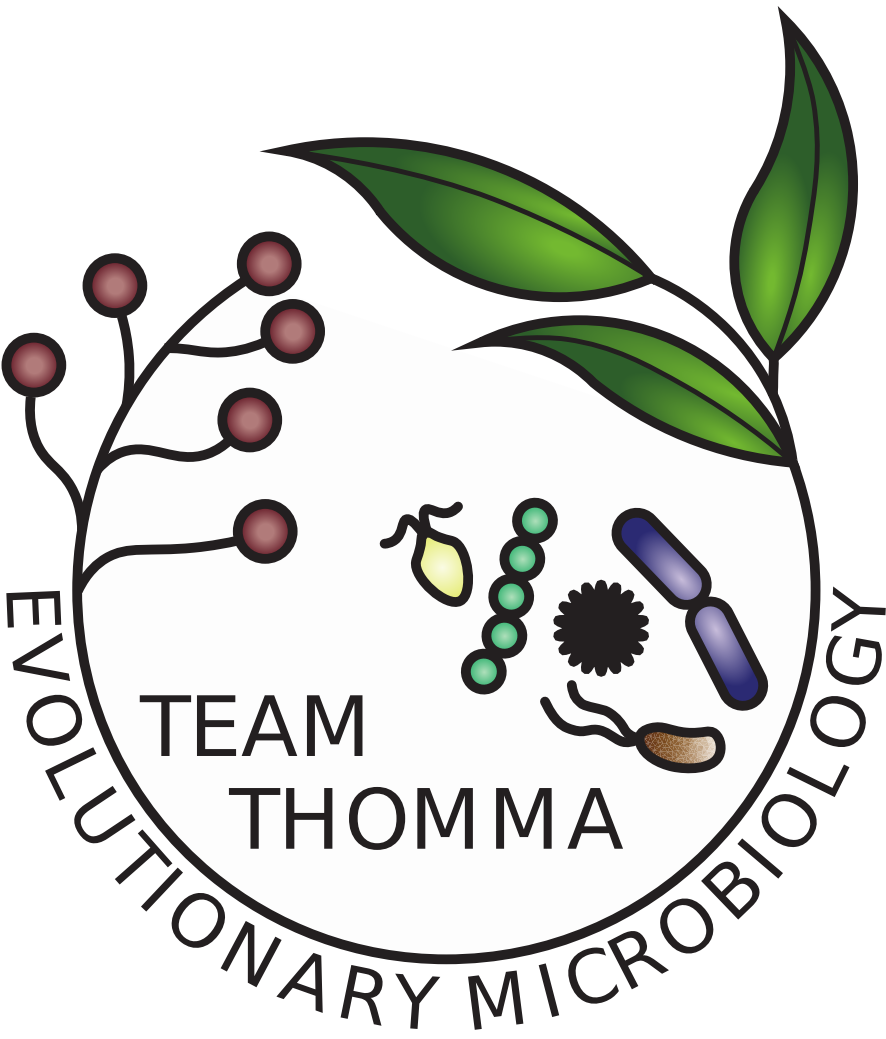


# Machine learning-assisted discovery of fungal effectors with antimicrobial activities

Fantin Mesny, Bart P.H.J. Thomma

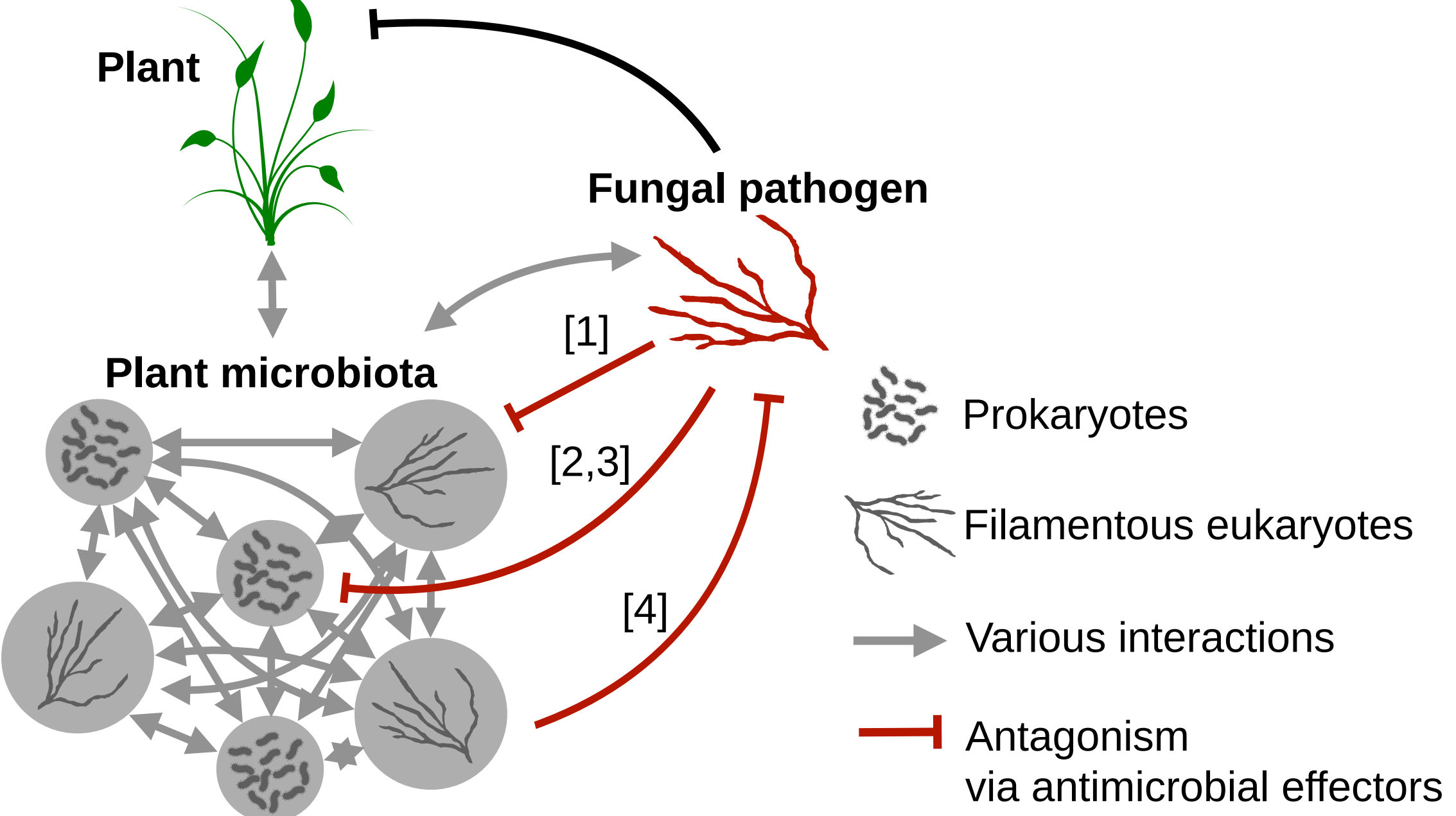
Institute for Plant Sciences, Cluster of Excellence on Plant Sciences (CEPLAS),  
University of Cologne, 50674 Cologne, Germany



**Contact:**  
fmesny1@uni-koeln.de  
Twitter: @f\_mesny

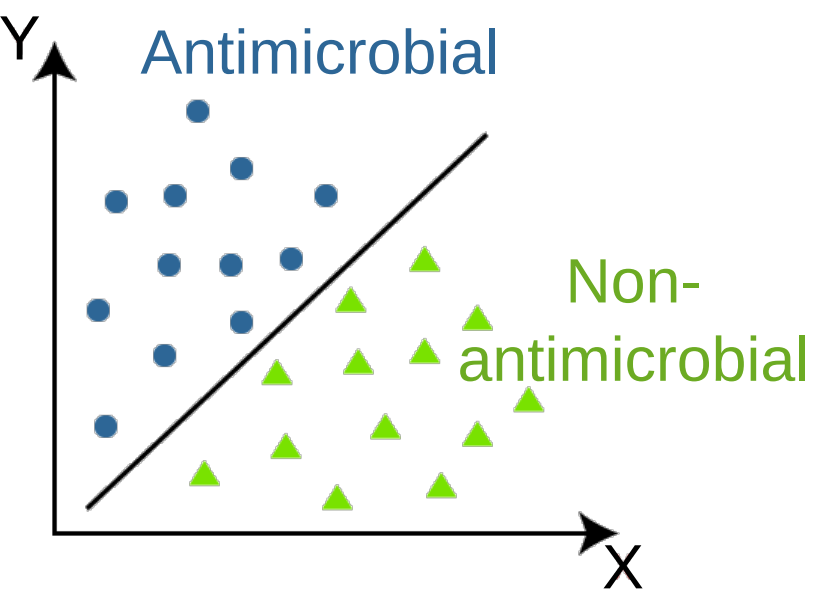
## 1. Antimicrobial fungal effectors

Plant-associated fungi, including pathogens as well as mutualists, secrete small proteins, typically referred to as effectors, to support their colonization of host tissues by targeting plant and/or microbial components. Effectors with antimicrobial activities were demonstrated to play a key role in both disease development and suppression. They are used by pathogens to antagonize protective microbiota members [1,2,3], and by mycobiota members to restrain pathogen colonization [4]. However, their occurrence and conservation throughout the fungal kingdom remain enigmatic.



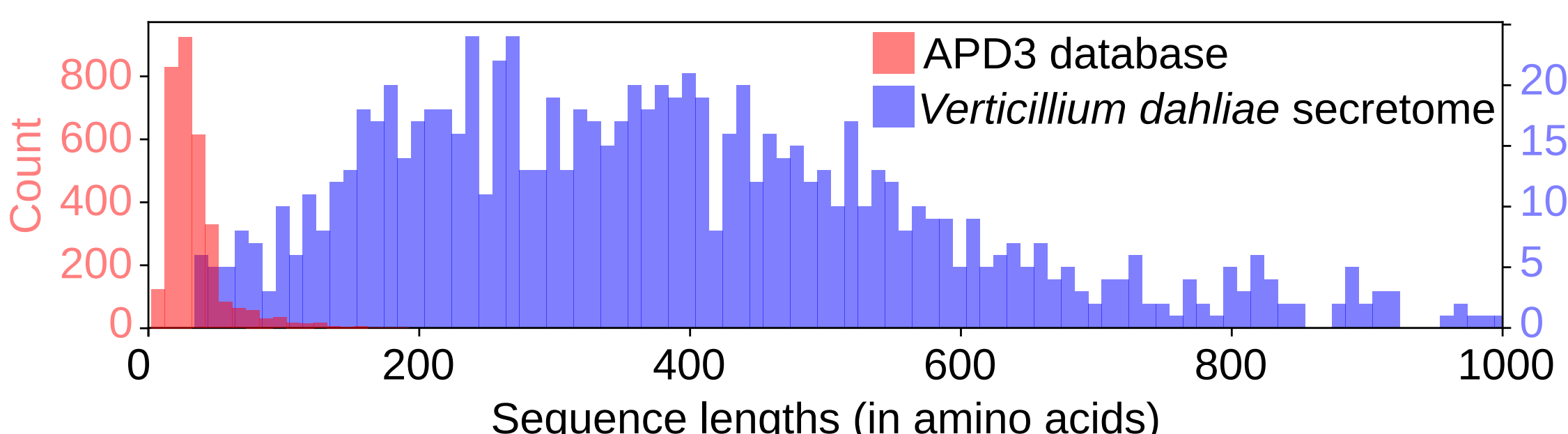
## 2. Predicting antimicrobial activity of proteins with machine learning models

**Aim:**  
(1) Training a model to classify proteins into two groups (●/▲), based on their sequence and structural properties (X, Y, Z,...).  
(2) Use this model to discover new candidate antimicrobial effectors in fungal genomes.



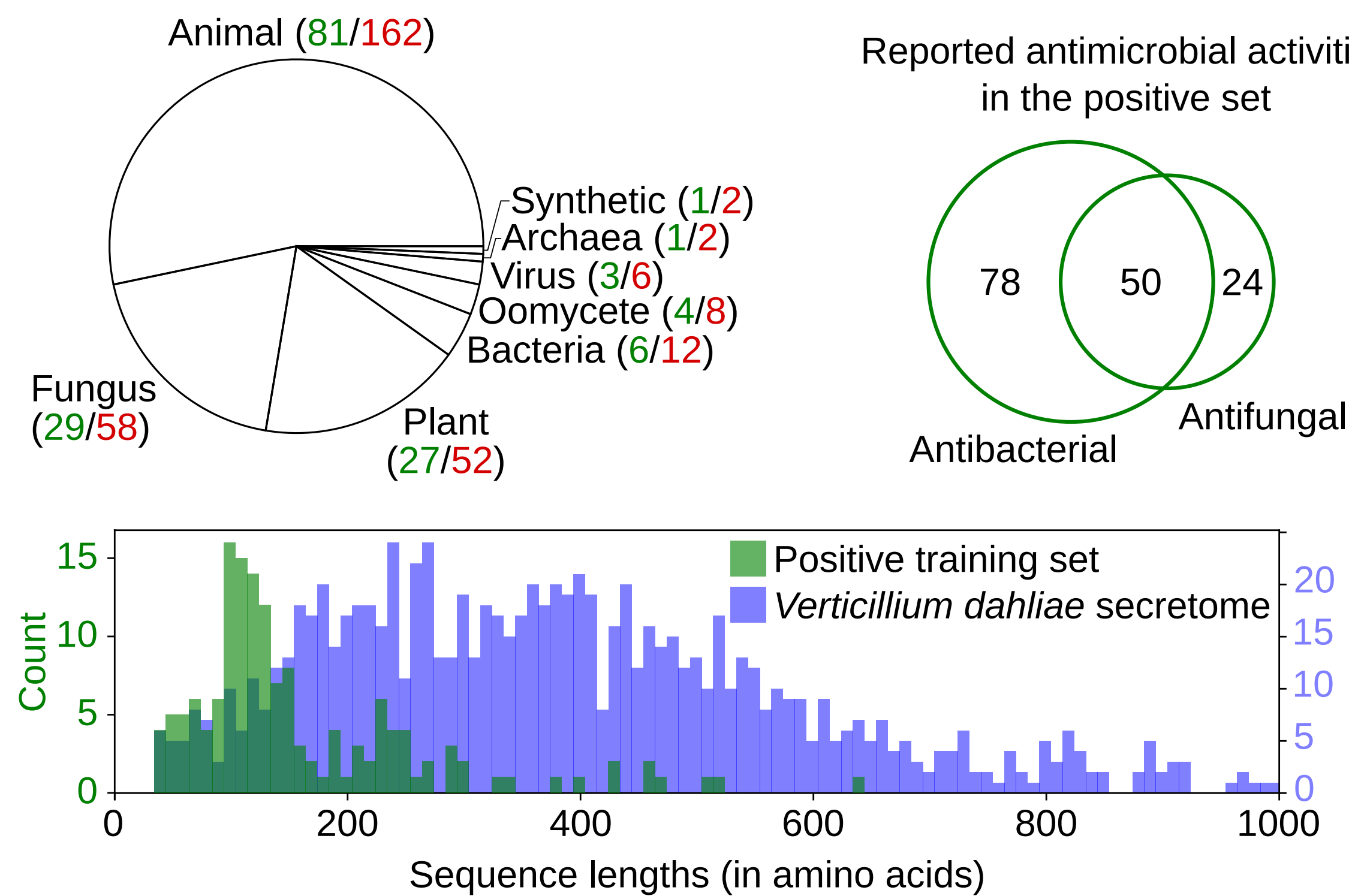
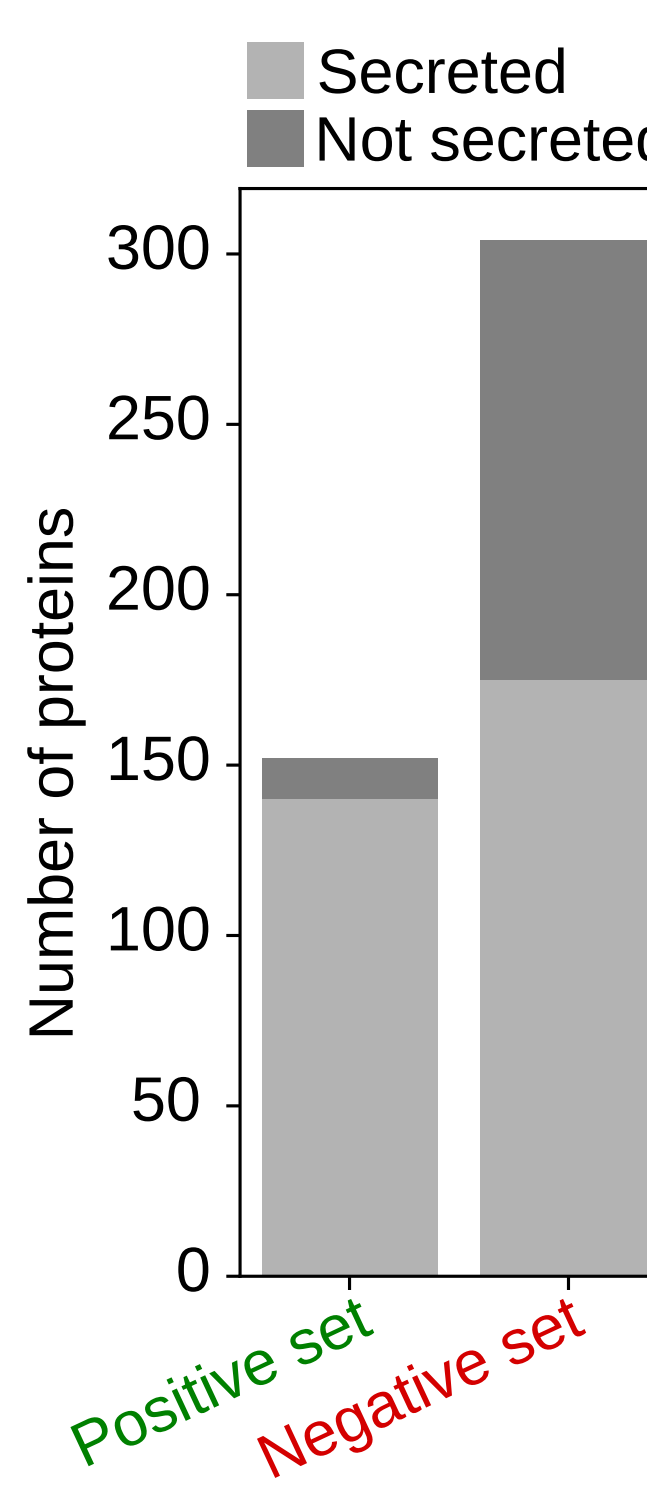
**Previously applied to short peptides**  
Antimicrobial activity predictors have previously been developed, but are dedicated to short peptides, and therefore not suitable for fungal effectors.

e.g. AMPScanner [5] was trained on the APD3 database [6]:

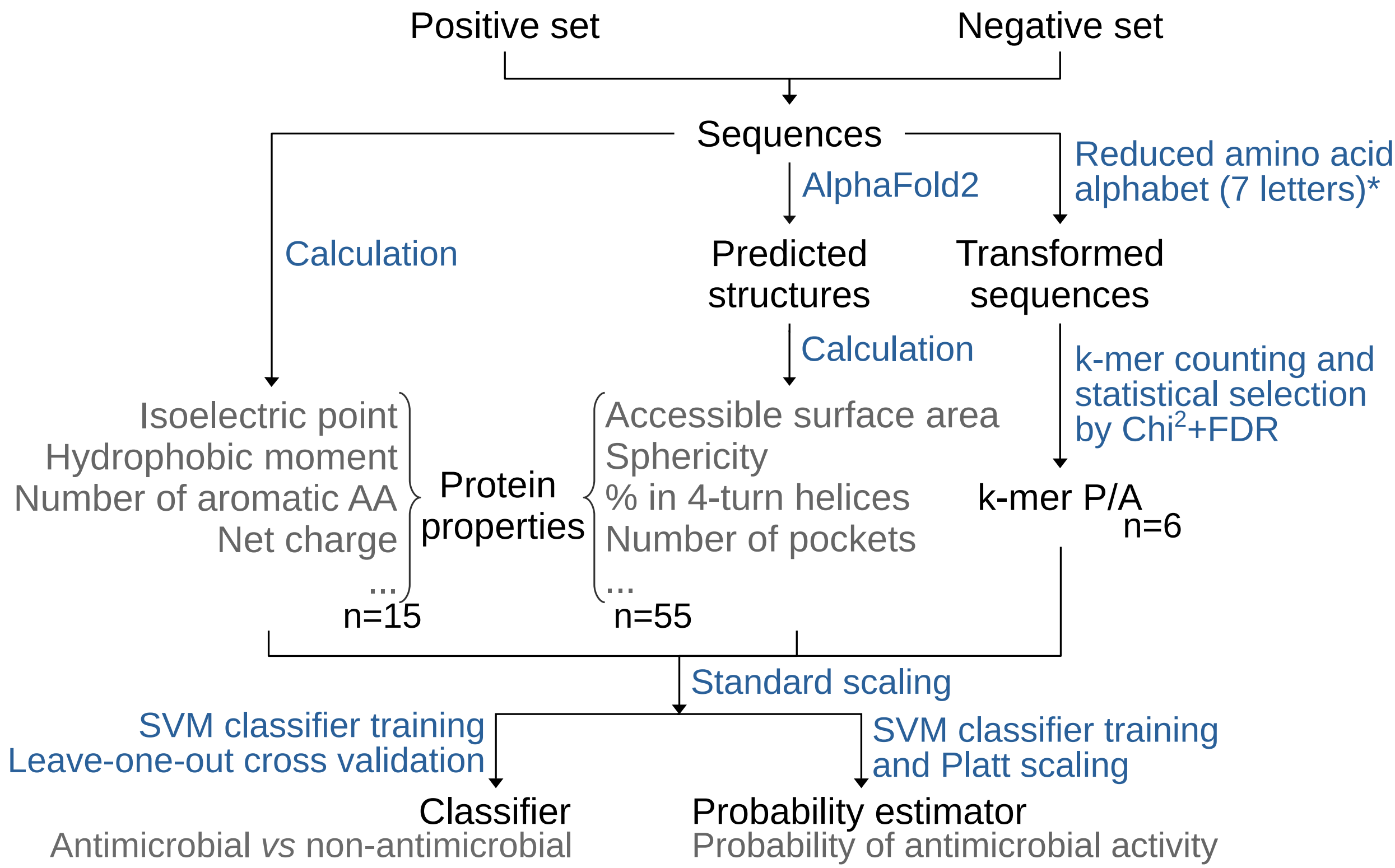


## 3. Assembling a training dataset by manual curation

**Positive training set:** 152 proteins demonstrated to have antimicrobial activity *in vitro*, manually curated based on literature.  
**Negative training set:** 304 proteins expected not to have any antimicrobial activity according to their functional annotation (same lengths and organisms as in the positive set).



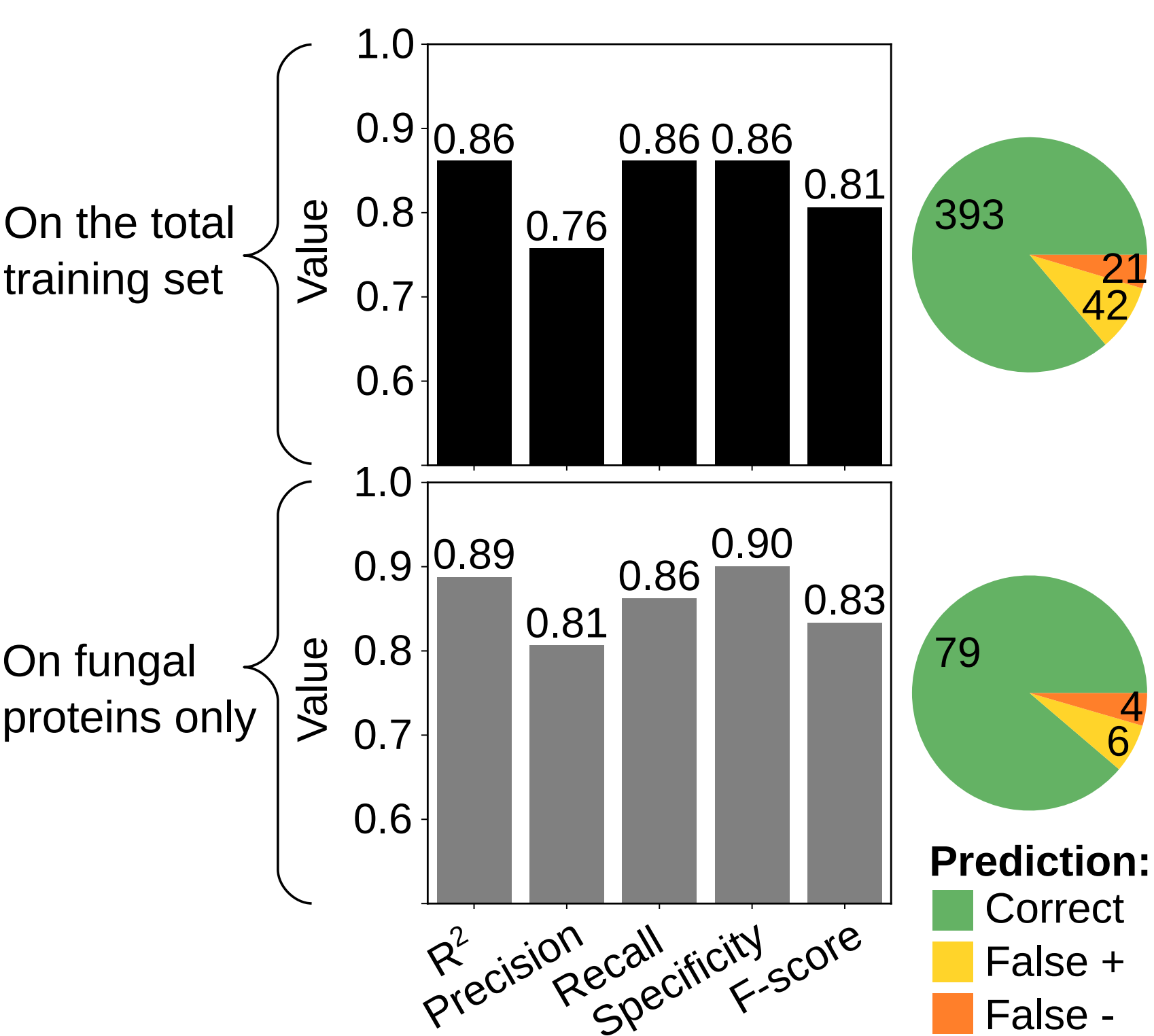
## 4. Training pipeline



AA= Amino acids  
FDR= False discovery rate (Benjamini-Hochberg correction)  
P/A= Presence/Absence  
SVM= Support Vector Machines

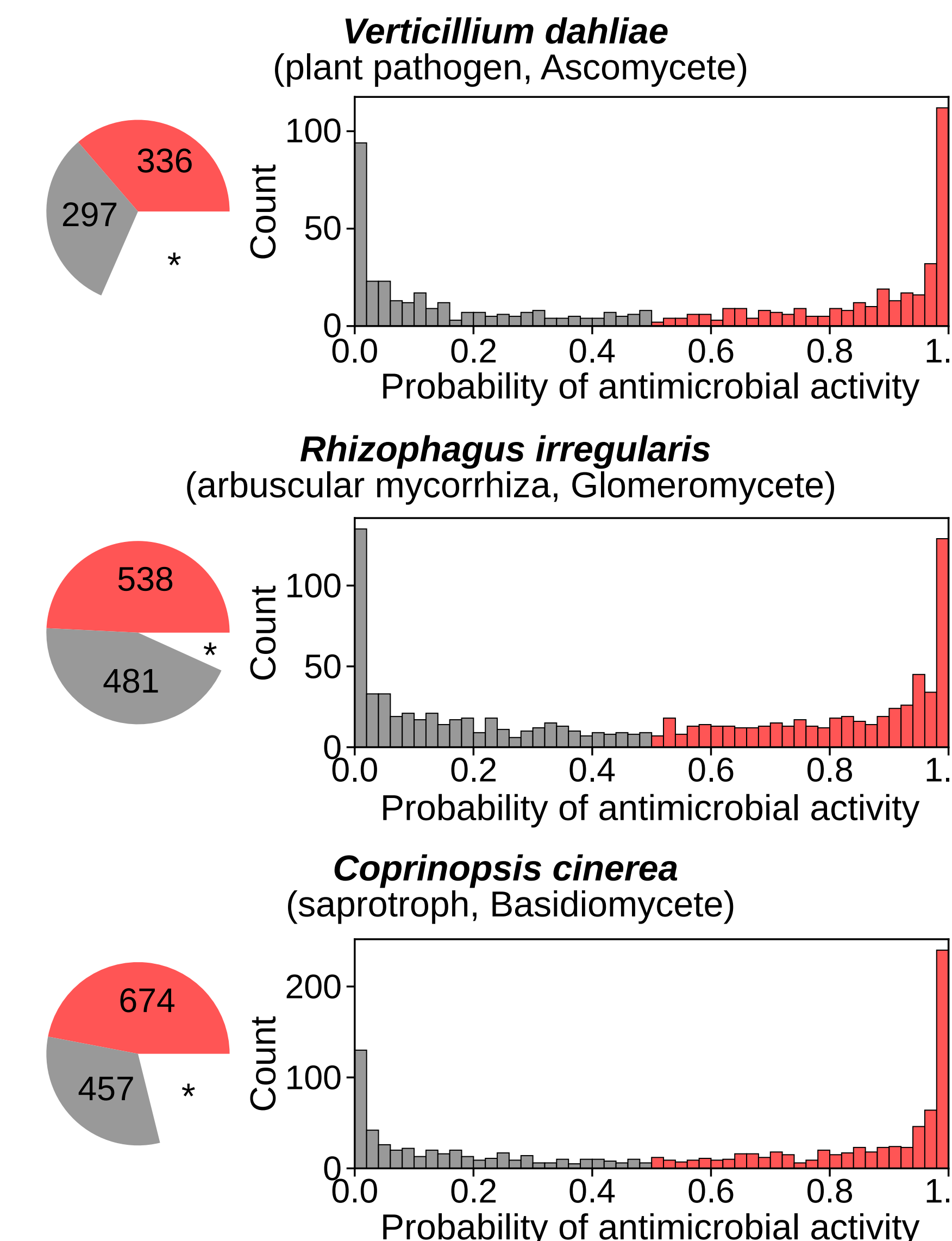
## 5. Classifier quality control

Classifier quality as estimated by leave-one-out cross-validation:



## 6. Prediction of antimicrobial activities in fungal secretomes

Our predictor reveals that one to two thirds of fungal secretomes may be composed of antimicrobial proteins.



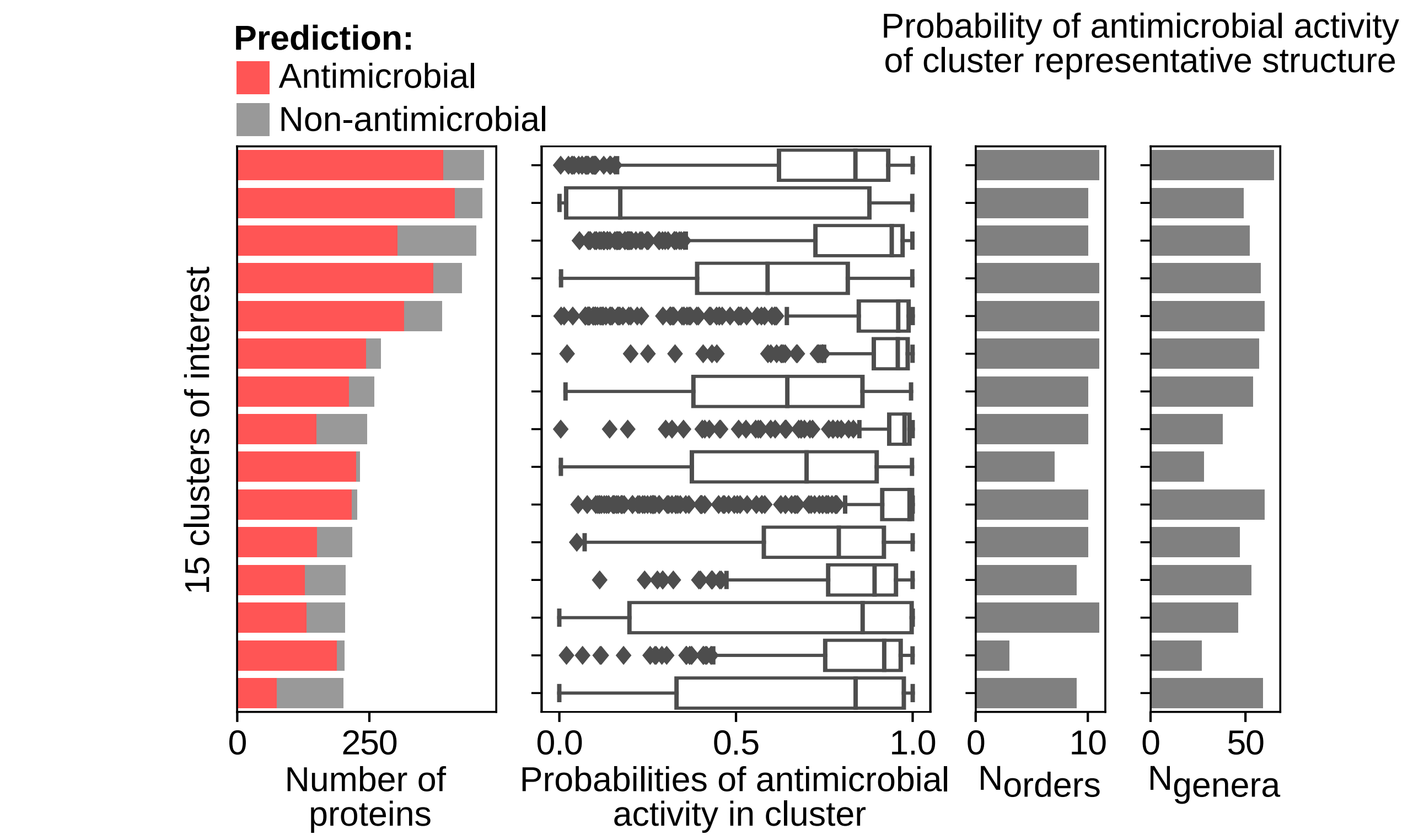
**Prediction:**  
Antimicrobial (red), Non-antimicrobial (grey)

\*: These sections correspond to carbohydrate-active enzymes, excluded from predictions since poorly represented in our training set.

## 7. Screening the AlphaFold database to discover novel antimicrobial effector families

We downloaded from the AlphaFold database [7] all the available structures for effector proteins produced by Sordariomycetes, then clustered them with Foldseek [8] using a 80% structural similarity threshold, and predicted their antimicrobial activity.

Biggest clusters with the highest probability values represent 15 candidate antimicrobial effector families with good conservation among Sordariomycetes, to further analyse and experimentally validate. These proteins are not (or poorly) functionally annotated.



[1] Snelders, Nick C., et al. "An ancient antimicrobial protein co-opted by a fungal plant pathogen for in planta mycobiome manipulation." *PNAS* 118.49 (2021): e2110968118.  
[2] Snelders, Nick C., et al. "Microbiome manipulation by a soil-borne fungal plant pathogen using effector proteins." *Nature Plants* 6.11 (2020): 1365-1374.  
[3] Snelders, Nick C., et al. "A highly polymorphic effector protein promotes fungal virulence through suppression of plant-associated Actinobacteria." *New Phytologist* 237.3 (d2023): 944-958.  
[4] Fardella, Patrick A., et al. "The *Epichloë festucae* antifungal protein Efe-AfpA protects creeping bentgrass (*Agrostis stolonifera*) from the plant pathogen *Clavireedia jacksonii*, the causal agent of dollar spot disease." *Journal of Fungi* 8.10 (2022): 1097.  
[5] Veltri, Daniel, et al. "Deep learning improves antimicrobial peptide recognition." *Bioinformatics* 34.16 (2018): 2740-2747.  
[6] Wang, Guangshun, et al. "APD3: the antimicrobial peptide database as a tool for research and education." *Nucleic Acids Research* 44.D1 (2016): D1087-D1093.  
[7] Varadi, Mihaly, et al. "AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models." *Nucleic Acids Research* 50.D1 (2022): D439-D444.  
[8] Barrio-Hernandez, I., et al. "Clustering-predicted structures at the scale of the known protein universe". *Nature* (2023).