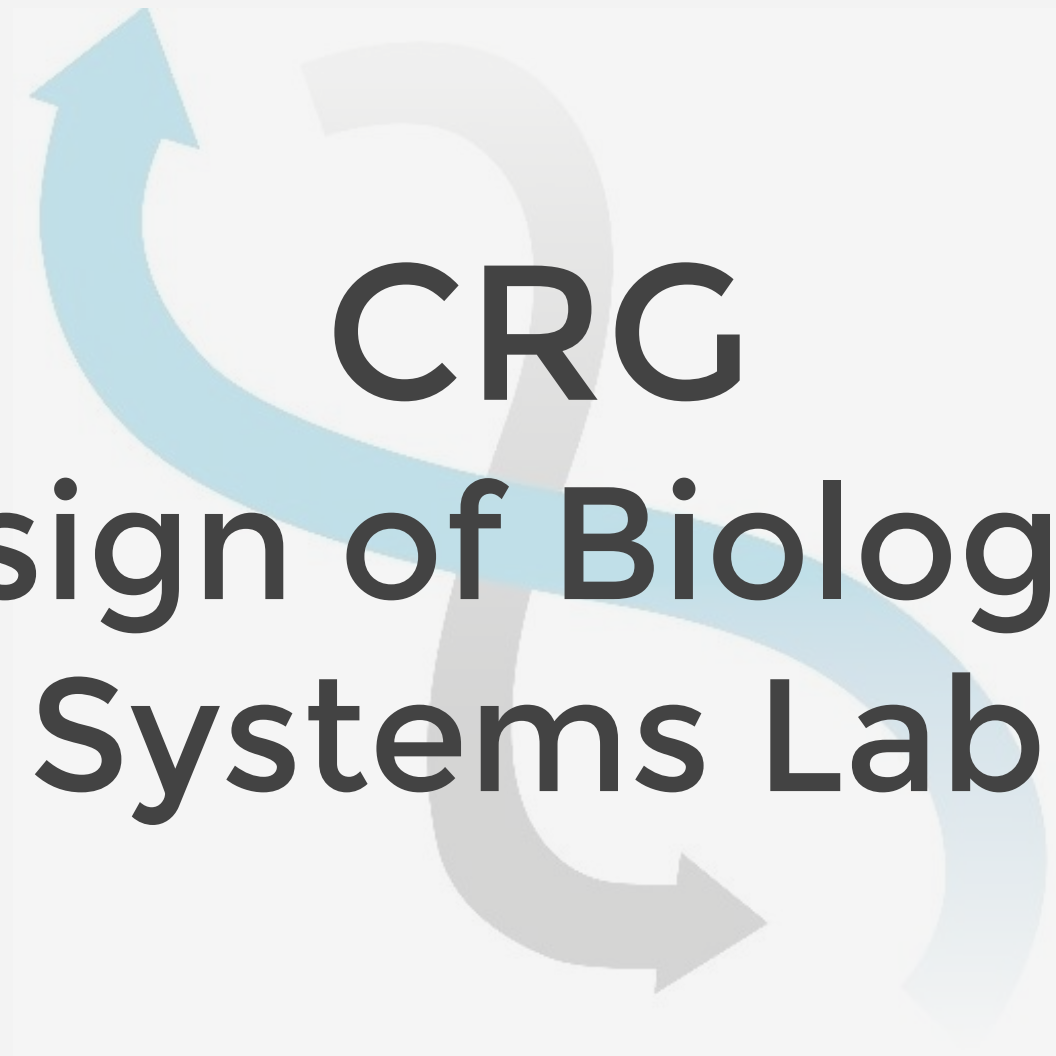


CRISpy & CRISPapp

A general python based algorithm for
CRISPR/Cas9 20-mers design presented in a Web
Application



CRG
Design of Biological
Systems Lab

Samuel Miravet Verde

Bioinformatics MSc Student - Pompeu Fabra University



Index



INTRODUCTION

MATERIAL & METHODS

RESULTS

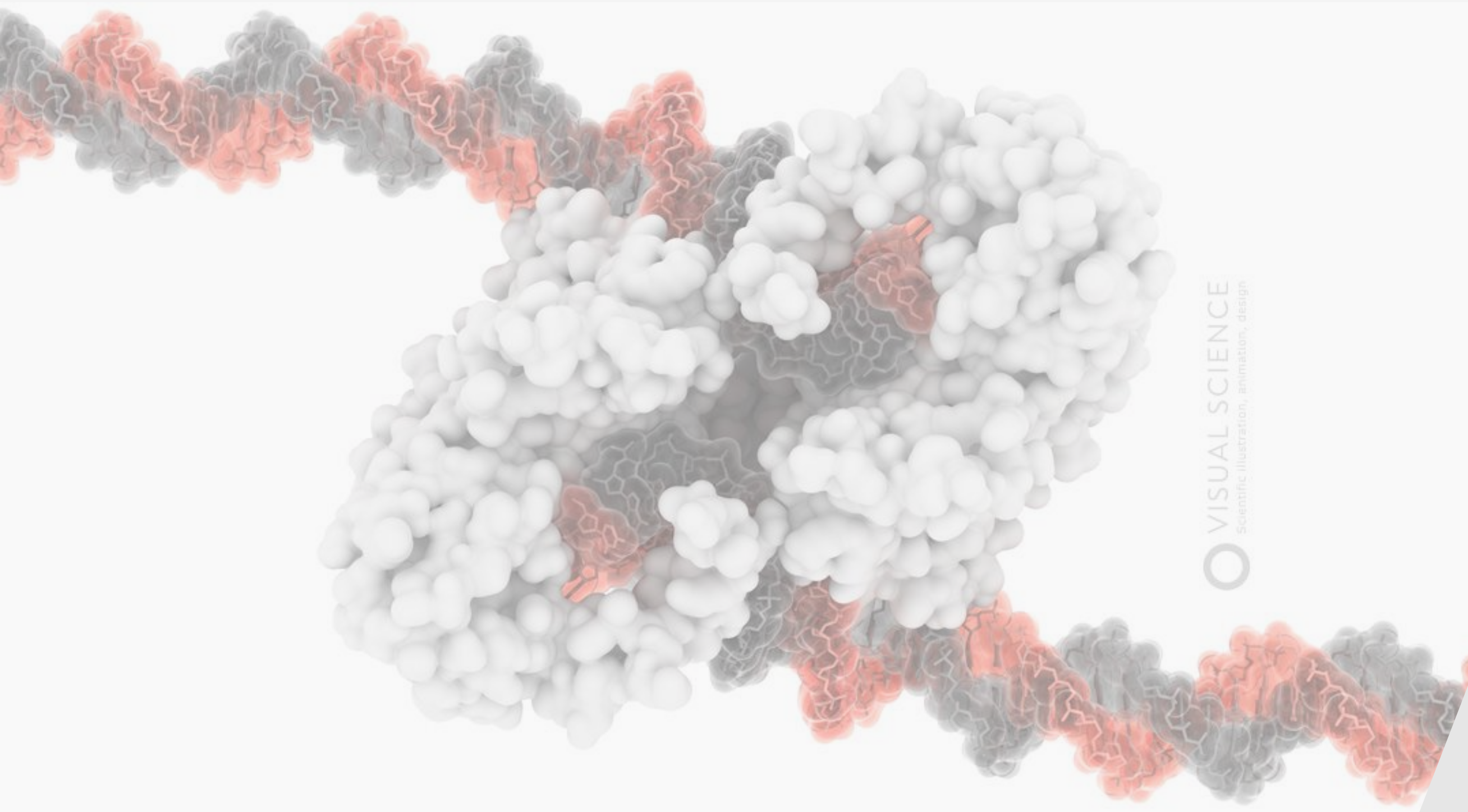
FURTHER IMPROVEMENTS

Introduction

Mycoplasma pneumoniae has **no repair machinery** → CRISPR/Cas9 **fails**

SOLUTION: **Transposons** allow time to succes

69994 InsTh41
237001 InsTh7



Project goals

- ★ **Library of oligos** (guided RNAs) + information about **genes** and **transposons**
- ★ **User-friendly** selection and comparison under a **WebApp**
- ★ **General implementation** applicable for organisms without repair machinery





Materials & Methods

Only 2 inputs required

```
1 >gi|13507739|ref|NC_000912.1| Mycoplasma pneumoniae M129 chromosome, complete genome
2 TATTTACCGACGAAATTAATACCATCAGGGTATTAAGATGCTACCAACGTGGTATTAATGTGCCCAAC
3 CGCGAAAAAGAAAGTGGTATATAGGAAAATGGAAAACAAATTTGTTTGCCGTTTACTATTAAATACTAAT
4 CTTCTATATAGTATAGAGAACTTTTTCTTTAACATAATATTATCTTAATATTATTTACCTACTAATAGC
5 TTAATATTATTAGTATTTATTTAGTATTATGCTAATACTATGCAGATATTATCTTAATATTATCTATAGT
6 ATTAGGCTAATATTATTCTTAATATTTATTAAGGTACTAAAGCATTACCTATAGGTGATATTATGACAAT
7 ACTAAAGTGGTTAGTATTATTAGGGTATTATTCAAAGTATTCTCCAACACTATTCCCTTAGCTCTAACTT
8 AATATCCCTTTAATACCAAGATGGTTAAACAATGAAGAAAATTAGTTGTTTGTTCTGGCCCTTTACTATA
9 TATGTACATAGTACATATATAAGAAAGGCTACCATGACATTATATATAATAGGCATCCAGAAATAAAAGC
10 CAAATTAATCTATAATTAAGTATTTATACTTAATTATAGTTATTTTATTAACAAAATTAGTGACATTTGG
11 CCATCCATAATAATAAAAATCAAATAAACAAAAACAAAAAGCAAAAAAACTATAAAAAGGGCATGAAAGTT
12 TTGATTAATAAGAATGAGTTGAACAAAATCCTCAAAAAACTCAACAATGTAATCGTATCTAACAATAAGA
13 TGAAACCATAACCACTCTTATTTATTAATAGAGGCTACAGAAAAGGAAATTAACCTTCTATGCTAACAACGA
14 GTACTTTTCTGCTAAATGTACCTTAGCCGAAAACATTGATGTACTTGAAGAAGGTGAAGTAATTGTTAAA
15 GGCAAAATCTTTAGCGAACTCATTAACGGAATTAAAGAAGACATCATTACTATTCAAGAGAAAGATCAAA
16 CTCTTTTAGTCAAAACAAAAAAAACAAACATTAACCTTAACACGATTGATAAGAAAGAATTCCCCAGAAT
17 CCGTTTCAACCAAAACGTTGATTTGAAGGAATTTGATGAACTTAAAATCCAACACAGCCTTTTAACTAAA
18 GGACTTAAAAAGATTGCCCATGCTGTTTCTACATTTAGAGAATCCACTAGAAAATTCAACGGGGTTAACT
19 TCAACGGTTCCAATGGTAAACAAATCTTTTATAGAGGCATCGGATTCTTATAAGCTCTCTGTTTATGAAAT
20 CAAACAAAAAACCGATCCATTTAATTTTATTGTCGAACTAATCTTTTGAGCTTCATCAATTCTTTTAAC
21 CCTGAAGGTGGTGATTTAATCAGTATCTTCTTCCGCAAAGAACACAAGGATGATTTAAGTACCGAATTAC
22 TGATTAAGTTAGATAACTTCTTAATTAACCTCAATTAACGAAAGCTTCCGCGGGTAATGCAGTT
```

50	InsTh41	59
51		/label="InsTh41_1"
52		/label=InsTh41_1
53	InsTh7	59
54		/label="InsTh7_2"
55		/label=InsTh7_2
56	InsTh7	62
57		/label="InsTh7_3"
58		/label=InsTh7_3
59	InsTh7	70
60		/label="InsTh7_4"
61		/label=InsTh7_4
62	InsTh7	75
63		/label="InsTh7_5"
64		/label=InsTh7_5

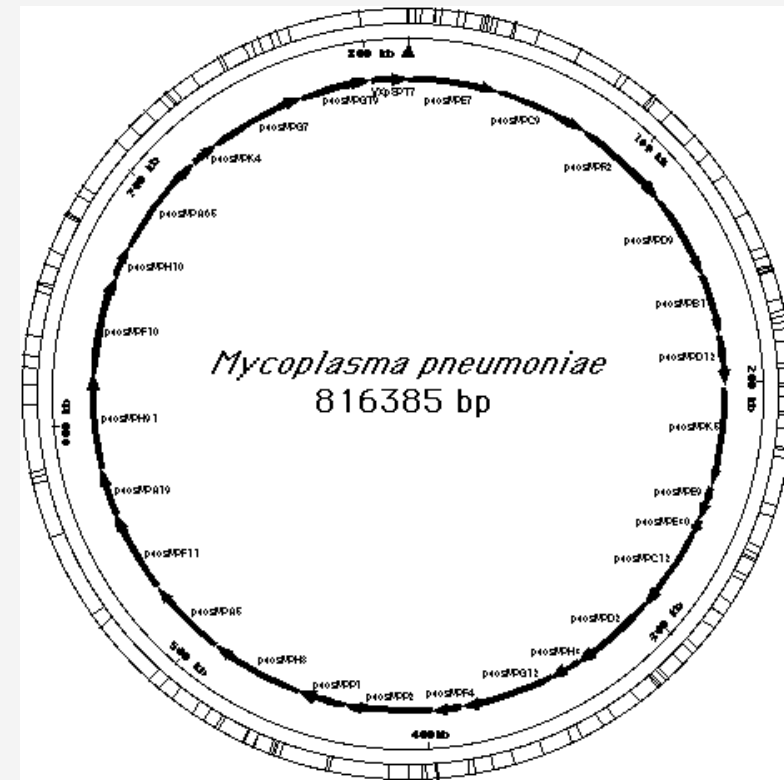
Transposons in genbank file

Target genome Sequence

Information already available for X organisms including: *Helicobacter pylori*, *Bacillus subtilis* and *Streptococcus pneumoniae*



Algorithm



NGG / CCN
Pattern
Matching

73516
PAMs +
oligos

Filtering

65674
unique
oligos

2521
multiple
oligos

oligos
start-
end

Intersection

genes

Intersection

InsTh41/7
(before and
after 17th
position)

Web application

Developer - friendly → **HTML5** + **Flask** (Python based Web framework)

- Data processing, requests and data downloading

User - friendly → **Javascript**

- Easy table manipulation and data exploration





Results for *Mycoplasma pneumonia*

65674 unique oligos:

- Specific genome editing with CRISPR/Cas9
- Information about genes they will interrupt
- Quantitative information about transposons in the oligo and their position

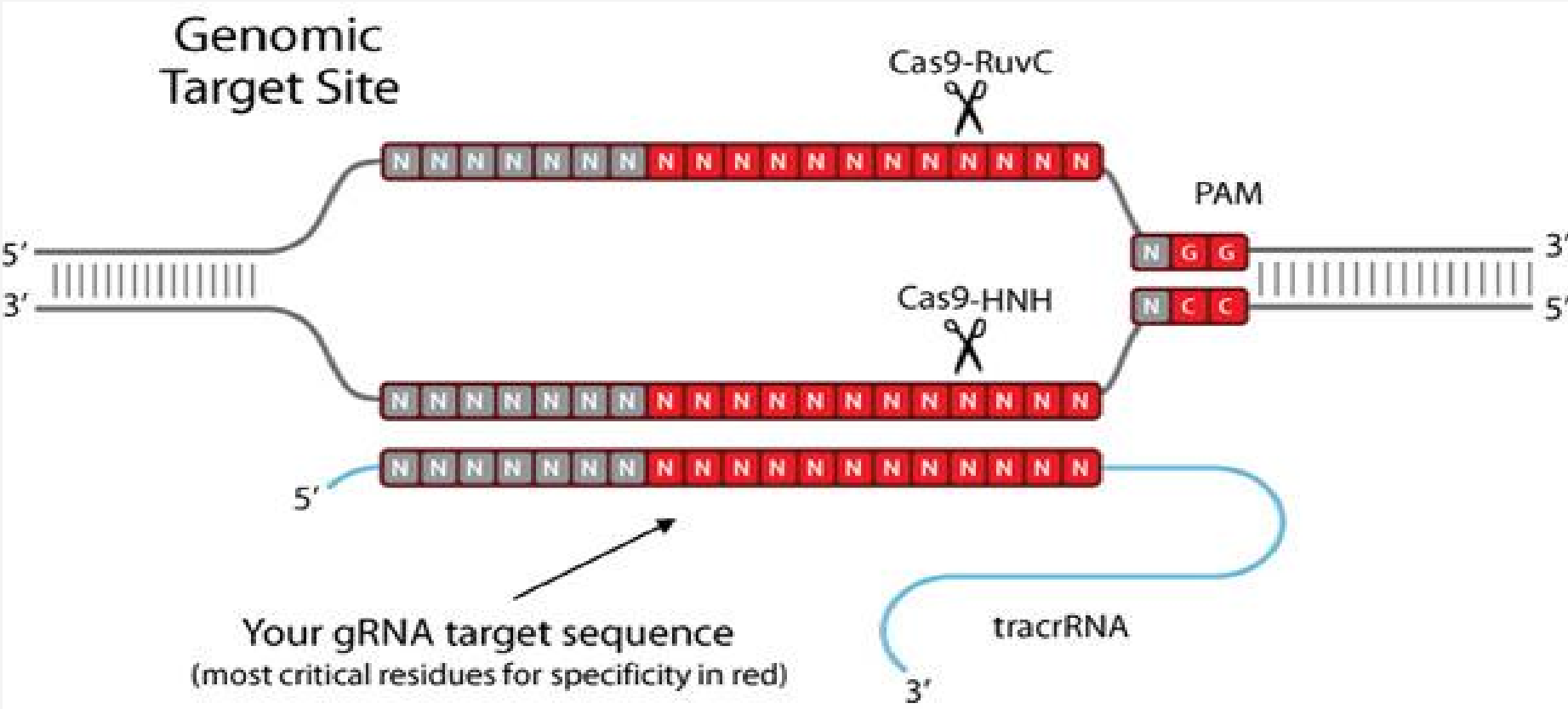


Figure from Sigma-Aldrich. Copyright © 2015 Sigma-Aldrich Co. LLC. All Rights Reserved.

200319	ATTCAAAATATAGTGCAAAA	MPN152	MPN152	2	0	2	11	2	13	15
200432	TTTTCTTTATCAATCTTCAA	MPN152	MPN152	6	1	7	15	3	18	25
200579	GCATCGTTACTATTGTCAGT	MPN152	MPN152	3	1	4	12	2	14	18
200614	CCCAAGTCTTTTCTATCGCC	MPN152	MPN152	2	0	2	10	2	12	14
200676	TTCAAACACACTGGCTTTGA	MPN152	MPN152	5	1	6	10	1	11	17
200685	GTGTCAGTTTTCAAACACAC	MPN152	MPN152	4	1	5	10	2	12	17
200766	AATATCGGTGCCCTTTTAT	MPN152	MPN152	4	0	4	11	1	12	16
200781	TCCCAGAACACCTTGAATAT	MPN152	MPN152	4	0	4	9	1	10	14
200816	ACAGAGGTAAATAAGGAGTT	MPN152	MPN152	4	0	4	12	2	14	18

Results for *Mycoplasma pneumoniae*

2521 multi-oligos:

- GCTTAGTTCGGCGCTTAGTT - 15 times in the genome with a PAM associated → repeated regions
- ACAATTAAACTTAACAAAC - 9 times along the genome.



Possible biosecurity system with application in Synthetic (not exclusively) Biology?

Results for *Mycoplasma pneumoniae*

Web Application:

<http://mycoplasma.tk/crispr/>

Crispr App Main Page

Bugs & Questions [here](#)

Select column to filter

pos bef PAM

Minimum value

10000

Maximum value

15000

Filter/Clear

Filter

Clear

Download Filtered Table

Showing rows 1-500 rows of a total of 320.

First Rows

Previous 500

Next 500

Last Rows

Click on a column name to order the table by this column

pos_bef_PAM	subsequence	genename	41_bef17	41_aft	total41	7_bef17	7_aft17	total7	total
14968	AAATTCCTCCTTTCTGCTCA	-	6	3	9	15	3	18	27
14998	TCCTTTTGGGAAATAACTTT	MPN013	4	3	7	13	3	16	23
13858	AGTTGAACTATAACTTTAAT	MPN012	6	1	7	11	3	14	21
14302	AAAATTTACCATTACCAAAA	-	5	2	7	9	3	12	19
13853	AACTATAACTTTAATTGGCA	MPN012	5	1	6	11	2	13	19

The web easily allows to:

- **Filter** by genome position / number of transposons
- **Order** by any of the columns
- **Explore** the results
- **Download** a .csv table with them

Ex: 14000 - 15000

Implementation results

Algorithm:

- **Oligo design** following a pattern for any sequence
- You can add **positional information** from any genbank file

Web App:

- Static table **editable** and **interactive**

- Implementation applicable in other organisms with/without repair machinery to design oligosequences for CRISPR/Cas9 or others
- Find oligos binding more than 1 region to kill the organism
- Results easily exposed in an interactive web... it just require to change the input table to show




Further Improvements



Further Improvements

- Add “**Flag**” values representing the **specificity** of each oligo
 - Random **Mutation** in different regions of the 20-mer
- **Filtering** of result by **gene name** interruption



**Thank you very
much for your
attention**

CRISpy

A general python based algorithm for
CRISPR/Cas9 20-mers design presented in a Web
Application



CRG

Design of Biological
Systems Laboratory

Samuel Miravet Verde

Bioinformatics MSc Student - Pompeu Fabra University