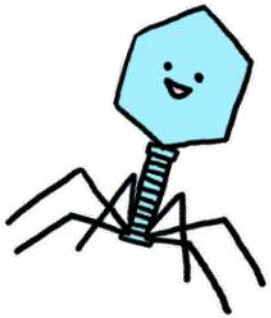
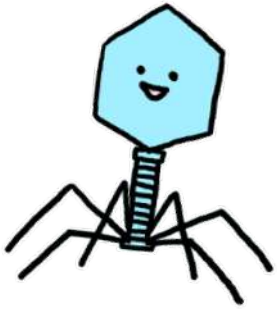
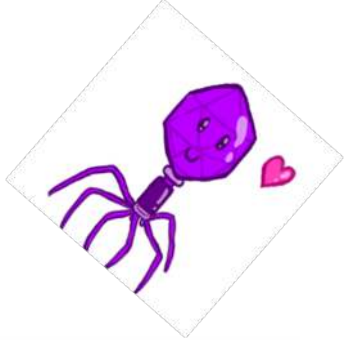
Plasmid host range prediction based on CRISPR arrays.
CRISPR Cas systems search.

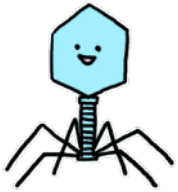
Students: Mikhail Kongoev
Iana Fedorova
Supervisor: Mikhail Rayko
CAB SpbU

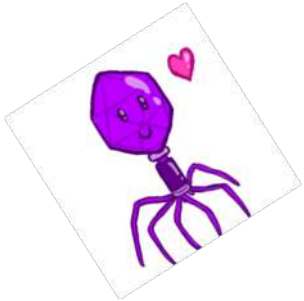
BIOINFORMATICS
INSTITUTE

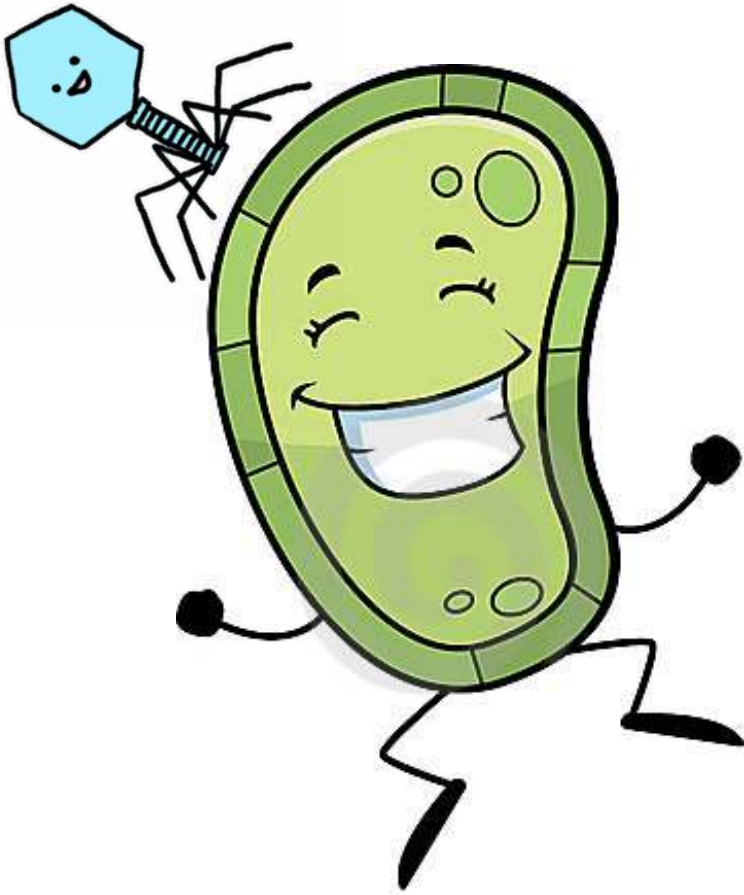


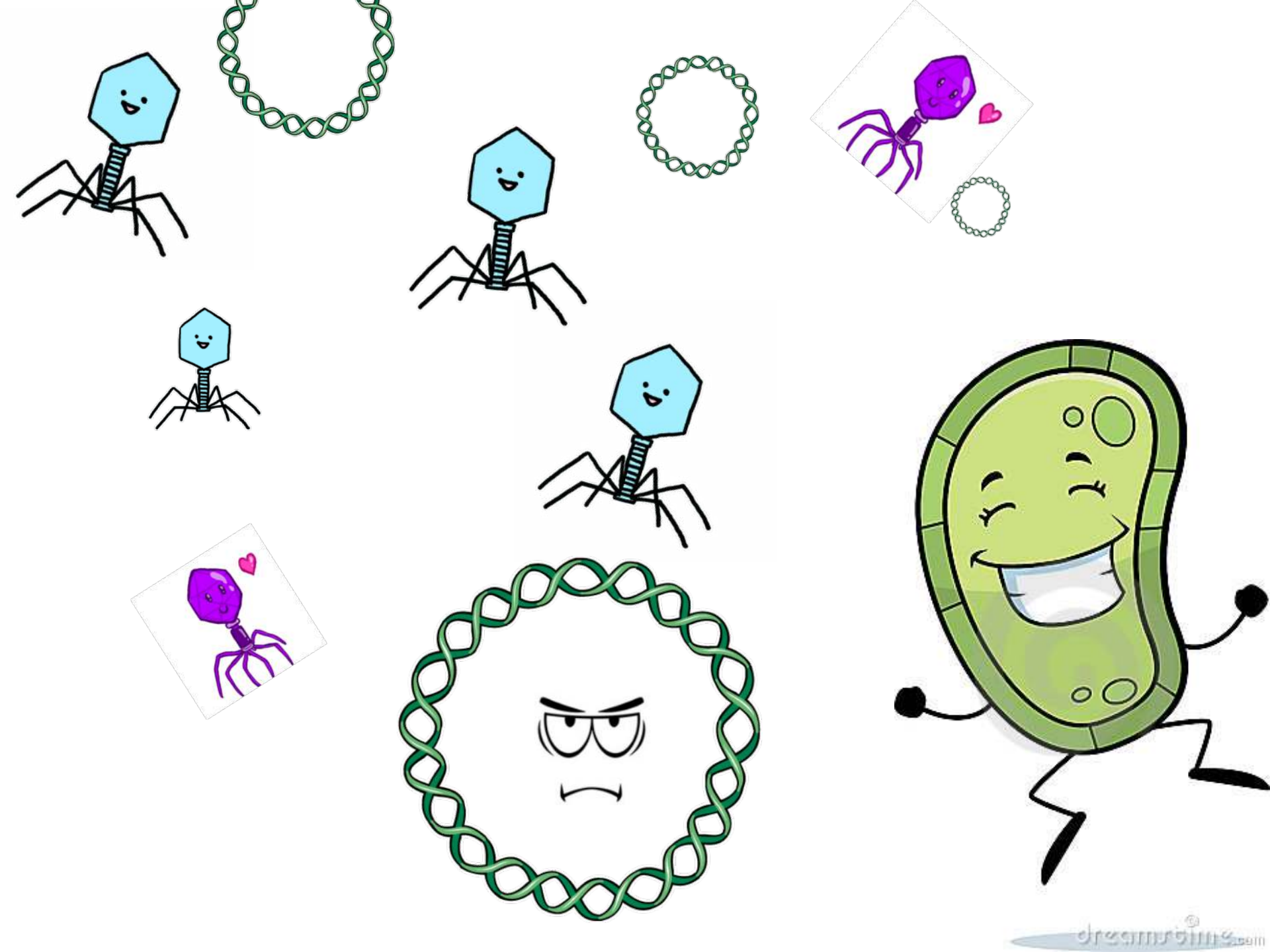


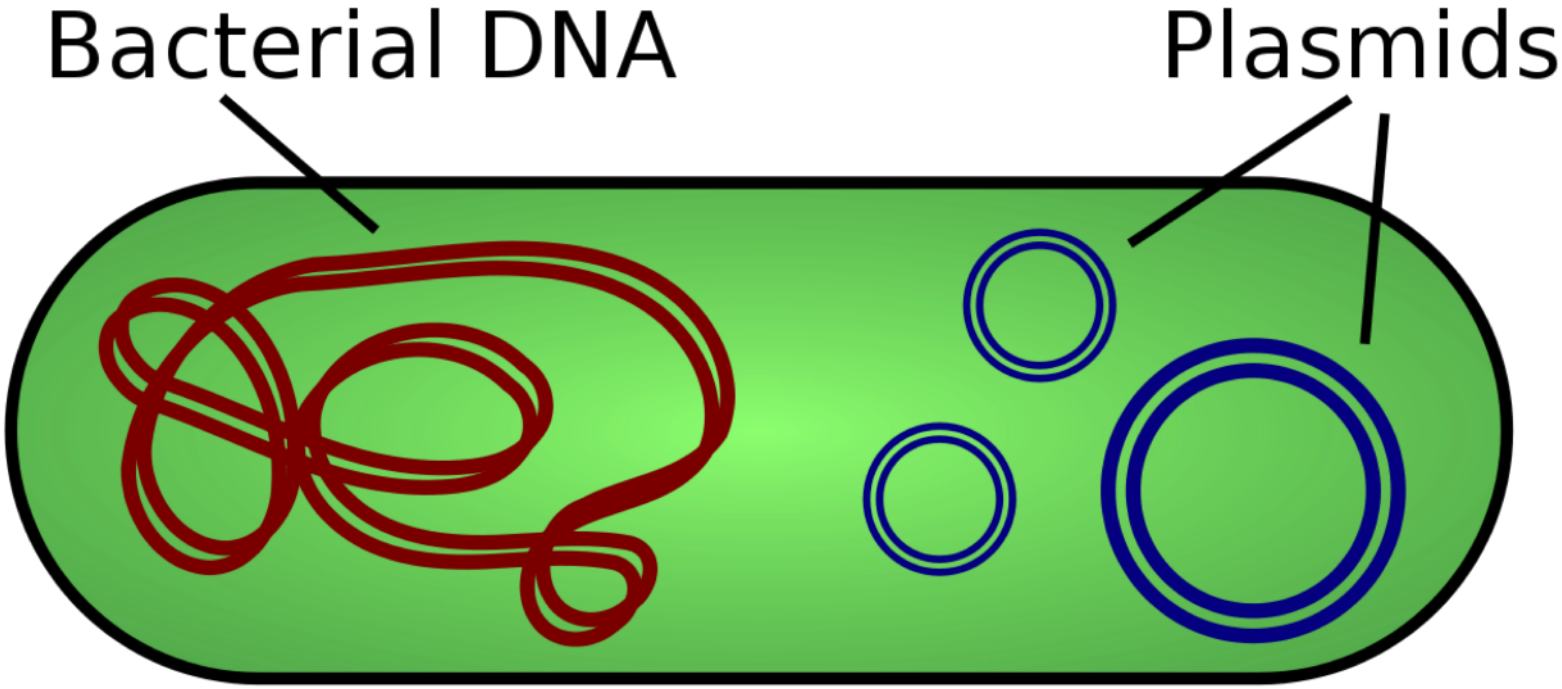




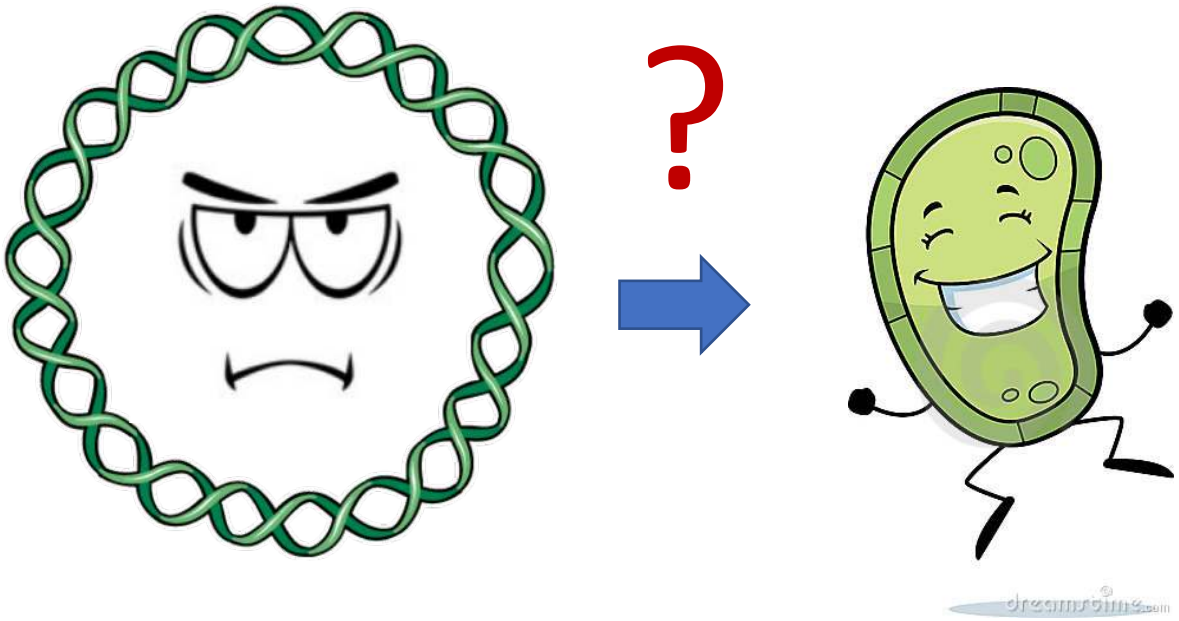




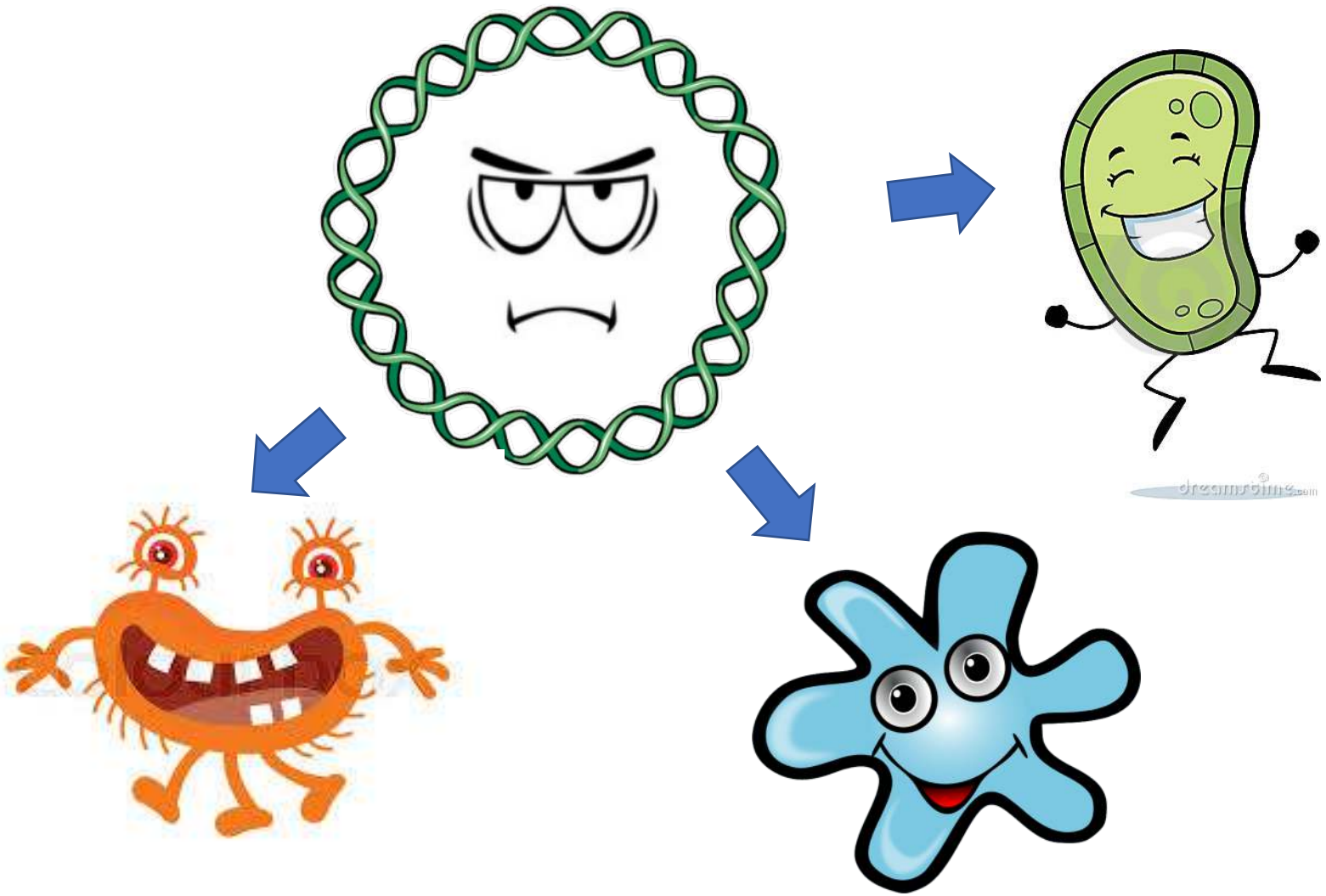



Bacterial DNA
Plasmids

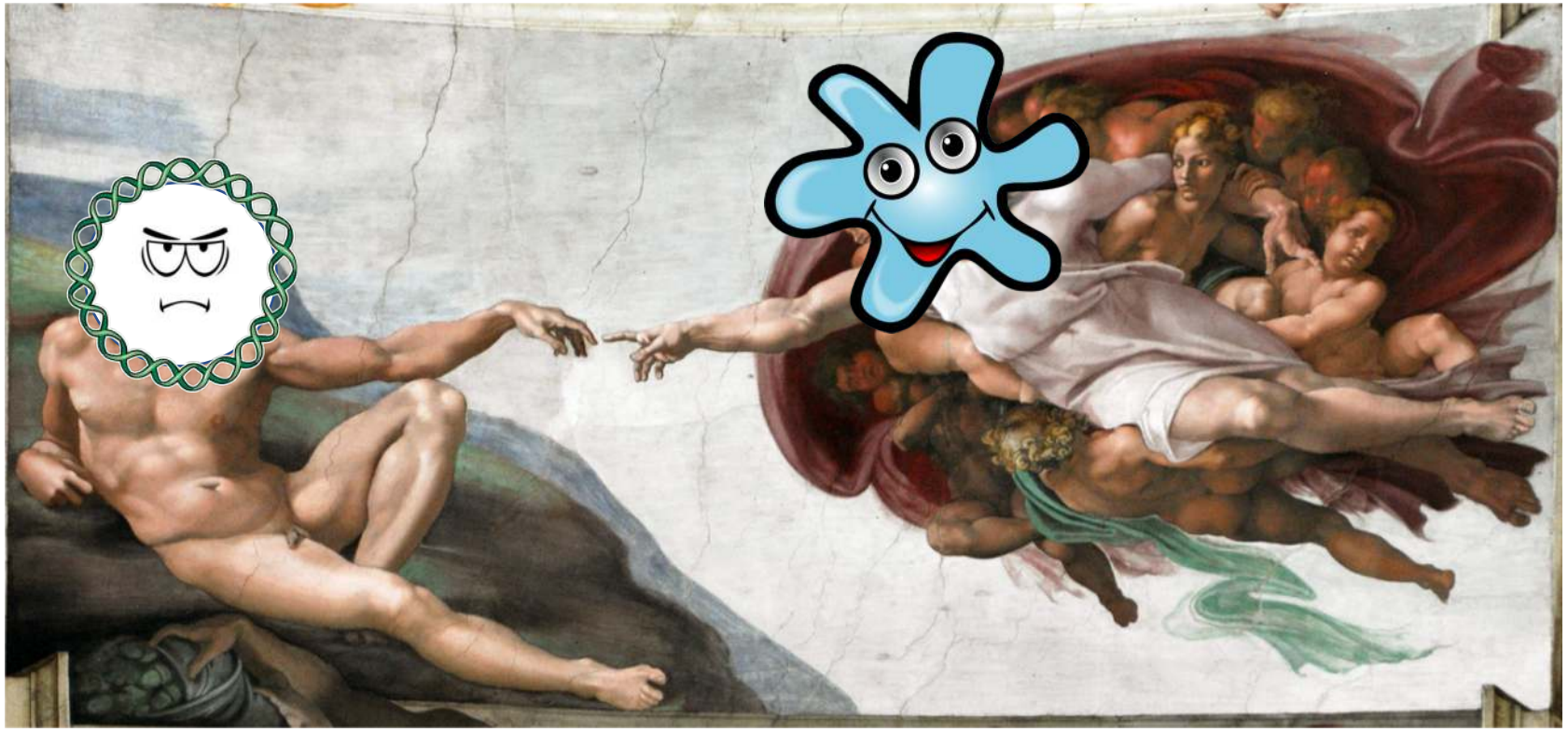
How to determine host range of the plasmid?

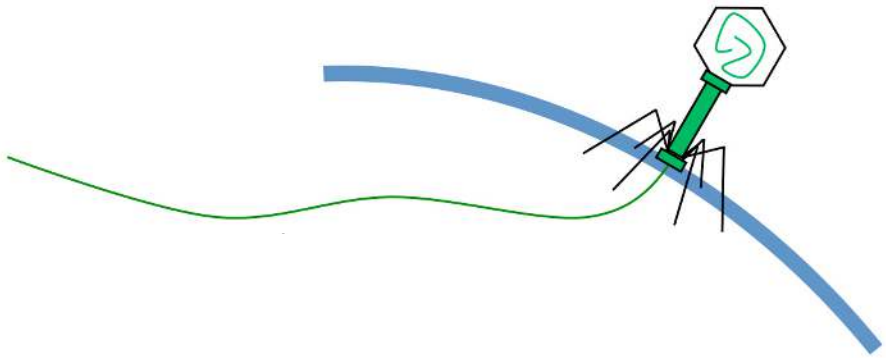


Can same plasmid infect several different hosts?

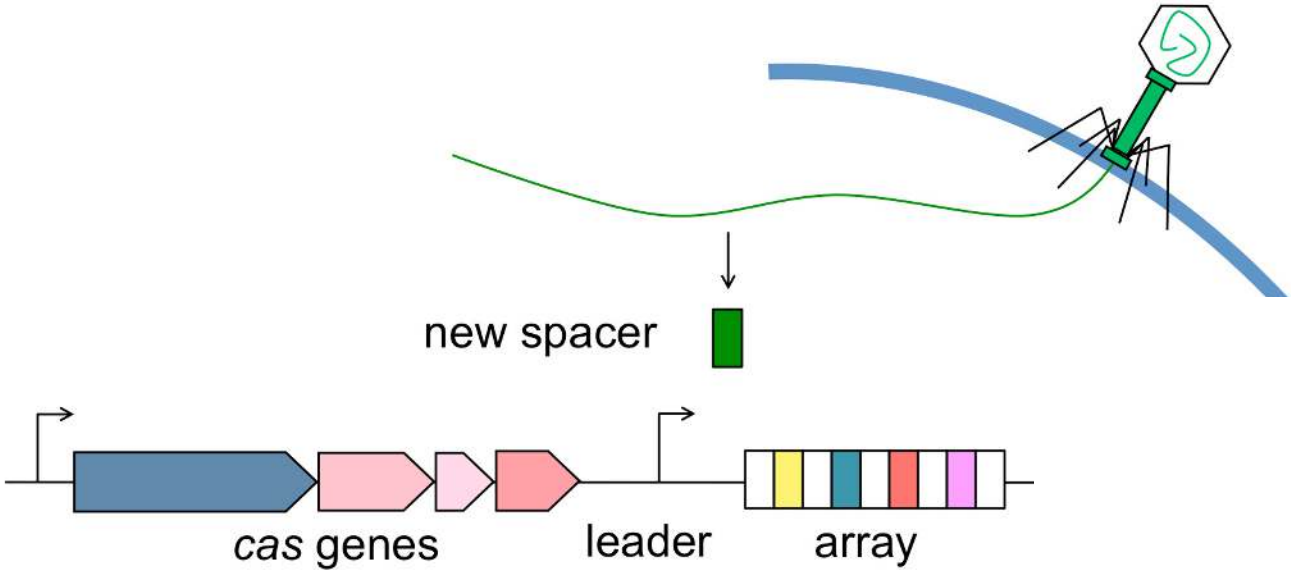


Is there any link between a host and it's plasmids?

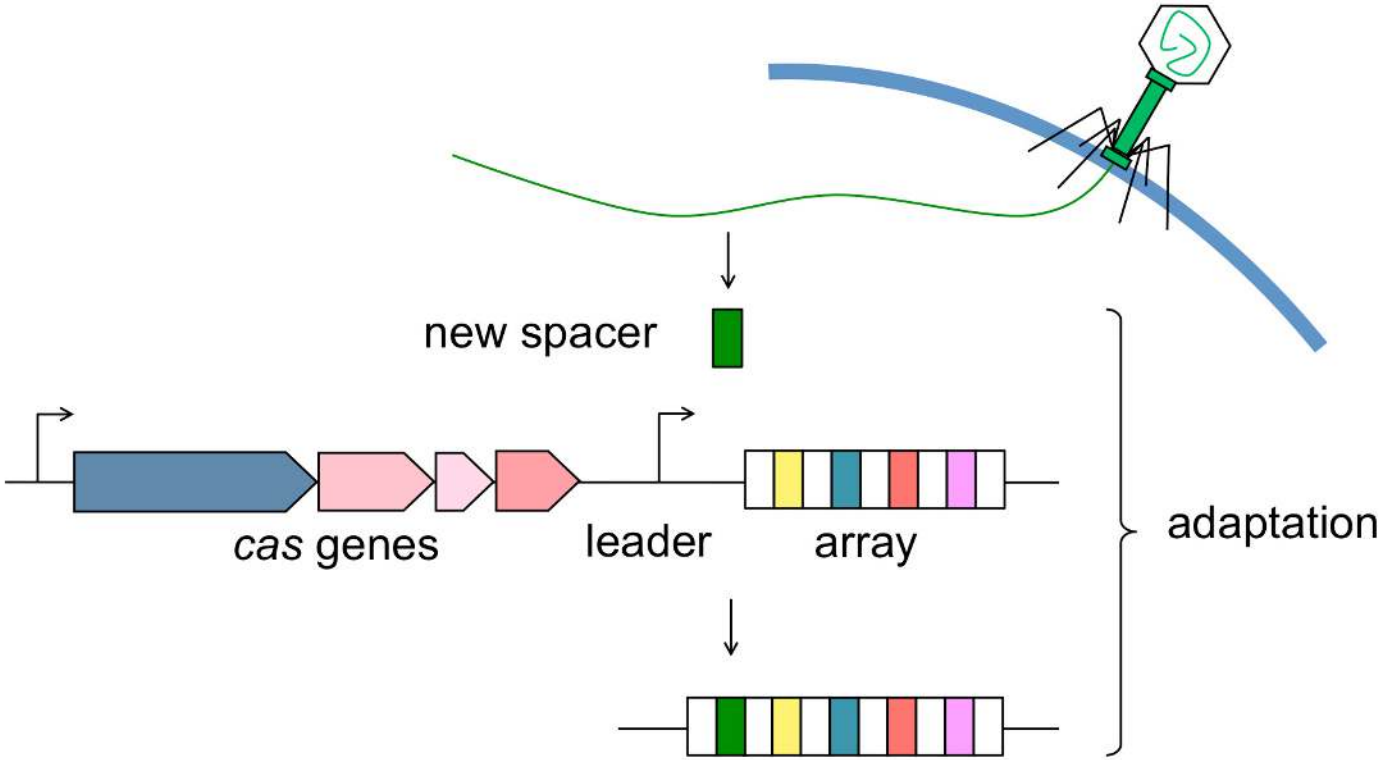




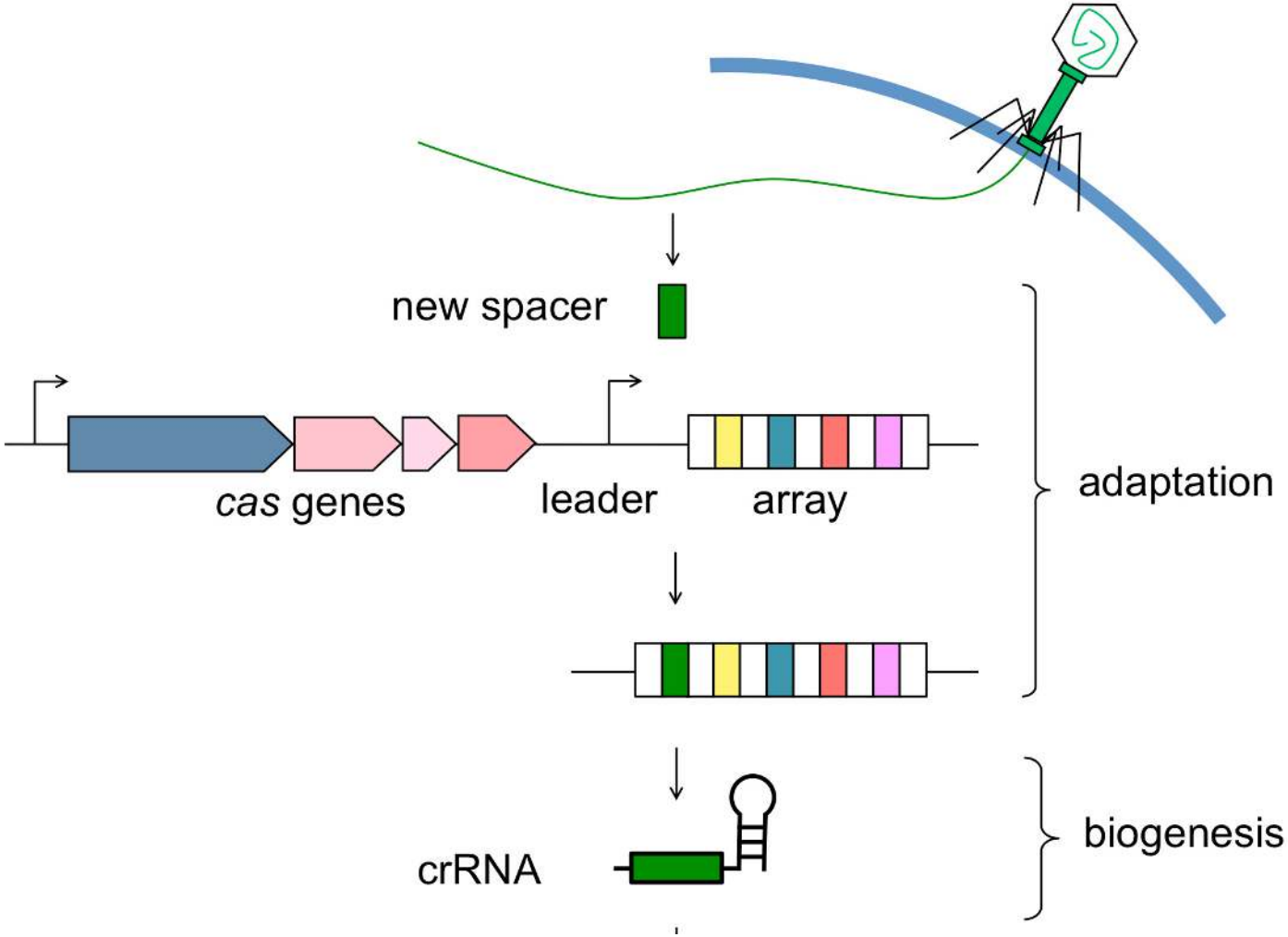
from Luciano Marraffini lab site



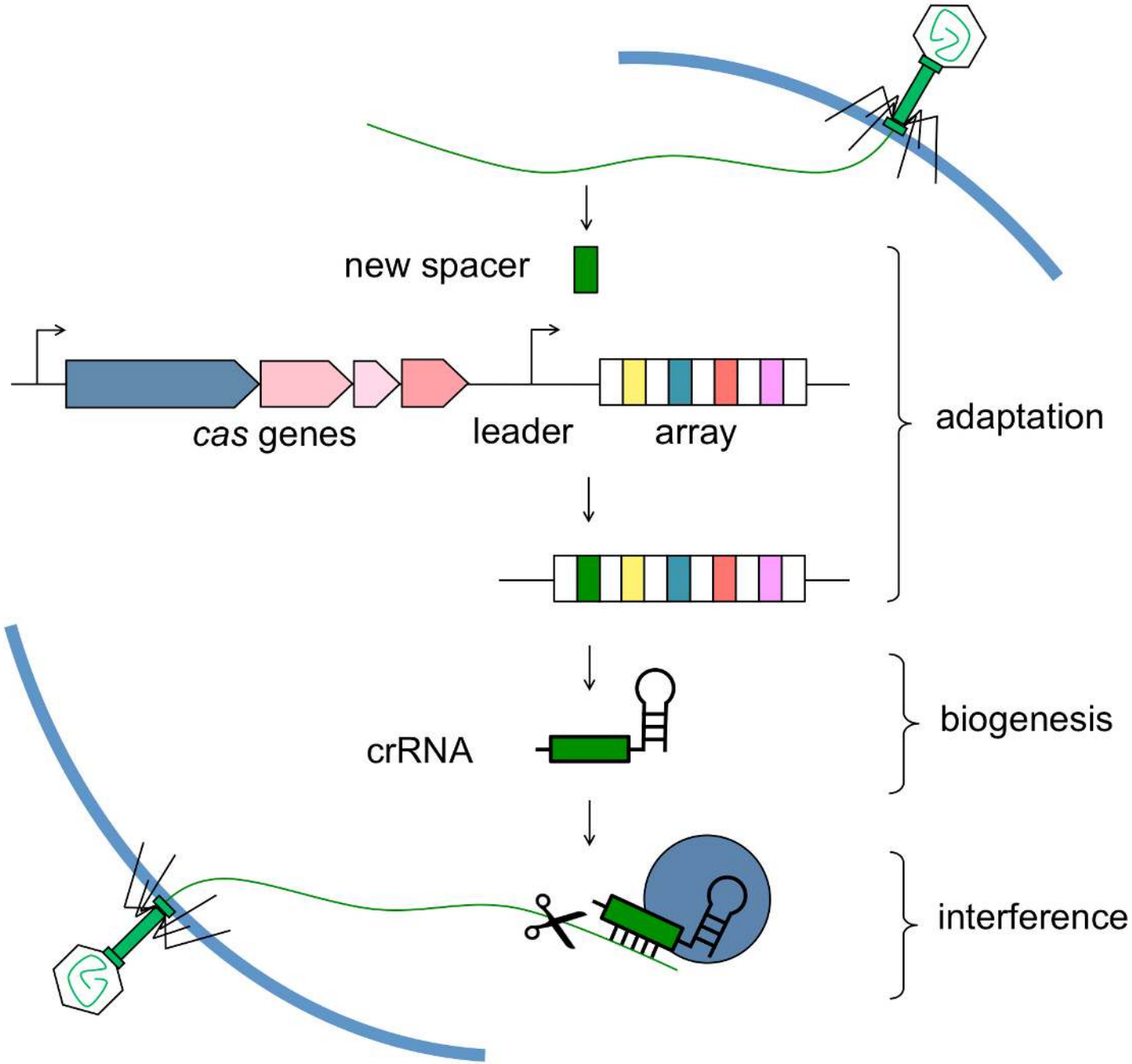
from Luciano Marraffini lab site



from Luciano Marraffini lab site

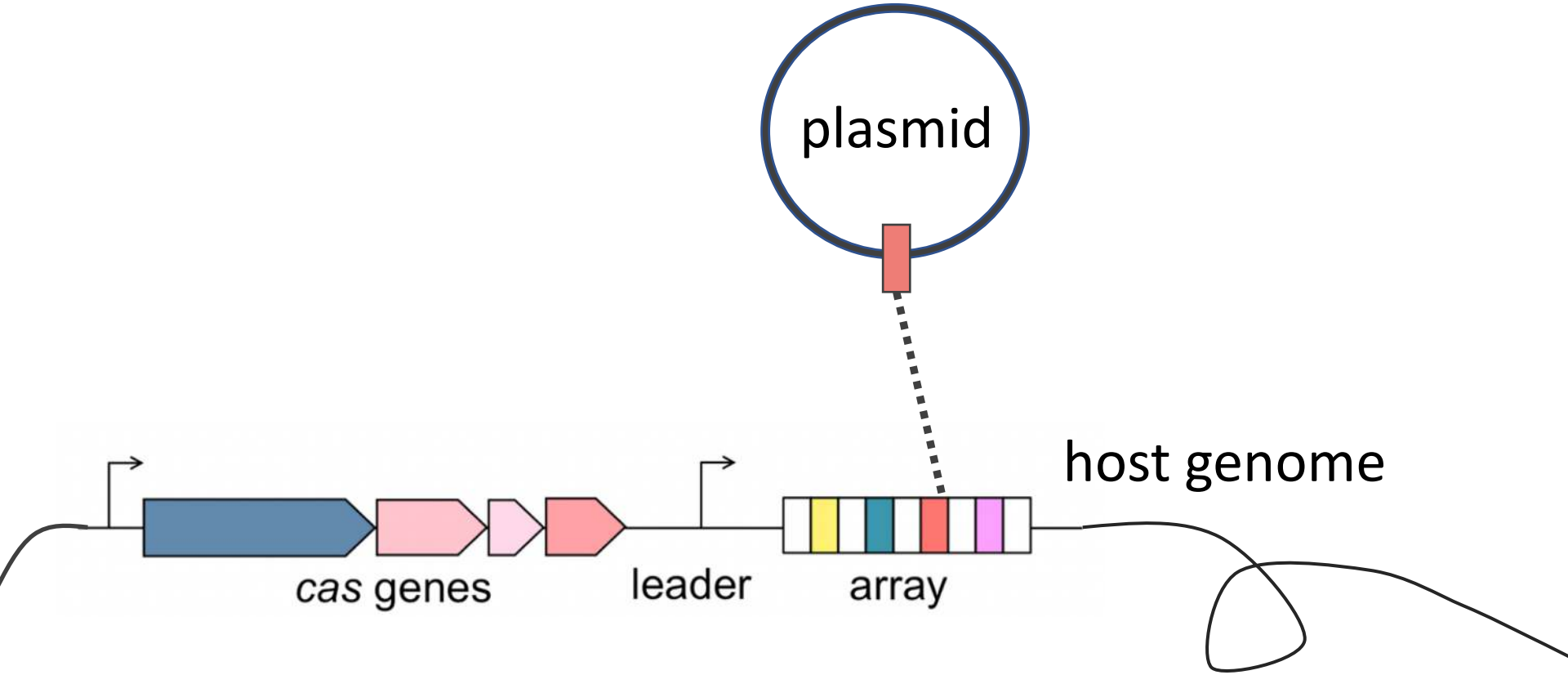


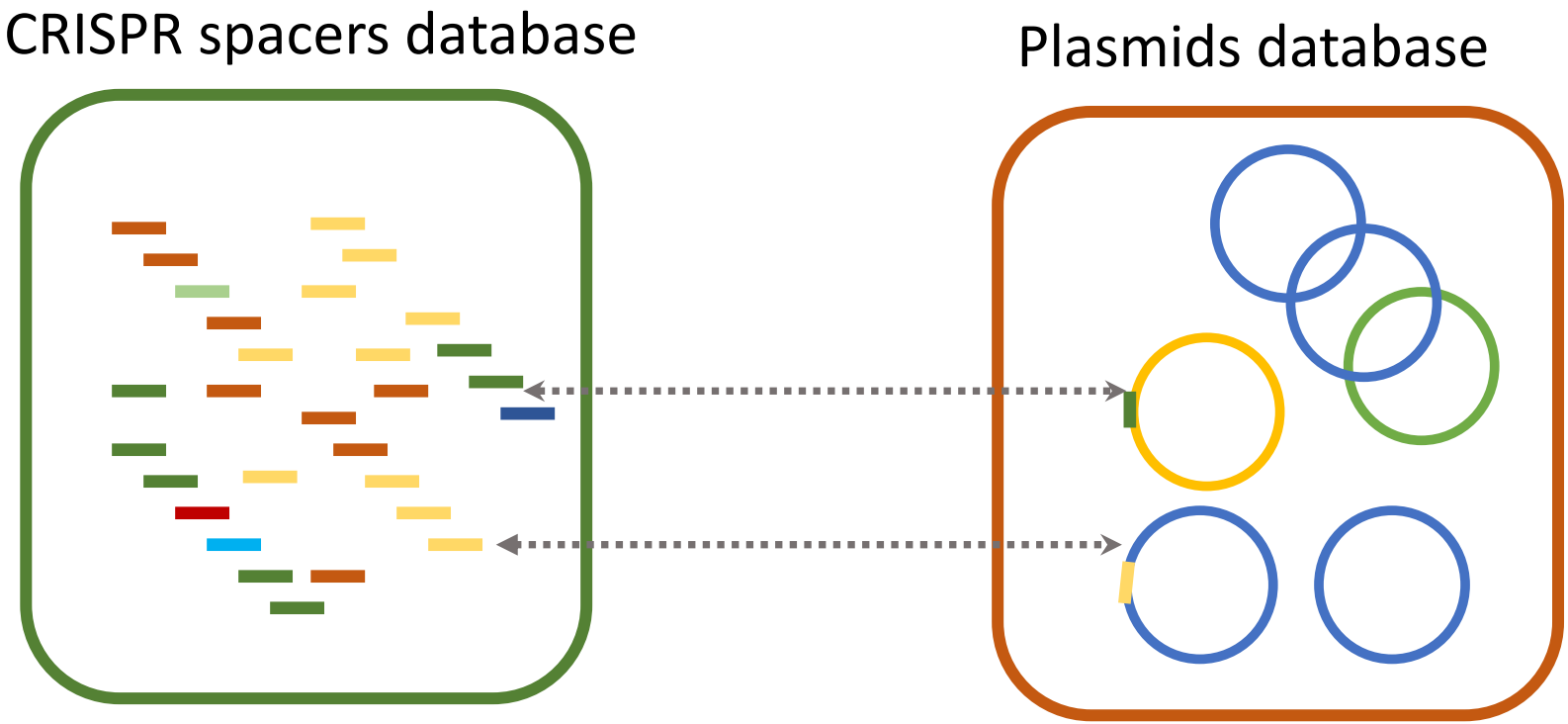
from Luciano Marraffini lab site



from Luciano Marraffini lab site

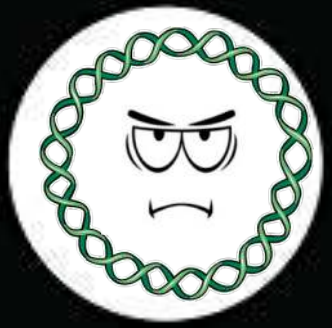
Plasmid host prediction by CRSIPR

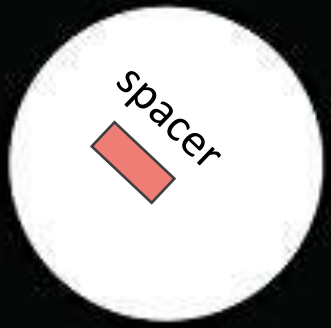



CRISPR spacers database
Plasmids database

It's a Match!

You guys have liked each other





Send Message

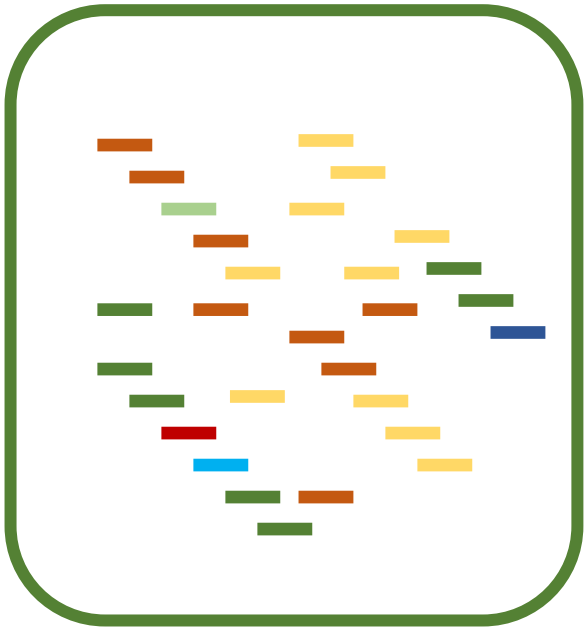
Keep Swiping

#mr&mrswilliams

The main goal: to determine all plasmids host range based on CRISPR arrays in bacterial genomes

- Download all plasmids known to date from databases
- Download CRISPR spacer database
- Align plasmids to spacers to define host
- Group results by taxonomy to define host range

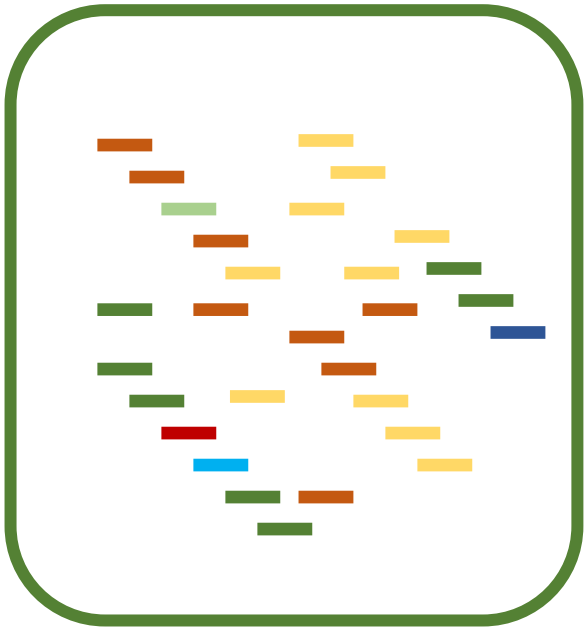
CRISPR spacers database



~ 141 705 spacers

CRISPR Finder database

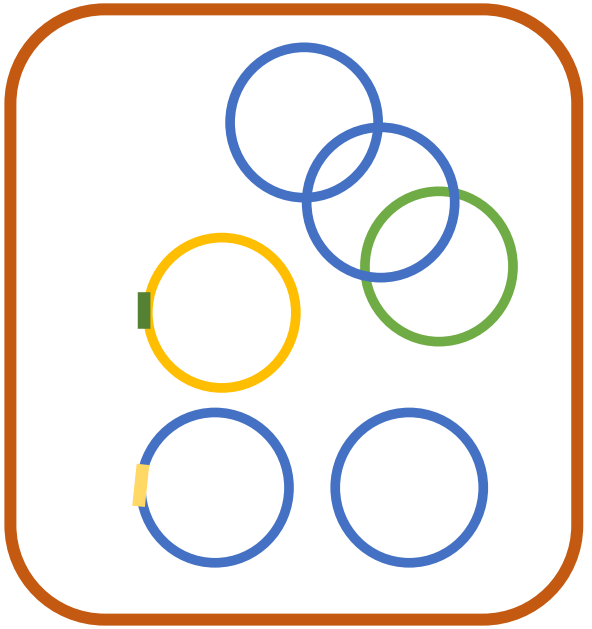
CRISPR spacers database



~ 141 705 spacers

CRISPR Finder database

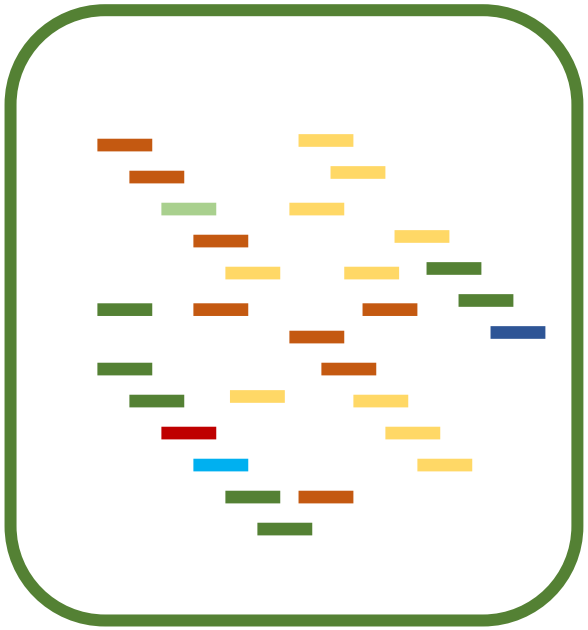
Plasmids database



~ 6954 plasmids

RefSeq database

CRISPR spacers database



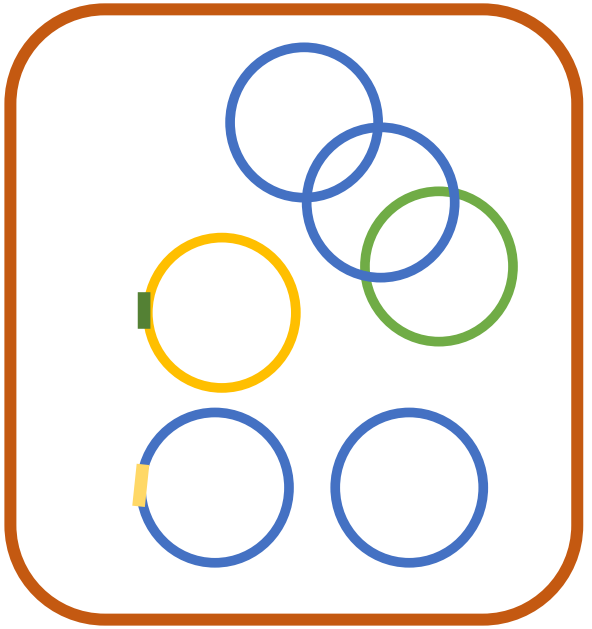
~ 141 705 spacers

CRISPR Finder database

BLAST

E-val $1x10^{-10}$
short blast

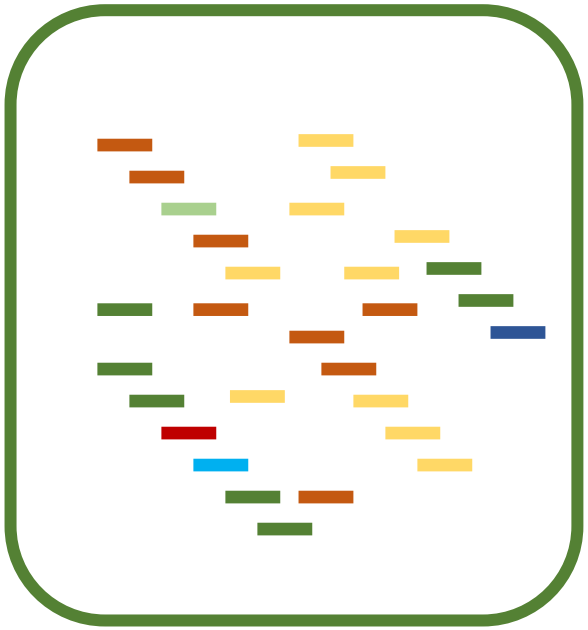
Plasmids database



~ 6954 plasmids

RefSeq database

CRISPR spacers database



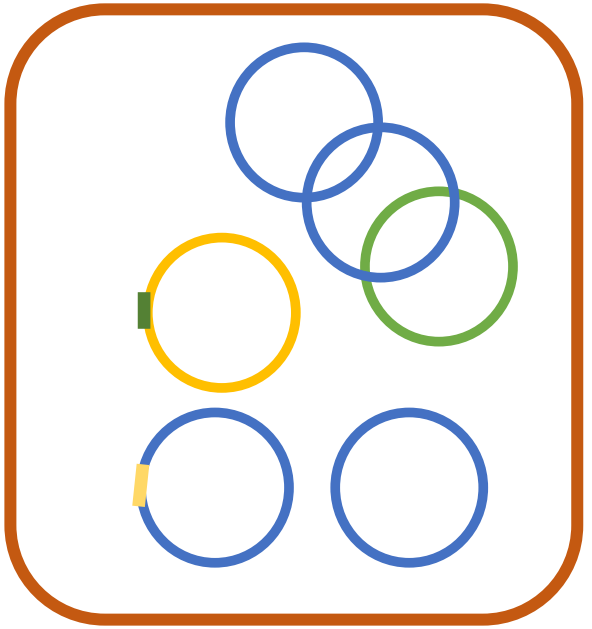
~ 141 705 spacers

CRISPR Finder database

BLAST

E-val $1x10^{-10}$
short blast

Plasmids database



~ 6954 plasmids

RefSeq database

141 705 spacers X 9000 plasmids → 13068 matches ☺

Do some plasmids have several hosts?

	With default E value	With E value 1E -10	With E value 1E -11
family	448	92	35
order	297	56	23
class	244	35	15
phylum	111	2	2

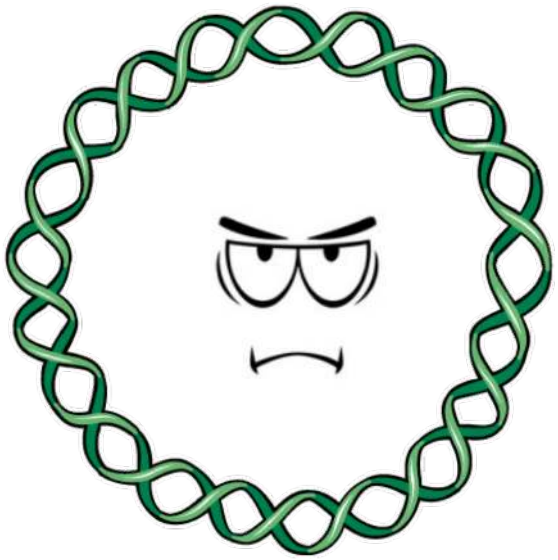
Do some plasmids have several hosts?

	With default E value	With E value 1E -10	With E value 1E -11
family	448	92	35
order	297	56	23
class	244	35	15
phylum	111	2	2

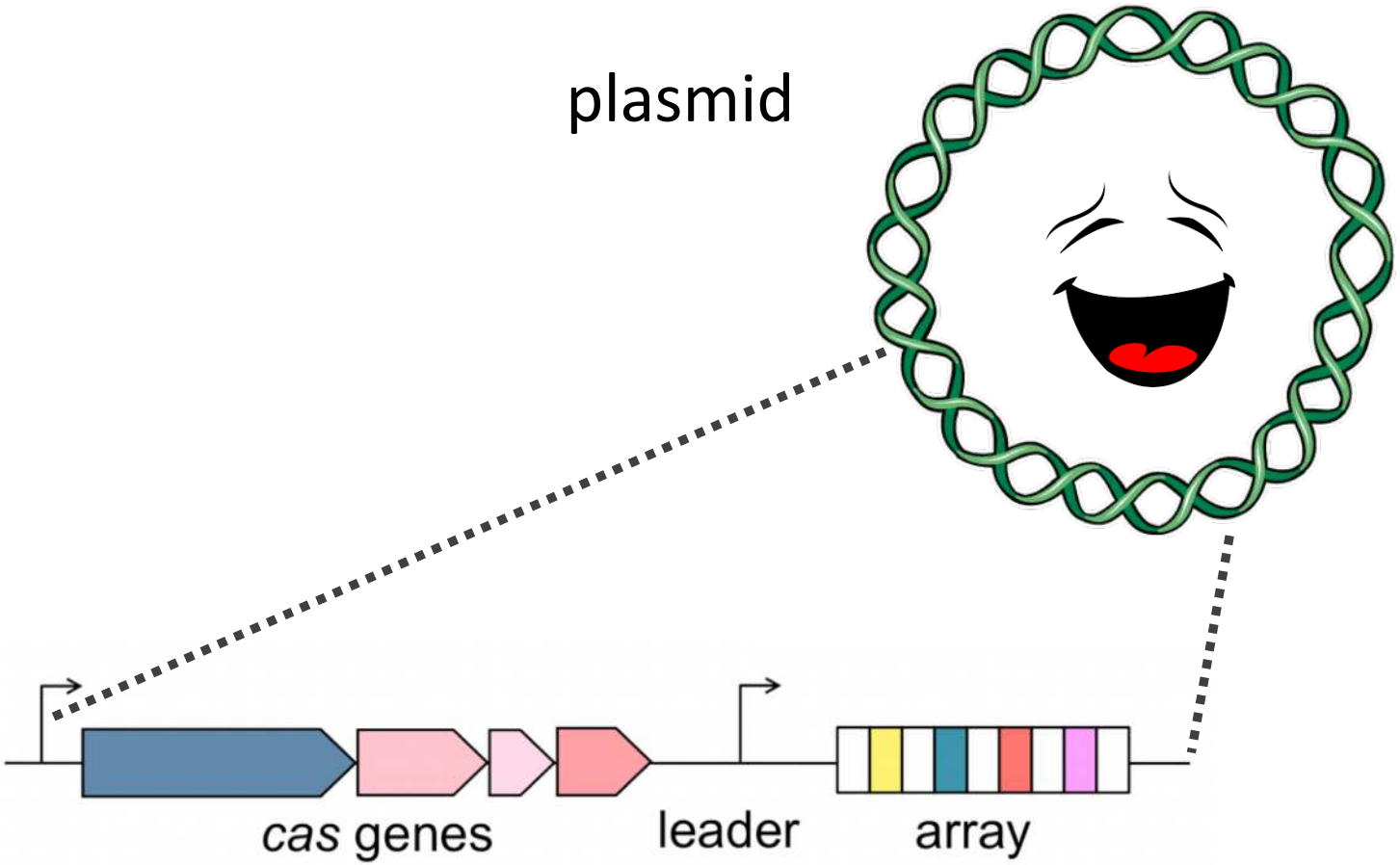
Yes, it is !!!

Is there any plasmid CRISPR Cas systems?

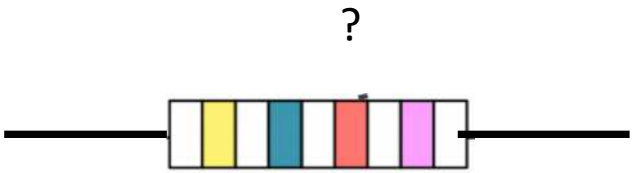
plasmid



Is there any plasmid CRISPR Cas systems?



Plasmid CRISPR cassettes search



CRISPR cassette

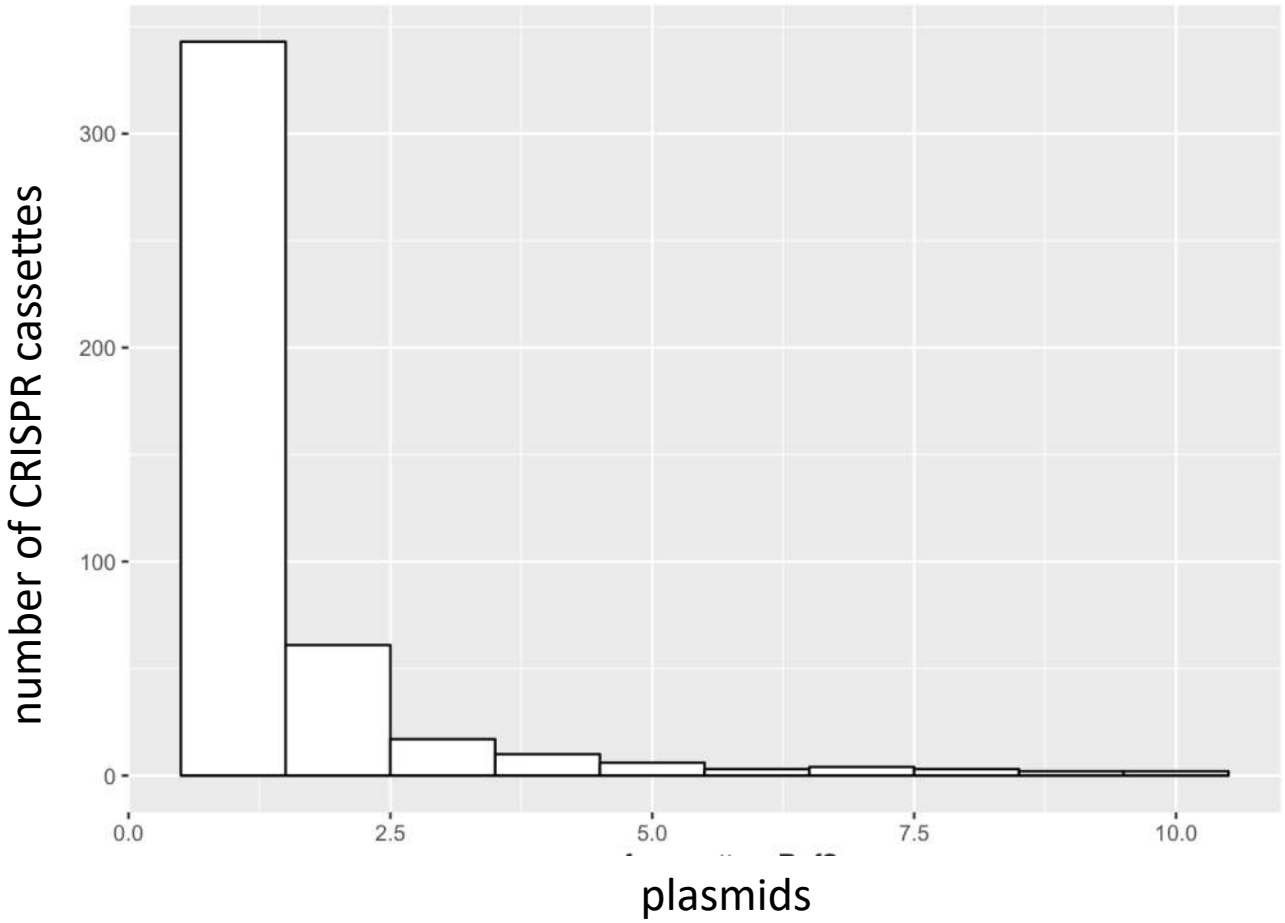
CRT CRISPR recognition tool:

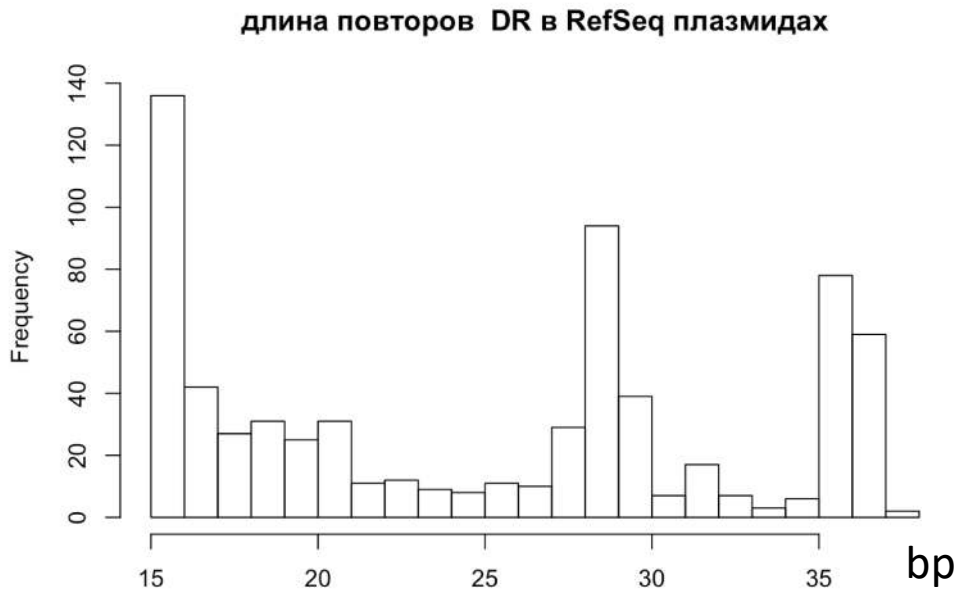
min 3 repeats

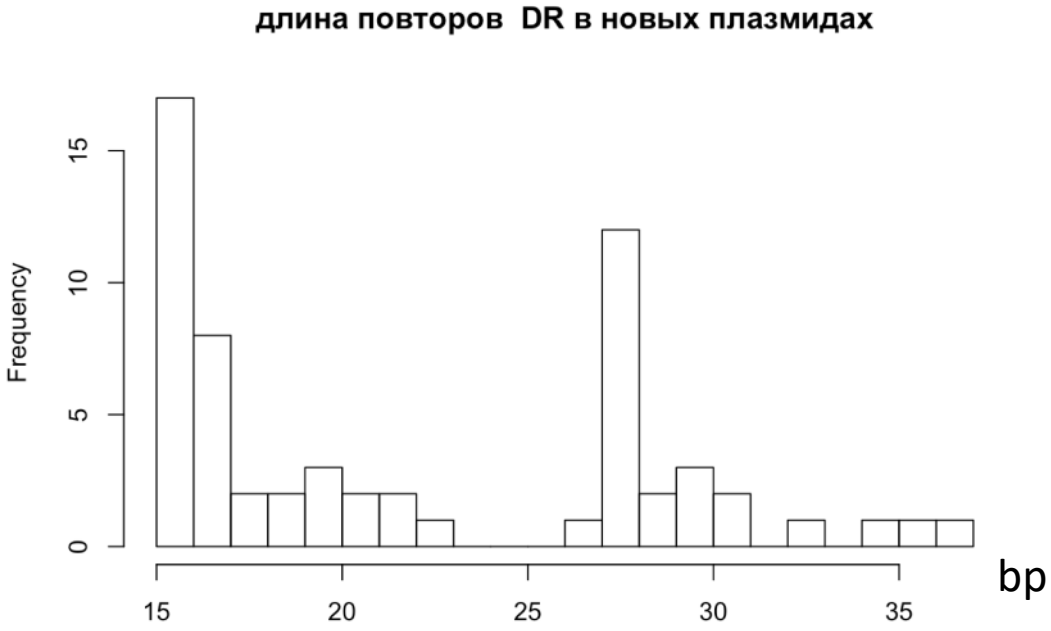
min repeat length 15 bp

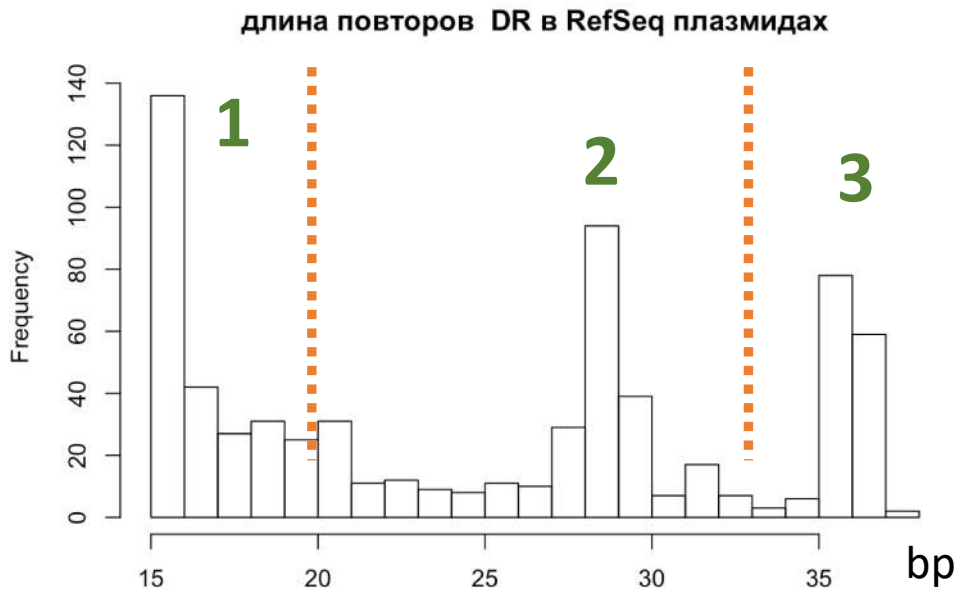
	RefSeq plasmids	New plasmids
Total number of plasmids	6954	2033
plasmids with CRISPR cassettes	451	55
Total number of CRISPR cassettes	695	62

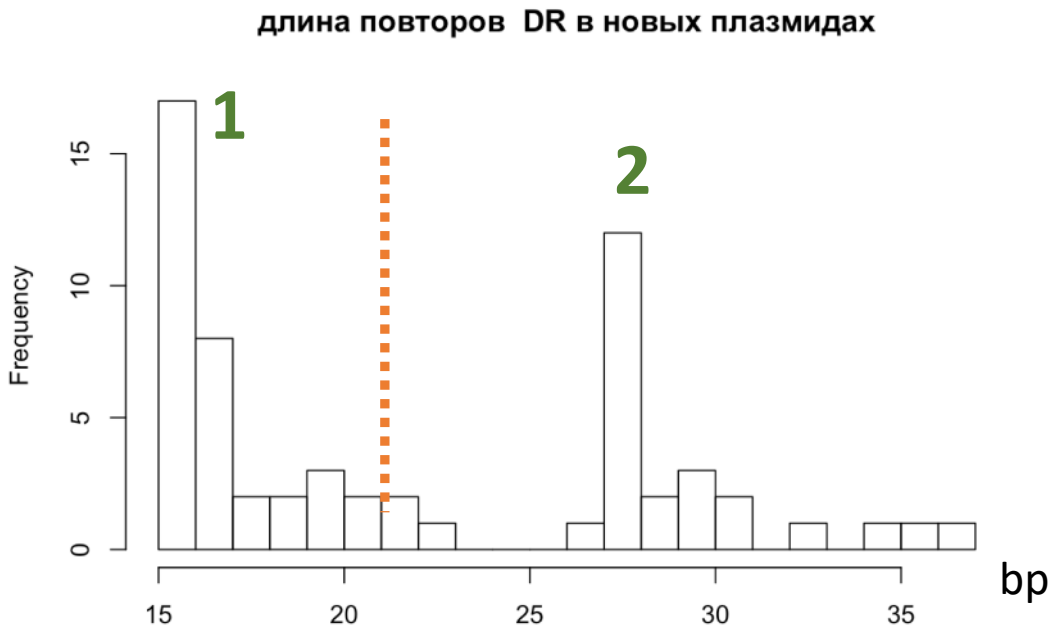
In most cases one plasmid carries only one CRISPR cassette



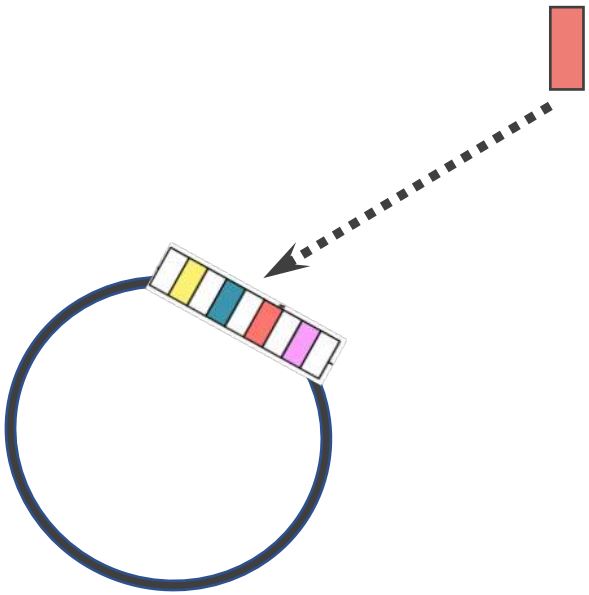

длина повторов DR в RefSeq плазмидах
Frequency
0
20
40
60
80
100
120
140
15
20
25
30
35
bp


длина повторов DR в новых плазмидах
Frequency
0
5
10
15
15
20
25
30
35
bp

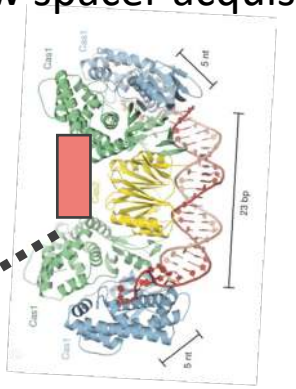

длина повторов DR в RefSeq плазмидах
Frequency
0
20
40
60
80
100
120
140
1
2
3
15
20
25
30
35
bp

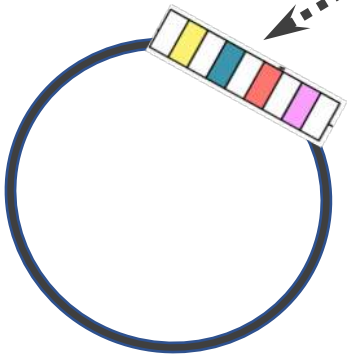

длина повторов DR в новых плазмидах
Frequency
0
5
10
15
1
2
15
20
25
30
35
bp

New spacer acquisition

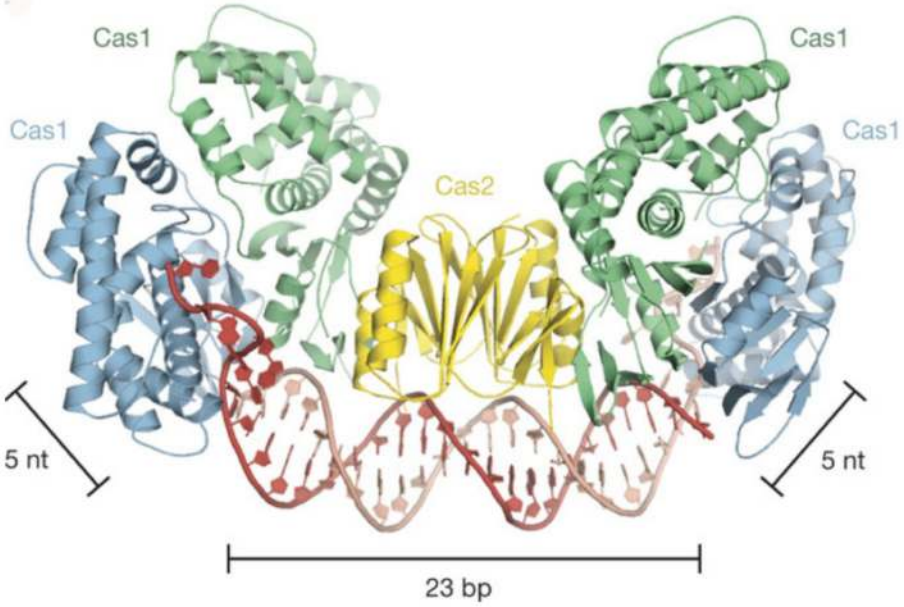


New spacer acquisition

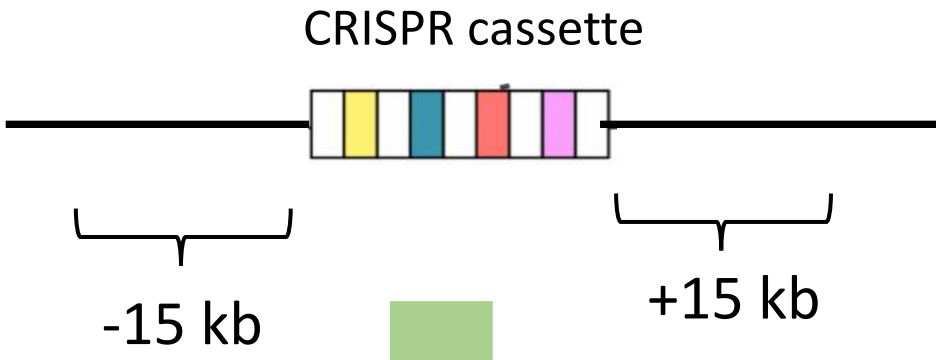




Cas1 – very conservative protein
Mark of putative active CRISPR Cas system



Let's search for **cas1** gene



ORF prediction

prodigal

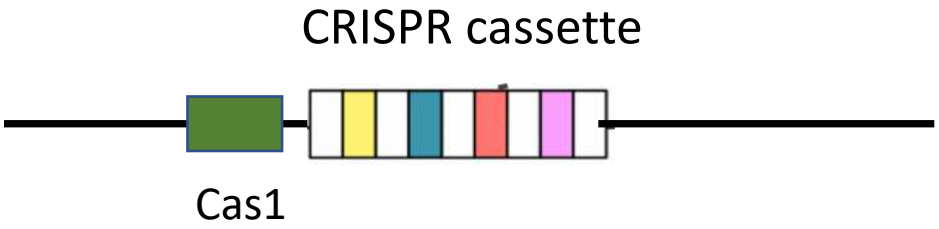
HMMER search by **Cas1 Pfam profile**

HMMER

	RefSeq plasmids	New plasmids
Total number of plasmids	6954	2033
plasmids with CRISPR cassettes	451	55
Total number of CRISPR cassettes	695	62
ORF	13687	1606
Cas1 hits	70	0

1/10 part of these CRISPR Cas systems could be active!

Are these CRISPR Cas systems active? – We don't know yet!



Are these CRISPR Cas systems active? – We don't know yet!

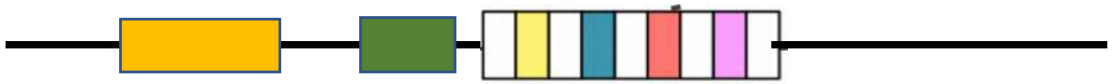
?

Host Cas proteins

?

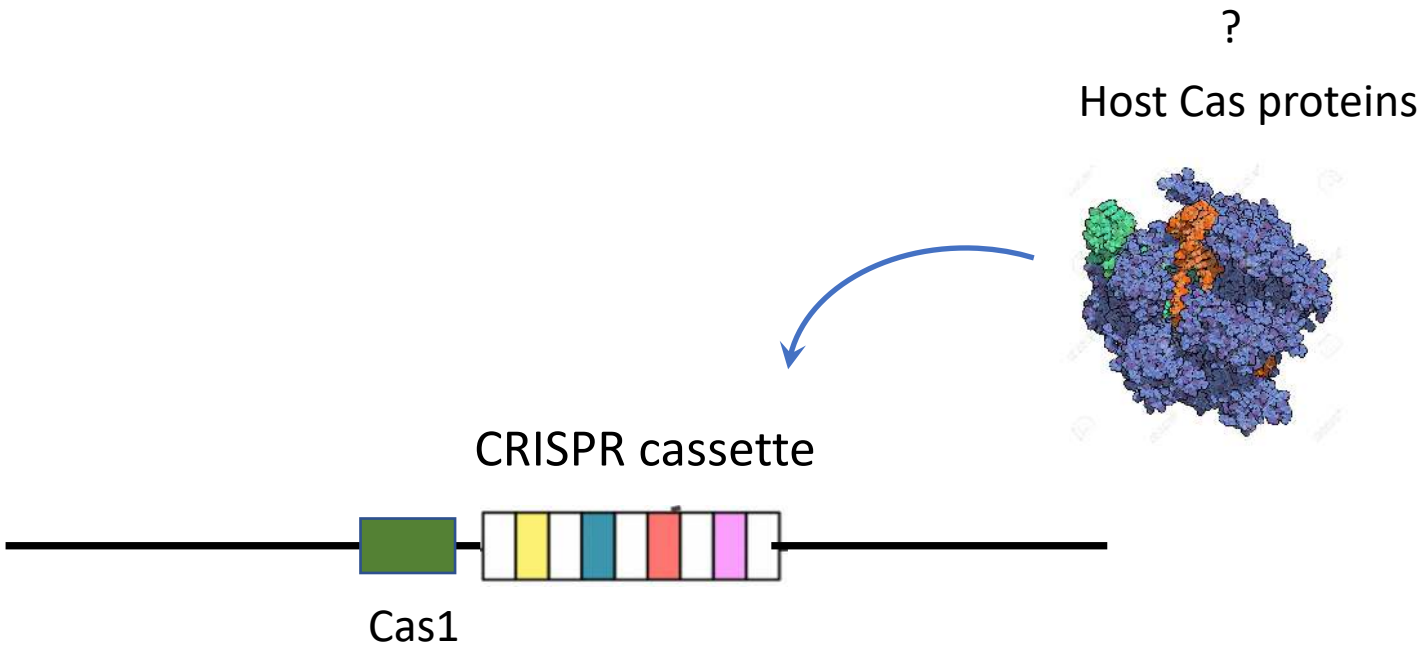
Cas effector genes

CRISPR cassette

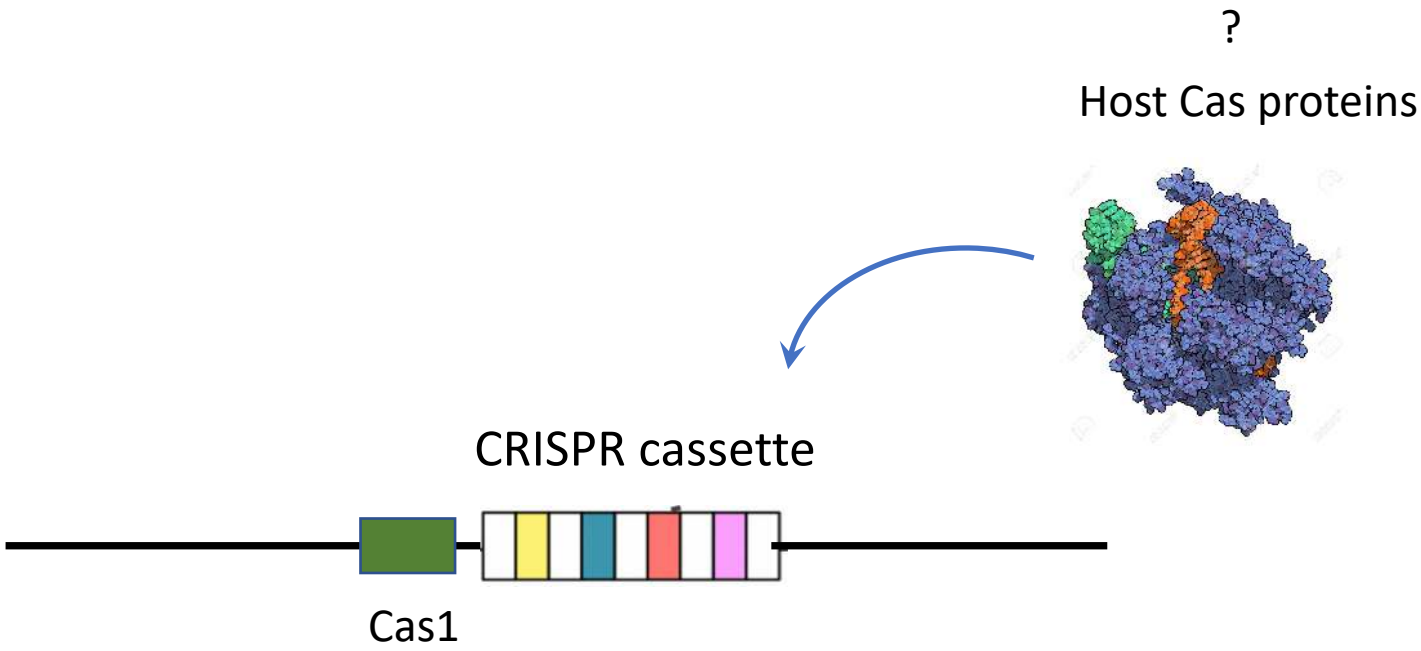


Cas1

Are these CRISPR Cas systems active? – We don't know yet!



Are these CRISPR Cas systems active? – We don't know yet!



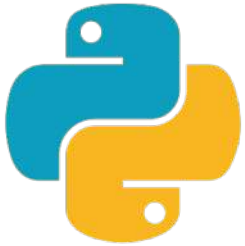
Further research – pending…!

Results summary

- Aligned plasmids to spacers to define host
- Defined plasmids host range => found plasmids which have several hosts

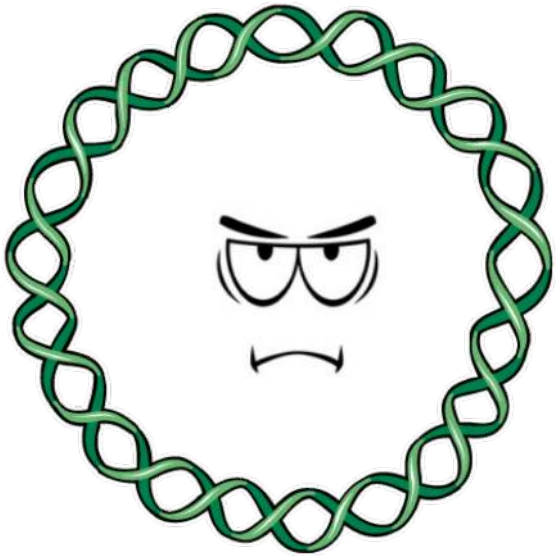
- Found and analyzed CRISPR cassettes on today known plasmids
- Identified Cas1 proteins – important part of active systems

During our work we used:



prodigal

CRT



Спасибо за внимание!