Application of Machine Learning and Natural Language Processing for Phage Therapy 2.0

Piotr Tynecki

with Yana Minina, Iwona Świętochowska, Joanna Kazimierczak and Arkadiusz Guziński co-op

PyWaw, 18.05.2020

Who Am I?

SLAVIC AI

PyStok

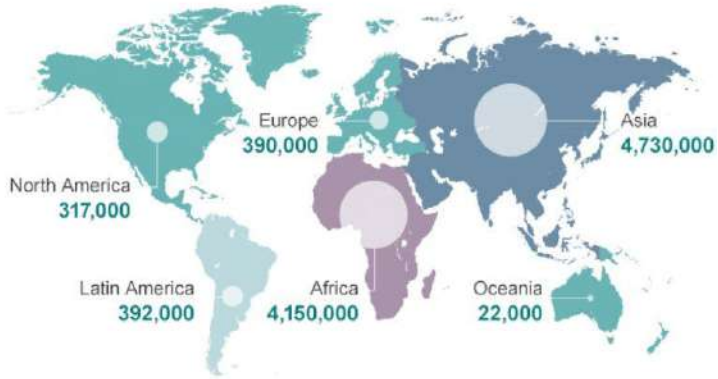
The future if we do not act now

GLOBAL A failure to address the problem of antibiotic resistance could result in:

10m deaths by 2050

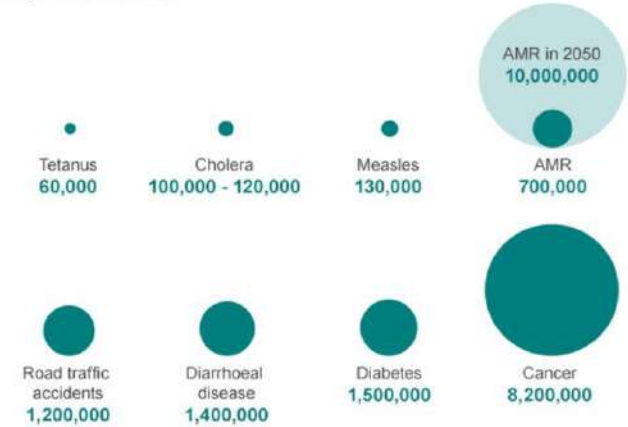
Costing £66 trillion

Deaths attributable to antimicrobial resistance every year by 2050



Source: Review on Antimicrobial Resistance 2014

Deaths attributable to antimicrobial resistance every year compared to other major causes of death



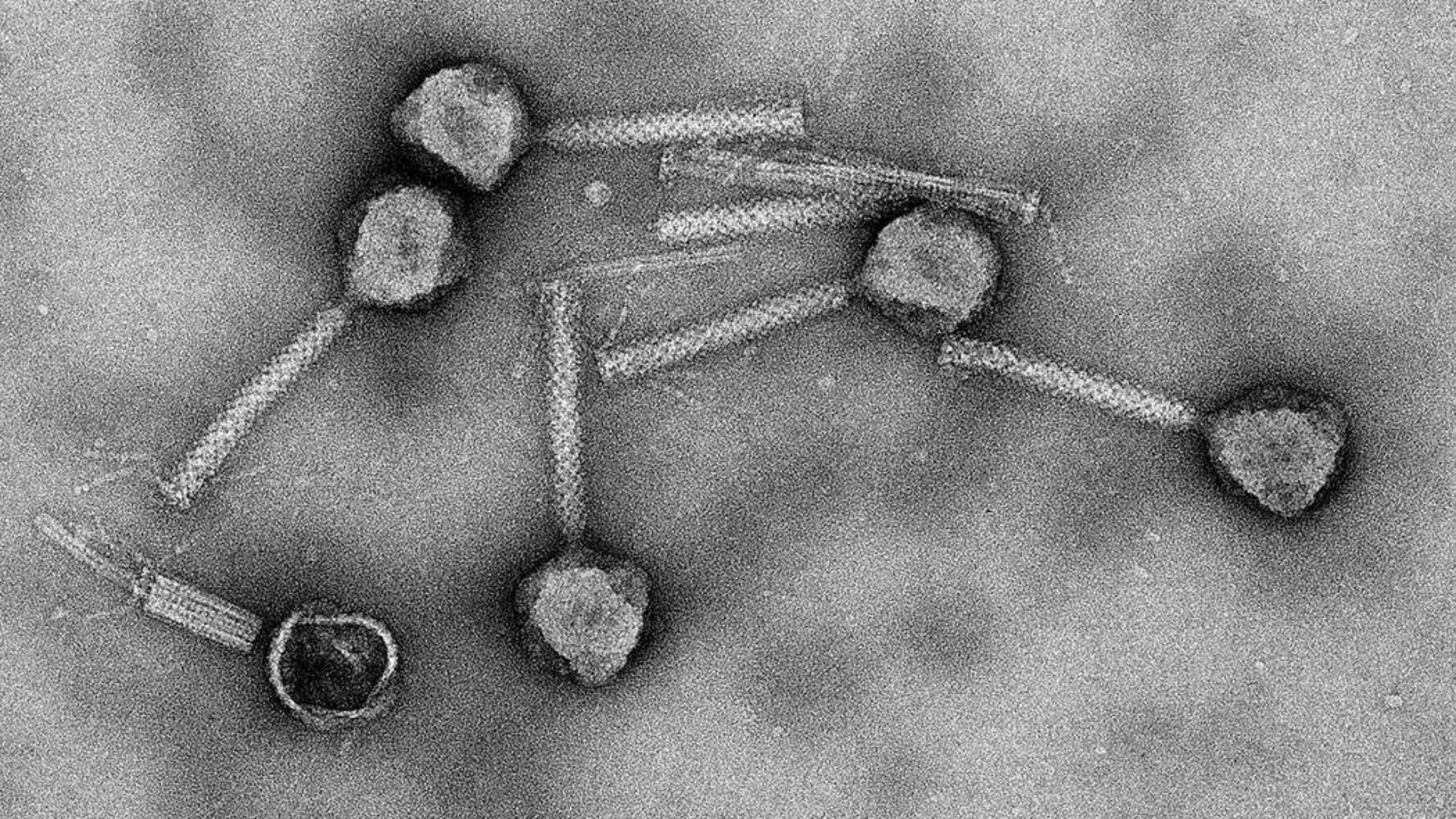
Source: Review on Antimicrobial Resistance 2014

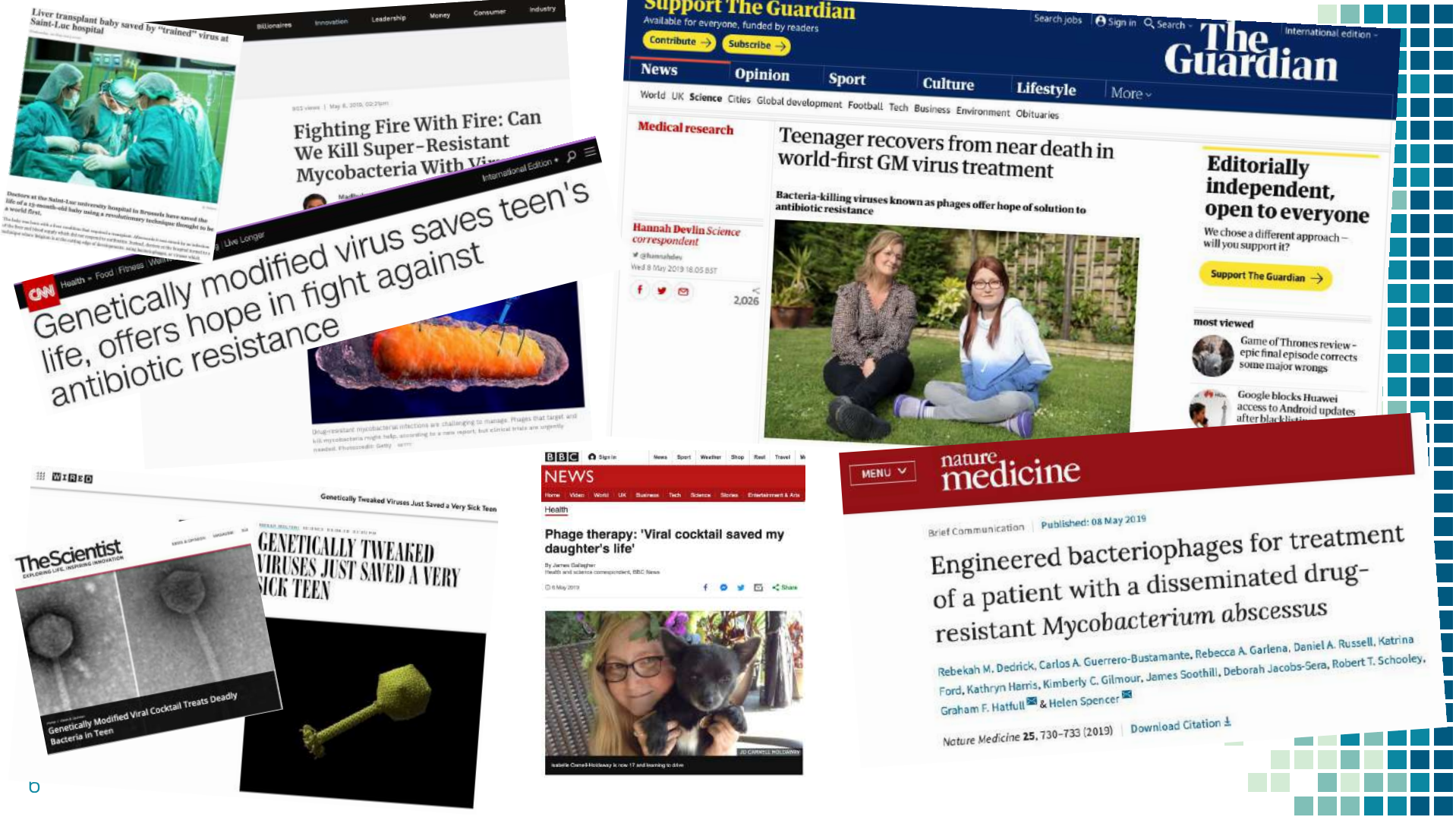
By 2050: more deaths from resistant infections compared to e.g. cancer

http://amr-review.org/

Review on Antimicrobial Resistance

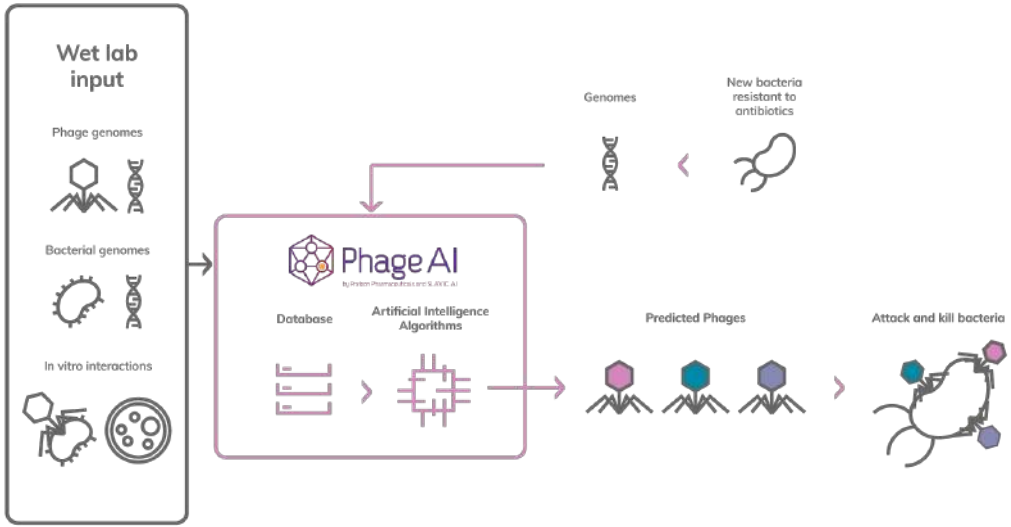
Tackling drug-resistant infections globally




Liver transplant baby saved by "trained" virus at Saint-Luc hospital
Billionaires Innovation Leadership Money Consumer Industry
Fighting Fire With Fire: Can We Kill Super-Resistant Mycobacteria With Vi
Doctors at the Saint-Luc university hospital in Brussels have saved the life of a 13-month-old baby using a revolutionary technique thought to be a world first.
CNN Health Food Fitness
International Edition
Live Longer
Genetically modified virus saves teen's life, offers hope in fight against antibiotic resistance
Drug-resistant mycobacterial infections are challenging to manage. Phages that target and kill mycobacteria might help, according to a new report, but clinical trials are urgently needed. Photocredit: Getty
Support The Guardian
Available for everyone, funded by readers
Contribute
Subscribe
Search jobs
Sign in
Search
The Guardian
International edition
News Opinion Sport Culture Lifestyle More
World UK Science Cities Global development Football Tech Business Environment Obituaries
Medical research
Teenager recovers from near death in world-first GM virus treatment
Bacteria-killing viruses known as phages offer hope of solution to antibiotic resistance
Hannah Devlin Science correspondent
@hannahdev
Wed 8 May 2019 18.05 BST
2,026
Editorially independent, open to everyone
We chose a different approach – will you support it?
Support The Guardian
most viewed
Game of Thrones review - epic final episode corrects some major wrongs
Google blocks Huawei access to Android updates after blacklist
WIRED
Genetically Tweaked Viruses Just Saved a Very Sick Teen
TheScientist
EXPLORING LIFE, INSPIRING INNOVATION
GENETICALLY TWEAKED VIRUSES JUST SAVED A VERY SICK TEEN
Genetically Modified Viral Cocktail Treats Deadly Bacteria in Teen
BBC
Sign in
News Sport Weather Shop Reel Travel
NEWS
Home Video World UK Business Tech Science Stories Entertainment & Arts
Health
Phage therapy: 'Viral cocktail saved my daughter's life'
By James Gallagher
Health and science correspondent, BBC News
8 May 2019
Share
Isabelle Cornish-Holloway is now 17 and learning to drive
JO CARNELL HOLLOWAY
MENU
nature medicine
Brief Communication
Published: 08 May 2019
Engineered bacteriophages for treatment of a patient with a disseminated drug-resistant Mycobacterium abscessus
Rebekah M. Dedrick, Carlos A. Guerrero-Bustamante, Rebecca A. Garlena, Daniel A. Russell, Katrina Ford, Kathryn Harris, Kimberly C. Gilmour, James Soothill, Deborah Jacobs-Sera, Robert T. Schooley, Graham F. Hatfull & Helen Spencer
Nature Medicine 25, 730–733 (2019)
Download Citation

How can we help?

Predict which bacteriophages could be applicate as alternatives to antibiotics in Clinical Care



Who support us

Business partners

Proteon Pharmaceuticals

Phages 4A

Academic partners

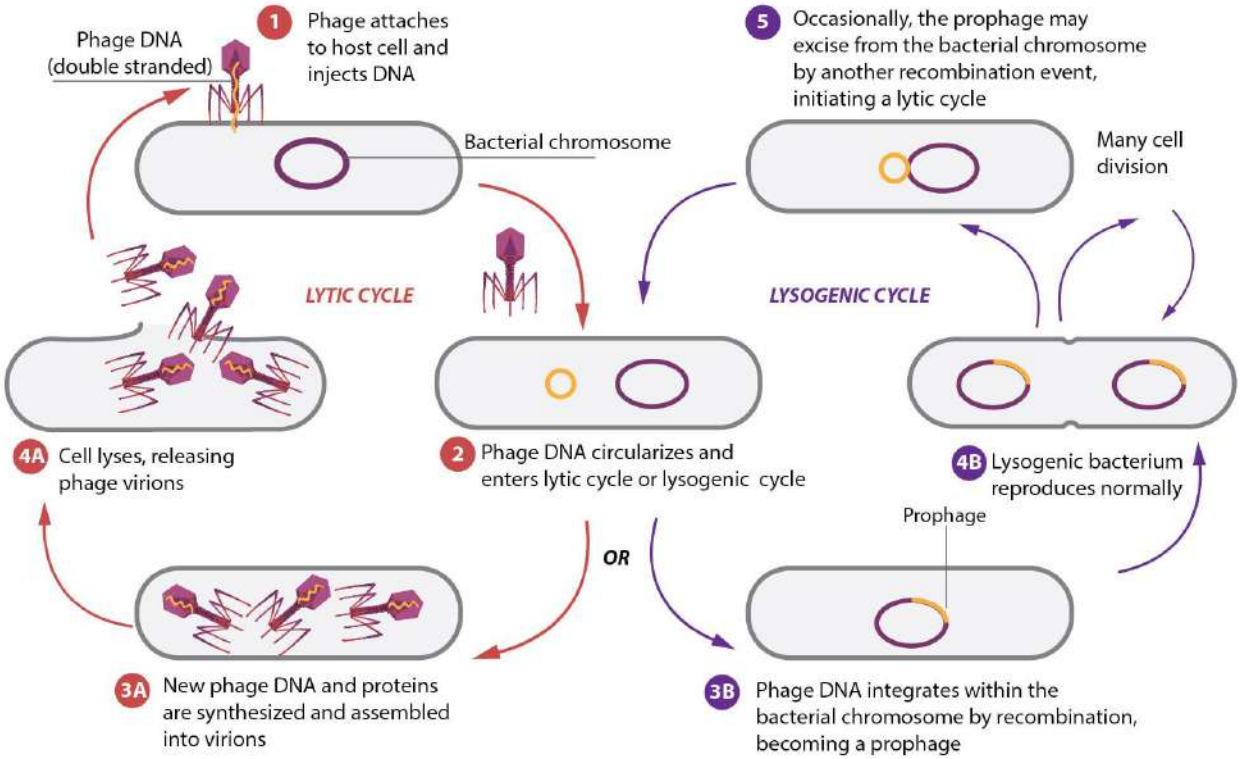
Ministerstwo Nauki i Szkolnictwa Wyższego

Politechnika Białostocka

NVIDIA

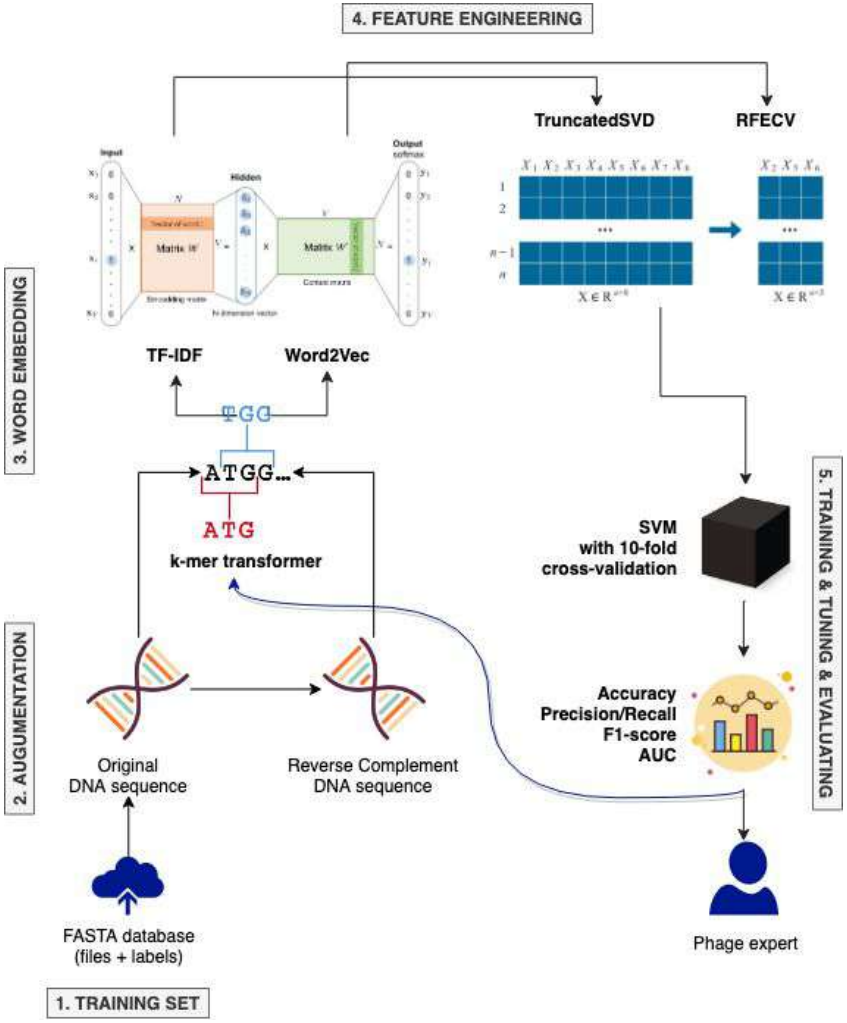
Microsoft

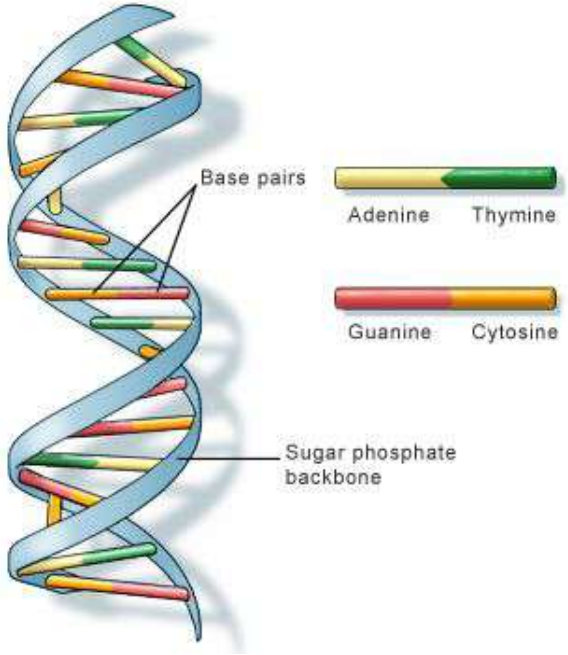
Phage Life Cycles – issue 1


1 Phage attaches to host cell and injects DNA
Phage DNA (double stranded)
Bacterial chromosome
LYTIC CYCLE
LYSOGENIC CYCLE
5 Occasionally, the prophage may excise from the bacterial chromosome by another recombination event, initiating a lytic cycle
Many cell division
4A Cell lyses, releasing phage virions
2 Phage DNA circularizes and enters lytic cycle or lysogenic cycle
4B Lysogenic bacterium reproduces normally
Prophage
OR
3A New phage DNA and proteins are synthesized and assembled into virions
3B Phage DNA integrates within the bacterial chromosome by recombination, becoming a prophage

98,90%

Life cycle recognition accuracy


4. FEATURE ENGINEERING
TruncatedSVD
RFECV
3. WORD EMBEDDING
TF-IDF
Word2Vec
ATGG...
ATG
k-mer transformer
SVM
with 10-fold
cross-validation
5. TRAINING & TUNING & EVALUATING
2. AUGUMENTATION
Accuracy
Precision/Recall
F1-score
AUC
Original
DNA sequence
Reverse Complement
DNA sequence
FASTA database
(files + labels)
Phage expert
1. TRAINING SET



Source: U.S. National Library of Medicine

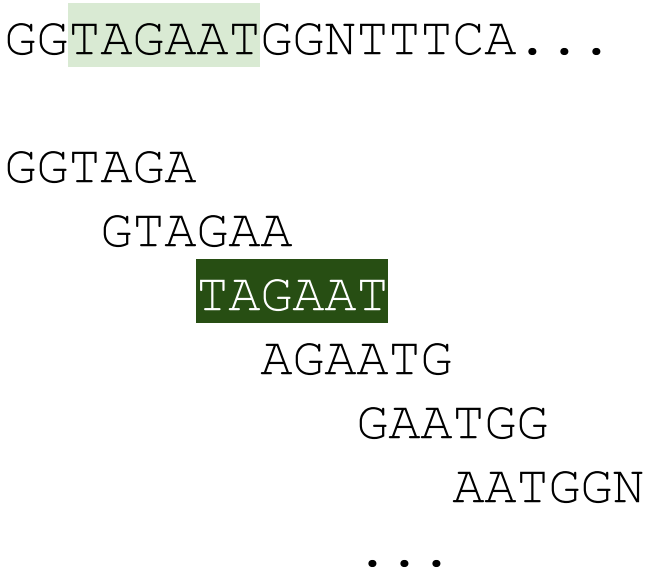
[2] 6-mer transformer

```
GGTAGAATGGNTTTCA...

GGTAGA
    GTAGAA
        TAGAAT
           AGAATG
               GAATGG
                   AATGGN
                  ...
```

[2] 6-mer transformer

```
GGTAGAATGGNTTTCA...

GGTAGA
    GTAGAA
        TAGAAT
           AGAATG
               GAATGG
                   AATGGN
                  ...
```

[2] 6-mer transformer

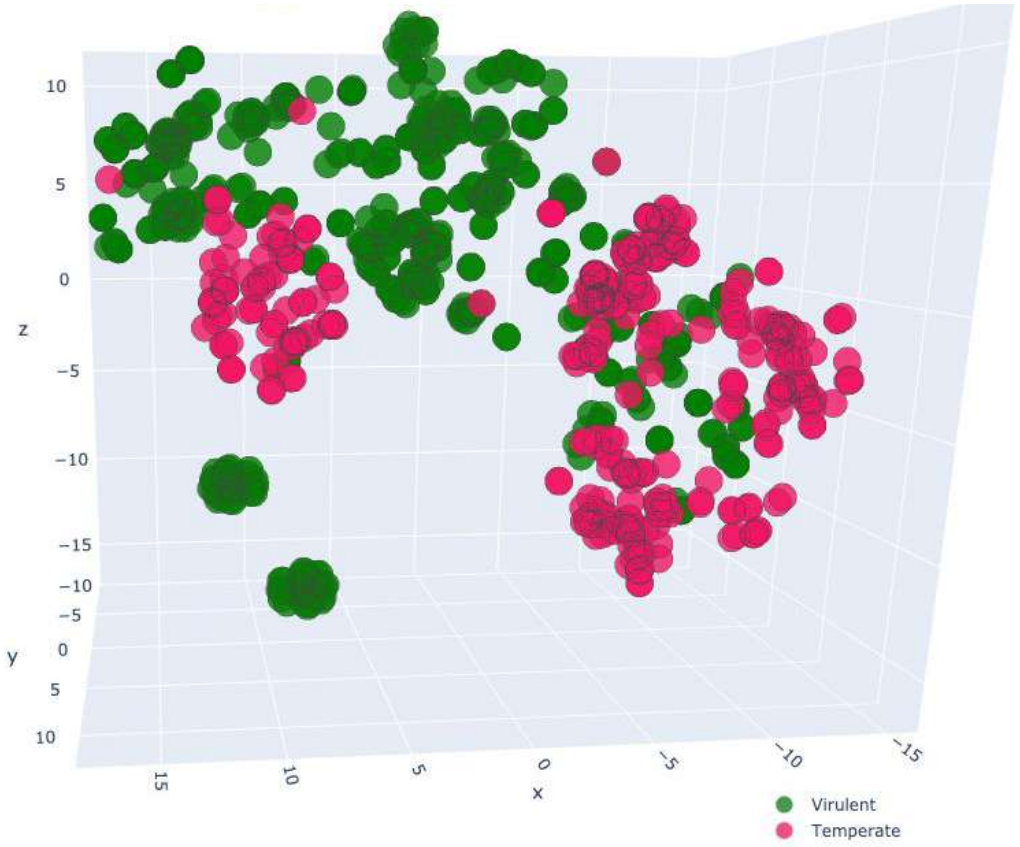


[2] 6-mer transformer

```
GGTAGAATGGNTTTCA...

GGTAGA
    GTAGAA
        TAGAAT
           AGAATG
               GAATGG
                   AATGGN
               ...
```

6-mers (bag of words)
Word2Vec Skip-gram + RFECV

```
[[ 0.15740727,  0.14283979,  0.01424173, ..., -0.04863179,
   0.36005523,  0.04962862],
 [ 0.14294244,  0.06846078,  0.03159813, ..., -0.02003489,
   0.29529446,  0.07867343],
 [ 0.14319768,  0.06886728,  0.03136309, ..., -0.01986326,
   0.29515907,  0.07877837],
 ...,
 [ 0.14686785,  0.10228563,  0.02458559, ..., -0.03324442,
   0.32741652,  0.04950592],
 [ 0.16520534,  0.14164333,  0.01523334, ..., -0.01981086,
   0.37183095,  0.02930221],
 [ 0.14716548,  0.05672845,  0.03785585, ..., -0.0188462 ,
   0.27017442,  0.0712469 ]]
```



Virulent and Temperate phages from training set after Word2Vec vectorization and t-SNE decompression.

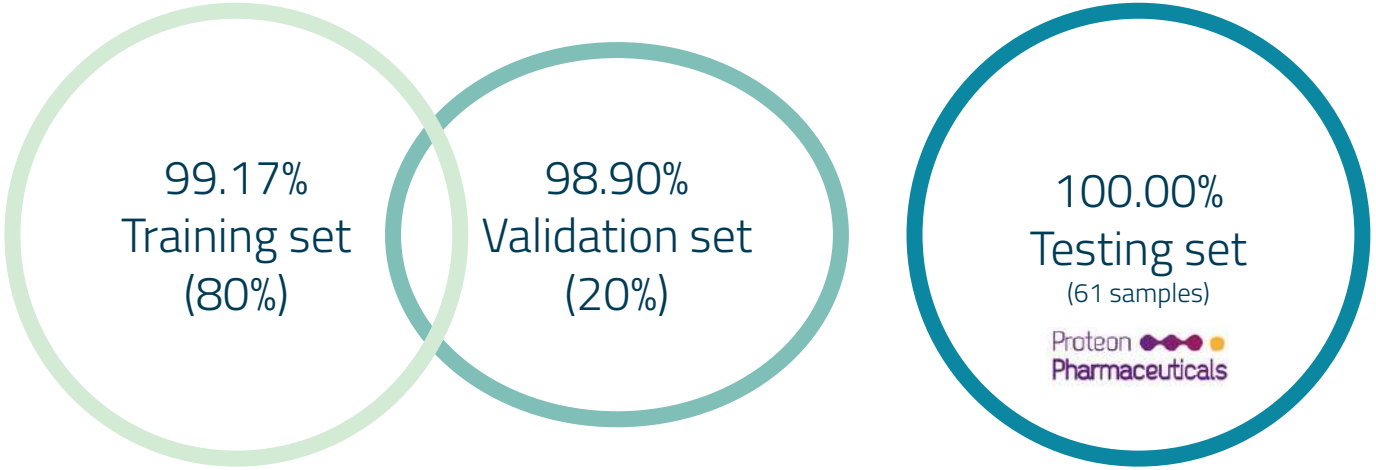
[5] Training & Tuning

- MultinomialNB
- RandomForest
- MLPClassifier
- LogisticRegression
- XGBoost
- **SVM**
- GradientBoosting
- SGDClassifier
- KNeighborsClassifier
- CatBoostClassifier
- LightGBM

- TF-IDF
- **Word2Vec** (Skip-gram/CBoW)
- fastText
- DNA2Vec
- fastDNA

BayesSearchCV

EVALUATION


99.17%
Training set
(80%)
98.90%
Validation set
(20%)
100.00%
Testing set
(61 samples)
Proteon
Pharmaceuticals

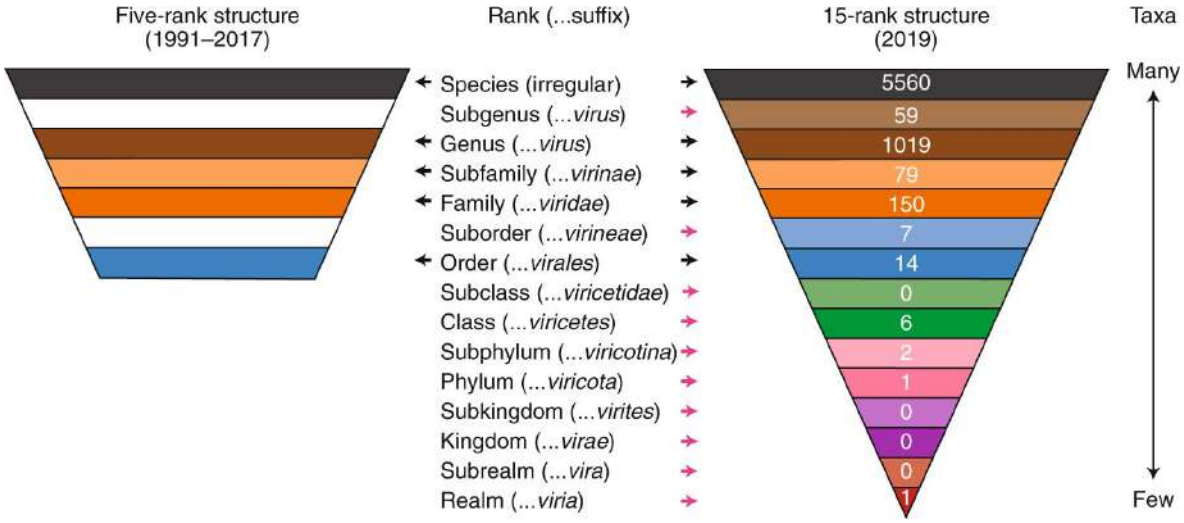
Article

BMC Bioinformatics

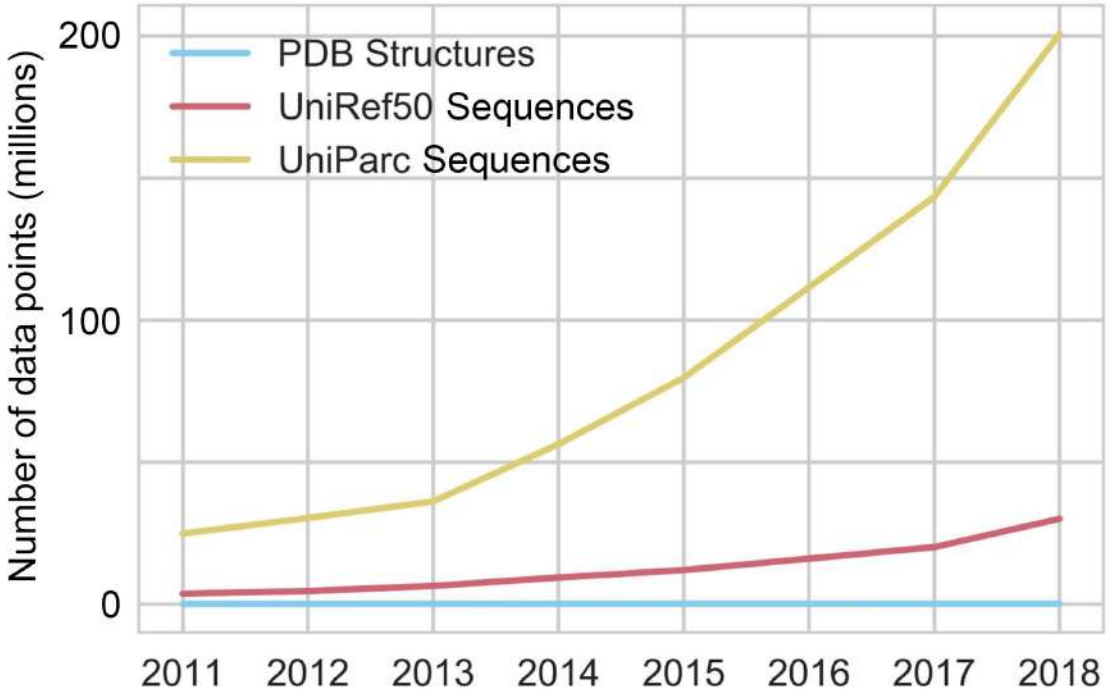
PhageAI - bacteriophage life cycle recognition with Machine Learning and Natural Language Processing

Q1 2020

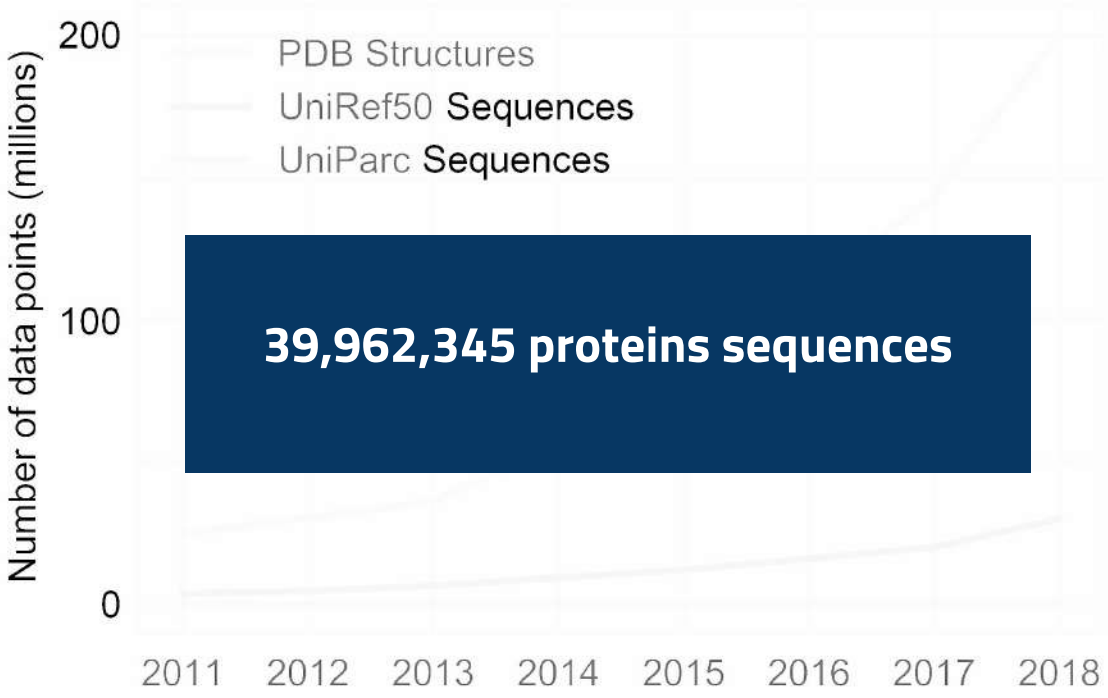
Taxonomy of Viruses - issue 2



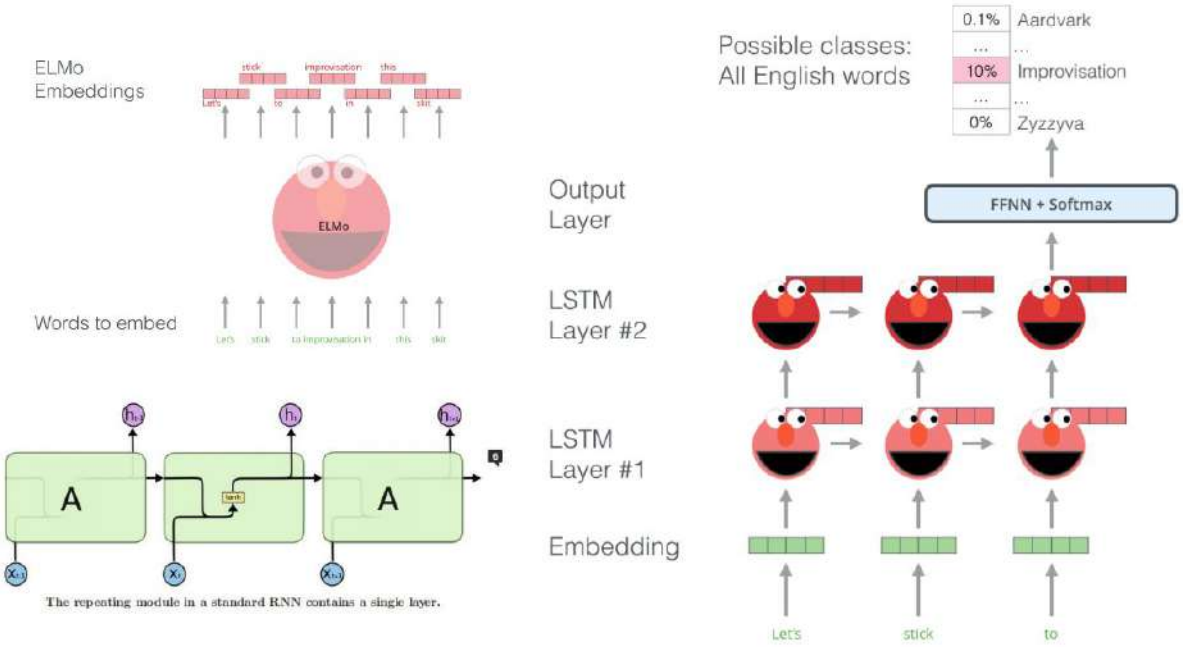
Source: nature.com/articles/s41564-020-0709-x



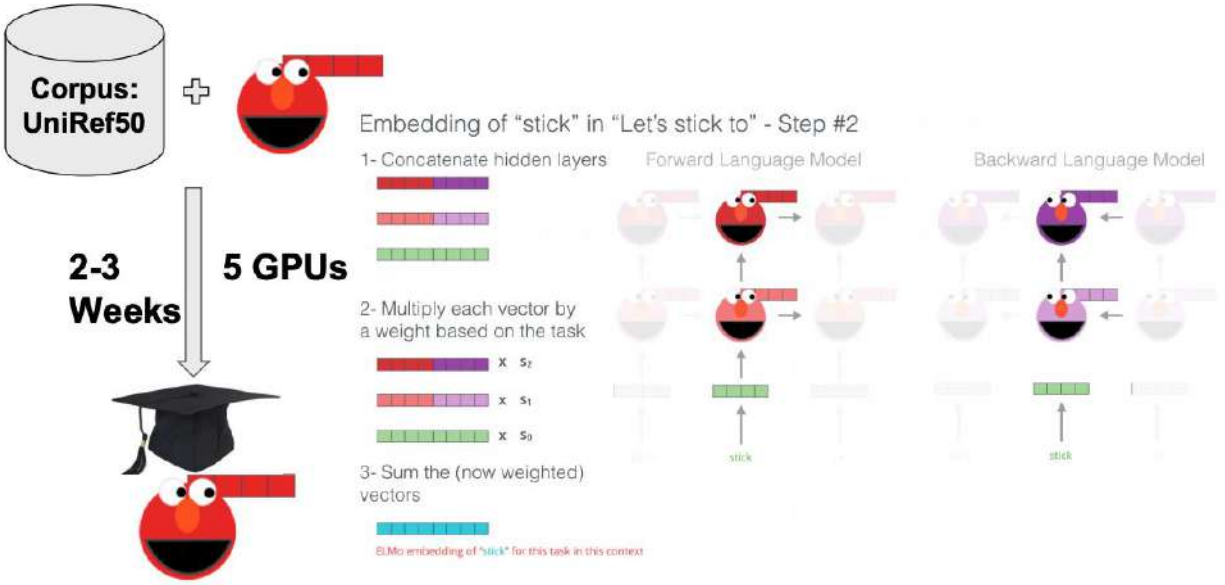
Source: Mohammed AlQuraishi



Source: Mohammed AlQuraishi

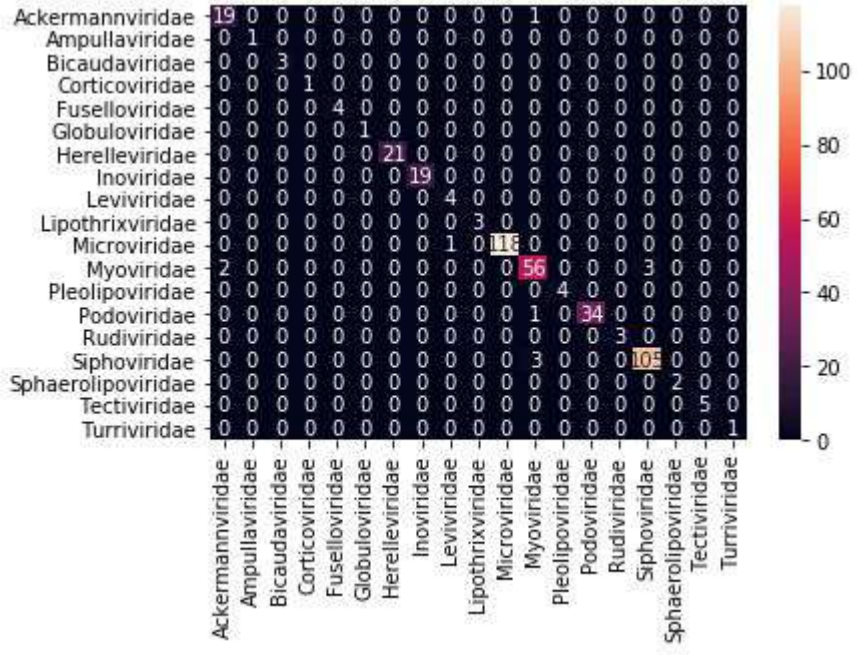


Source: Peters, Matthew E., et al. "Deep contextualized word representations." arXiv preprint arXiv:1802.05365 (2018).



Source: M Heinzinger, et al. "Modeling the Language of Life-Deep Learning Protein Sequences" (2019)

Family Taxonomy: ELMo + SVM



Accuracy: 97.35%

AUC: 99.57%

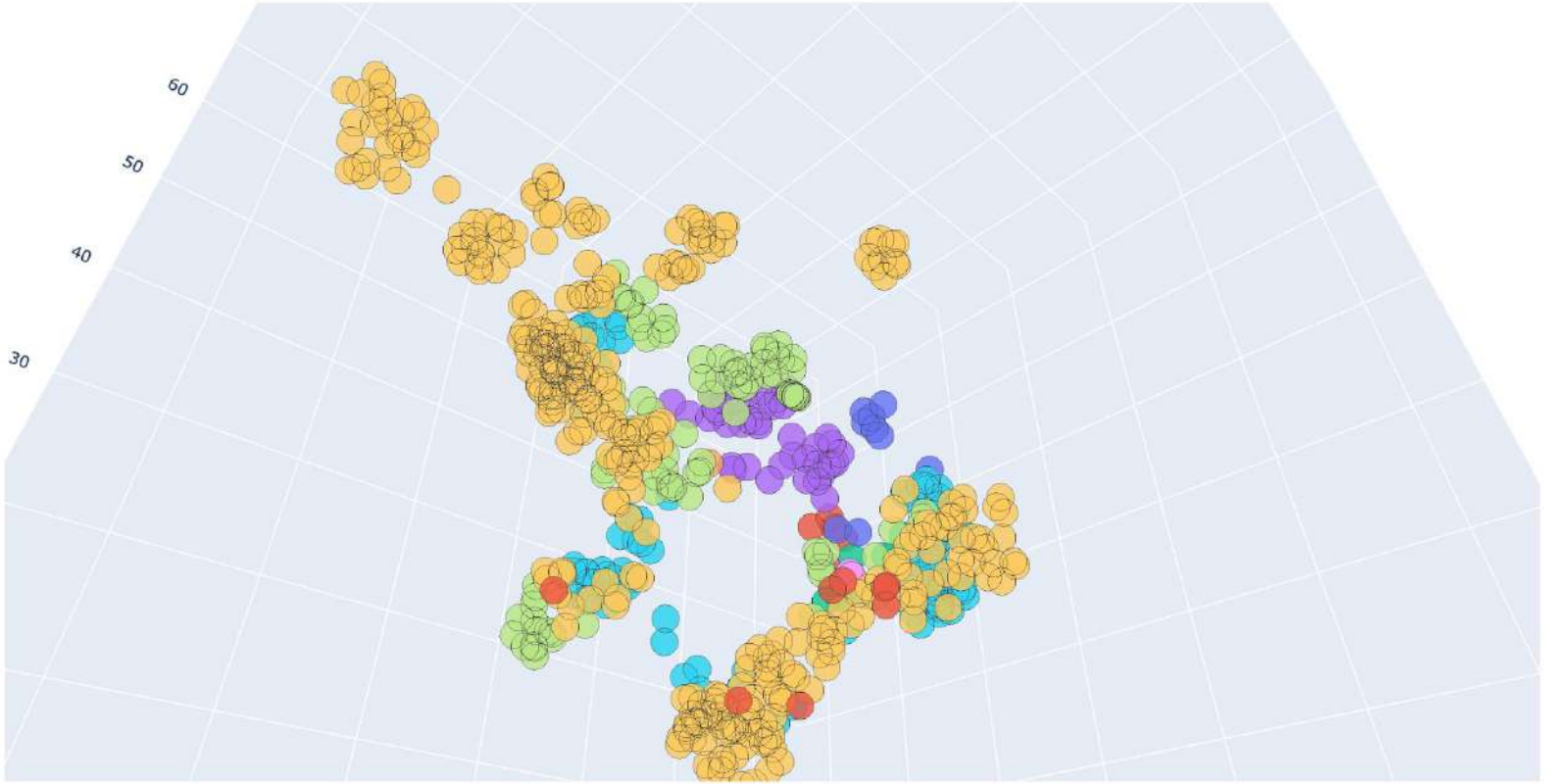
Classification report:

	precision	recall	f1-score	support
0	0.90	0.95	0.93	20
1	1.00	1.00	1.00	1
2	1.00	1.00	1.00	3
3	1.00	1.00	1.00	1
4	1.00	1.00	1.00	4
5	1.00	1.00	1.00	1
6	1.00	1.00	1.00	21
7	1.00	1.00	1.00	19
8	0.80	1.00	0.89	4
9	1.00	1.00	1.00	3
10	1.00	0.99	1.00	119
11	0.92	0.92	0.92	61
12	1.00	1.00	1.00	4
13	1.00	0.97	0.99	35
14	1.00	1.00	1.00	3
15	0.97	0.97	0.97	108
16	1.00	1.00	1.00	2
17	1.00	1.00	1.00	5
18	1.00	1.00	1.00	1
accuracy			0.97	415
macro avg	0.98	0.99	0.98	415
weighted avg	0.97	0.97	0.97	415

Training set score: 99.90%

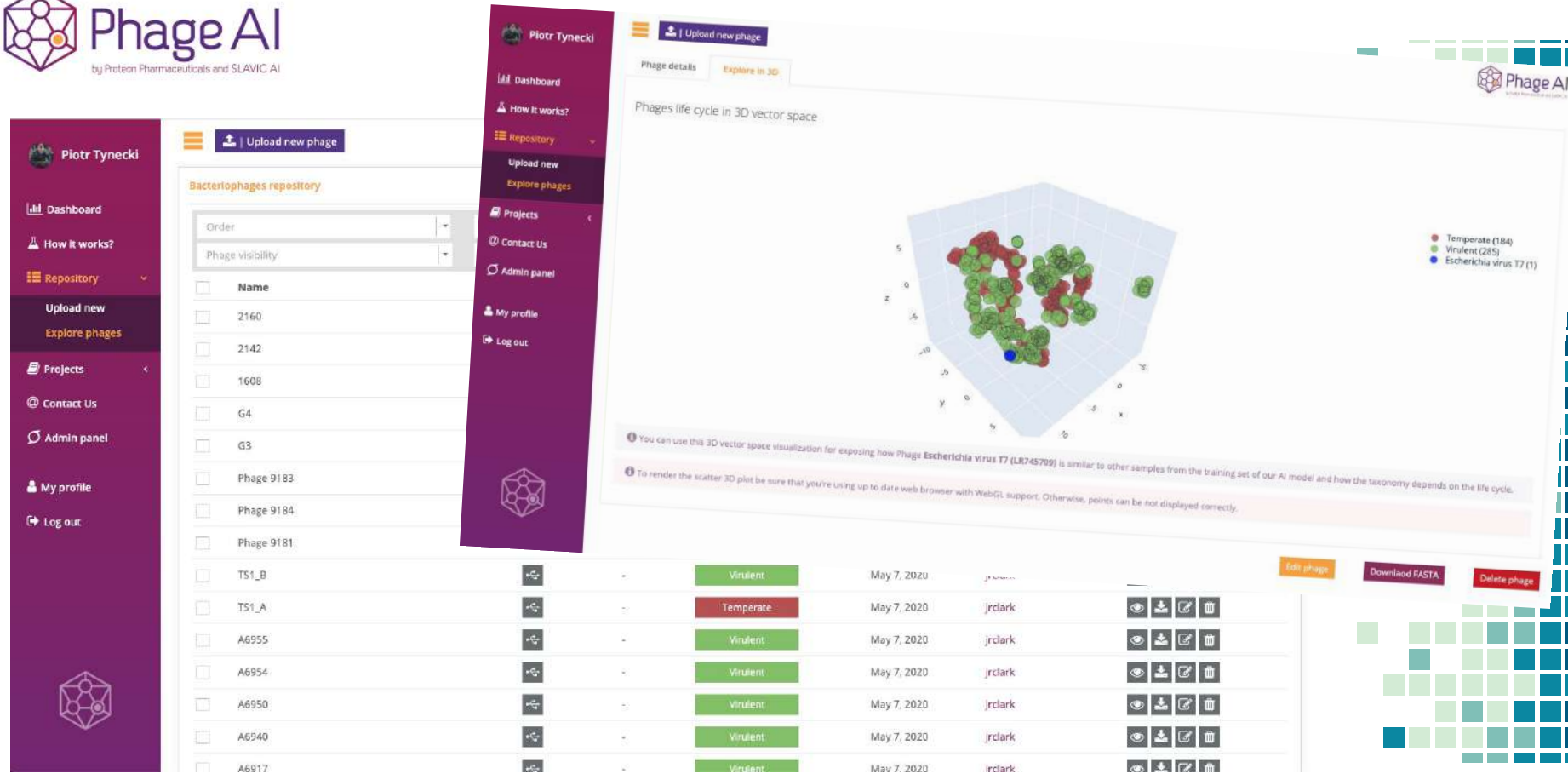
Validation set score: 97.35%

Family Taxonomy: ELMo + SVM (PCA(50) -> UMAP)



Phage AI

by Proteon Pharmaceuticals and SLAVIC AI



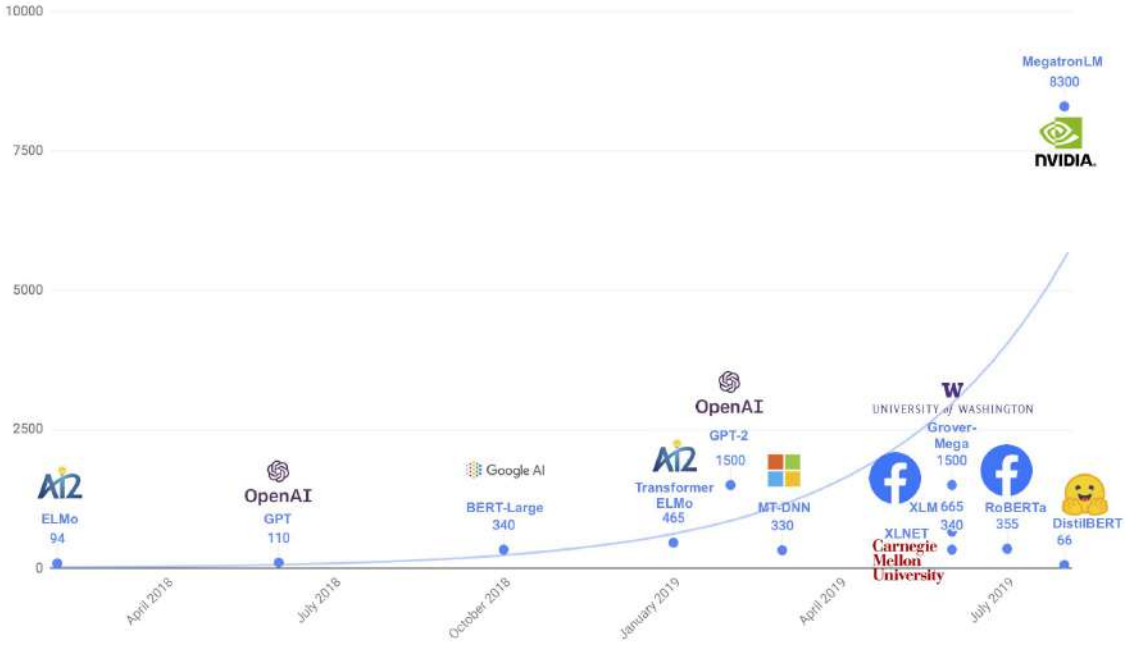
What else...?

The Structure and Function of Proteins
- issue 3

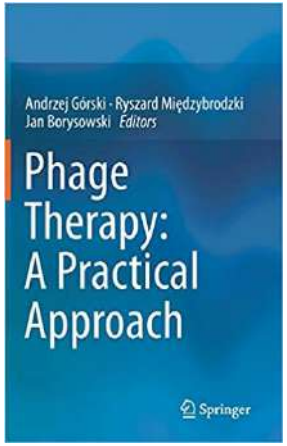
Phage-Host matching
- issue 4

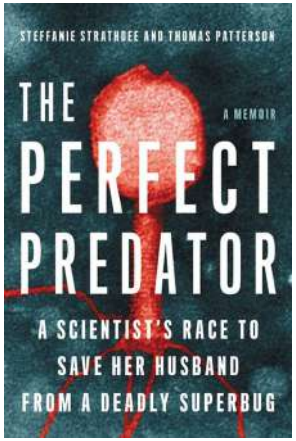
Deep Generative Networks for Bacteriophages Genetic Edition
- issue 5

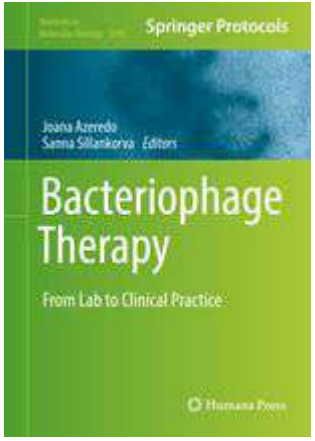
The Future of Phages Science will not be Supervised...



Must see & read







Bacteriophages: the cure for antibiotics resistance

Phage Therapy: An Effective Alternative to Antibiotics?

Using Viruses to Fight Antibiotic-Resistant Infections

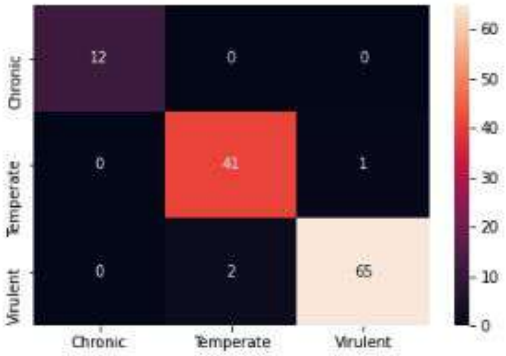
Data sources

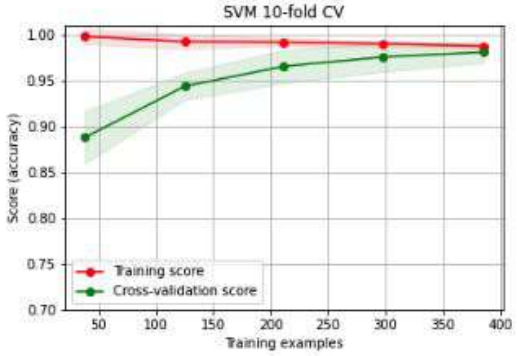
Thank you for your attention

Any questions?

Twitter: @ptynecki
LinkedIn: piotrtynecki
E-mail: p.tynecki@doktoranci.pb.edu.pl

[5] Evaluation





SLAVIC AI

Virus Activity Detector

for Education and Research