Skoltech

Skolkovo Institute of Science and Technology

MSc Program

Evaluation of CRISPR spacers diversity in *Clostridium difficile* clinical isolates

Student: *Polina Muzyukina*

Research Advisor: *Ekaterina Savitskaya*

Co-Advisor *: Konstantin Severinov*

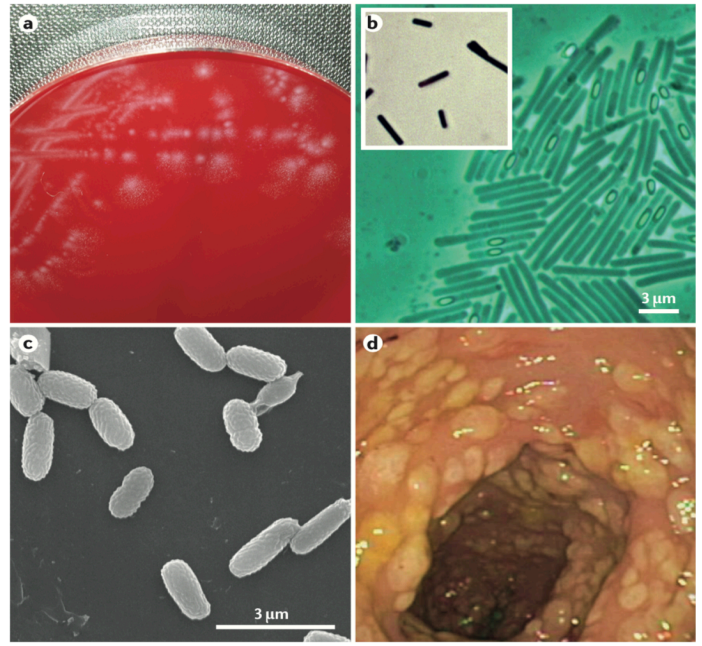
Center of Life Sciences

June, 2019

General problem

Clostridium difficile

- Gram-positive spore-forming anaerobic enteropathogen
- Cause of hospital-acquired diarrhea, pseudomembranous colitis, perforation of the colon, sepsis
- Toxin producing
- Possess multiple antibiotic resistance



Smits at al., 2016

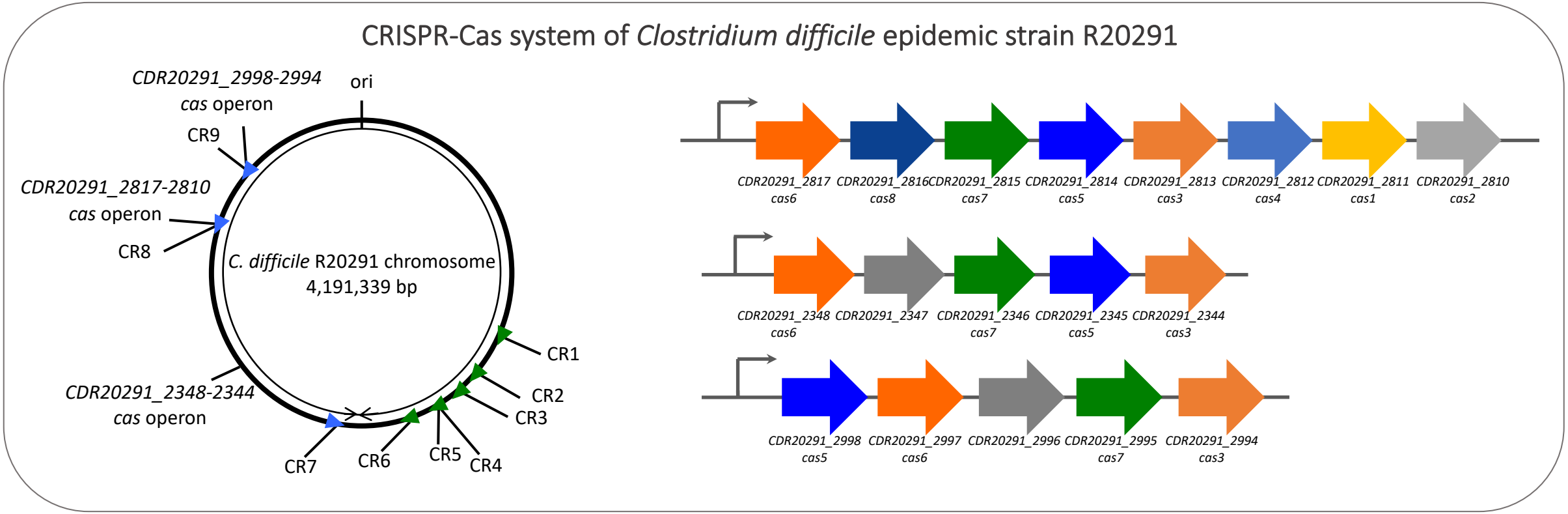
Clostridium difficile infections bacteriophage treatment

- Efficient treatment of *Clostridium difficile* infection by phage cocktails in mammalian models (Ramesh et al., 2016, Nale et al., 2018)

But…

Highly active in CRISPR-Cas defense system

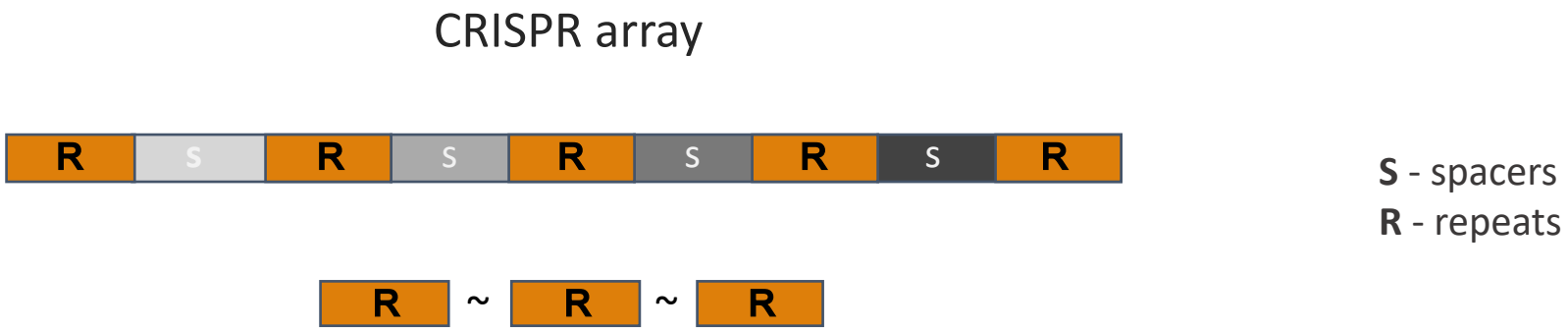
- CRISPR-Cas I-B type
- Abundance of CRISPR arrays (the average 8 array/genome)
- Conserved 5’ PAM motif CCT/A and 8nt seed region



Boudry et al., 2015. (modified)

CRISPR array composition

- Spacers determine bacterial immunity to phages and plasmids
- Repeats partially share sequences similarity



Project aim

To develop the procedure for fast assessment of CRISPR spacers repertoires in clinical samples of *Clostridium difficile*-infected people to predict phages that would not be subject to CRISPR-Cas immunity and thus could be used for therapy

Repeat-based spacer amplification procedure

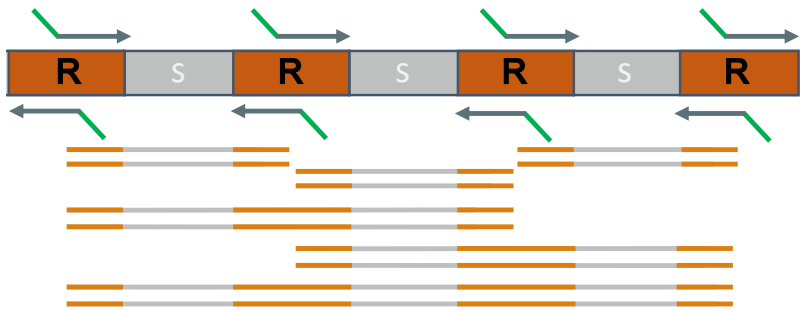
Primer for repeat — Illumina Adapter

Repeats

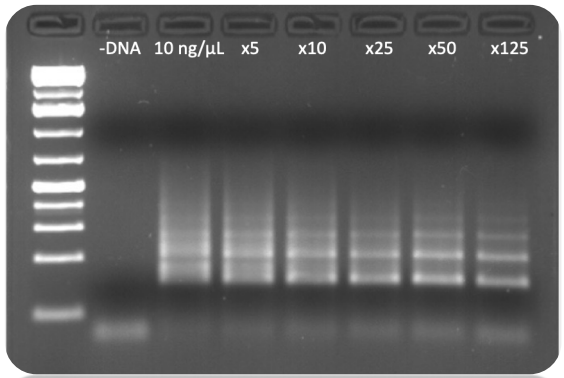
GTTTTATATTAACTAAGTGGTATGTAAAG
GTTTTATATTAACTAAGTGGTATGTAAAT
GTTTTATATTAACTATATGGAATGTAAAT
GTTTTAGATTAACTATATGGAATGTAAAT
GTTTTATATTAACTATATGGAATGTAAAT
GTTTTATATTAACTATGTGGTATGTAAAT
GTTTTATATTAACTATATGGAATGTAAAG
GTTTTATATTAACTAAGTGGTATGTAAAT

1. Known *C.difficle* repeats alignment

2. Primers design based on consensus repeat sequence



3. Repeat-based spacer amplification



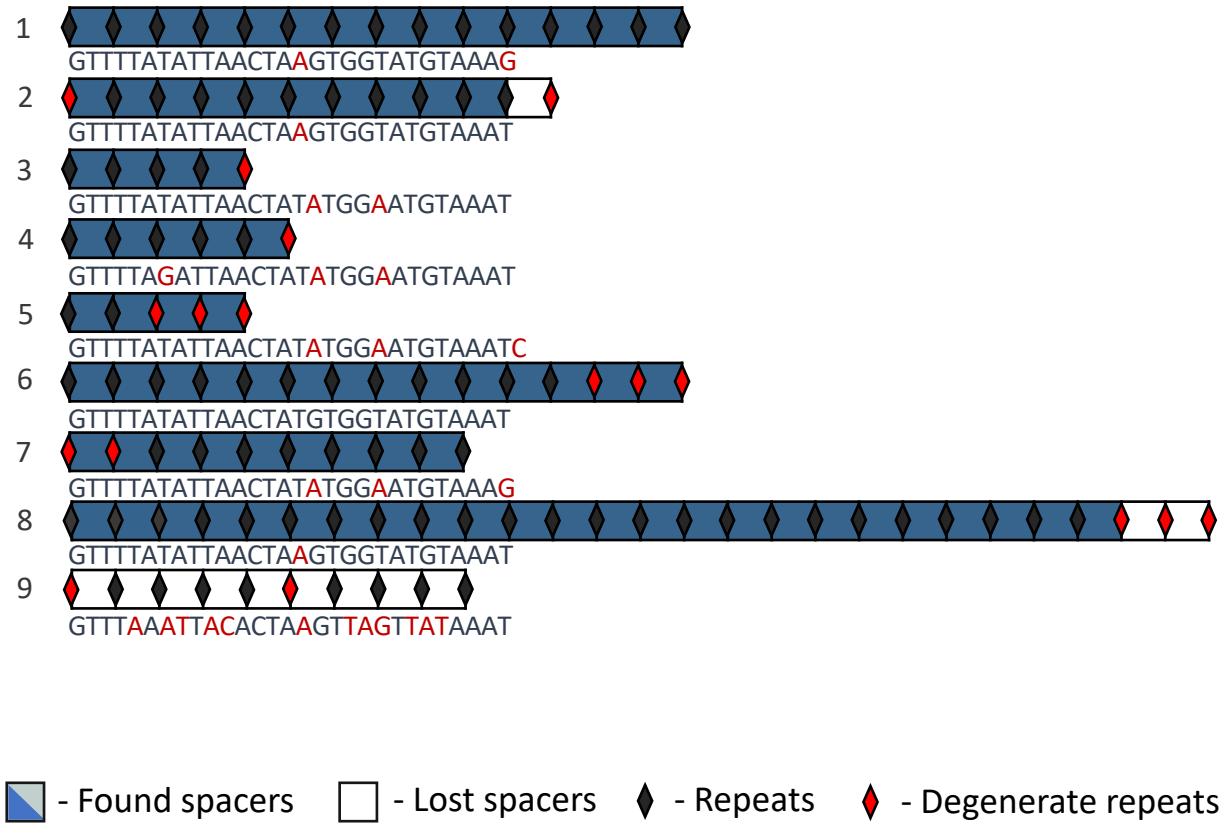
4. Gel extraction

5. NGS and data analysis

Validation of Repeat-based spacer amplification procedure on model strains

630 strain
GTTTTATATTAACTATATGGAATGTAAAT

R20291 strain
GTTTTATATTAACTATGTGGTATGTAAAT



- Found spacers
- Lost spacers
- Repeats
- Degenerate repeats

Validation of Repeat-based spacer amplification procedure on clinical isolates

Isolate	Number of spacers found by Repeat-based spacers amplification method	Number of spacers found by WGS	% of spacers found by Repeat-based spacers amplification method comparing to WGS	Phage genome hits of amplified spacers
CD77	64	77 (10 CRISPR arrays)	83%	6
CD79	95	100 (11 CRISPR arrays)	95%	19
CD210	144	168 (8 CRISPR arrays)	85,7%	25
CD211	81	107 (7 CRISPR arrays)	75,7%	12
CD221	100	105 (7 CRISPR arrays)	95,2%	16
CD648	120	162 (6 CRISPR arrays)	74%	27

Validation of Repeat-based spacer amplification procedure on stool samples

Stool sample ID	Number of spacers found by Repeat-based spacers amplification method	Phage genome hits of amplified spacers
CD25	79	11
CD27	362	51
CD42	181	18
CD43	90	9

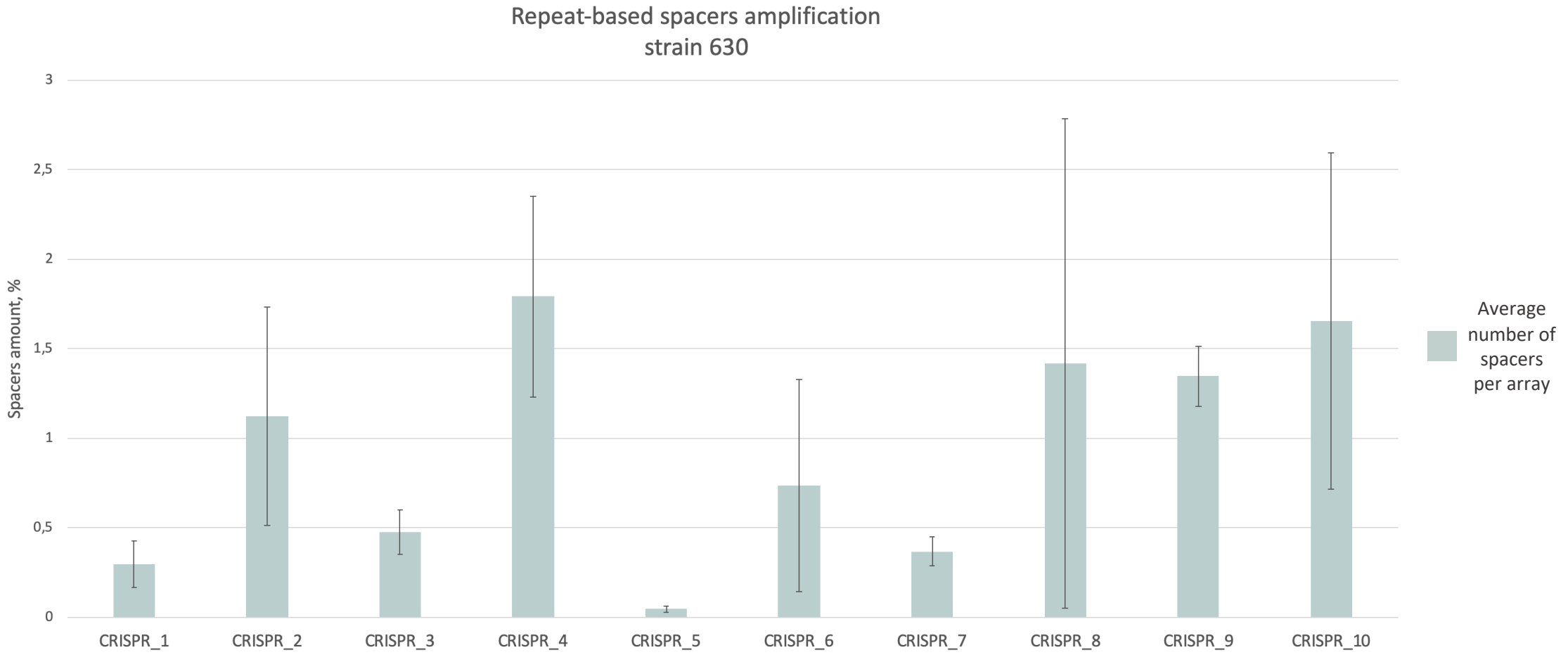
Problem:

C. difficile difficult to isolate and cultivate, it is extremely sensitive to even low levels of oxygen

Solution:

For the assessment of CRISPR spacers no difficult isolation and cultivation of *C. difficile* is needed

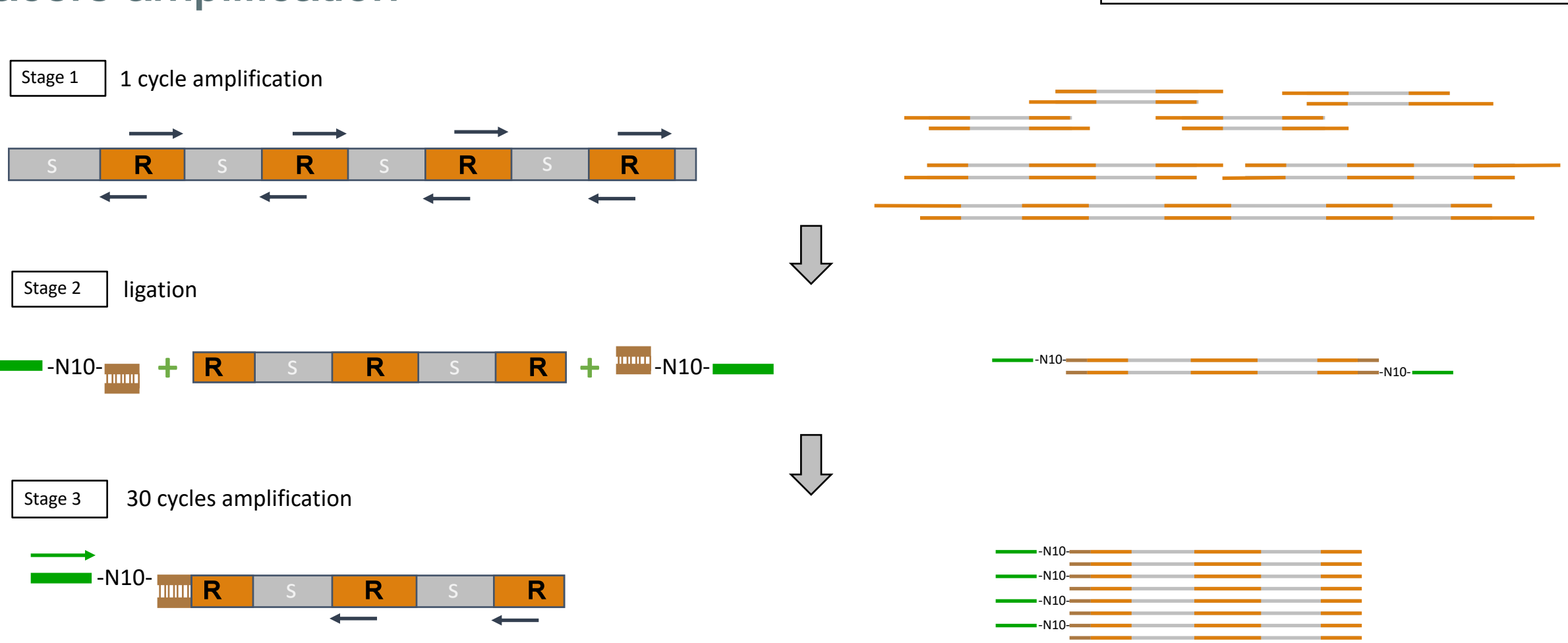
Spacers are sampled with unequal efficiency



Possible reasons of biased amplification:

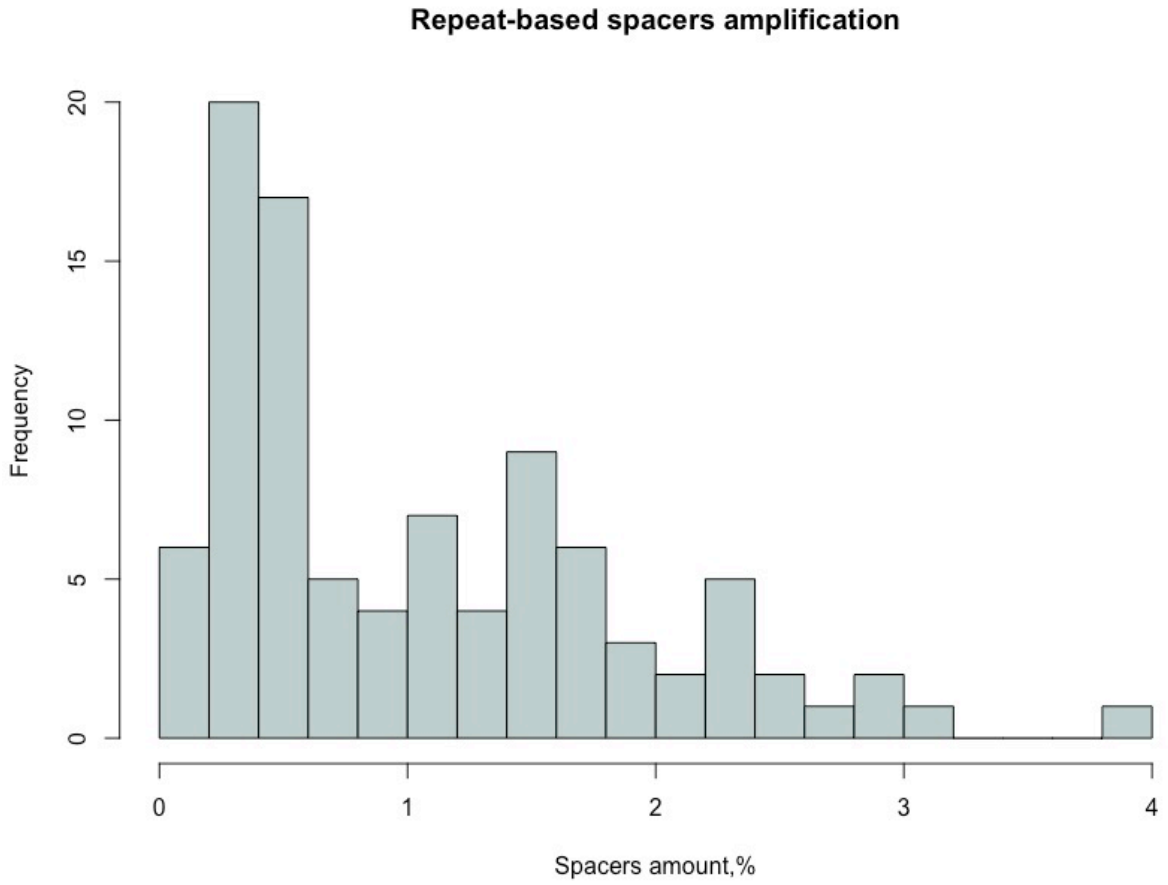
- Primers anneal unevenly due to repeats polymorphism
- Difference in nucleotide composition of spacers

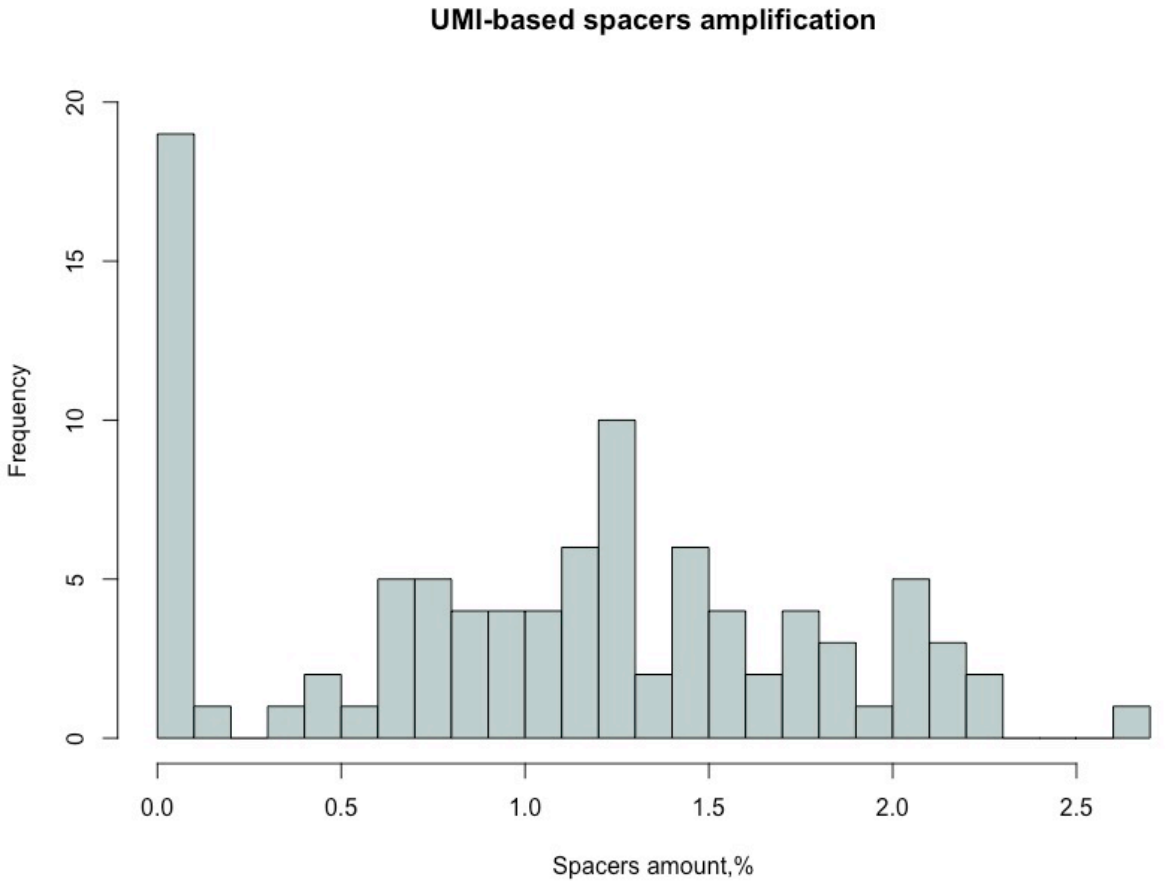
UMI (unique molecular identifiers) repeat-based spacers amplification



Data analysis: MIGEC software
Total number of reads: $13*10^6$
UMI number: $1*10^5$

Comparing the result of UMI Repeat-based spacers amplification

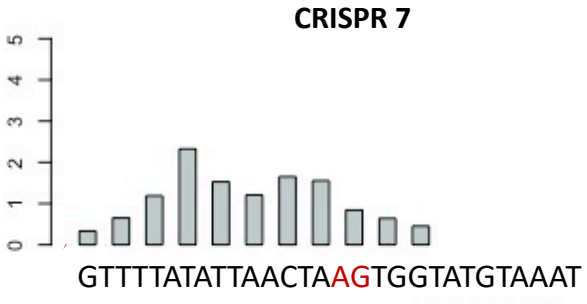


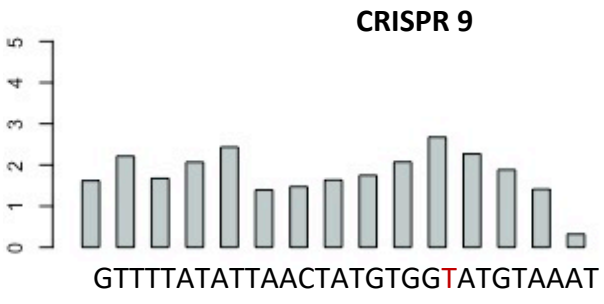


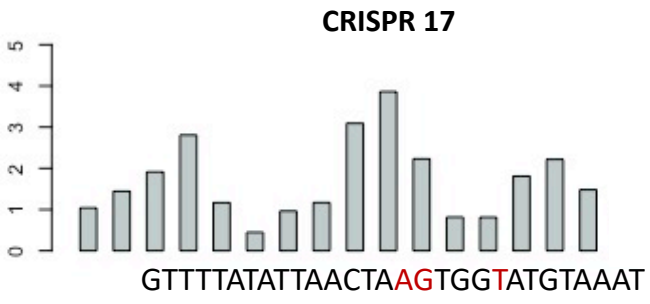
Comparing the result of UMI Repeat-based spacers amplification

Strain 630

Repeat-based spacers amplification







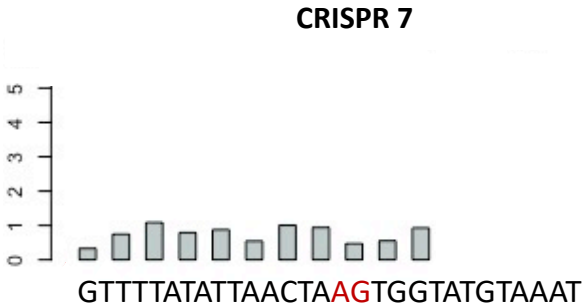
F-test

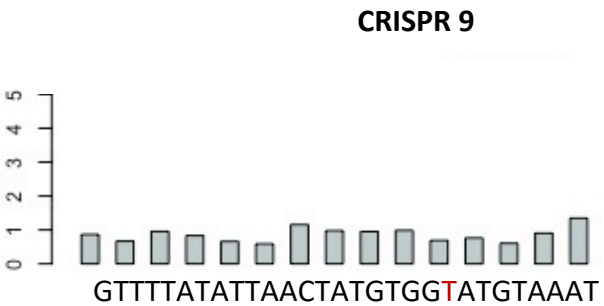
P-value=0.06

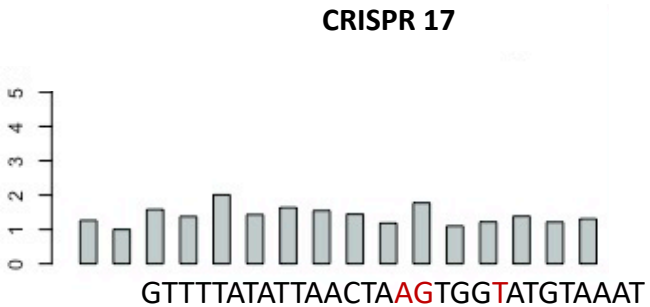
P-value=0.005

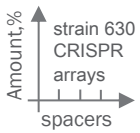
P-value=0.0001

UMI-based spacers amplification









Conclusions

- Repeat-based amplification of *C. difficile* spacers allows fast recovery of spacer associated with different versions of repeats
- Spacers flanked by repeats with more than 6 mismatches to the consensus repeat sequence tend to be lost during amplification
- Repeat-based spacers amplification procedure does not require preliminary labor-consuming strain isolation or laboratory enrichment
- UMI-based procedure does not affect amplification biases associated with repeat polymorphism, yet, it partially eliminates biases of unequal spacer representation **within** CRISPR array

Acknowledgments

- Dr. Ekaterina Savitskaya, Skoltech
- Prof. Konstantin Severinov, Skoltech
- Prof. Olga Soutourina, Paris-Saclay University, I2BC Institute
- Prof. Dmitry Chudakov, Skoltech
- Anton Rykachevskiy, PhD Skoltech
- Sonya Medvedeva, PhD Skoltech
- Anton Shkaruta, Ms Skoltech

In addition to CRISPR-Cas what are other bacterial defense mechanisms against phage infection that should be considered ?

In case of *Clostridium difficile*:

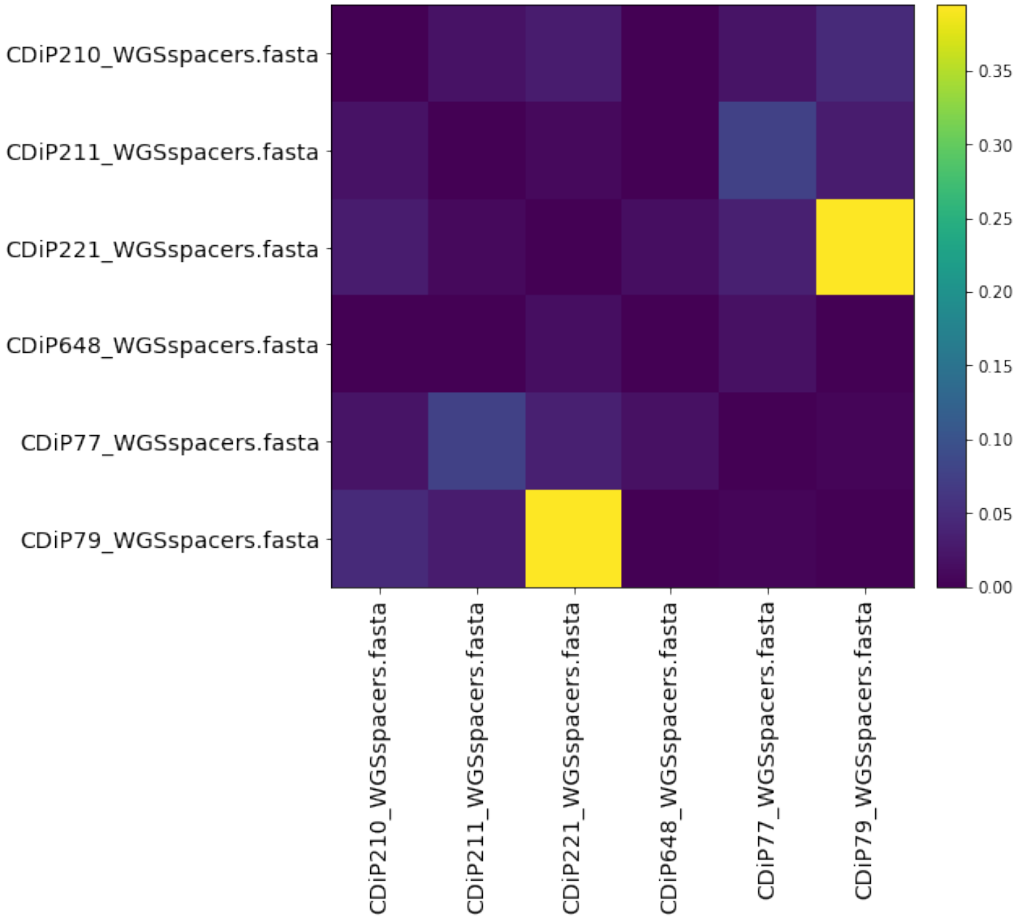
- CwpV - cell wall protein
- Prophages with CRISPR arrays

Other defense systems:

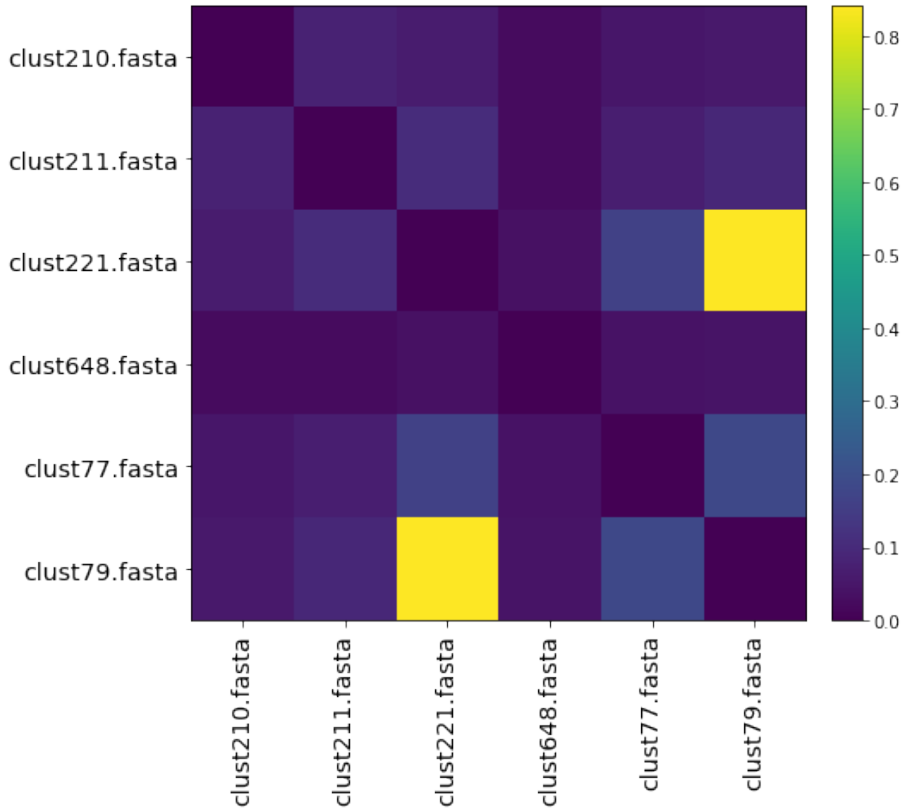
- Restriction-modification
- BREX
- Abi (abortive infection)
- Sie (superinfection)
- Receptors modifications

CRISPR spacers diversity in clinical isolates

A

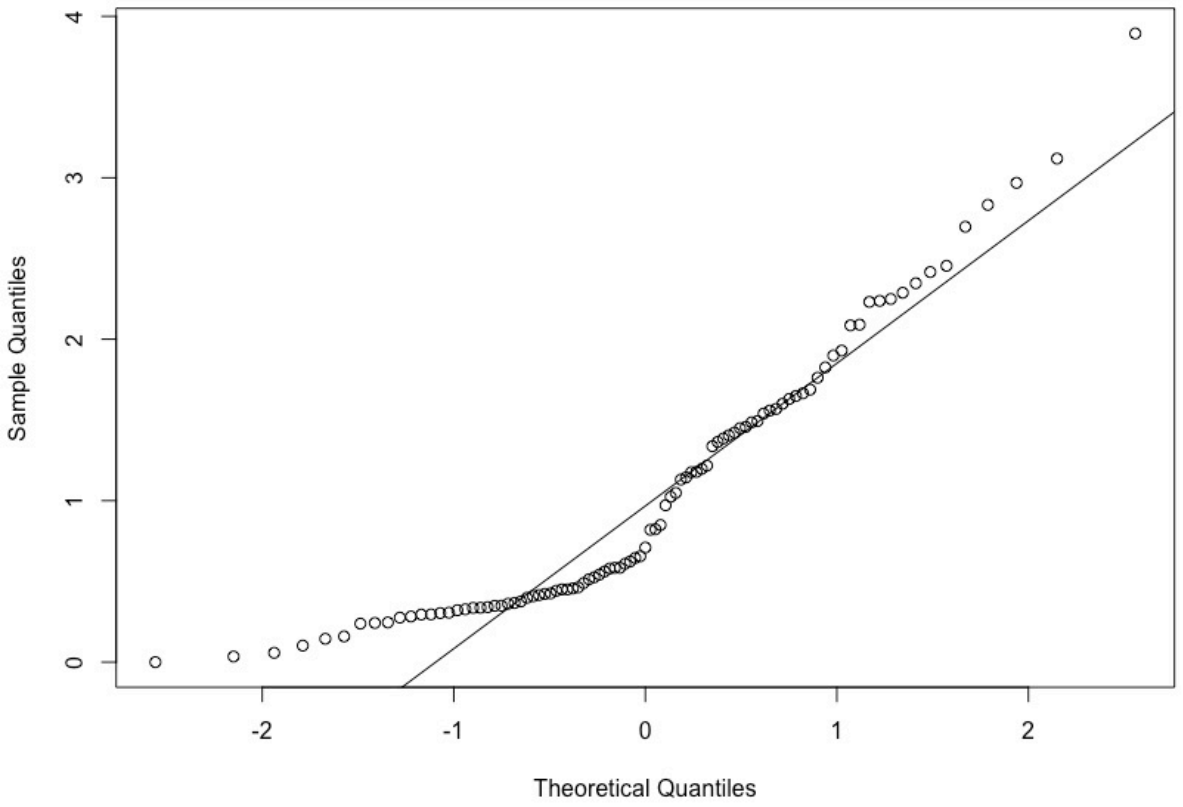


B

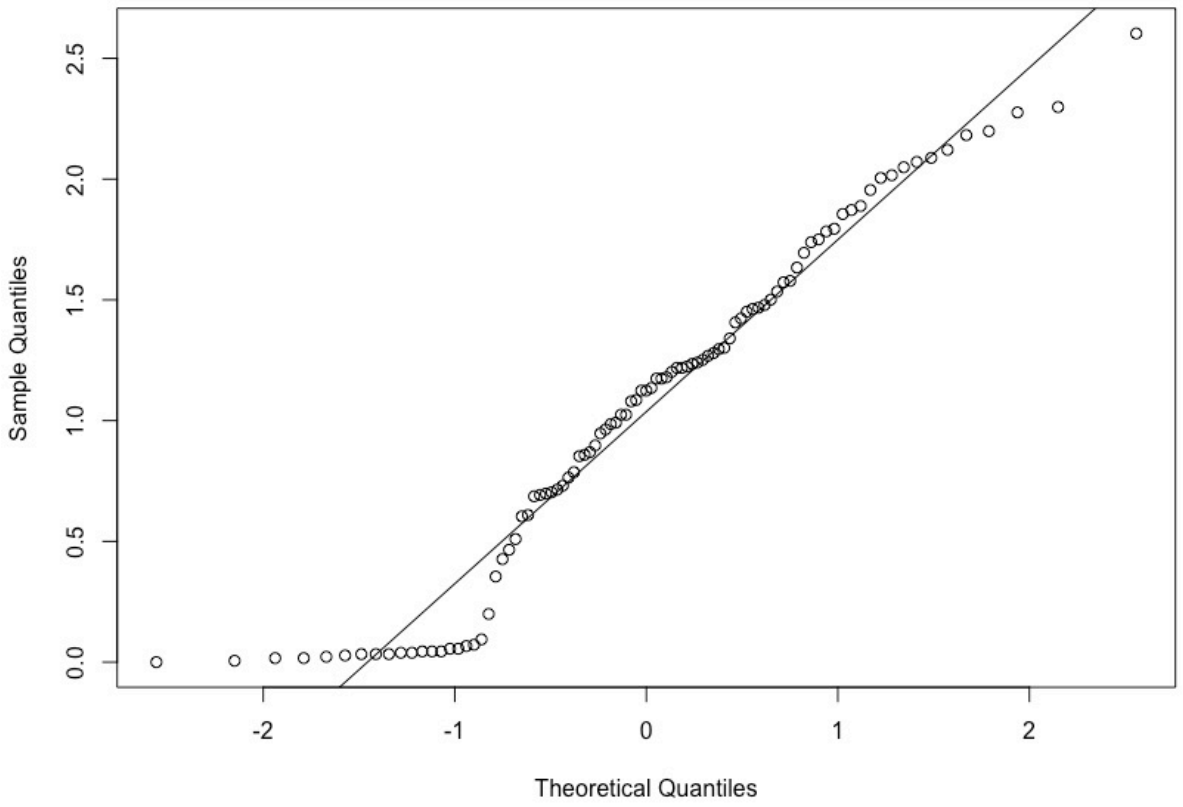


Q-Q plot

Repeat-based spacers amplification



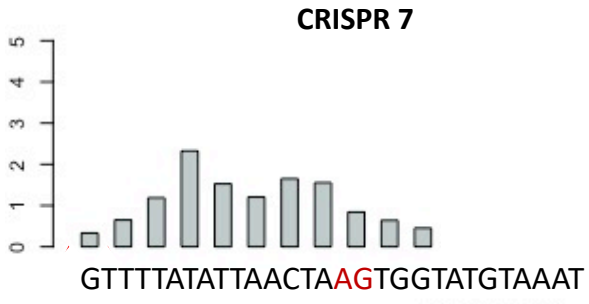
UMI-based spacers amplification

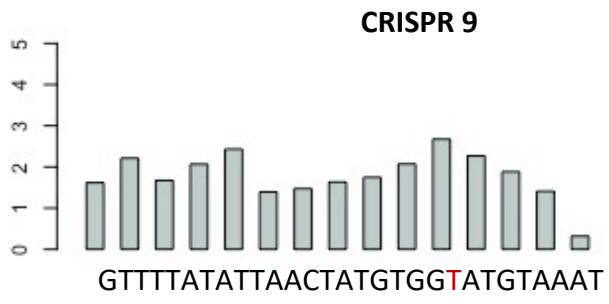


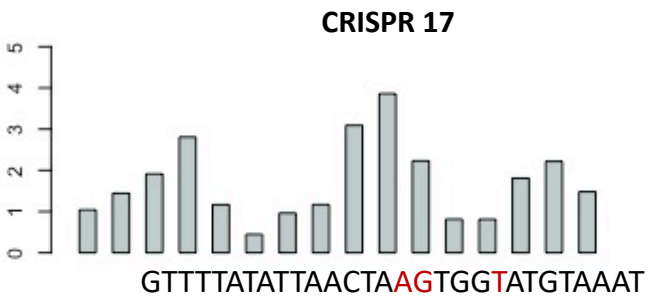
Comparing the result of UMI Repeat-based spacers amplification

Strain 630

Repeat-based spacers amplification







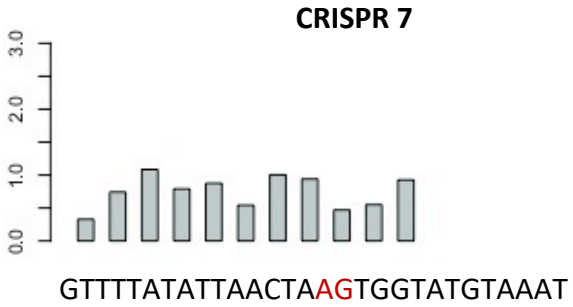
F-test

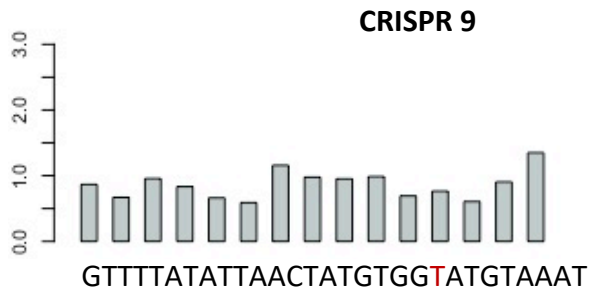
P-value=0.06

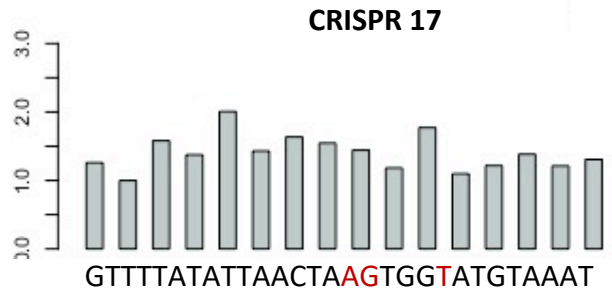
P-value=0.005

P-value=0.0001

UMI-based spacers amplification







UMI (unique molecular identifiers) repeat-based spacers amplification