Восстановление пути эволюции белкового семейства major CAPSID NCLDV

Проект по базам данным в биоинформатике
Декабрь 2020

Цели и задачи проекта

Цель проекта

Восстановить возможный путь эволюции белкового семейства CAPSID NCLDV и предложить эксперименты “мокрым” биологам, изучающим представителей данного семейства, в лаборатории

Задачи проекта

1. Планирование и подготовка проекта:
   - ❏ работа с выравниванием и его техническая обработка;
   - ❏ добавление в него последовательностей, относящихся к CAPSID NCLDV, из организмов, геномы которых еще не собраны;
   - ❏ поиск дальнего гомолога белков CAPSID NCLDV, который станет аут группой для укоренения.
2. Построение филогенетического дерева:
   - ❏ построение филогенетическое дерева на основе полученного выравнивания в программе MEGA;
   - ❏ «схлопывание» ветвей в программе TreeGraph;
   - ❏ проработка визуальной составляющей дерева в онлайн-инструменте iTOL.
3. Восстановление эволюционного пути белкового семейства:
   - ❏ составление функциональной характеристики семейства;
   - ❏ оценка консервативности аминокислотных остатков важных для выполнения основных функций белков CAPSID NCLDV;
   - ❏ оценка значимости аминокислотных остатков на основе анализа 3D структуры и пространственного выравнивания;
   - ❏ описание возможного эволюционного пути на основании работы;
   - ❏ предложение экспериментов “мокрым биологам” изучающим CAPSID NCLDV.

Предлагаемый подход к реализации проекта

01.

Планирование и подготовка проекта

1. Скачать выравнивание белков семейства с базы Pfam в формате fasta (full alignment) и провести техническую обработку выравнивания.
2. Добавить в него последовательности, относящиеся к белковому семейству, из организмов, геномы которых пока еще не собраны.
3. Провести поиск гомологов с помощью разных видов BLAST (tblastn и tblastx), указывая в качестве запроса наименее выделяющийся белок семейства и базу поиска контигов.
4. Добавить найденные последовательности к имеющимся и построить выравнивание заново с помощью Jalview.
5. Найти с помощью BLAST дальнего гомолога белков CAPSID NCLDV, который станет аут группой для укоренения.

02.

Построение филогенетического дерева

1. Построить филогенетическое дерево на основе полученного выравнивания в программе MEGA.
2. Схлопнуть ветви в программе TreeGraph.
3. Проработать визуальную составляющую дерева в онлайн-инструменте iTOL.

03.

Восстановление эволюционного пути белкового семейства

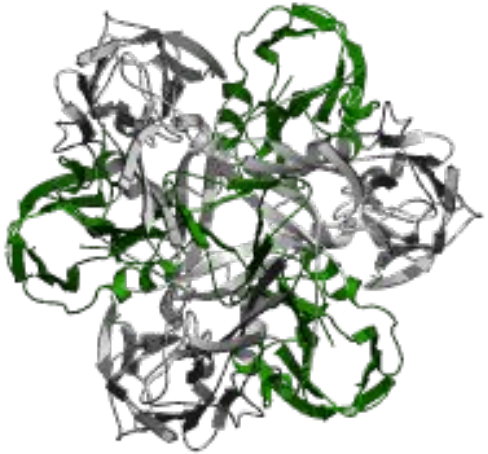
1. Составить функциональную характеристику семейства.
2. Оценить консервативность аминокислотных остатков важных для выполнения основных функций белков CAPSID NCLDV.
3. Оценить значимость аминокислотных остатков на основе анализа 3D структуры и пространственного выравнивания.
4. Описать возможный эволюционный путь на основании проделанной работы.
5. Предложить эксперименты “мокрым биологам”, изучающим CAPSID NCLDV.

Наше понимание ситуации

Факты:

- NCLDV* capsid protein - самый распространенный структурный белок, на долю которого может приходиться до 45% белка вириона.
- NCLDV составляют огромную группу эукариотических вирусов, состоящую из семейств Poxviridae, Asfarviridae, Iridoviridae, Ascoviridae, Phycodnaviridae, Marseilleviridae, Pithoviridae и Mimiviridae.
- Все эти вирусы имеют двухцепочечную ДНК, которые варьируются в размерах от примерно 100 килобаз (кб) до более чем 2,5 мегабаз.
- > 40 основных генов, которые являются общими для большинства NCLDV

*Nucleocytoplasmic Large DNA Viruses



PDB entry 6KU9: Structure of the African swine fever virus major capsid protein p72

Работа с выравниванием (1/2)

- fasta sorting
- getting proteins names from Uniprot
- BLAST searching
- DNA translation
- lineage from Uniprot to .csv
- creating color annotation for iTOL

Работа с выравниванием (2/2)

Search for model organism: MUSCLE, sorting by Muscle_Order

Model organism:

- ❏ Megavirus G5CQ21_9VIRU
- ❏ Acanthamoeba polyphaga mimivirus CAPS3_MIMIV

Homologs found: 244

BLAST options: algorithm tbastn, base of Whole Genome Sequences(wgs), matrix BLOSUM45, word size 3.

Построение филогенетического дерева (1/9)

- Vir only (bootstrap 40) - только вирусы семейства (iTOL)
- Euk Vir (bootsrap 5) - всё семейство вирусы+эукариоты (iTOL)
- Vir Homo (bootstrap 30) - вирусы и их гомологи (iTOL)
- Euk Vir Homo (bootstrap 5) - всё семейство вирусы+эукариоты и гомологи вирусов (iTOL)
- **Euk Vir Bac (bootsrap 5) - всё семейство вирусы+эукариоты и найденные бактерии** (iTOL)

Virus - Host

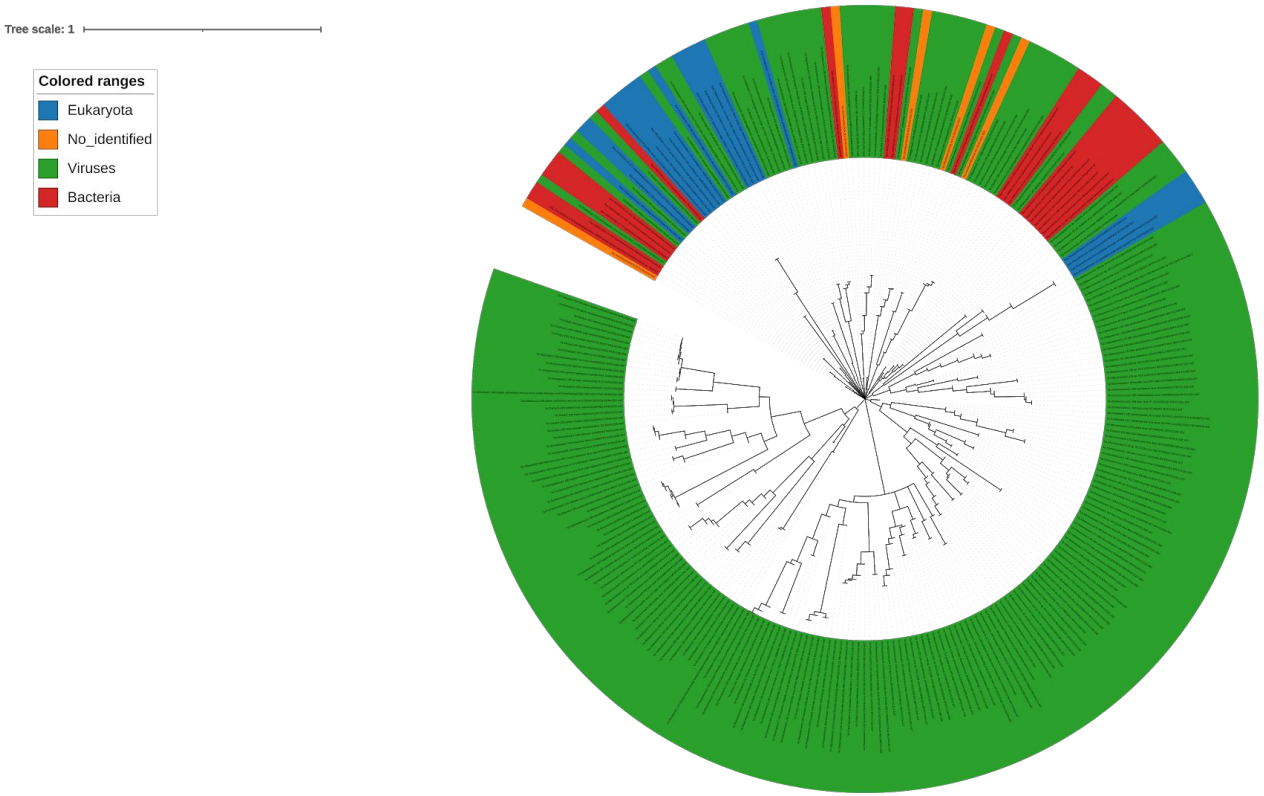
	Eukaryote	Virus / Homologue	Dist
1	Acanthamoeba castellanii str. Neff/1-185	Aureococcus anophagefferens virus	0,71
2	Acanthamoeba castellanii str. Neff/238-298	Uncultured/Marine virus DNA	0,41
3	Acanthamoeba castellanii str. Neff/105-294	Aureococcus anophagefferens virus	0,70
4	Acanthamoeba castellanii str. Neff/217-261	Aureococcus anophagefferens virus	0,49
5	Catenaria anguillulae PL171/233-408	Aureococcus anophagefferens virus	0,65
6	Chlamydomonas eustigma/459-666	Afrovirus	0,59
7	Chlamydomonas eustigma/401-665	Afrovirus	0,45
8	Ectocarpus siliculosus Brown alga Conferva siliculosa/244-431	Ectocarpus siliculosus virus	0,03
9	Gonapodya prolifera strain JEL478/230-405	Moumouvirus	0,64
10	Gonapodya prolifera strain JEL478/166-340	Yasminevirus	0,83
11	Klebsormidium nitens Green alga Ulothrix nitens/713-888	Polyphaga lentillevirus	0,50
12	Klebsormidium nitens Green alga Ulothrix nitens/264-439	Polyphaga lentillevirus	0,48
13	Klebsormidium nitens Green alga Ulothrix nitens/238-410	Aureococcus anophagefferens virus	0,62
14	Klebsormidium nitens Green alga Ulothrix nitens/171-266	Afrovirus	0,56
15	Phytophthora cactorum/8-140	Megavirus	1,03
16	Phytophthora cactorum/286-503	Pacmanvirus A23	0,77
17	Phytophthora nicotianae Buckeye rot agent/284-502	Pacmanvirus A23	0,79
18	Phytophthora parasitica strain INRA-310/284-502	Pacmanvirus A23	0,87

—> search for new hosts

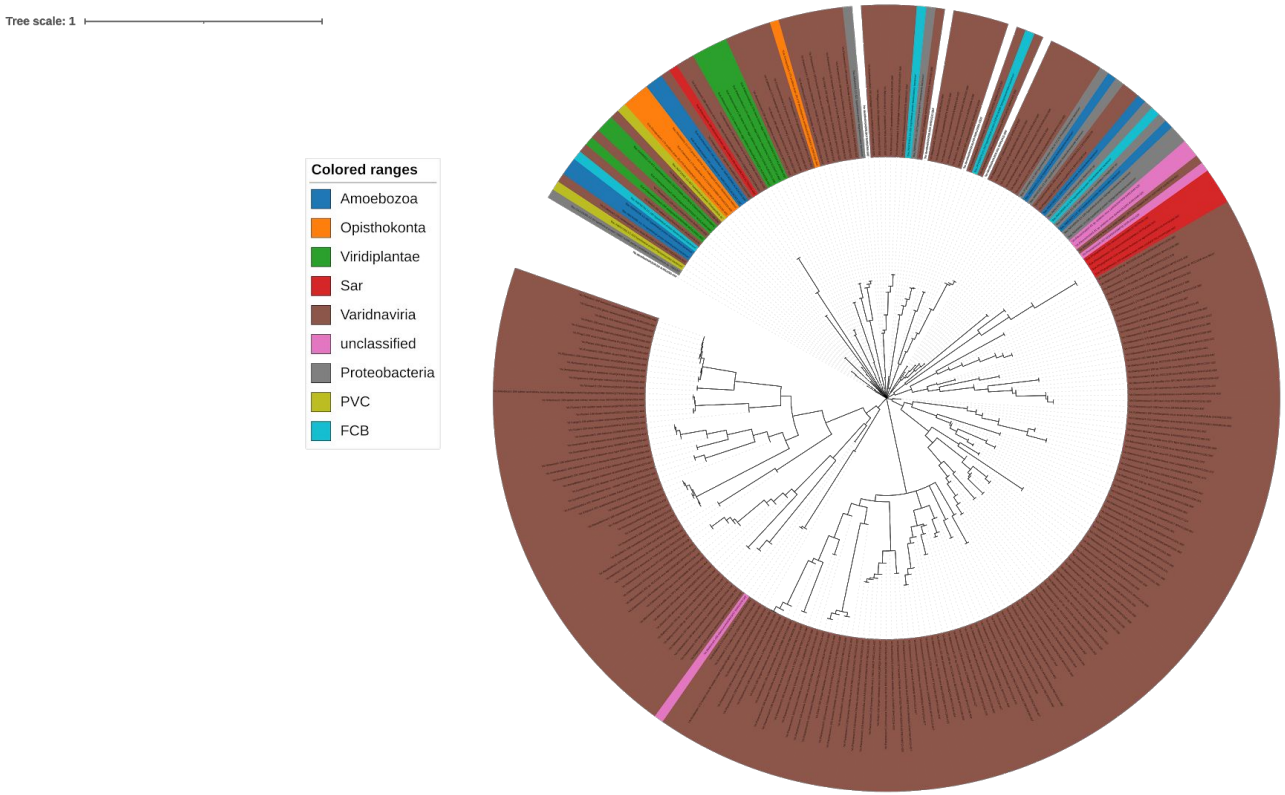
Virus - Host

	Bacteria	Virus / Homologue	Dist
1	Actinobacteria bacterium	Chrysochromulina ericina virus CeV01	0,34
2	Actinobacteria bacterium/1-118	Phaeocystis globosa virus	0,28
3	Actinobacteria bacterium/1-334	Phaeocystis globosa virus	0,36
4	Actinobacteria bacterium/1-495	Aureococcus anophagefferens virus	0,37
5	Alphaproteobacteria bacterium/1-109	Tetraselmis virus 1	0,48
6	Alphaproteobacteria bacterium/1-585	Paramecium bursaria Chlorella virus NY2A	0,61
7	Bacteroidetes bacterium	Aureococcus anophagefferens virus	0,52
8	Bdellovibrionaceae bacterium	Phaeocystis globosa virus	0,42
9	Blyttiomyces helicus (Euk)	Chrysochromulina ericina virus CeV01	0,54
10	Candidatus Endolissoclinum sp.	Aureococcus anophagefferens virus	0,37
11	Candidatus Pelagibacter sp.	Paramecium bursaria Chlorella virus NY2A	0,63
12	Chlamydomonas eustigma (Euk)	Aureococcus anophagefferens virus	0,62
13	Flavobacteriales bacterium	Chrysochromulina ericina virus CeV01	0,35
14	Magnetococcales bacterium	Phaeocystis globosa virus	0,78
15	Marinovum sp.	Paramecium bursaria Chlorella virus NY2A	0,61
16	Micrococcales bacterium	Paramecium bursaria Chlorella virus NY2A	0,61
17	Planctomycetes bacterium	Aureococcus anophagefferens virus	0,56
18	Proteobacteria bacterium	Phaeocystis globosa virus	0,41
19	Rhodobacteraceae bacterium	Chrysochromulina ericina virus CeV01	0,27
20	Rhodopirellula sp.	Aureococcus anophagefferens virus	0,52
21	Sphingobacteriales bacterium	Moumouvirus australiensis	0,61

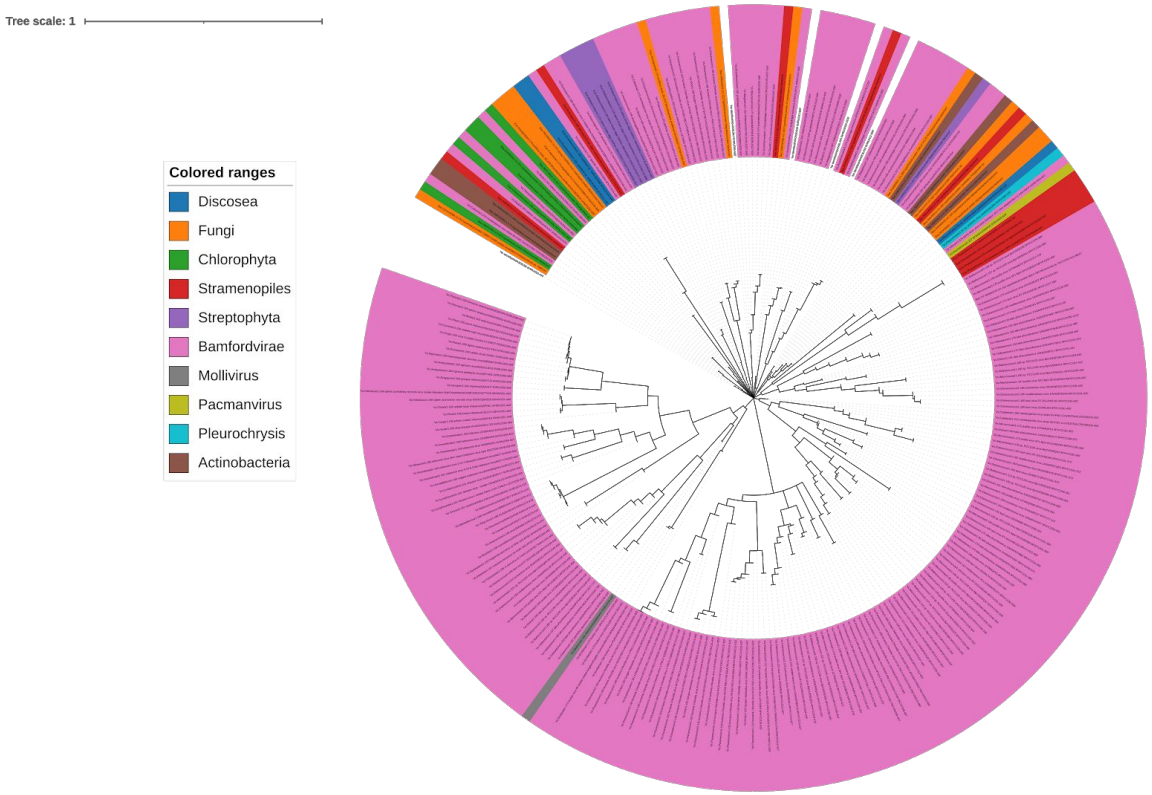
Tree: Vir+Euk+Bac (level 1 - superkingdom)



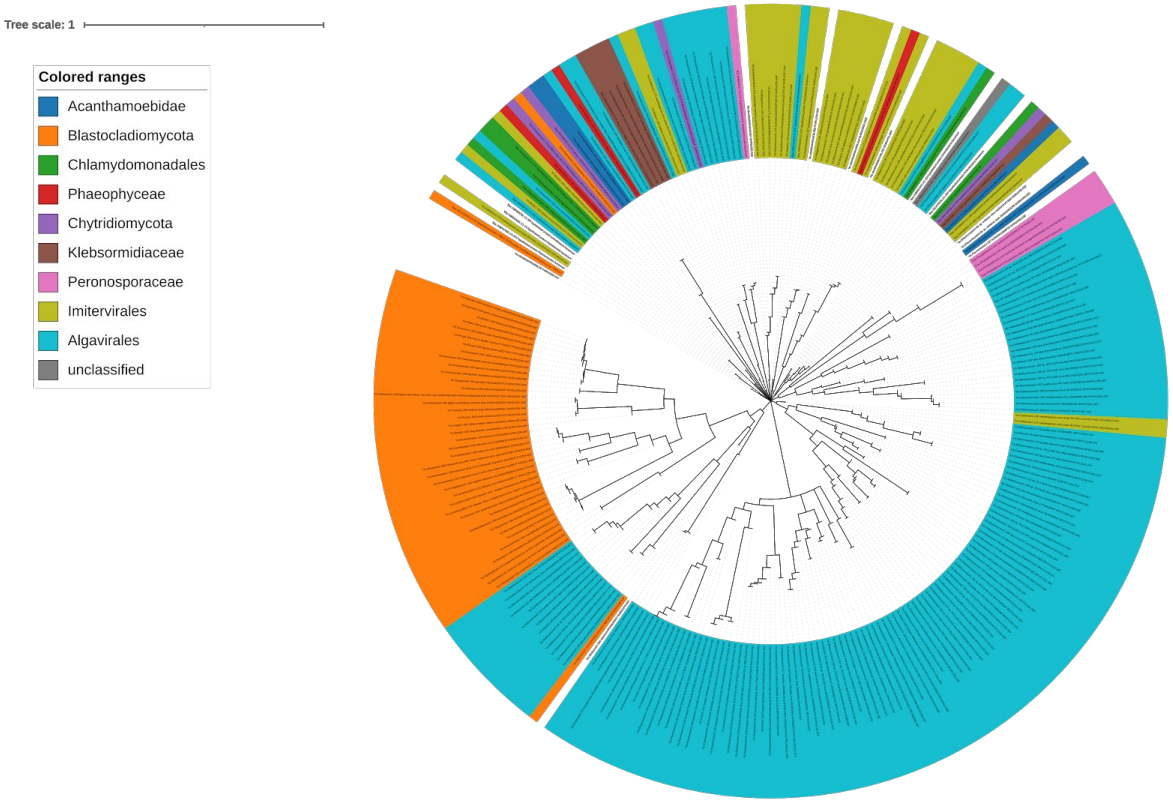
Tree: Vir+Euk+Bac (level 2 - kingdom)



Tree: Vir+Euk+Bac (level 3 - subkingdom)

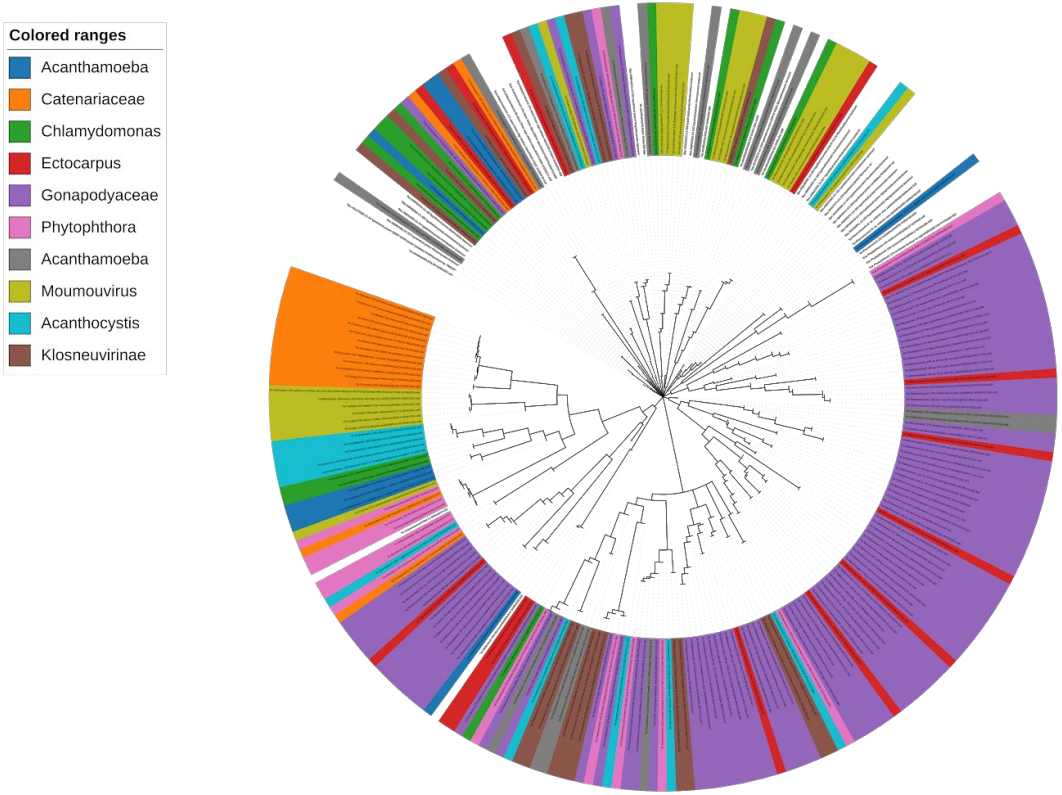


Tree: Vir+Euk+Bac (level 6 - subphylum)



Tree: Vir+Euk+Bac (level 9 - subclass)

Tree scale: 1

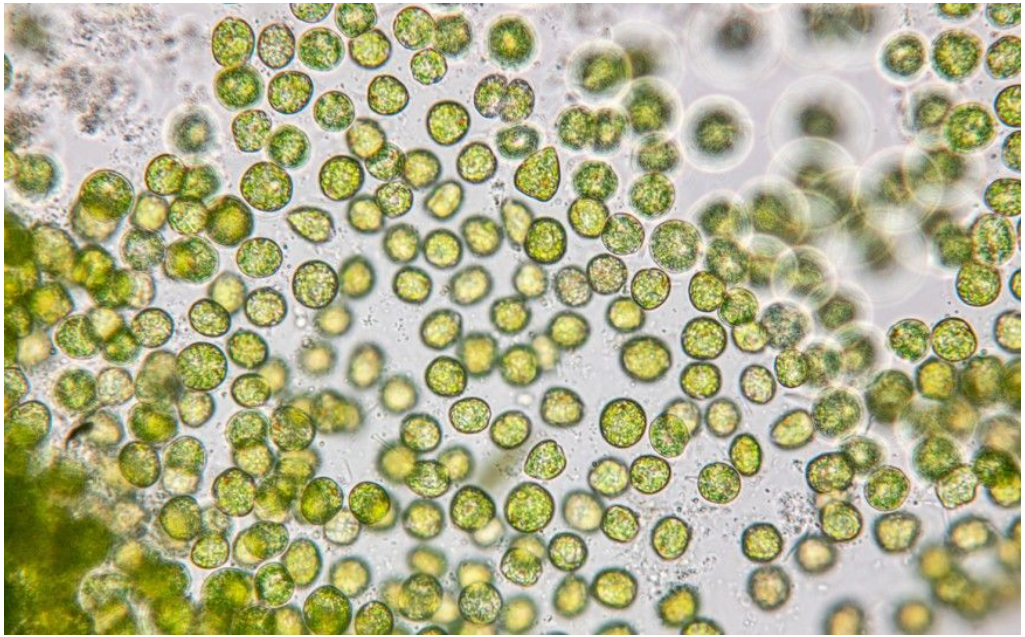


Содержание

Results analysis

-We have found 2 sequences of green algae

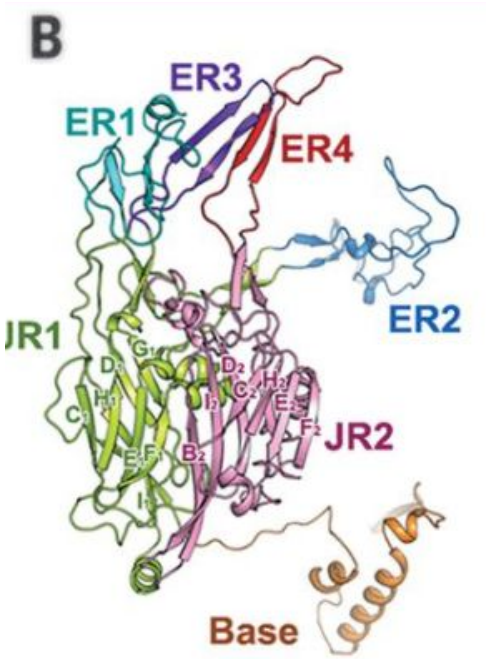
It was recently shown that Giant DNA Viruses encodes almost 10% of green algae genome.

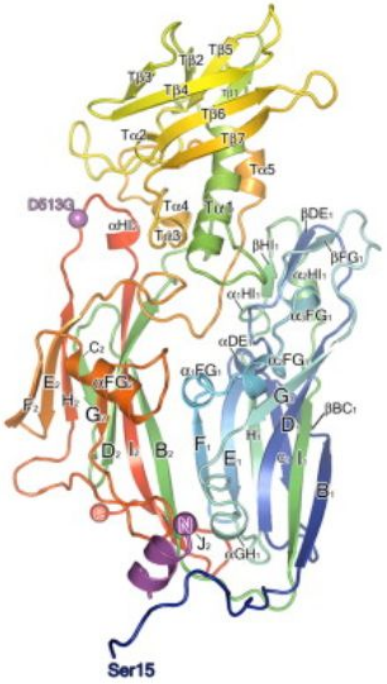


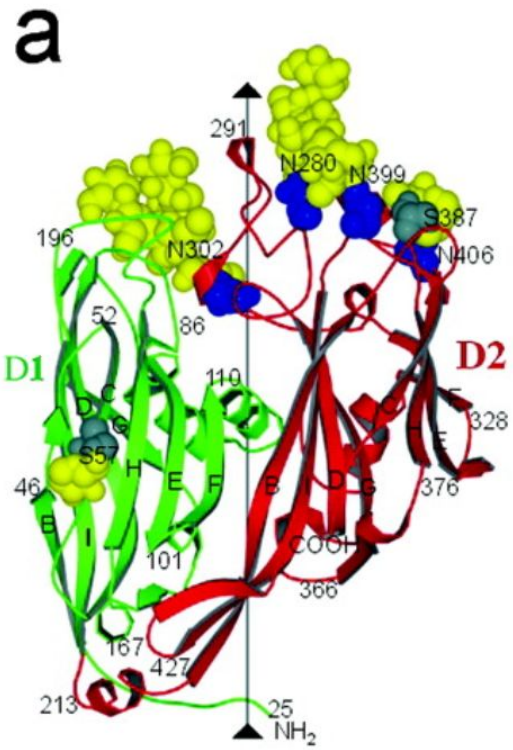
Article | Published: 18 November 2020

Widespread endogenization of giant viruses shapes genomes of green algae

Main 3D feature - double jelly-roll







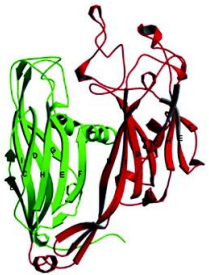
p72 monomer of ASF

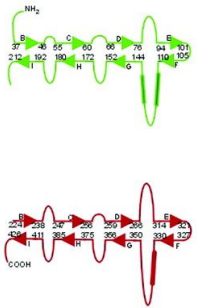
D13 monomer of VACV

Vp54 monomer of PBCV-1[17]

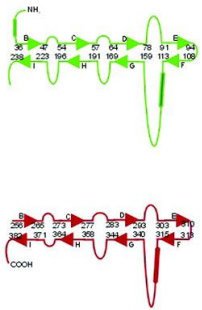
Evolution of jelly-roll

PBCV-1 Vp54

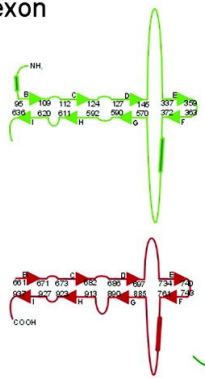




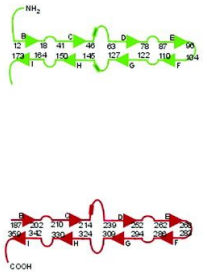
PRD1 P3



Adenovirus hexon



CpMV large subunit



PM2

PDB: 2VVF

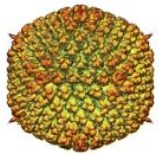
Adenovirus

PDB: 6B1T

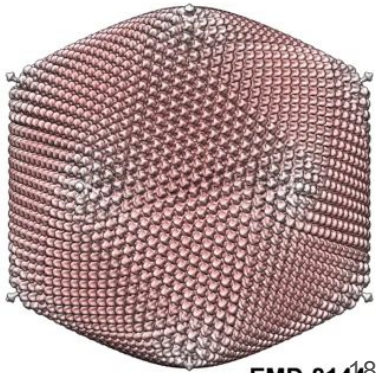
Faustovirus

PDB: 5J7V

EMD-1082
T=21d

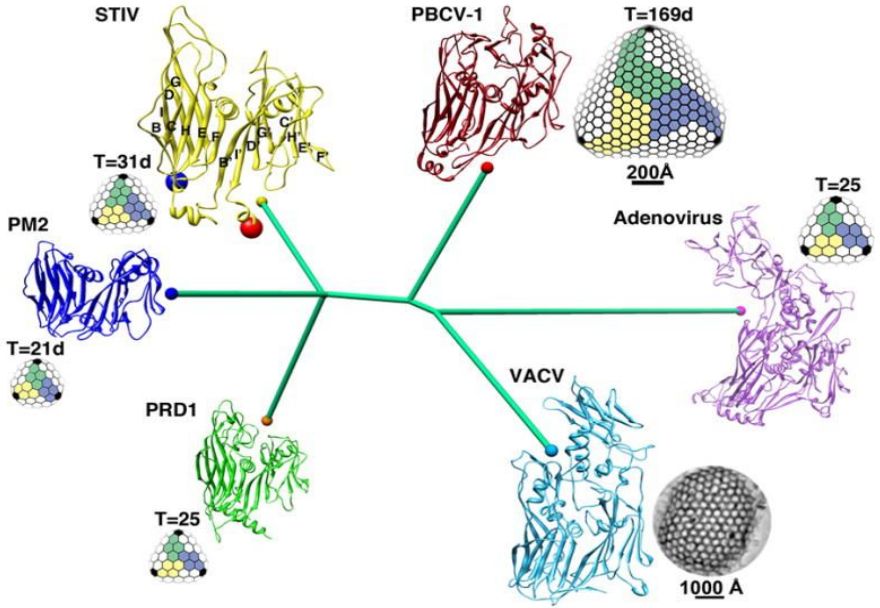


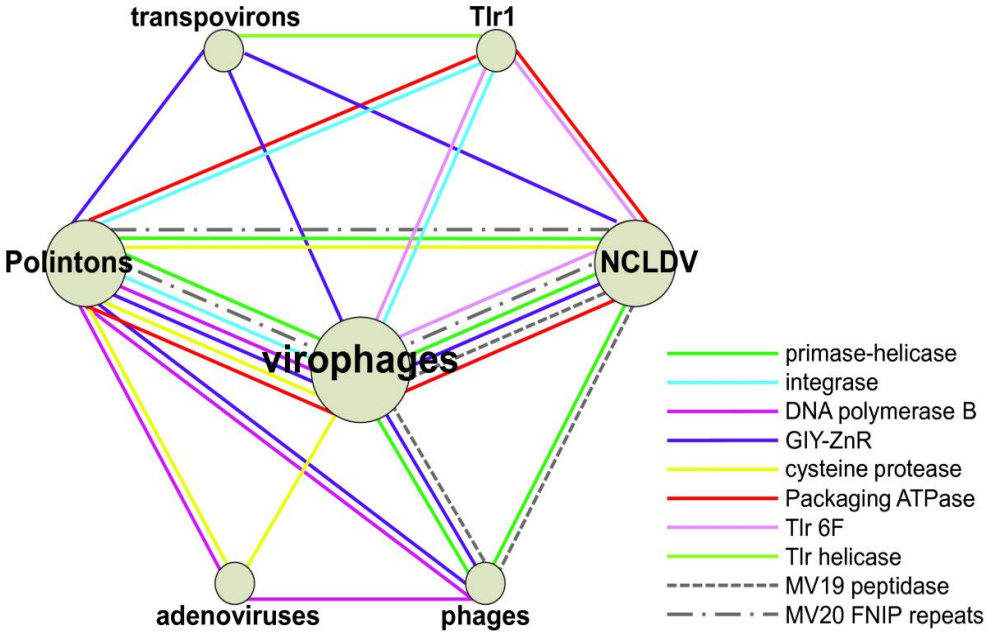
EMD-1586
T=25



EMD-8144
T=277

Evolution of jelly-roll - polinton-like proteins





Example of structural alignment

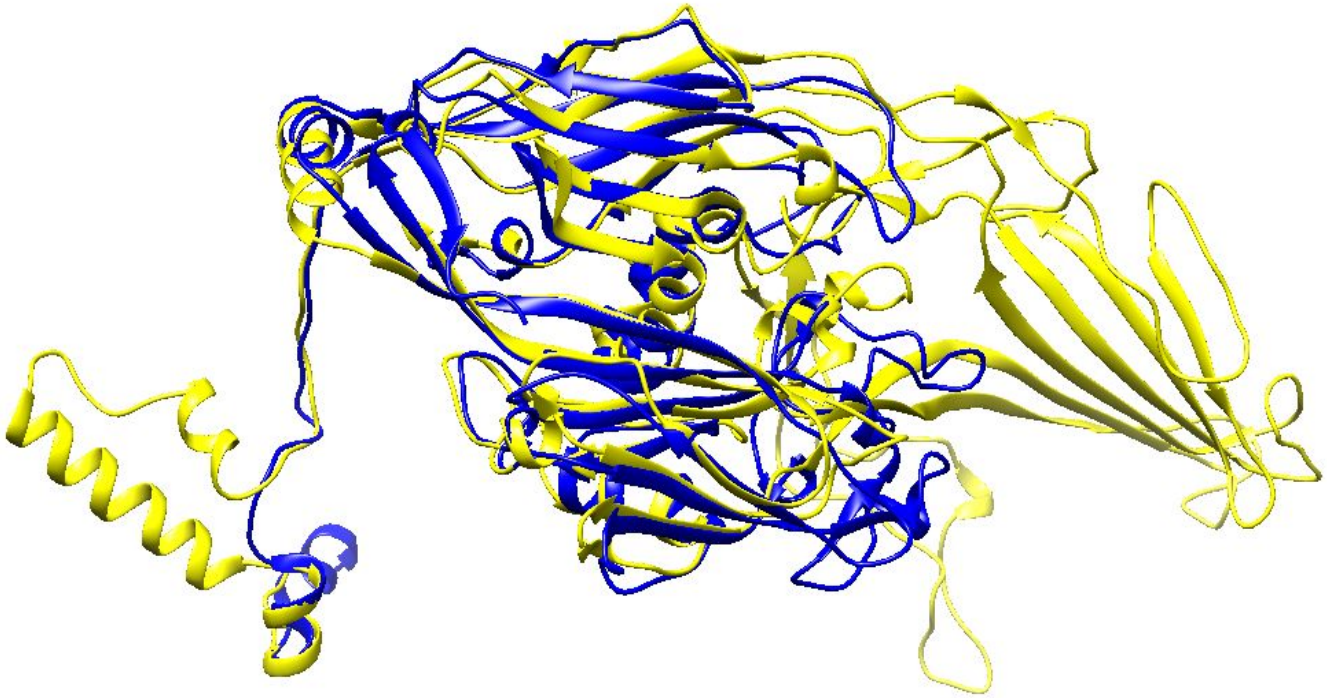
Structures:

6OJN

5TIP

6KU9

5J7U



5TIQ:A - Major Capsid protein of PBCV-1
5J7U:B - Faustovirus major capsid protein

Summary

1. Found jelly-roll structure => earliest steps of evolution
2. Found viral genome parts in both eukaryotic and prokaryotic organisms
3. We assume we have discovered horizontal gene transfer
4. Visualized the evolution of the family as a tree

Further analysis

1. Check Fungi contamination
2. Find connections between all DNA viruses, check for jelly-roll.
3. Find ways to treat ASFV in pigs
4. Analyze why bacterias contain parts of:
   Chrysochromulina ericina virus CeV01, Phaeocystis globosa virus, Aureococcus anophagefferens virus, Tetraselmis virus 1, Paramecium bursaria Chlorella virus NY2A, Moumouvirus australiensis.
5. Check in labs infecting other organisms by:
   Aureococcus anophagefferens virus, Uncultured/Marine virus DNA, Afrovirus, Ectocarpus siliculosus virus, Moumouvirus, Yasminevirus, Polyphaga lentillevirus, Megavirus, Pacmanvirus A23

Save pigs!!!!

Контакты

Иван Тараскин

taraskin.ia@phystech.edu

Юнона Поспелова

pospelova.iu@phystech.edu

Дмитрий Пустошилов

pustoshilov.dv@phystech.edu

NAPOLEON IT

Декабрь 2020

github.com/pustoshilov-d/BioinfDB

Literature

1) https://science.sciencemag.org/content/366/6465/640.long
2) https://www.pnas.org/content/99/23/14758
3) https://doi.org/10.1016/j.str.2011.03.023
4) https://virologyj.biomedcentral.com/articles/10.1186/s12985-018-1097-1
5) https://www.pnas.org/content/114/12/E2401#F2
6) https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3242167/