

# CrisprVi: a software for visualizing and analyzing CRISPR sequences of prokaryotes

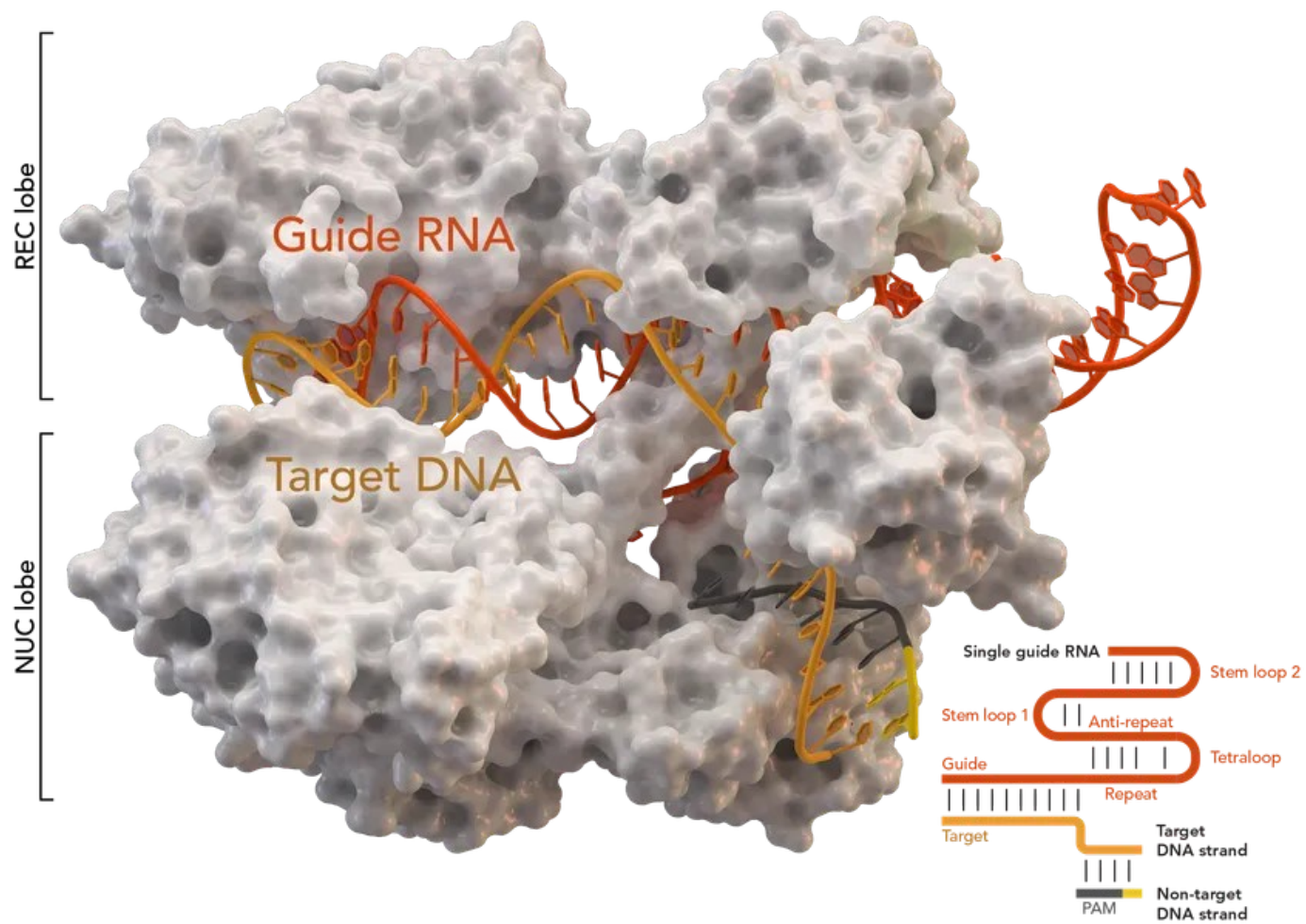
John Wesley Mathis  
ECE/SSE 591

# What is CRISPR?

- Clustered Regularly Interspaced Short Palindromic Repeats
- Natural defense mechanism in bacterial immune system
- CRISPR regions contain short, repetitive DNA sequences interspaced with unique sequences called spacers
- CRISPR is used to store genetic info from past viral infection and recognize and destroy invading viruses in subsequent encounters

# What is Cas?

- CRISPR-Cas System
  - CRISPR-associated (Cas)
  - Proteins essential for bacterial adaptive immunity



# The Problem with CRISPR Visualization and Analysis

- Comprehensive Tools
  - Investigate loci and sequences of DRs and spacers
  - Existing tools: CRT, PILER-CR, CRISPRFinder, etc.
- Limitations of Current Tools:
  - Lack interactive and user-friendly visualization
  - Complexity and non-interactivity of Excel macros
  - Over-reliance on specific detection tools
  - Confusing visual outputs with complex datasets
- Gap in the market
  - Need for an all-in-one tool to visualize, manipulate, and analyze CRISPR arrays effectively

# Why use Python?



- Extensive Bioinformatics libraries
  - Biopython: sequence analysis and data manipulation
  - Pandas: efficient data handling and analysis
  - Matplotlib & Seaborn: powerful data visualization tools
- Seamless integration with existing bioinformatics tools
- Large, active community
- Capable of handling large datasets efficiently

# Python Libraries Used

- **Biopython:** sequence analysis and manipulation
- **Pandas:** efficient data organization and management
- **NumPy:** handling large datasets
- **Matplotlib:** creating visuals
- **Seaborn:** high-level statistical graphics

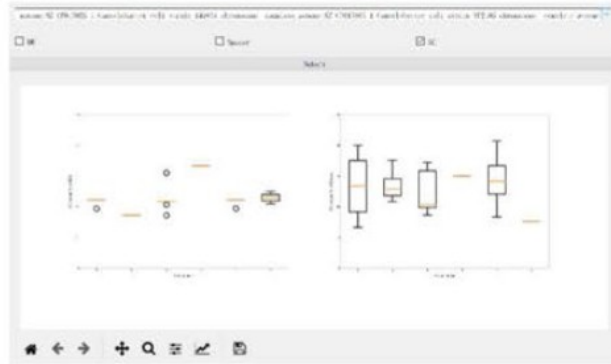


# Results



# GUI Overview

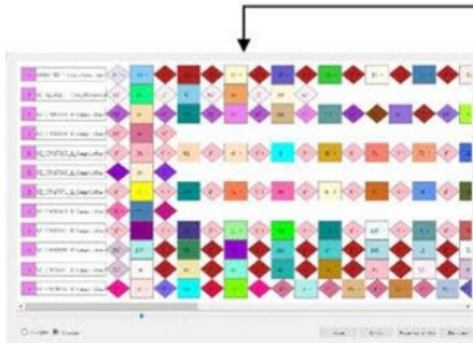
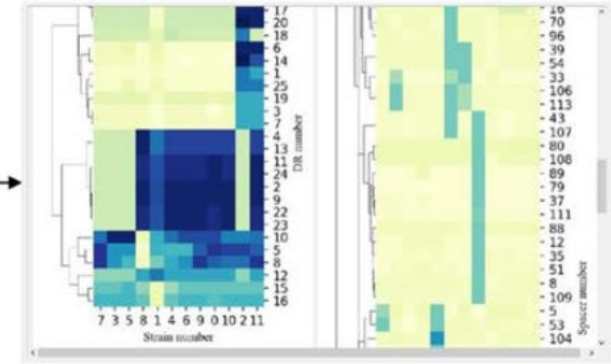
Statistics



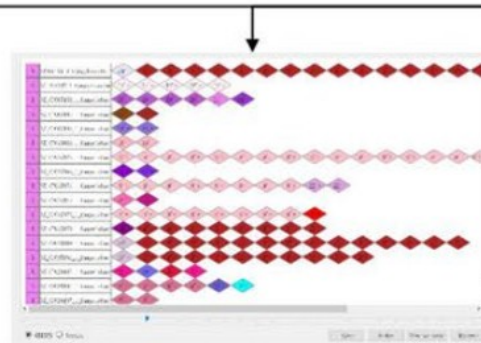
Input wizard



Clustering heatmap



DRs and spacers



DRs



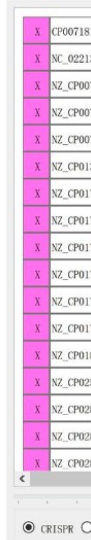
Spacers



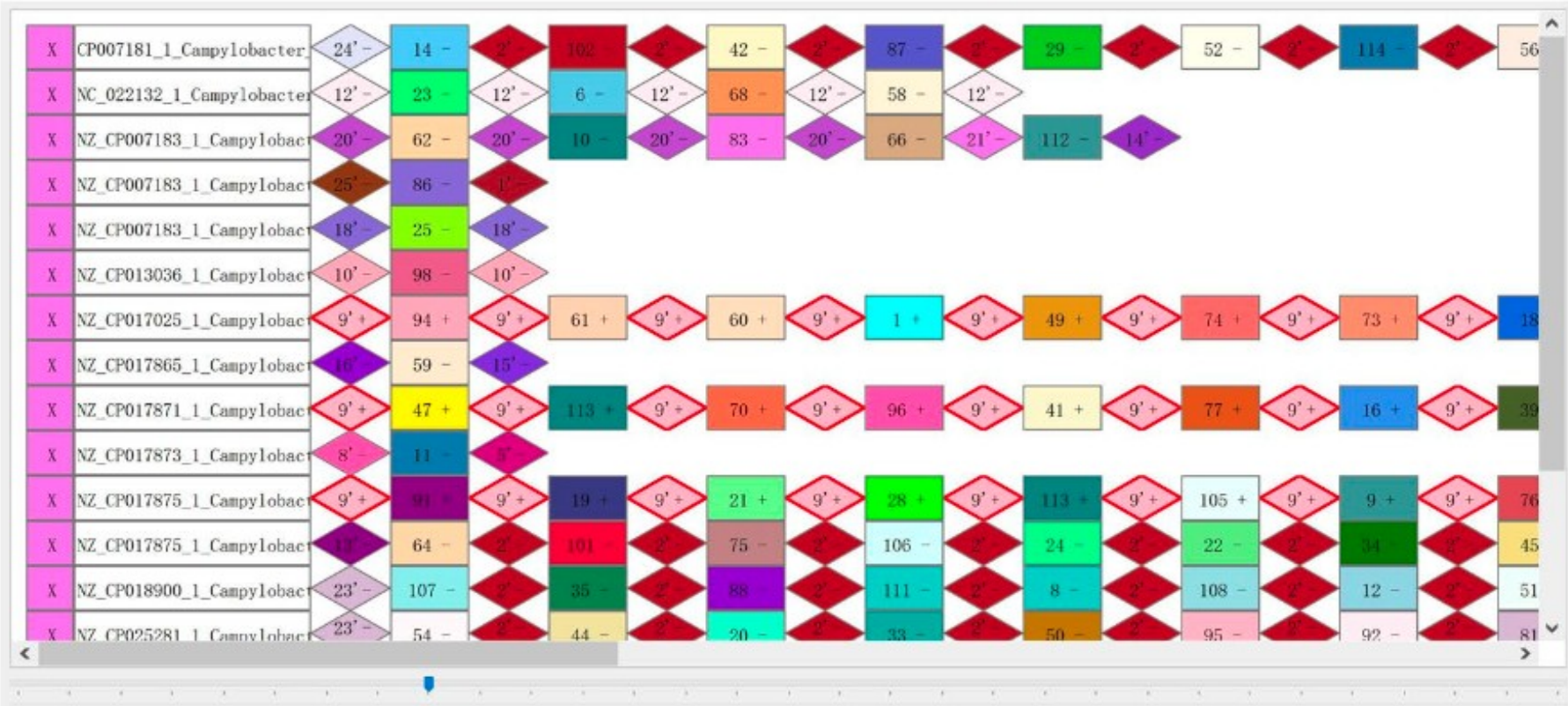
Alignment of spacer arrays

# User can customize graphics for DRs/spacers

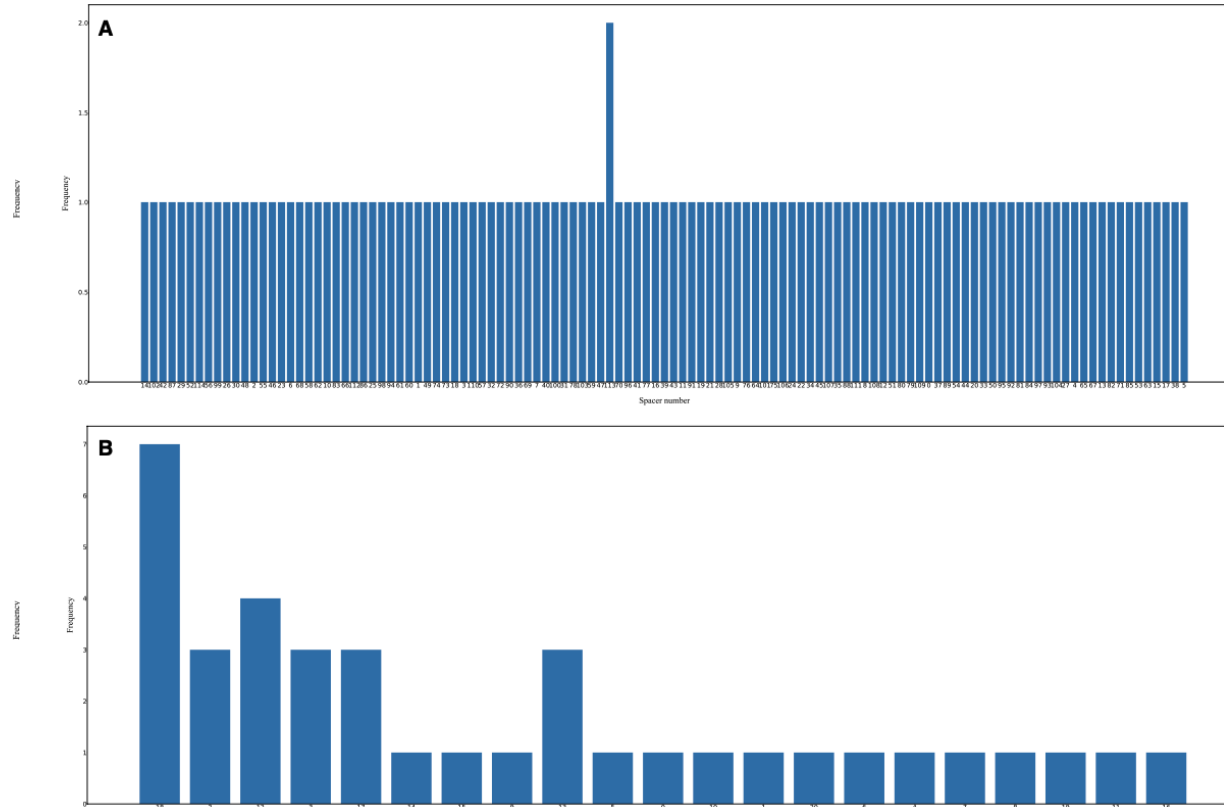
A



C



# Perform statistical analysis of the DRs/spacers on selected strains



# More Detailed Visualization

A.



B.



C.



# My Thoughts

Feature	CrisprVi	CRISPRviz	CRISPRStudio
Visualization of DRs and Spacers	Yes	Yes	No (Only Spacers)
Customization	Yes	Limited	No
Interactive GUI	Yes	Yes	No
Statistical Analysis Tools	Yes	No	No
Accepts outputs from various CRISPR Finding tools	Yes	No (MinCED only)	No
Flexibility in CRISPR detection methods	Yes	No	No
Ability to change visualization settings	Yes	No	No
Efficiency (Data Loading Speed)	Moderate	Fast	N/A

# References

- Biopython. (n.d.). Biopython logo. Retrieved from <https://biopython.org/wiki/logo>
- Jia, B., Zhang, X., Jiang, Q., Wang, Z., Zhang, Y., Hu, Y., & Wang, Y. (2022). CrisprVi: a software for visualizing and analyzing CRISPR sequences of prokaryotes. *BMC Bioinformatics*, 23(1), 257. <https://doi.org/10.1186/s12859-022-04716-9>
- Matplotlib developers. (n.d.). Matplotlib logo. Retrieved from <https://matplotlib.org/stable/gallery/misc/logos2.html>
- NumPy developers. (n.d.). NumPy logo. Retrieved from <https://github.com/numpy/numpy/tree/main/branding/logo>
- pandas development team. (n.d.). Pandas logo. Retrieved from <https://pandas.pydata.org/about/citing.html>
- Python Software Foundation. (n.d.). Python logo. Retrieved from <https://www.python.org/community/logos/>
- Seaborn developers. (n.d.). Seaborn logo. Retrieved from <https://seaborn.pydata.org/citing.html>
- Splettstoesser, T. (n.d.). Cas9 protein structure [Image]. Wikimedia Commons. Retrieved from [https://en.wikipedia.org/wiki/Cas9#/media/File:Cas9\\_5AXW.png](https://en.wikipedia.org/wiki/Cas9#/media/File:Cas9_5AXW.png)

# Questions?