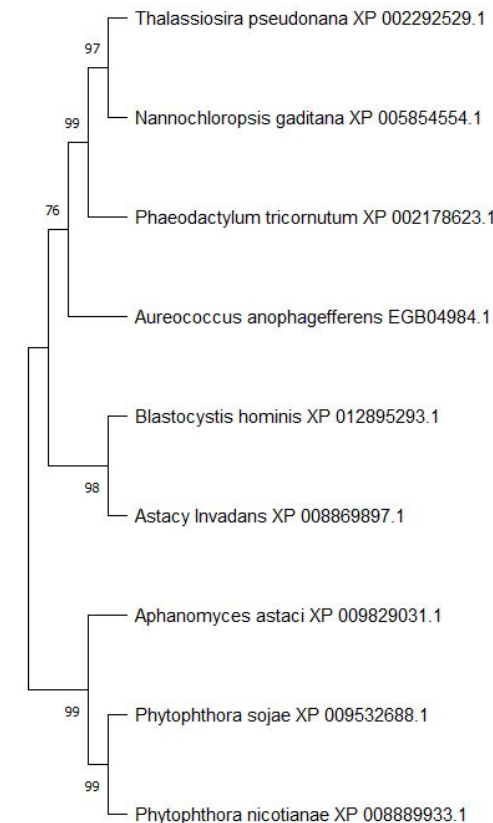




1. Aureococcus anophagefferens EGB04984.1	F	Y	M	V	I	E	L	M	E	G	G	E	L	F	E	R	I	V	K	K	T	F	Y	N	E	K	E	A	R	D	L	I	R	I	L	L	D	A	L	A	Y	L	H	H	R	H	I	V	H	R	D	L	K	P	E	N	L	L	L	K	S	P	Y	N	D	F	I	K	L	A	D	F	G	F	A	K	K	V	
2. Thalassiosira pseudonana XP 002292529.1	F	Y	L	V	T	E	Q	M	R	G	G	E	L	F	D	R	I	V	S	K	S	Y	N	E	K	E	A	R	D	V	C	K	I	L	F	E	S	I	G	F	C	H	S	K	S	V	A	H	R	D	L	K	P	E	N	L	L	L	R	A	E	D	N	D	S	D	I	K	I	A	D	F	G	F	A	K	K	V	
3. Phytophthora sojae XP 009532688.1	L	Y	M	M	M	E	L	A	D	T	-	D	L	H	R	L	I	Q	S	S	C	P	L	T	E	G	H	I	R	V	I	M	Y	Q	V	L	S	G	V	K	A	M	H	D	N	G	V	L	H	R	D	L	K	P	G	N	L	L	L	N	K	D	-	-	-	C	E	L	K	I	T	D	F	G	L	A	R	M	M
4. Blastocystis hominis XP 012895293.1	F	F	L	V	Q	D	L	M	L	G	G	E	L	F	H	A	I	I	N	A	H	S	F	S	E	A	Q	A	Q	R	C	V	M	T	L	L	Y	T	L	H	Y	C	H	Q	R	G	I	V	H	R	D	L	K	P	E	N	L	L	L	A	E	Q	N	N	L	D	S	I	R	I	A	D	F	G	L	A	N	E	I
5. Aphanomyces astaci XP 009829031.1	L	Y	M	M	M	E	L	A	D	T	-	D	L	H	R	L	I	Q	S	S	C	P	L	H	E	G	H	I	R	V	I	M	Y	Q	V	L	C	G	V	K	T	M	H	D	N	G	V	L	H	R	D	L	K	P	G	N	L	L	V	N	K	D	-	-	-	C	E	L	K	I	T	D	F	G	L	A	R	M	V
6. Astacy Invadans XP 008869897.1	D	Y	V	V	C	D	L	E	E	G	N	D	L	F	D	V	L	V	K	K	G	R	F	S	E	T	E	A	K	A	L	M	K	V	L	L	E	T	V	G	A	C	H	A	L	G	M	V	H	R	N	L	K	P	E	S	V	R	M	P	S	E	-	D	P	A	S	L	M	L	S	N	F	S	L	C	S	S	I
7. Phaeodactylum tricornutum XP 002178623.1	F	Y	L	V	M	E	R	M	R	G	G	E	L	F	D	R	I	V	A	K	A	Y	N	E	K	E	A	R	D	T	C	K	I	V	L	E	A	V	G	Y	C	H	Q	N	H	V	A	H	R	D	L	K	P	E	N	L	L	L	L	S	E	H	D	D	S	A	V	K	I	A	D	F	G	F	A	K	K	V	
8. Phytophthora nicotianae XP 008889933.1	L	Y	M	M	M	E	L	A	D	T	-	D	L	H	R	L	I	Q	S	S	C	P	L	T	E	G	H	I	R	V	I	M	Y	Q	V	L	S	G	V	K	A	M	H	D	N	G	V	L	H	R	D	L	K	P	G	N	L	L	L	N	K	D	-	-	-	C	E	L	K	I	T	D	F	G	L	A	R	M	M
9. Nannochloropsis gaditana XP 005854554.1	F	Y	L	V	L	E	I	C	L	G	G	E	L	F	D	R	I	V	E	K	K	S	Y	N	E	K	E	A	R	D	L	V	R	I	L	F	D	T	L	R	Y	C	H	S	T	K	V	V	H	R	D	L	K	P	E	N	L	L	L	E	D	E	R	D	D	A	N	I	K	L	A	D	F	G	F	A	Q	R	M

PF00069

В этом выравнивании есть совпадающие участки. Фитофторы снова попали в одну кладу на дереве. Сверху на дереве 4 водоросли попали в одну кладу, снизу 5 паразитных организмов попали ближе друг к другу. Странно, что Arphanomyses далеко друг от друга на дереве.



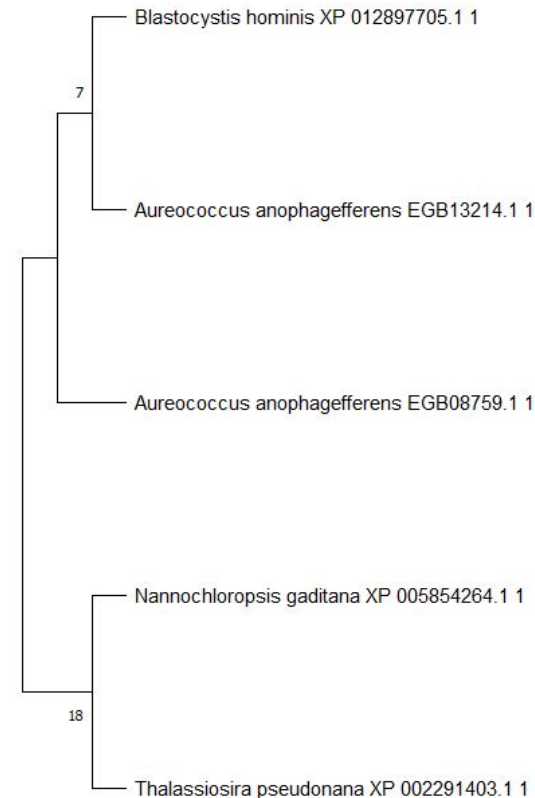
Выводы

- У генов-ортологов есть протяженные области генов с совпадающими участками. Можно сделать вывод, что эти участки консервативные и имеют какую-то функциональную роль.
- Организмы из 1 семейства часто попадают в одну кладу - их последовательности очень схожи.
- Похожими друг на друга оказываются белки водорослей и белки паразитов.

1. Blastocystis hominis XP 012897705.1 1	-	G	G	G	T	A	G	A	T	G	G	G	A	T	T	T	G	G	G	G	A	C	C	C	T	G	G	G	G	-	-	-	-	
2. Nannochloropsis gaditana XP 005854264.1 1	-	-	-	-	-	G	G	G	G	A	A	G	T	T	C	C	G	G	G	G	A	G	G	A	G	G	G	A	A	A	A	G	G	G
3. Aureococcus anophagefferens EGB08759.1 1	G	G	G	T	G	C	G	G	G	G	G	G	G	G	G	C	G	G	C	G	G	G	C	G	C	G	G	G	-	-	-	-	-	
4. Aureococcus anophagefferens EGB13214.1 1	G	G	G	G	C	C	G	G	G	G	A	T	C	G	A	C	G	G	G	G	C	C	G	C	C	T	C	G	G	G	G	-	-	-
5. Thalassiosira pseudonana XP 002291403.1 1	-	-	-	-	-	-	G	G	G	A	G	G	G	T	G	T	C	T	G	G	G	A	G	G	G	G	G	G	-	-	-	-	-	-

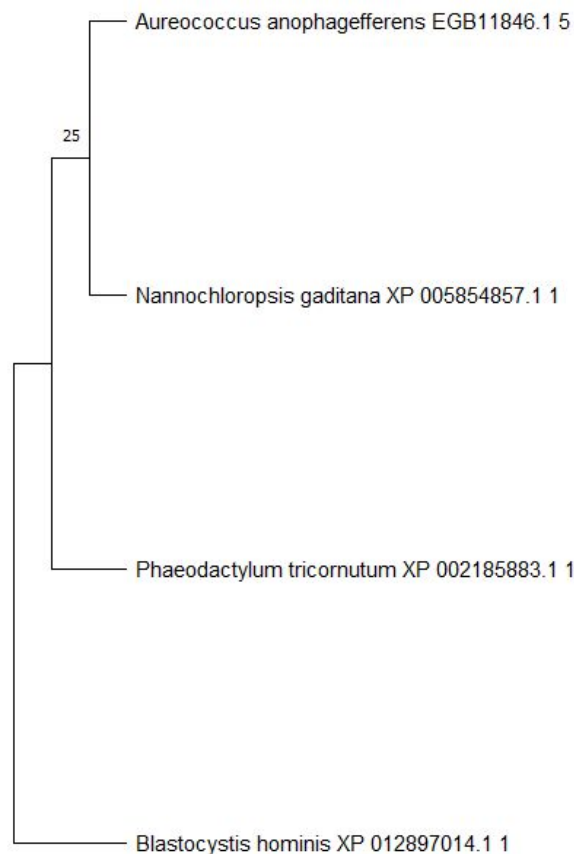
G-квадруплексы

Квадруплексов нашлось мало: в каждом промотере всего 1 квадруплекс у одного организма. На выравнивании эти квадруплексы заметно отличаются. Возможно, квадруплексы в промотерах у наших организмов не консервативны, потому что не имеют важных функций.

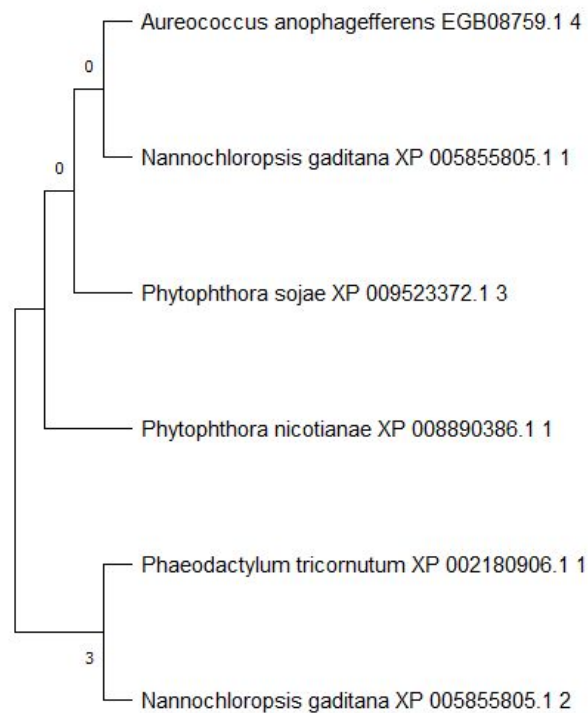


1. <i>Aureococcus anophagefferens</i> EGB11846.1 5	-	-	-	-	-	C	C	G	C	G	T	G	T	A	C	G	G	C	C	-
2. <i>Blastocystis hominis</i> XP 012897014.1 1	G	A	G	A	C	G	T	G	C	G	C	G	T	G	C	G	G	A	G	-
3. <i>Phaeodactylum tricornutum</i> XP 002185883.1 1	-	-	-	-	-	T	C	A	C	A	C	G	C	A	C	A	C	A	C	A
4. <i>Nannochloropsis gaditana</i> XP 005854857.1 1	-	-	-	-	G	C	C	G	T	G	C	G	C	A	A	G	-	-	-	-

Z-DNA

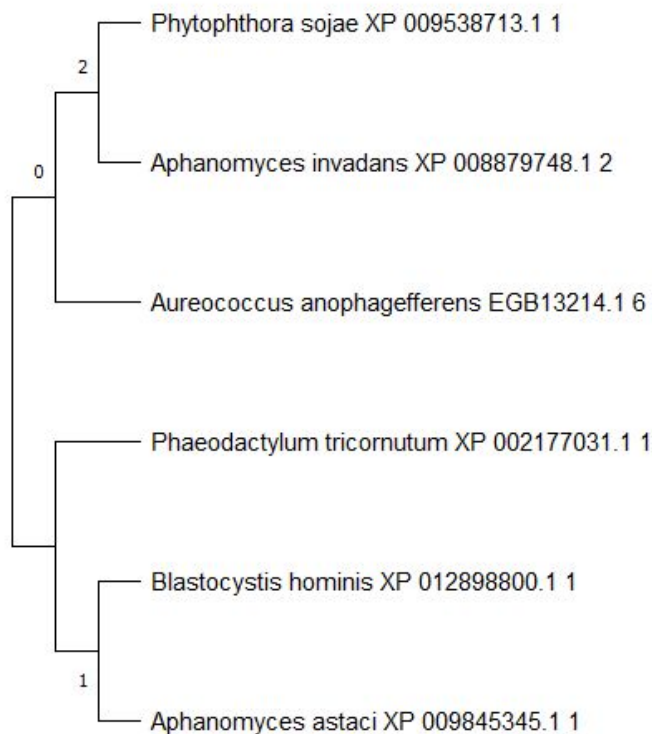


1. <i>Aureococcus anophagefferens</i> EGB08759.1 4	-	-	-	C	G	T	C	G	A	C	G	C	G	C	G	C	C	C	T	C	G	C	G
2. <i>Phytophthora sojae</i> XP 009523372.1 3	-	-	-	-	T	C	C	A	G	C	G	C	G	C	G	C	C	C	G	T	T	-	-
3. <i>Phaeodactylum tricornutum</i> XP 002180906.1 1	G	G	C	T	G	C	G	C	G	T	G	T	G	T	G	T	G	T	G	T	-	-	-
4. <i>Phytophthora nicotianae</i> XP 008890386.1 1	-	-	-	-	-	C	T	A	G	T	G	C	G	C	G	C	C	C	A	T	-	-	-
5. <i>Nannochloropsis gaditana</i> XP 005855805.1 1	-	-	-	G	A	C	G	G	G	T	G	C	A	C	G	C	C	C	T	T	-	-	-
6. <i>Nannochloropsis gaditana</i> XP 005855805.1 2	-	-	-	-	-	C	T	T	G	T	G	T	G	C	G	T	G	T	G	-	-	-	-



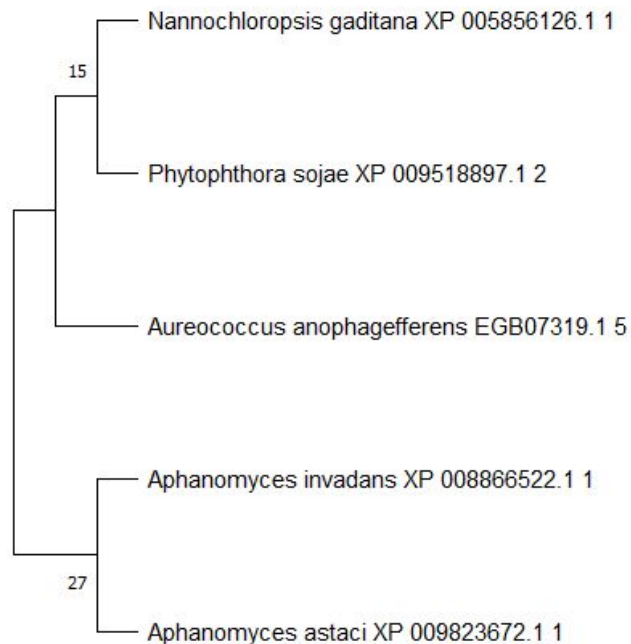
Z-DNA

1. <i>Aureococcus anophagefferens</i> EGB13214.1 6	-	-	T	C	G	C	C	G	C	G	C	G	C	G	T	T	C	A	G
2. <i>Phaeodactylum tricornutum</i> XP 002177031.1 1	-	-	-	G	G	A	C	G	C	A	C	G	T	G	T	-	-	-	-
3. <i>Phytophthora sojae</i> XP 009538713.1 1	-	-	T	T	T	C	C	A	C	G	C	G	C	G	G	C	-	-	-
4. <i>Blastocystis hominis</i> XP 012898800.1 1	T	G	C	C	G	G	T	G	C	G	C	G	G	G	T	T	-	-	-
5. <i>Aphanomyces invadans</i> XP 008879748.1 2	-	-	T	T	C	T	C	G	C	G	C	G	C	A	C	T	C	G	-
6. <i>Aphanomyces astaci</i> XP 009845345.1 1	-	-	-	C	A	G	T	G	C	G	T	G	C	A	T	G	A	-	-



Z-DNA

1. <i>Aureococcus anophagefferens</i> EGB07319.1 5	-	C	C	G	G	C	C	C	G	C	G	C	G	T	A	C	G	T	G	T	T	T
2. <i>Aphanomyces invadans</i> XP 008866522.1 1	-	-	C	A	G	C	A	C	A	C	G	T	G	C	A	T	G	A	A	-	-	-
3. <i>Aphanomyces astaci</i> XP 009823672.1 1	C	A	C	G	G	C	G	C	G	C	G	C	C	A	G	-	-	-	-	-	-	-
4. <i>Nannochloropsis gaditana</i> XP 005856126.1 1	C	T	C	T	G	T	G	C	G	C	G	T	G	C	T	C	G	-	-	-	-	-
5. <i>Phytophthora sojae</i> XP 009518897.1 2	-	-	A	G	A	C	G	C	G	T	G	C	A	C	C	T	G	T	-	-	-	-



Z-DNA

Из выравниваний z-днк тоже сложно сделать какие-то выводы. Видимо, последовательности z-днк не консервативны у наших организмов.